

地方科技管理大数据平台建设的问题 与对策研究*

王翔^{1,2} 侯威¹ 唐丽娟¹

(1. 安徽省科学技术情报研究所, 合肥 230011; 2. 合肥工业大学计算机与信息学院, 合肥 230009)

摘要: 本文以地方科技管理大数据平台建设研究为背景, 分析当前地方科技管理大数据平台建设过程中存在的问题。在此基础上, 以安徽省科技管理大数据平台模型构建及创新服务模式实践为例, 提出地方科技管理大数据平台建设对策建议, 以期地方科技管理大数据平台发展提供有益参考。

关键词: 科技管理; 大数据; 科技计划; 平台建设

中图分类号: G203

DOI: 10.3772/j.issn.1673-2286.2018.08.006

随着国务院《关于印发促进大数据发展行动纲要的通知》(国发[2015]50号)发布, 大数据已经上升为国家战略资源, 习近平总书记强调推动实施国家大数据战略, 加快完善数字基础设施, 推进数据资源整合和开放共享, 保障数据安全, 加快建设数字中国, 更好地服务我国经济社会发展和人民生活改善。国家发改委、生态环境部、交通运输部、工信部等部委相继出台了各自领域的大数据发展规划和指南^[1-4], 教育部发布的2017年度普通高等学校本科专业备案和审批结果中, 近300所国内高校成立了“数据科学与大数据技术专业”^[5], 国民经济的各行各业都正在逐渐由数据来驱动。国家大数据战略对科技管理的深度与广度都提出了更高要求, 以往粗放式、经验式的管理模式, 将很难应对大数据带来的科技管理工作变革, 难以适应创新主体对科技管理工作的需求。因此, 把大数据技术、思想应用于科技管理部门的决策过程及公共服务模式中, 通过深化科技管理大数据应用, 使科技管理模式从主观经验向科学循证转变, 有着十分重要的意义。

本文以笔者参与的安徽省科技管理大数据平台建设项目为实践背景, 用大数据思维分析地方科技管理领域大数据平台建设与应用中存在的问题, 通过地方科技管理大数据平台模型构建与服务模式研究, 提出

地方科技管理大数据平台建设对策建议, 为大数据技术应用于地方科技管理决策部门提供有益参考。

1 地方科技管理大数据的内涵与国内研究现状

1.1 地方科技管理大数据的内涵

本文所指科技管理主要是科学技术研发与应用的过程管理。传统的地方科技管理数据是与科技管理相关各类数据集合, 包括科技项目、研发机构、科研成果、专家、企业(高新型或科技型)、财政投入、政策法规等信息。从大数据概念上讲, 传统的地方科技管理数据体量有限, 类别也很难称为复杂, 而且大多数属于关系型数据, 无论是项目还是论文、专利、成果, 价值密度都比较高, 这些特点都不符合大数据的特征^[6]。

本文认为, 如果运用大数据思维去分析, 地方科技管理涉及的人力、物力、财力资源最终都会以数据的形式存储、整合, 这些数据分散于长期积累的各类电子政务系统、科技管理信息系统及其他业务系统中, 如能将这些系统的数据整合与互联网实时数据进一步扩展, 此时的地方科技管理数据无论是体量还是类型复杂

*本研究得到国家自然科学基金项目“面向多源高维数据流的在线特征选择与分类方法研究”(编号: 61673152)资助。

度、数据价值密度都将接近大数据特征。如传统的专家库包含的信息相对有限,如果加入专家画像技术,通过收集专家的项目、成果、专利、论文、合作关系、同行评价、社交媒体引用热度等,对一位专家的综合评价会更加准确。同样的,单纯的科技计划项目数据难以称为大数据,如果按照国务院办公厅《关于印发科学数据管理办法的通知》(国办发[2018]17号)规定的,将每个项目涉及的科学数据与项目进行匹配,综合财政提供的项目资金使用情况、项目中期检查情况、财务预算及开支情况等信息,再辅以科技报告、科技成果登记等文献资源,将在项目立项查重、评审专家遴选、项目实施过程管理、项目验收及成果共享等科技管理流程发挥重要作用。

