

# 大数据时代科技情报服务的挑战与思考\*

裴 雷 孙建军 肖 璐

(南京大学信息管理学院 江苏南京 210023)

**摘 要:** 数据处理是科技情报机构服务创新的动力。大数据的离散分布和开放可得、用户信息需求向深度广度发展以及资源服务一体化,弱化了科技情报机构的资源优势,提高了信息分析的难度,给科技情报机构的服务能力提出了挑战。同时,也为大规模资源保障体系和信息服务模式的创新提供了可能。文章从大数据科技环境的时代特征着手,分析了大数据环境对科技情报工作的主要挑战,并结合当前实践提出了相应的对策建议。

**关键词:** 大数据; 科技情报工作; 挑战; 对策建议

**中图分类号:** G250.2 **文献标识码:** A **DOI:** 10.11968/tsyqb.1003-6938.2015123

## Challenges and Rethinking of S&T Information Service in Big Data Era

**Abstract** Data processing improvement is one of the main driven forces of S&T information service innovation. It becomes more and more difficult for S&T information institutes to face the challenges of scattered data resources, deepened and widened users' demands, and the complexity of big data analysis technologies. Meanwhile, these challenges make it possible for the innovation and sharp turn to new information service models. This paper analyzed the technical environment of big data, deconstructed the main challenges of intelligence work, and gave several suggestions as countermeasures.

**Key words** big data; S&T information service; challenges; suggestions

科技情报服务领域普遍将当前面临的海量数据处理和服务能力的矛盾,视为泛在信息环境所涌现出的、特有的“大数据”问题,并产生了隐隐的“大数据忧虑”。随着国家大数据战略的提出,各界对数据问题的关注和对数据服务的需求激增,学术界和科技情报服务实践中对“大数据”应用也呈现出多样化的解读,比如王飞跃提出的“基于 ACP 方法的开源情报解析理论框架”和“四熵服务”理念<sup>[1]</sup>,李广建等提出的情报研究五个转变<sup>[2]</sup>,贺德方、乔晓东、曾建勋等提出的“事实型数据+专用工具+专家智慧”的 FAST4Wisdom 服务理念<sup>[3-4]</sup>和情报学四维度拓展变革<sup>[5]</sup>,以及杨善林等提出的情报内涵转变<sup>[6]</sup>等。因此,消解科技情报机构的大数据忧虑,洞察科技情报大数据发展特征与发展趋势,是当前科技情报实践领域的当务之急。

### 1 科技数据挑战与科技情报服务的发展

科技情报服务领域大数据问题,实质是全球数

据自发式、爆发式增长以及数据处理高效率要求的不对称体现<sup>[7]</sup>。而科技情报工作很早就关注到知识增长趋势及其在管理和服务领域带来的信息爆炸与信息过载问题<sup>[8]</sup>。科技情报工作从诞生之初起就一直面临数据或信息资源的加工处理和分析利用的挑战,并在不同发展阶段呈现出不同的问题表征。在某种意义上,科技情报机构就是为解决“大数据”问题而诞生的专门机构,而数据挑战也成为科技情报服务创新的内在动力。

#### 1.1 科技情报初创时期的数据挑战与服务创新

1944 年,美国图书馆学家 F.Rider 关注到图书馆馆藏大约每 16 年增长 1 倍,并以此估计耶鲁大学图书馆将在 2040 年拥有超过 2 亿册图书,大约 2PB 的信息存储,而这些图书将占据 6000 英里长的书架,至少需要 6000 名图书馆员完成编目和维护工作<sup>[9]</sup>。以当时的数据处理条件而言,2 亿册图书的存储空间和集中管理维护就是可能的“大数据”问题。与之类似,战后美国政府收缴的 40 吨德国和日本技术资

