

美国的大数据研发计划及对我国的启示

李健¹ 王丽萍² 刘瑞¹

(1. 国家科技基础条件平台信息技术中心, 北京 100191; 2. 上海研发公共服务平台, 上海 200235)

摘要: 2012年3月美国政府公布了大数据研发计划。这项计划重点是增强美国联邦政府收集海量数据, 分析萃取信息的能力。美国科学基金会等6个联邦部门为此宣布投资2亿美元, 主要用于提前获取、处理、共享海量数据的工具和技术, 文章在重点分析美国政府大数据研发计划背景的基础上, 结合我国科技平台资源共享现状, 借鉴美国经验, 提出制定科技平台大数据发展计划, 促进资源共享的必要性以及相关具体措施建议。

关键词: 大数据; 科技平台; 资源共享; 大数据研发计划; 美国

中图分类号: G203

文献标识码: A

DOI: 10.3772/j.issn.1674-1544.2013.01.004

Revelation of US Government “Big Data Research and Development Initiative” for China’s Scientific Platform Resource Share

Li Jian¹, Wang Liping², Liu Rui¹

(1. Information Technology Center for National Science & Technology Infrastructure (ITC-NSTI), BeiJing 100191; 2. Shanghai R & D Public Service Platform (SGST.CN), Shanghai 200235)

Abstract: In March 2012, the U.S. government announced “Big Data Research and Development Initiative”. The program is focused on improving the U.S. Federal government’s ability to extract knowledge and insights from large and complex collections of digital data, the initiative promises to help solve some of the Nation’s most pressing challenges. The program includes several federal agencies including NSF, HHS/NIH, DOE, DOD, DARPA and USGS who pledge more than \$200 million in new commitments that they promise will greatly improve the tools and techniques needed to access, organize, and glean discoveries from huge volumes of digital data. This article will focus on analyzing the background of the Big Data program, combined with current situation of china’s science & technology platforms and resource sharing, proposed technology platform learn from American experience to develop big data plan, as well as the related specific measures.

Keywords: big data, S&T platform, resource share, revelation, Big Date Research and Development Initiative, USA

1 引言

随着互联网、物联网、云计算等的快速兴起和普及, 当前人类社会的数据增长比以往任何一个时期都要快, 数据的规模越来越大, 数据变得越来越多样化, 也变得越来越复杂。在这种情形下, “大数据”这一全新的概念应运而生。大数据时代已经到来。为了在大数据时代抢得先机, 保持自己

在科技、经济等多个方面的领先地位。2012年3月29日, 奥巴马政府公布了“大数据研发计划”(Big Data Research and Development Initiative), 以增强联邦政府收集海量数据、分析萃取信息的能力, 迎接新的挑战。美国科学基金会、卫生福利部/国立卫生研究所、能源部、国防部、国防部高级研究计划局、美国地质调查局等6个联邦部门为此宣布投资2亿美元^[1-2], 以提高收集、储存、保留、管理、

第一作者简介: 李健(1983-), 女, 数据工程师, 主要研究方向: 科技资源共享。

收稿日期: 2012年10月18日。

分析和共享海量数据所需的核心技术和先进性。同时,奥巴马政府还计划与工业界、大学研究界、非营利性机构与管理者一起利用大数据所创造的机会,形成“众人拾柴火焰高”的局面。这是美国政府在政策层面上将“大数据”上升到国家意志的重要举措,其影响将极为深远。

如果能够借鉴美国实施大数据研发计划的经验,制定符合我国自身特色的大数据战略,实现科技资源共享,将会有效地提升我国科技资源的加工和共享利用水平,提高我国科技管理决策的科学性,更好的支持科技创新。

2 美国启动大数据研发计划的背景分析

美国政府将数据定义为“未来的新石油”。认为一个国家拥有数据的规模、活性及解释运用的能力将成为综合国力的重要组成部分,甚至对数据的占有和控制将成为继陆权、海权、空权之外的另一个国家核心资产。毫无疑问,21世纪,数据将成为下一个大的战略资源。在这个背景下,美国启动大数据研发计划,可谓是寓意深远。