1.2 地方科技管理大数据的国内研究现状

在大数据与科技管理的理论研究方面,乔晓东等^[7]提出科技大数据概念,分为科技知识大数据与科技活动大数据,分别对应文献类资源和科技活动中的实体及其关系之间的数据,进而提出一种包括资源发现、网络检测采集、科研分析评价、科技情报服务等综合科技管理的大数据服务体系;袁伟等^[8]从科技资源的挖掘与利用角度,提出借助大数据管理方法与模式对科技资源数据进行标准化整合汇聚,面向科技创新发展与管理决策需求深度挖掘与利用科技资源数据的对策建议;徐迪威^[9]对大数据与科技管理的关系进行研究,提出基于大数据的科技管理存储层、处理层、应用层的层次体系。在大数据与科技管理的实际应用方面,北京市科学技术委员会提出打造“首都科技大数据平台”的口号,整合分散于各政府部门、科研机构等单位的科技资源并向社会数据需求方开放^[10];贵州省借助“云上贵州”建设,提出“科技云”这一科技数据资源开发与利用的新范式^[11];上海市研发公共服务平台整合专家人才、大型仪器、科技文献、研发基地等科技创新数据资源,建成“上海科技创新资源数据中心”^[12];“内蒙古科技创新大数据平台”分类实现服务云、管理云、数据云,实现“一站式”创新资源平台服务^[13]。

2 地方科技管理大数据建设的现存问题

由于我国大数据环境下的科技管理研究尚处于起

步阶段,虽取得一定成果,但还存在以下问题。

2.1 缺少成体系的理论研究支撑

以科技管理为主题的大数据建设研究,公开的文献报道多以平台搭建、数据中心建设为主;对科技管理数据资源整合的标准,跨体系各分散系统的科技管理数据如何汇交,科技管理大数据平台的基础模型如何构建,社交媒体等互联网数据如何融入科技管理,以及平台安全保障及运行维护模式等方面的深入研究不多;同时,虽然对大数据技术在科技管理某些环节的应用开展了研究,但对于科技管理整体过程的体系化研究不够全面,未能覆盖科技管理的各个环节,如针对应用大数据优化科技管理的关键决策环节、优化科技管理公共服务模式等问题,缺少相关理论研究。

2.2 平台重建设,轻服务

在科技管理领域,各地都纷纷开展各自的大数据平台建设,如“首都科技大数据平台”“内蒙古科技创新大数据平台”“贵州科技云”“青岛科技大数据平台”“上海科技创新资源数据中心”等,重点多在于基础环境建设及平台搭建,在服务内容方面,大部分停留在信息整合发布、数据初步分析等粗加工服务模式^[14]。

“上海科技创新资源数据中心”提供以各类情报分析产品为代表的科技管理数据精加工服务,而在时下热门的精准化定制及智能决策支撑等创新服务模式方面未见过多涉及。

2.3 缺少特色数据,数据应用能力较弱

已有研究成果在研究或实现科技管理大数据时,多将专利、期刊论文、成果等科技文献元数据作为大数据的主要来源,虽然在一定程度上保证了科技管理大数据的体量,但缺少深度挖掘(如基于科技文献的专家画像、科研合作关系发现、专家及人才评价等),也缺少特色的科技管理数据(如项目管理过程数据、科技成果登记数据、财政支付财务数据等),最终实现的科技管理服务仅是常规科技管理的信息化展示。除此之外,各大数据平台提供的数据开放、创新辅助等公共服务能力较弱,社会公众对科技管理大数据的获取及应用热情不高,影响平台的社会价值。

2.4 存在数据主权争议、涉密数据安全难以保障

由于科技管理业务的条块化分割及其他方面的原因,大多数数据处于碎片化存储阶段,各省很难有同一机构保存科技管理的全部或者大多数原始数据,也很难在所有权层面实现科技资源的重新整合。虽然有学者提出“科技云”这一概念,以期在资源供给层面减少科技信息资源服务的复杂性,但实际操作中总在一些技术层面存在无法逾越的壁垒,需要通过研究相关理论、制定相关政策法律等以保障实施。科技管理部门并非不愿意主动开放科技管理大数据平台,很大一部分原因是地方科技管理部门对科技管理数据的研究并不充分,对相关数据哪些可以公开、哪些包含隐私信息、是否涉密等问题,缺少专门机构梳理和研究,在一定程度上影响了科技管理数据的有效利用。