\* 本文系国家社科重大基金项目“面向学科领域的网络资源深度聚合与服务研究”(编号:12&ZD221)研究成果之一。

收稿日期:2015-12-20;责任编辑:魏志鹏

料的处理,也超出了当时的文献处理能力。科技情报界提出了“信息爆炸”的理念,并主张信息分析和处理工作的专门化和体系化,也视之为科技情报工作的开端。依托业务部门进行体系化的树状业务分解和专门信息的保存、交流和利用,构建跨部门协作保障体系,极大纾解了信息总量增长困难,并通过在1950年代引入缩微技术和数字存储技术进一步缓解了信息存储空间问题。

### 1.2 数字资源建设时期的数据挑战与服务创新

20世纪60年代数据存储技术的引入有效缓解了数据存储和数据集中管理问题,但对原始数据的加工处理和分析技术短板却再次成为“大数据”问题。例如20世纪50年代美国国家安全局雇佣了超过12000名安全技术人员来处理每年侦听或加密的17000卷磁带<sup>[10]</sup>;1965年,美国国家数据中心的600个数据集存储了1亿张打孔卡片和3万份存储磁带,但对其拥有的1.75亿枚指纹记录、每年7.42亿份纳税记录的分析束手无策<sup>[11]</sup>。因此,1965年Dunn报告认为,“面临的最大问题就是不知如何发现不同机构、不同数据之间的内在关联”<sup>[11]</sup>,即数据分析的“大数据问题”。Kaysen委员会提出的集中存储和数据共享方案,推动数据库技术、信息检索技术和数据科学的产生。这一时期,科技情报服务机构也引入了数字存储和数据库管理体系,并引入了信息检索技术,比如1966年《化学文摘(CA)》引入的自动处理系统和计算机可读数据库、1967年Roger Summit主持开发的Dialog在线数据访问。科技情报领域产生了以数据加工和信息系统管理为特征的海量科技情报服务,出现了从文献保障服务向以数据库、信息检索为代表的技术服务的融合,提升了科技情报处理的数据数量和处理速度。

### 1.3 数据分析和处理时期的数据挑战与服务创新

20世纪80年代数据库技术日益成熟,以数据挖掘和高阶数据分析为代表的数据分析和处理技术得到广泛应用,数据分析进入智能化阶段。但这一时期的数据分析主要局限于封闭计算环境的中央数据存储、局限于结构化数据的分析和处理。因而,在20世纪80年代面临的新的“大数据”问题是海量外部数据的采集、加工和分析,以及如何降低数据分析成本和重复建设的“数据黑洞”,因而以竞争情报、趋势跟

踪等定向或定题的专项信息分析为代表的创新服务得以引入。科技情报服务机构完成了从信息数据产品向情报产品的升级,具备了情报监测和智能分析功能,光盘数据库和应用软件产品得到推广,使得信息分析和预测水平有了大幅提升,比如1987年发行的Medline光盘产品、1988年的科技信息网络STN Express<sup>®</sup>软件以及90年代初期的大量面向个人计算机的数据应用软件。

### 1.4 开放网络和自动加工处理时期的数据挑战与服务创新

20世纪90年代以来网络技术、个人数据终端普遍采纳,数据存储成本快速下降,科技信息资源的分布发生了巨大变化。首先,以互联网企业和技术服务企业为代表的新生态迅速崛起,以图书档案和科技情报机构为中心的科技信息服务体系受到挑战或颠覆;其次,数据可得性大大增加,数据规模和数据多样性大大增加,以搜索引擎为代表的自动抓取、分析和推送技术极大增强了数据分析和处理能力,数据挖掘和机器学习技术得到广泛应用;再次,信息服务提供商更加积极地关注用户个性需求和反馈,对用户交互、行为数据和个性化推荐更加关注。这一时期,“总量巨大(Volume)”、“类型多样(Variety)”的两个“V”的问题已经凸显,“大数据”问题初现端倪。相对于互联网企业而言,传统科技情报服务机构逐步通过兼并联合、扩展数据来源,提升数据应用能力予以应对。这一时期,面向网络的数据开放服务、数据集成与融合技术、数据应用分析工具与“数据+应用”的深度服务模式快速发展。比如SciFinder、SciVal、Scopus等数据分析工具的开发和WorldCat全球协作机制和协同服务模式的发展。