2.1 增强科研能力,推进科研方式转变

未来的科学技术创新将越来越依靠科学数据的优势,以及将科学数据通过数据挖掘、集成、分析和可视化工具等转换为信息和知识的能力^[3]。数据采集、数据整理、数据分析及数据可视化等都是大数据所包含的内容,最终实现目标是科学家、科研人员、决策者和公众提供决策支撑,新型科学研究需要从原始数据向决策信息转变^[4]。伴随着大科学工程装置、物联网技术在科研活动中应用,海量科学数据的获取、传输、处理、应用成为新挑战。数据是信息的原材料,也是科研的源头活水,谁掌握了这一资源,谁就掌握了发展的主动权。

2.2 提高政府管理水平和决策能力

2012年5月,MeriTalk发布了题为《大数据鸿沟》的报告。根据之前对美国151个政府首席信息官和信息技术经理调查得出的数据,仅是未来两年,政府机构将添加1000万亿字节(PB)存储数据^[5]。在这些数据中,不仅包括结构化数据,也包括大量松散的非结构化数据。由于信息堆积、缺乏数据存储和获取能力以及受制于计算能力和人员等原因,政府部门难以从大数据中受益,迫切需要通过大数据技术改善整体部门效率,改善/加速精确决策和提高预测水平^[6]。

2.3 抓住技术机遇,推进经济发展

不仅是政府部门,企业也开始认识到大数据所蕴含的巨大价值。是否能够充分利用和把握大数据的机遇,将是未来决定企业成败的关键^[7-8]。美国总统委员会的科学技术顾问Stephen Brobst曾表示:“过去3年里产生的数据量比以往4万年的数据量还要多,大数据时代的来临已经毋庸置疑。我们即将面临一场变革,新兴大数据将成为企业发展的当务之急,而常规技术已经难以应对PB级的大规模数据量。这一变化所带来的挑战,是成功企业在未来发展过程中必须面对的。只有那些能够运用这些新数据形态的企业,方能打造可持续的重要竞争优势。^[9]”

全球IT、互联网巨头都已经意识到数据的重要性。包括EMC、惠普、IBM、微软等在内的企业都纷纷通过收购与数据厂商来实现技术整合^[10]。

随着企业对能处理和分析海量结构化和非结构化数据工具的需求增加,“大数据”领域已经吸引了投资者的注意,唤起了他们的投资热情。有业内人士预测,10年内,大数据工具市场会从2011年的90亿美元增长到860亿美元。到2020年,大数据工具花费将占企业所有IT支出的11%^[11]。

3 美国大数据研发计划目标和内容

白宫科技政策办公室设定的大数据研发计划的目标主要包含以下3个方面内容:(1)改进现有人们从海量和复杂的数据中获取知识的能力,开发能对大量的数据进行收集、存储、维护、管理、分析和共享的最先进的核心技术。(2)利用这些技术加快科学和工程领域探索发现的步伐,并加强国防安全,转变现有的教学和学习方式,解决美国面临的最具挑战性的问题。(3)扩大从事大数据技术研发利用工作的人员数量,增加大数据技术开发和应用所需人才的供给^[12]。

在美国联邦大数据研发计划中,首批参与的6个部门都制定了详细的计划,为此白宫办公室发表了相关的大数据计划一览表(表1)。每个机构大数据研发计划侧重点也有所不同,主要包括数据仓库、数据存储、数据挖掘、数据分析、可视化、网络安全、虚拟实验室、观测系统、协作工具、超级计算机、网络基础设施等方面,既包括数据中心也包括软硬件系统,并且部分计划内容涉及了详细的学科领域,是科研信息化与专业领域相结合。