2.5 缺少专业化人才团队

科技管理大数据建设不仅需要数据科学、软件开发等方面的技术人才,也需要相关政策理论研究人员、科研管理人员参与其中,缺少复合型的专业化人才团队已成为地方科技管理部门开展大数据平台建设遇到的突出问题之一,管理决策部门还停留在经验决策阶段并不擅长应用大数据环境下的分析,数据分析部门并不了解底层数据逻辑无法有效建模,平台开发人员不理解科技管理业务逻辑无法有效设计并开发数据模型及业务系统接口,科研管理人员对科技管理大数据应用的需求也很难转化为大数据平台开发者需要的模型。

3 地方科技管理大数据平台模型及服务模式实例

在科技管理大数据平台建设方面,已开展了一些共性方面的研究,比较一致的做法是将科技管理大数据平台按科技管理服务内容划分为若干子平台。如许秀梅^[15]提出的基础条件子平台、科技研发服务子平台、成果转化与交易子平台、产业科技大数据服务平台;李欢^[16]提出的科技战略子平台、技术产业化平台、科技研发子平台等。这些平台构建模型,根据科技管理各项业务进行划分,是对现有科技管理体系的改良,将原先分散在各个不同机构的科技管理业务整合到一个平台

上,从而解决原先各个科技管理业务系统间相互关联性不足的问题。2017年,安徽省科学技术情报研究所提出“十三五”期间要建成“一个智库、一个中心、一个服务基地”,其中“一个中心”即建成科技管理大数据中心。本文以此为实例探讨科技管理大数据平台建设实践问题。

3.1 数据融合策略

通过安徽省科技管理大数据平台建设项目实践,本文提出科技管理大数据整合的策略:一是尽可能升级系统,拓展数据接口,互通互享公开信息;二是某个机构承接科技管理数据的使用权;三是通过制定统一的数据汇交标准、数据汇交管理办法等搜集科技管理数据;四是必须与有知识产权的数据商合作。通过上述策略,实现从离散数据到大数据,从大数据到大情报,从大情报到大决策的转变。

在大数据管理及数据融合方面,采取行政手段与数据交换相结合的方式,一是将原有安徽省科技管理信息系统升级,并预留与其他业务系统的接口,如安徽省政务服务系统、科技文献服务平台、科技成果登记系统、科技报告服务系统、成果奖励系统、各类科技统计系统、财政支付平台等;二是将不适宜重构的其他科技管理大数据相关业务系统的业务、管理与后台支撑分离,不影响业务的开展,同时将数据的使用权集中于一个机构;三是对于极少部分存在数据主权争议的系统,通过行政手段制定数据交换制度,虽然在一定程度上影响了数据的时效,但保证了科技管理大数据的完整性;四是采购专业数据商的可信社交媒体数据,并与专业数据商合作,开展全部科技管理大数据整合加工工作。

3.2 平台模型与服务模式

在尽可能获取科技管理相关大数据的基础上,借鉴吴信东等^[17]提出的“从大数据到大知识”的HACE+BigKE模型,提出安徽省科技管理大数据平台数据融合框架及平台整体架构(见图1)。该模型主要展现地方科技管理大数据池的数据来源,包括科技计划项目、科技成果登记、科技统计数据、地方专家库、专利信息服务及科技成果奖励等;单位自建的科技文献平台及可交换数据的大型仪器共享平台等;采购来自专业数据商的社交媒体数据。