## 2 大数据时代科技情报工作面临的主要挑战

大数据的4V特征,即超大规模和快速增长的数据体量(Volume)、异构和多样性的数据结构(Variety)、大量不相关信息的低价值密度(Value)和大量实时分析应用工具的使用(Velocity),是数据分析和利用的主要障碍,在科技情报领域也同样如此。此外,从整个社会 and 用户需求角度看,科技情报机构的资源角色和地位进一步弱化,用户需求向更深更广信息分析领域拓展,情报服务与创新服务的融合以

及资源-服务的一体化对科技情报机构的服务能力也提出了更高挑战。

### 2.1 科技情报服务机构的资源优势弱化

在数据开放趋势下,可得可用的信息资源更加丰富,科技情报服务机构资源独占优势不复存在。以往科技情报服务过于依赖其资源优势,而对信息加工、信息分析的优势没用充分重视,存在短期的“转型瓶颈”。而政府数据公开、研究机构数据公开获取以及数据集市商务模式(GitHub)的出现,加剧了科技情报机构作为数据枢纽角色的弱化。以政府数据公开为例,美国政府目前可提供192440个数据集<sup>[12]</sup>,英国政府可提供1353个部门和机构开放的20688个数据集、386个APP应用<sup>[13]</sup>。数据开放可得,意味着业务机构可直接跳过科技情报服务中介而直接存取和利用信息,对科技情报机构的资源业务服务带来了挑战。如美国NTIS提供的科技报告服务,因政府部门科技报告的免费公开,在1999年以后连续多年亏损。因而,单纯提供内容获取或计量的服务模式并不能完全支持业务决策的需要。

### 2.2 科技情报服务的用户需求转变

大数据环境下,用户不再满足于信息资源的整序获取,不再满足于以文献单元为主要特征的加工整理和存取分析,而是对信息分析深度和广度提出了更高要求:在深度上突出碎片化信息加工和计算化服务<sup>[14-15]</sup>,包括数据资源快速评价推荐、知识单元的抽取和分析、多维数据融合、细粒度数据分析以及可视化、计算化的数据呈现与分析,力争将大数据去冗分类、去粗存精、去伪存真;在广度上以全局性和宏观战略性情报服务为特征<sup>[16]</sup>,包括动态监测服务、态势分析研究服务和前瞻预测研究服务三种类型,需要对多源数据、异构数据、随机动态数据进行收割、融合、跟踪和监控。同时,用户数据或信息资源利用也不同于以往数据阅读、数据参考和数据统计等浅层利用,而是在决策分析和学术研究中寻求高阶数据分析、寻求多维数据的降维理解、寻求专家智慧的介入矫正、寻求从数据分析向情报解析的升级。因此,科技情报服务不可避免地将由传统的依托资源数据的数据分析服务向多源多方协作的情报解析和计算分析服务转变,是对现有科技情报机构的分析技术能力的挑战。

### 2.3 资源服务一体化下的服务能力不足

科技情报机构过去一直强调资源保障能力,服务协作能力并未上升到战略地位。当大数据开放环境使资源获取相对容易时,反而发现“要读的太多,要挖掘、发现和分析的隐性交织的内容太多太复杂,要掌握的方法、技术与工具太多太复杂,有效分析和利用信息成为一个日益复杂、负担沉重的问题”<sup>[17]</sup>。对科技情报机构而言,提出了两个要求:第一,要求全面介入,无处不在,并注重分析方法和分析技术的全面协助;第二,要求在科技情报服务流程中更早更深地介入用户服务,提供更加精细化的服务内容,并提供更好的内容质量控制水平。

相比于传统的委托服务模式,新的科技情报服务团队将由行业专家、数据分析员、情报分析员3个角色构成<sup>[18]</sup>,并在选题阶段更早介入服务流程,同时融合数据服务、信息服务和知识服务,提供资源与服务一体化的服务体系(见图1)。依托大数据的科技情报服务要求与创新主体高度融合,以用户解决问题为中心,融入解决问题的过程,支持对问题解决方案的探索、构建和测试等服务机制<sup>[19]</sup>,从选题确立到任务分解、任务实施、数据解析,均需双方和行业专家频繁交互与协作,以应对分析需求的易变性和不可预见性<sup>[20-21]</sup>。在服务过程中,大数据分析和可视化技术提升了用户对数据的认知和理解,也可能将“情报服务”提升到知识协作、协同创新的高度,这也意味着科技情报服务人员的能力要求更高,可能需身兼行业专家、数据分析员、情报分析员中的多种角色。