表1 美国相关联邦机构大数据研发计划总结

机构	项目目标侧重点	部分项目名称内容	部分项目名称内容
美国国家科学基金会与国立卫生研究院	推进大数据科学和工程的核心方法及技术研究, 管理、分析、可视化以及从大量的多样化数据集中提取有用信息的核心科学技术 ^[12]	神经科学计算合作研究(CRCNS)	互联网病人控制医学图像共享
		推进大数据科学与工程技术与核心技术 (BIGDATA)	神经影像信息工具和资源信息中心 (NITRC)
		21世纪科学与工程学网络基础设施框架 (CIF21)	扩展影像学档案工具箱(XNAT)
		引文数据	解剖学计算和多维建模资源
		开放科学数据和软件保存 (DASPOS)	洛杉矶神经成像实验室(LONI)
		计算先行者计划	电脑辅助功能性神经外科数据库
		随机网络模型重点研究组	生物信息科学与技术倡议(BISTI)
		创意实验室	神经科学信息框架(NIF)
		数学与统计中计算和数据处理 (CDS&E) (CDS&E-MSS)	人类大脑联络图工程
		关键转换点监测 (MCTP)	国家生物医学计算中心(NCBC)
		激光引力波干涉观测站(LIGO)	病人报告结果测量信息系统(PROMIS)
		开放科学网格(OSG)	传染病代理研究模型(MIDAS)
		理论和计算天体物理学网络 (TCAN)	结构型基因组计划
		数据挖掘挑战计划	全球蛋白质数据银行 (wwPDB)
		癌症成像存档(TCIA)	生物医学信息研究网(BIRN)
		癌症基因组图谱(TCGA)	生物学及病床信息集成(i2b2)
心血管病研究网格(CVRG)	国家老龄计算机数据存档(NACDA)		
集成数据的分析、匿名与共享 (iDASH)	人口研究数据共享(DSDR)		
国防部以及国防部高级研究局	推进大数据辅助决策, 集中在情报、侦查、网络间谍等方面, 汇集传感器、感知能力和决策支持建立真正的自治系统, 实现操作和决策的自动化 ^[12]	警惕网络: 培育和测试网络防御情境意识规模竞争	网络内部威胁 (CINDER) 计划
		情报共同体 (IC)	对加密数据的编程计算(PROCEED)研究
		NSA/ CSS商业解决方案中心(NCSC)	视频和图像的检索和分析工具(VIRAT) 计划
		多尺度异常检测 (ADAMS) 项目	XDATA项目: 分析大量的半结构化和非结构化数据的计算技术和软件工具
美国能源部	改善数据集研究, 通过先进的计算进行科学发现, 建立可扩展的数据管理、分析和可视化研究所 ^[12]	高级科学计算研究中心(ASCR)	生物和环境研究计划 (BER)
		高性能存储系统(HPSS)	系统生物学知识库 (Kbase)
		千万亿次数据分析处理计划	高级计算科学发现计划(SciDAC)
		下一代网络方案支持工具	高能物理计算项目
		BES科学用户设施	美国核数据计划(USNDP)
美国地质调查局	给科学家提供深入分析的场所和实践、最高水平的计算能力和理解大数据集的协作工具, 催化在地理系统科学的创新思维 ^[12]	约翰韦斯利鲍威尔分析及综合中心	

4 美国政府、高校、企业和非营利组织的任务与合作

美国政府大数据研发计划的特点是以美国各级政府为主导,依托高校和科研机构,鼓励广泛参与(既包括联邦政府机构和各州政府,也包括高校、非政府组织和企业等),并以多种形式充分合作。组织形式上主要以具体项目为纽带,研究目标明确,任务具体。

4.1 联邦政府机构

白宫科技政策办公室不但发布了大数据研发计划,并组织了大数据高级监督小组协调和拓展政府在这一重要领域的投资^[12]。以美国国家科学基金会等为首的六大美国政府联邦机构积极开展计划促进大数据应用。其中的一些具体做法和进展如下。

(1)利用数据技术获得实时地震信息。2012年4月,美国地质调查局利用大数据技术,使earthquake.usgs.gov网页更具特色,可以获得实时地震信息^[13]。

(2)提出纳米知识基础设施倡议。2012年5月,美国NNI提出纳米知识基础设施倡议,该倡议将在4个重要领域实施:①多种多样的科学家、工程师和技术人员协作团体,支持纳米技术研究、开发和应用,满足国家需求。②用于多科目智能协作的灵敏模型网络,有效耦合试验性基础研究、建模和应用开发。③足以支撑的网络工具箱,有效应用纳米设计相关模型和知识。④强大的纳米技术数据和信息基础社会,支持不同学科和应用领域的有效数据共享、协作和创新^[14]。