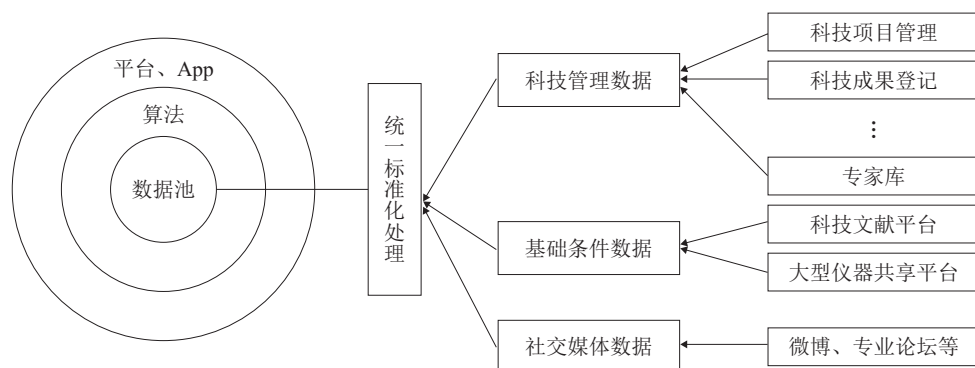


图1 “地方科技管理大数据平台”模型

同时，该模型还实现了安徽省科技管理大数据平台主体架构。

数据池外层是各项数据挖掘算法，如最基础的中文分词系统，专家画像所需的聚类及关联规则算法，项目评审所需的基于深度学习的专家推荐系统，成果转化推荐所需的协同过滤算法等。

最外层是整个科技管理大数据的统一服务平台，对外主要分为四大类服务，即面向社会公众的开放数

据服务、基础条件服务、统计分析服务等，面向宏观科技管理决策的深度分析与情报支撑服务，面向科技研发的嵌入式信息服务，面向成果产业化的技术转移与成果推介服务等。同时，还需要有移动App及微信小程序等移动端服务的支撑，以适应当前用户的使用习惯。

本文以面向科技研发的嵌入式信息服务为例，说明地方科技管理大数据平台如何嵌入科研项目流程，提供面向科技研发的嵌入式信息服务（见图2）。

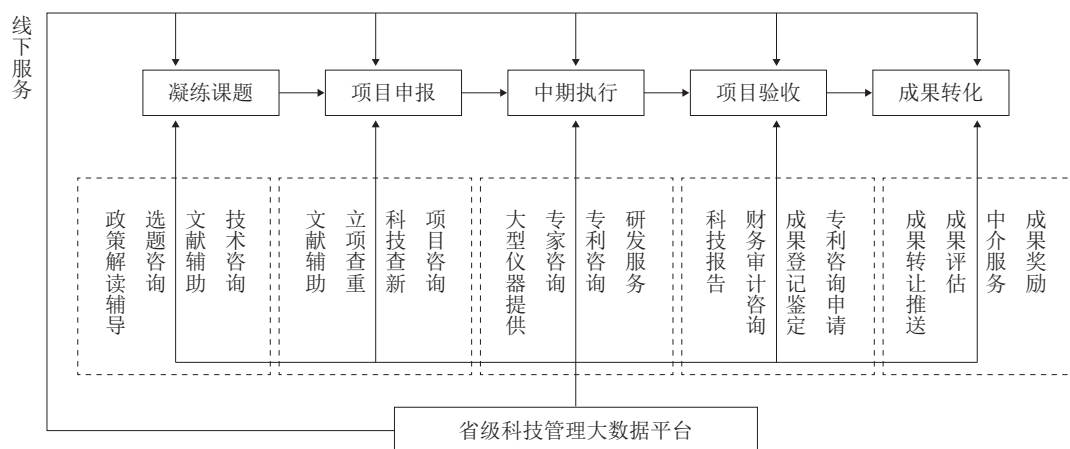


图2 “地方科技管理大数据平台”服务模式举例

面向科技研发的嵌入式信息服务需要结合创新主体的科技研发需求，主动对接、参与重大科研项目的研发活动，体现对重大科研项目的支撑作用，主要实现科技文献辅助、专利分析支持、创新资讯检索、文献资源检索、科技政策咨询、技术难题找专家或团队等业务功能，实现以科技项目研发活动为中心，嵌入科技研发链各个环节，面向重大科研项目的精准化、链条化信息服务模式。

4 推进地方科技管理大数据建设的建议

大数据平台建设是一项由问题驱动的工作，涉及数据科学、算法设计、管理科学、软件开发等多方面工作，本文通过安徽省科技管理大数据平台建设的实践，提出地方科技管理大数据建设相关对策。