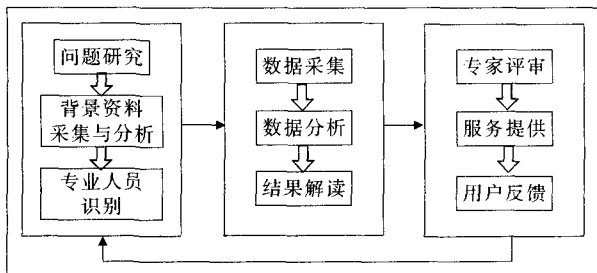


图1 资源服务一体化的科技情报服务流程<sup>[18]</sup>

## 3 大数据时代科技情报工作的资源保障体系建设

资源保障是信息服务的基础,建立全面、高质、开放的可得资源体系仍是科技情报工作的首要任务。

### 3.1 建立信息资源保障的大数据理念

大数据环境下,信息资源保障的内涵有所不同:第一,信息资源利用形式更开放快捷,大量信息资源来不及建立知识产权屏障就已过时,使得信息资源的价值实现更依赖资源利用和分析的溢出效应,而非知识产权变现;第二,信息资源的边界越来越模糊,以往受限于机构边界和知识产权保护而构建的信息资源保障“优选精选”理念,正面临低廉存储、海量来源的资源体系冲击;第三,信息资源的表现形式更多样,质量参差不齐,直接利用困难,资源保障体系的重心逐渐由收藏保障向利用保障转移;第四,随着独有信息资源优势的弱化,信息资源保障的开放性和协作利用将更为突出。

### 3.2 扩展多元数据来源和新型数据存储

科技信息机构应扩展资源的广度,关注片段信息或社会网络数据等低价值密度信息资源或新型信息资源的采集和保存<sup>[2]</sup>。在具体的采集和开发过程中,有两种典型做法:一种主张抢救性保存,因担心数字信息资源的易逝性,如不加以保存以后将无法利用,因而突出优先存储,逐步利用的原则,并不一定要优先设定明确和具体的应用目标。典型的是美国国会图书馆启动的网络存档项目,其已采集网络数据 500TB 以上,尤其是 2010 年启动的 Twitter 存档的特种馆藏项目已存储 2006~2010 年的 1700 亿条 133.2TB 的 Twitter 记录<sup>[22]</sup>,并以每天 5 亿条的规模快速增长。目前虽然尚未完全开放服务并且也面临数据解析整理的技术困难,但历史学和人文学者均期待其资源的开放利用<sup>[23]</sup>。一种主张纵向驱动和片段利用驱动,突出以若干利用和主题推动项目的存储,最终实现数据汇集和大数据。如联合国在雅加达设立全球脉动(Globe Pulse)大数据利用<sup>[24]</sup>和大多数的政府开放数据集建设都采用纵向驱动方式。

### 3.3 推动深层数据存储和核心资源质量

相对于文献资源、规范元数据等资源形式,推动深层数据类型的采集和保存将有效提升科技情报机构的信息分析能力。以往科技情报机构的资源主体是文献信息、描述信息、事实信息以及部分浅层知识资源,现在需要进一步加强内容实体信息、科学数据、深层知识资源以及用户行为信息的保存,继续发

展和完善知识库、本体库、情景库、资料库、实体库等新型数据库类型<sup>[16]</sup>。比如中信所在专利分析中对人名、机构名和技术术语、领域深层主题及其相互关系的构建<sup>[3]</sup>,中科院科学数据中心对科学数据与科技文献集成服务关键技术研究与应用<sup>[25]</sup>等。