(3)创建可扩展数据管理与可视化机构。美国能源部部长朱棣文2012年3月29日宣布,将耗资500万美元创建可扩展数据管理与可视化机构,其目标是改善美国从大量复杂数字化数据集合中获取知识,形成深刻见解^[15]。

(4)实施《推进大数据科学与工程的核心技术与科技》规划。美国国家科学基金会联合美国国立卫生研究院发布了《推进大数据科学与工程的核心技术与科技》规划,其目标是:挖掘和利用从海量数据集收集到的知识,加速科学与工程研究的进程。该规划将为旨在改善数据收集、管理、数据分析以及科研信息化协作环境的开发和评估新算法、统计方法、技术和工具研究提供资金支持^[16]。

4.2 州政府

在联邦政府部门的支持下,美国各州政府也

积极参与大数据研发计划或者指定符合自身特点的大数据研发计划,其中麻省是代表之一。2012年5月30日,麻省公布了几个新州政府计划,目标是支持新兴的大数据研究。麻省将组建大数据联盟的小组。这是一个由麻省学术界和行业界领导的委员会。该联盟将与非营利的麻省技术协作组织合作,发展相匹配的大数据项目资助规划。一些大数据公司和风险投资公司也加入协作。即将完工的投资1.63亿的麻省绿色高性能计算中心将成为麻省大数据研究的一个资源^[17]。

4.3 高校

作为技术创新的重要力量,高校也积极投身到大数据研发计划中来,一方面努力开展相关技术的研究工作,另一方面则积极开展大数据相关人才的培养。2012年5月30日,麻省理工学院举办新闻发布会,公布bigdata@CSAIL计划。英特尔公司也宣布将在麻省理工学院人工智能实验室(CSAIL)建立英特尔科学和技术中心^[17]。bigdata@CSAIL计划实施的研究重点领域包括诸如金融,医药,社交媒体和安全等领域。其中一个项目将开发更精致的工具,深入彻底的处理医学信息,从而形成更精确地治疗技术,为患者提供更好的治疗方案。包括布朗大学等其他6个教育机构的研究人员将远程参与该计划。另外,赖斯大学研究人员创建了一个以网页为基础的计算器,预测休斯顿某一特定地点飓风带来的风险。该工具借鉴历史数据和气象数据,为该城市居民提供实时风险预测^[18]。此外,在一些大学里也开始开设一门全新的研究型课程,培养下一代的“数据科学家”。位于奥斯汀的启动公司Civitas Learning获得了410万美元资金,将凭借大数据概念开发一个用于教育决策的数字化平台。学院和大学拓展资源以尝试和满足多种多样和日益增加的学生需求。从课程教学计划、资助、奖学金、测验和教学风格,高校都可以获益于大数据收集和分析,并且这可能潜在地降低美国学生中途退学数量^[19]。

4.4 企业和非营利组织

大数据最先被互联网企业所重视,如Facebook等大型科技公司巧妙地利用用户遗留在网络上的数码痕迹分析用户需求,然后向用户推荐观看电影或者与某人联系的建议。大数据已经成为知名咨询公司麦肯锡重要决策的必备因素。美国一些大型公司已经开始赞助大数据相关的竞赛,并且在为高等院校的大数据研究提供资金。EMC、惠普、IBM、微

软等IT巨头纷纷通过收购“大数据”相关企业来实现技术整合^[10]。

美国非营利机构也在利用大数据提供公共服务。例如“数据无边界”通过无偿的数据收集、分析，以及可视化为非营利性组织提供帮助。

4.5 广泛合作

在大数据的研究上，美国政府以及各机构还积极采取措施开展各种形式的合作。例如，2012年6月5日，美国国家科学基金会主任萨布拉·苏瑞喜和日本文部省科学大臣平野博文在日本东京会面，双方原则同意支持计算机科学家、工程师、社会科学家、生物学家、地学科学家、物理学家和数学家等专家研究协作，通过大数据技术，增强对灾难了解，提高社会韧性。