4.1 加强理论研究，形成体系方案

理论模型对于科技管理大数据平台的建设至关重

要。“首都科技大数据平台”的招标中,关于平台架构及其服务模式的理论研究预算几乎是平台软硬件建设预算的2倍^[18]。在开展安徽省科技管理大数据平台建设项目前,通过申请软科学专题研究项目、内设调研课题等方式,梳理涉及数据方面的问题,主要包括科技管理大数据覆盖范围、如何解决涉密和隐私数据问题、如何建立科技管理大数据平台模型、如何保障跨系统数据汇交及大数据平台的创新服务模式等方面问题,并在取得的理论研究成果基础上,成功申请省级重点研发项目,开展科技管理大数据平台建设。

4.2 做好顶层设计,开展精准服务

在设计科技管理大数据平台时,充分调研各类用户对科技管理服务需求,社会公众、政府决策、科学研究(智库)、企业科技攻关等,各自需要什么样的数据服务,最终实现四大类精准服务。一是面向社会大众提供基础性信息与数据公开服务,通过分析用户的检索词、页面停留时序分析等技术,将科研成果向潜在用户进行关联推送;二是面向科技管理决策支撑的情报数据服务,如提供低人力成本的智能化产业分析报告、计划项目管理关键环节的数据统计支撑服务等;三是面向科学研究的支撑服务,如科研院所引进人才时的人才评价数据服务、相关科研领域的科技报告数据服务等;四是面向重点科技攻关项目的嵌入式服务,通过用户分析,主动推介科技管理相关数据,如项目申报阶段,提供相关文献、政策指南等公共服务及项目查新等定制化有偿服务等,实现对重点研发项目的支撑作用。

4.3 融合地方特色,深化服务内涵

数据只有用起来,才更有价值。在需求基础上,还要进行拓展研究,对共性需求的数据服务需要做精做深,最大限度地保证科技管理大数据平台源于需求又高于需求。在地方特色方面,根据地区创新发展情况,全部或部分使用平台数据,编制并向社会公众发布《安徽省各城市创新能力评价》等,实现大数据平台对创新的宏观辅助功能;面向相同产业研发机构的共性需求,与地方情报服务机构共建特色产业知识服务平台等,将大数据平台服务延伸至市县的创新型园区,既有线上的数据服务,也有线下的需求反馈及服务支撑。

4.4 确保数据可信,减少主权争议

科技管理大数据既包括科技管理过程产生的数据,也包括科研项目涉及的科学数据、文献类资源,以及来自社交媒体的相关互联网数据,任何一个机构独立做好科技管理数据的集成整合都是非常困难的,必须强调协同合作。一是与数据主权单位的协作,二是与数据加工商的协作,保证数据源的可信度,在此基础上大数据分析结果才能更加准确。

解决数据主权争议,一是在技术实现层面,借助科技管理信息系统升级,预留与其他相关系统接口,互通共享数据;二是主动承接机关科技管理事务性工作,做好各类科技管理、科技统计、专利服务等相关信息系统后台网络支撑服务,确保数据能实时可用;三是通过行政手段,制定数据交汇相关政策,保障数据更新频次;四是与国内知名数据商合作,避免文献类资源的知识产权争议。

4.5 “走出去,请进来”,共建科研团队

为保障科技管理大数据建设持久开展,必须坚持“走出去,请进来”的原则,一是对对已有团队进行扩充,与省内相关大学共同培养科研人才;二是邀请省内大数据相关知名学者参与大数据平台研发工作,通过修炼内功与外部辅助,打造专业的大数据研发服务团队。

5 结语

地方多年来不断积累的各类科技管理信息系统及电子政务系统在某种意义上决定了开展科技管理大数据平台建设很难脱离现有基础,需要兼顾历史基础、现实需求与未来发展。本文认为,地方开展科技管理大数据平台建设离不开政府主导,可以有效解决地方范围内科技管理领域可用、可信数据相对匮乏等问题,保证地方科技管理数据的连续性,是地方科技管理从治理向服务职能转变的重要手段,也是响应国务院《促进大数据发展行动纲要》的有益实践。地方科技管理大数据平台建设,不仅需要政府将其应用到决策管理中,也需要包括科技研发、成果转化、社会公众等多方共同深化大数据应用,真正实现从大数据到大知识、大情报。