### 3.4 科学规划参考资源架构体系

大数据环境下,大多数科技情报机构都需要考虑自建数据源、开放数据源与商业数据源的混搭利用。美国中央情报局的首席技术官 Gus Hunter 在推进 OpenStack 项目时,提出了参考资源架构(Reference Architecture)的规划理念。相对于规范的学术资源和科技信息,大数据更零散、更原始、更复杂,巨量、快速和多样的数据存在明显的碎片化特征,并不能直接利用<sup>[2]</sup>。因而,科技情报机构需要建立更加具有层次性、协调性的数据来源和数据参考规划,突出海量数据支撑与核心资源质量控制相结合的资源保障策略。一方面需要加强对科技信息获取渠道及其内容的质量评价,构建多维度科技信息质量评价指标体系和综合运用多种质量控制策略<sup>[26]</sup>;一方面需要加强对信息资源的深度加工与科学解析,提升数据的可理解性和可用性。

### 3.5 加强多源科技情报资源的集成与聚合

大数据环境下,科技信息服务需要面对方法、工具、数据的离散化分布<sup>[3]</sup>,需要融合多种来源结构异质的数据,集成多种数据分析方法,并且通过融合多源数据发现资源之间潜在关联<sup>[27]</sup>。关联性或弱信息的挖掘,被认为是大数据服务区别以往科技情报服务的典型差异<sup>[28-29]</sup>。而通过把不同渠道、利用多种采集方式获取的具有不同数据结构的信息汇聚到一起,形成具有统一格式、面向多种应用的数据集合,这一过程称为多源数据融合<sup>[2]</sup>。相对于以往图书情报机构基于目录的 OPAC、基于元数据收割协议的数字馆藏、基于协议标准的跨库检索等聚合方法,当前的数字资源聚合更注重内容层面的知识发现,尤其是数据驱动的弱信息关注<sup>[30]</sup>以及注重内容片断进行搜索、过滤、浏览、链接、注释和分析的战略性阅读<sup>[31]</sup>。

资源融合过程也是数据处理和分析技术的应用过程。李广建等统计了国际上 143 个数字资源整合项目<sup>[32]</sup>,归纳出数据仓库、Mediator/Wrapper(中介器

或封装器)、Agent 和 P2P 等四种整合机制,而且约占 85% 的项目或系统采用前两种方法,其中涉及海量存储技术、ETL 技术、信息源监控与更新技术、信息源选择技术、信息抽取技术、查询处理技术、结果整合技术、语义整合技术等。尤其针对云计算环境下多源、异构、大规模、动态信息资源特征和人们日益发展的个性化信息服务需求,基于语义的本体资源聚合、分布式协作资源聚合成为当前关注的焦点<sup>[33]</sup>。

### 3.6 积极推动资源开放获取与协作

大数据环境下,科技情报机构还应充分利用开放数据,加强机构间的协作共享。具体需要从业务模式和技术手段层面加以规划,改变传统科技情报获取方法,降低用户情报获取难度,促进资源有效利用。从业务模式上看,要加大政府开放信息源的有效利用和商业机构的技术力量协作,以提高数据使用的内在效率为原则,盈利动机为补充,引入社会力量参与<sup>[34]</sup>。以英国政府开放数据为例,目前已有多家数据创业公司与 data.gov.uk 合作,如 Open Corporates 的全球企业信息查询、Placr 的整合位置与数据查询、Locatable 的综合税率和交通情况的居住地与购房决策查询等。从技术手段看,则需增加数据的可得性。通常采用关联数据集(Linked Data Set)技术支持开放共享,以 RDF(资源描述框架)构建数据模型,利用 URI(统一资源标示符)定义数据实体,通过 HTTP 协议浏览并获取数据<sup>[35]</sup>。

## 4 大数据时代科技情报工作的服务模式创新

资源、技术与服务是科技情报工作的核心要素,资源的改变、技术的提升必然需要匹配服务的创新与升级。大数据环境下,科技情报服务模式创新强调以用户需求为中心,注重运用大数据技术分析方法,并建立拓展的资源保障体系与用户需求之间的关联<sup>[29]</sup>。结合大数据本身特征,科技情报服务将进一步突出轻量化服务、细粒度服务、技术服务和一体化服务的趋势。