双方认为具有研究潜力的主题包括：(1)处理灾难中产生的大数据，提高分析、建模和计算能力，应用诸如概率性灾难模型。(2)改善信息技术的顺应力和反应能力，实现实时数据感知、可视化、分析、检验和预测，这对具有时效性的决策至为关键。(3)提升基础知识和创新，支撑民用基础设施以及分布式基础设施网络。(4)获取大数据并改善人类、社会和国际范围内关于灾难预防和反应的宽泛知识，包括人类、社会、经济和环境规模等。(5)整合多个学科专业知识，利用突发事件准备以及社区响应等相关的各来源大数据，为终端使用者提供信息^[20]。双方赞成制定一个行动方案，将在2012年底前宣布更多的细节协议。

此外，白宫科技政策办公室将会支持建立一个跟大数据相关论坛，包括最新的公私组织之间的合作。政府机构还将拨款资助一些公私合营项目。

5 我国采用大数据实现科技资源共享服务的建议

国家科技基础条件平台是根据《“十一五”国家科技基础条件平台建设实施意见》和《国家中长期科学和技术发展规划纲要（2006年-2020年）》等文件的精神，由科技部、财政部共同启动建设的。它由大型科学仪器与实验基地、自然科技资源、科学数据、科技文献、网络科技环境、科技成果转化基地等23家国家科技平台构成，具有基础性、公益性、共享性等显著特性，其宗旨是充分运用现代技术，推动科技资源共享，促进全社会科技资源优化配置和高效利用，提高我国科技创新能力。

经过“十一五”期间的努力，国家科技基础条件平台已整合28大类科技资源信息数据库，各类科技资源信息近600万条，涉及科技数据总量1000余TB；涵盖科技文献、自然科技资源、科学数据、大型仪器、科研成果等各领域，其中大型仪器近4万台，自然科技资源800多万份，科技图书文献20多万种，可共享科学数据库（集）160TB以上。与此同时，各省市也建成2000余个公共研发平台和近3000个企业创新研发平台；建立了300余个产业共性技术服务平台和近400个技术转移转化服务平台；建立了200余个公益性地方科技资源共享平台。

由此可见，目前我国在科技平台的建设上已经取得了很大的进展，各科技平台已经搜集了建国以来我国科学研究所产生的科研成果，资源种类多样，数据量大，不仅包括论文、标准、数据库（集）等实体数据，还包括植物种质、微生物菌种等实物资源。但同时也必须承认，在这些资源的充分利用、加工和对外服务方面，与国外先进水平相比还存在着较大的差距，数据的价值还远远没有得到开发和体现。在这种情况下，借鉴美国大数据研发的成功经验，利用大数据相关的先进技术，对于提高我国科技平台的建设水平，推进平台由传统信息服务向现代知识服务转变具有重要的意义。

(1) 完善标准规范，建立数据生态系统

科技资源数据的标准化和质量控制工作是实现科技资源数据共享的基本前提和保证。但我国科技平台建设还没有形成完整的规范体系，数据质量的控制机制也不够完备，部分数据资源还存在着质量问题。所有这些问题都不同程度地影响了科技资源的共享及价值的充分发挥。为此，迫切需要在现有工作的基础上，结合大数据研究的先进成果，制定一套完整的科技资源数据汇交、加工和保藏、共享的规范化和标准化体系，并开发相应的软件工具，建立对数据的多级质量审核控制机制，保证数据的质量。最终，建立一个数据生态系统，让各种科学数据都能在统一的标准和协议的框架下交流和操作，而不会发生语义扭曲；使得不同平台之间的数据和服务能够无缝整合，具有互操作性，保障不同学科的科学家们协作和跨领域研究。

(2) 建立健全的数据汇交、保藏机制

面对层出不穷的科研成果，应当建立健全的数据汇交机制，制定相应的管理办法，将科技资源数据的汇交纳入到科研管理的全过程中来，使其成为

课题申请和课题结题等的必要条件。同时,利用大数据技术建立相关的软件系统,实现对海量科技资源数据的采集、保藏、溯源、分类、检索,完成对国家财政资助的各级各类科研项目的成果数据的搜集、保藏,实现平台科技资源的持续更新和完善。