参考文献

- [1] 国家发展改革委关于印发“十三五”国家政务信息化工程建设规划的通知[EB/OL]. (2017-08-24) [2018-01-01]. http://www.gov.cn/xinwen/2017-08/24/content_5220193.htm.
- [2] 关于印发《生态环境大数据建设总体方案》的通知[EB/OL]. (2016-03-08) [2018-01-01]. http://www.zhb.gov.cn/gkml/hbb/bgt/201603/t20160311_332712.htm.
- [3] 交通运输部办公厅关于推进交通运输行业数据资源开放共享的实施意见[EB/OL]. (2016-09-02) [2018-01-01]. http://zizhan.mot.gov.cn/zfxgk/bnssj/kjs/201609/t20160902_2083604.html.
- [4] 工业和信息化部关于印发大数据产业发展规划(2016—2020年)的通知[EB/OL]. (2017-01-17) [2018-01-01]. <http://www.miit.gov.cn/n1146295/n1652858/n1652930/n3757016/c5464999/content.html>.
- [5] 教育部关于公布2017年度普通高等学校本科专业备案和审批结果的通知[EB/OL]. (2018-03-15) [2018-07-01]. http://www.moe.gov.cn/srcsite/A08/moe_1034/s4930/201803/t20180321_330874.html.
- [6] 杜义华. 大数据背景下中科院科研管理数据中心建设探讨[J]. 计算机系统应用, 2015(1): 79-85.
- [7] 乔晓东, 朱礼军, 李颖, 等. 大数据时代的技术情报工程[J]. 情报学报, 2014(12): 1255-1263.
- [8] 袁伟, 石蕾. 大数据背景下科技资源信息挖掘与利用的思考[J]. 中国科技资源导刊, 2017(6): 1-5.
- [9] 徐迪威. 大数据与科技管理[J]. 科技管理研究, 2013(24): 216-218, 232.
- [10] 施永孝. 大数据视角下北京市科技创新平台运营机制研究[D]. 北京: 北京理工大学, 2015.
- [11] 贵州省科技云[EB/OL]. [2017-08-02]. <http://www.gzst.gov.cn/>.
- [12] 上海科技创新资源数据中心[EB/OL]. [2018-02-26]. <http://www.sstir.cn/>.
- [13] 内蒙古科技创新大数据平台[EB/OL]. [2017-03-28]. <http://www.cxnmg.gov.cn/>.
- [14] 张苗苗, 赵捧未, 范晓玉. 云环境下科技管理数据服务模式创新研究[J]. 情报理论与实践, 2018, 41(1): 8-42, 60.
- [15] 许秀梅. 科技大数据服务平台系统架构与运作模式探析[J]. 财会月刊, 2017(20): 75-80.
- [16] 李欢. 大数据背景下科技管理创新平台构建研究[J]. 科学管理研究, 2014, 32(3): 44-48.
- [17] 吴信东, 何进, 陆汝钤, 等. 从大数据到大知识: HACE+BigKE[J]. 自动化学报, 2016, 42(7): 965-982.
- [18] 首都科技大数据平台研发与应用项目招标公告[EB/OL]. (2016-11-11) [2018-01-01]. <http://www.bidchance.com/info.do?channel=calgg&id=17201148>.

作者简介

王翔, 男, 1982年生, 博士研究生, 助理研究员, 研究方向: 数据挖掘、科技情报, E-mail: 27808697@qq.com。

侯威, 男, 1963年生, 副研究员, 研究方向: 科研管理、科技情报。

唐丽娟, 女, 1984年生, 硕士, 助理研究员, 研究方向: 科研管理、科技情报。

Research on Problems and Countermeasures in Big Data Platform Construction of Provincial Science and Technology Management

WANG Xiang^{1,2} HOU Wei¹ TANG LiJuan¹

(1. Institute of Scientific and Technical Information of Anhui, Hefei 230011, China;

2. School of Computer and Information, Hefei University of Technology, Hefei 230009, China)

Abstract: Based on the research for big data platform construction of the provincial management of science and technology, analysis of the current problems in construction of provincial science and technology management big data platform were executed. Furthermore, model of big data platform for Anhui science and technology management was set up and a new service pattern carried out by the big data platform model was illustrated. After that, some suggestions on the construction of provincial science and technology management big data platform were proposed. It will provide reference values for the development of provincial science and technology management big data platform.

Keywords: Science and Technology Management; Big Data; Science and Technology Project; Platform Construction

(收稿日期: 2018-07-23)