### 4.1 轻量数据分析服务

轻量数据分析的服务设计主要基于动态数据和片段数据的快速增长,较小地牺牲信息分析准确性而开展的快速信息分析方式,即“短、平、快”的服务模式。动态数据的迅速更新与快速增长导致全量数

据分析的难度与时间复杂度增加,大而全的分析模式反而会降低情报分析结果的时效性,降低情报价值。因而,轻量化科技情报服务模式要突出三个要点:一是强调数据的快速解析与知识发现,一般采用即时跟踪和动态监控的服务方式<sup>[36]</sup>;二是突出最大可能降低信息分析的效度损耗,建立高效的数据筛选机制,从大量信息中准确选择目标数据,尽可能通过较小数据吞吐发现问题;三是突出用户价值,通过用户定制、用户交互实时获取用户需求,满足用户多变需求。

### 4.2 细粒度情报解析服务

细粒度情报解析主要面向用户的深层信息需求以及对海量数据信息的降维与浓缩需求。细粒度情报解析服务的实现方式主要有:第一,对文献实体知识单元的提取和分析,对文献主题、章节结构、文献元素(如图表、数据、化学式、专利号)等进行有效提取和聚合分析,典型如学术本体、知识地图、知识扩散与突变等侦测;第二,对文献资源之间的弱关联关系的发现<sup>[37]</sup>,尤其是数据背后关系网络的发现,发现数据或知识之间的相关关系;第三,对内容片段进行搜索、过滤、浏览、链接、注释和分析的战略性解析。

从服务表现形式上,细粒度情报服务导致从单一渠道获取数据的数量与质量下降,为了保证情报分析的有效性需要扩大数据来源,提高多源数据分析能力。一方面,针对不同渠道采集的异构数据进行有效解析和聚合;另一方面要关注泛在网络知识的利用,以及将开放数据内在的交互性、开放性转化为可以利用资源,并利用海量数据的可知识对象化、可计算化开展服务。

### 4.3 技术应用即服务方式

技术应用即服务方式是对资源保障分析的延伸,将信息分析过程技术化、工具化,并以服务的形式提供给用户。典型的应用如万方创新助手、百度数据开放平台、百度开题助手、腾讯云分析等平台,在数据服务基础上,将数据解析和处理工具也提供给用户,用户不仅利用数据,而且也能享受服务方提供的高度定制化的技术工具。相对一般的 SaaS 服务,科技情报机构的技术应用服务本身不存在数据委托存储的数据安全问题,同时更能体现科技情报机构资源与分析技术的协同效应。同时,技术服务也为科

技情报机构之间提供了资源和服务合作的桥梁。欧美学术图书馆就建议情报服务机构提供从资源到资源管理的完整服务体系<sup>[38]</sup>。

#### 4.4 全流程一体化服务

全流程一体化服务强调科技情报服务对业务过程的全程参与与支持,满足用户的多元需求。大数据时代数据准确采集与精确加工难度增加,业务部门或用户对情报机构的依赖性将愈发明显,科技情报可将服务链扩展到情报采集、加工、服务、服务反馈及延伸应用的全生命周期链。典型应用是 OCLC 提出的学术图书馆研究支撑服务<sup>[39]</sup>。OCLC 发现研究者对情报机构的需求不仅仅是文献提供与保障,还有研究机会发现(研究项目趋势)、研究者的知识产权管理与商业价值实现、研究合作者发现、管理和存储文本与数据集、大量文本或数据分析、改进信息检索或管理能力、引文管理、寻找论文发表机会、职位晋升或工作机会发现、学术声誉管理、出版前出版中和出版后的文献管理等。