此外,还应当建立完善的知识产权保护和激励机制,解除科技人员在共享方面存在的顾虑,并鼓励其他类型的科研项目的科技资源汇交。

(3) 开展资源共享共性技术研究

通过网络特别是互联网,提供基于海量数据的各类互联网服务或信息服务是信息社会发展的趋势。虽然各领域平台大都建立了自己的信息系统和网站,但由于种种原因,各个科技平台采用的技术、标准等各有不同,水平参差不齐。这不仅影响了各平台的服务水平,也使得跨平台的数据利用面临很大的困难。

必须注意的是,虽然各大平台承载的内容各不相同,但其基本功能都涉及数据保藏与整合、信息分类和检索、数据共享、数据分析挖掘和可视化、信息安全、科研协同等多个方面等。因此,在国家科技平台大数据应用中,应该针对这些共性需求,在现有技术和工具的基础上,基于利用大数据的理念和技术,开展自主设计和研发,形成相应的核心技术和产品,并进行工程化推广,从而解决各平台面临的共性问题,提升平台整体的技术水平和服务能力,减少重复开发和建设。

(4) 建立完备的平台服务评价体系

“大数据”正在对每个领域产生影响,在商业、经济和其他领域中,决策行为将日益基于数据分析做出,而不是像过去更多凭借经验和直觉。而在公共卫生、经济预测等领域,“大数据”的预见能力已经开始崭露头角。

为了逐步建立和完善国家科技平台体系,实现科技平台的可持续发展,推进科技资源的合理布局,有效利用,服务于科技、经济与社会发展,可以通过搜集有关平台科技资源、平台运行服务、用户反馈等多方面的数据,设计相应的指标体系,利用大数据技术,通过统计分析和挖掘,对平台的建设和共享服务情况进行客观的监督和评价,从而为平台建设的科学决策提供依据。

(5) 积极开展国际协作和交流

近年来,科技资源的全球化已经成为经济全球化的重要组成部分,它使传统科研组织结构和创新

方式发生了重大变化。开展国际合作交流是提高我国科学资源共享服务水平,加速我国科技平台建设进程,充分利用国际科技资源的必由之路。

科技平台应当积极开展与国外的科技管理部门、大数据研究部门以及科技资源数据中心等相关单位的合作交流,包括:参与国际合作项目、互派研究人员交流、定期组织召开国际学术会议等多种形式。通过与世界一流研究机构和研究人员的交流合作,跟踪国际先进水平,引进先进技术,促进自主技术研究与创新能力的迅速提升,带动国内本领域技术发展总体水平的不断提高,缩小与国际先进水平的差距。

(6) 大力开展人才培养

科技资源的共享服务是一项长期性的工作,它并不是简单地把原始数据交付给用户使用,只有对数据进行大量的加工、分析、整理,才能形成可用的数据产品进行共享。要完成这些工作,不仅要熟悉相关的领域知识,而且需要熟悉相关的信息技术和方法,这就对从事科技共享的人员提出了很高的要求。

坚持尊重劳动、尊重知识、尊重人才、尊重创造的重大方针,统筹大数据人才发展,加强人才的引进和培养,建设一支规模适度、结构合理、素质优良,既了解大数据技术,又了解相关行业领域特点的大数据人才队伍,满足科技平台资源共享对大数据人才的需求,为大数据技术应用提供后备军。

6 总结

大数据技术的发展,为平台的建设和共享服务提供了新的机遇。积极开展大数据技术的研究和应用工作,必将提升平台的技术水平和服务能力,为我国的科技创新、成果转化和科学决策提供有力支撑。

参考文献

- [1] Big Data Across the Federal Government[EB/OL]. [2012-03-29]. http://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_fact_sheet_final_1.pdf.
- [2] The White House is Spending Big Money on Big Data [EB/OL]. [2012-05-15]. <http://www.forbes.com/sites/reuvencohen>.
- [3] Chinese Academy of Social Sciences, Ministry of

- Education of the People's Republic of China. China's e-Science Blue book[M]. Beijing: Science Press, 2011:33-45.
- [4] Big Data Could Remake Science--And Government [EB/OL]. [2012-05-02]. <http://www.nationaljournal.com/tech/big-data-could-remake-science-and-government-20120502>.
- [5] Big Data Still a Big Challenge for Government IT[EB/OL]. [2012-05-09]. <http://www.eweek.com/c/a/Government-IT/Big-Data-Still-a-Big-Challenge-for-Government-IT-651653/>.
- [6] Report Says US Government Agencies Making Little Progress in BigData[EB/OL]. [2012-05-10]. <http://www.techzone360.com/.../289893-report-says-us-government-agencies-making-little-progress.htm>.
- [7] Big Data Equals Big Business Opportunities: Survey [EB/OL]. [2012-05-18]. http://www.arnnet.com.au/article/424974/big_data_equals_big_business_opportunities_survey/.
- [8] As Data Demands Speed up, Information Management Tools Pave Roadahead[EB/OL]. [2012-05-29]. <http://searchdatamanagement.techtarget.com/news/2240150285/As-data-demands-speed-up-information-management-tools-pave-road-ahead>.
- [9] Digital Universe to Add 1.8 Zettabytes in 2011[EB/OL]. [2011-06-28]. <http://www.datacenterknowledge.com/archives/2011/06/28/digital-universe-to-add-1-8-zettabytes-in-2011/>.
- [10] Open Source Powers Big Data Index[EB/OL]. [2012-06-18]. http://news.cnet.com/8301-1001_3-57451537-92/open-source-powers-big-data-index/.
- [11] Investors Are Pouring Funds Into Big Data[EB/OL]. [2012-05-04]. http://www.computerworld.com/s/article/9226840/Investors_are_pouring_funds_into_big_data.
- [12] Obama Administration Unveils "Big Data" Initiative: Announces \$200 Million In New R&D Investments[EB/OL]. [2012-03-29]. http://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_press_release.pdf.
- [13] Enhanced Real-time Earthquake Information Now Online [EB/OL]. [2012-04-17]. http://www.usgs.gov/newsroom/article.asp?ID=3171&from=rss&utm_source=feedburner&utm_medium=feed&utm_campaign=Feed%3A+UsgsNewsroom+%28USGS+Newsroom%29.
- [14] NNI Announces New Signature Initiative: Nanotechnology Knowledge Infrastructure[EB/OL]. [2012-05-17]. <http://www.nanowerk.com/news/newsid=25262.php>.
- [15] Secretary Chu Announces New Institute to Help Scientists Improve Massive Data Set Research on DOE Supercomputers[EB/OL]. [2012-03-29]. <http://energy.gov/articles/secretary-chu-announces-new-institute-help-scientists-improve-massive-data-set-research-doe>.
- [16] NSF Leads Federal Efforts In Big Data[EB/OL]. [2012-03-29]. http://www.nsf.gov/news/news_summ.jsp?cntn_id=123607&WT.mc_id=USNSF_51&WT.mc_ev=click
- [17] Mass. and MIT launching Big Data Initiatives [EB/OL]. [2012-06-05]. <http://bostonglobe.com/business/2012/05/29/mit-mass-launch-big-data-initiatives/QOasOp8Wnnrm75eJraLCyH/story.html>.
- [18] Turning Big Data Into Action: Know When to Evacuate[EB/OL]. [2012-06-11]. <http://gigaom.com/cloud/turning-big-data-into-action-know-when-to-evacuate/>.
- [19] The Growing Industry of Higher Education Big Data [EB/OL]. [2012-06-06]. www.zdnet.com/.../the-growing-industry-of-higher-education-big-data/16324.
- [20] MEXT and NSF Statement on Big Data and Disaster Research Collaboration From NSF Director Dr. Subra Suresh and MEXT Minister Mr. Hirofumi Hirano[EB/OL]. [2012-06-08]. http://www.nsf.gov/news/news_summ.jsp?cntn_id=124398&WT.mc_id=USNSF_51&WT.mc_ev=click.