## 5 结语

在文献服务、技术服务、应用服务和决策服务的演进过程中,不仅科技情报工作的服务对象逐渐从文献资料、系统信息过渡到覆盖全生命周期数据、文献知识单元、分析与应用工具以及决策与咨询建议的立体服务结构,而且用户需求和受众的情报利用方式也发生了变化。以学术领域的情报工作服务为例,知识大数据的兴起不仅推动了学术知识工程和数据存储服务在科技领域的深度应用<sup>[40]</sup>,科学家和人文学者对知识数据的利用方式改变也催生了知识大数据服务的诞生<sup>[41]</sup>。在商业领域,Chen 和 Chiang 通过梳理情报分析技术进展,发现情报分析从面向数据库的结构化数据分析向面向网络的非结构化数据分析、移动网络和感知数据分析的 BI&A3.0 快速转变<sup>[42]</sup>。虽然科技情报机构面临着大数据在资源保障和信息分析技术上的挑战,但另一方面科技情报的应用领域越来越广泛、用户对科技情报机构分析能力的依赖将越来越强,面临巨大的发展机遇。

#### 参考文献:

- [1] 王飞跃. 知识产生方式和科技决策支撑的重大变革——面向大数据和开源信息的科技态势解析与决策服务[J]. 中国科学院院刊, 2012, 27(5): 527-537.
- [2] 李广建, 杨林. 大数据视角下的情报研究与情报研究技术[J]. 图书与情报, 2012(6): 1-8.
- [3] 贺德方. 基于大数据、云服务的科技情报工作思考[J]. 数字图书馆论坛, 2013 (6): 2-9.
- [4] 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, 33(12): 1255-1263.
- [5] 曾建勋, 魏来. 大数据时代的情报学变革[J]. 情报学报, 2015, 34(1): 37-44.
- [6] 王晓佳, 杨善林, 陈志强. 大数据时代下的情报分析与挖掘技术研究[J]. 情报学报, 2013, 32(6): 564-574.
- [7] 张斌, 马费成. 大数据环境下数字信息资源服务创新[J]. 情报理论与实践, 2014(6): 28-33.
- [8] Santovena A Z. Big data: evolution, components, challenges and opportunities[D]. Massachusetts Institute of Technology, 2013.
- [9] Rider, F. The scholar and the future of the research library: A problem and its solution[M]. New York city: Hadham Press, 1944.
- [10] Kraus R. Statistical déjà vu: The National Data Center Proposal of 1965 and its descendants [J]. Journal of Privacy and Confidentiality, 2013, 5(1): 1
- [11] U.S. Bureau of the Budget, Office of Statistical Standards, Review of Proposal for a National Data Center, Statistical Evaluation Report No. 6, December 1965, reprinted in Edgar S. Dunn, Jr., Social Information Processing and Statistical Systems—Change and Reform[M]. New York: John Wiley & Sons, 1974: 204.
- [12] The Home of the U.S. Government's Open Data[EB/OL]. [2015-10-10]. <https://www.data.gov/>.
- [13] Opening Up Government UK[EB/OL]. [2015-10-10]. <https://data.gov.uk/data/search>.
- [14] 李国杰, 程学旗. 大数据研究: 未来科技及经济社会发展的重大战略领域——大数据的研究现状与科学思考[J]. 中国科学院院刊, 2012, 27(6): 647-657.
- [15] 邬贺铨. 大数据时代的机遇与挑战[J]. 求是, 2013(4): 47-49.

- [16] 赵凡,冷伏海.科技情报研究与服务的发展现状与趋势[J].数字图书馆论坛,2006(10):52-60.
- [17] 张晓林.研究图书馆 2020:嵌入式协作化知识实验室?[J].中国图书馆学报,2012(1):11-20.
- [18] 刘如,吴晨生,李梦辉.大数据时代科技情报工作的机遇与变革[J].情报理论与实践,2015,38(6):35-39.
- [19] 张晓林.走向知识服务:寻找新世纪图书情报工作的生长点[J].中国图书馆学报,2000,26(5):32-37.
- [20] 钱力,张晓林,李春旺,等.利用 OSGi 的科技情报分析集成服务架构研究与应用 [J].现代图书情报技术,2014(12):62-70.
- [21] 毛晓燕.大数据环境下图书馆信息服务走向分析[J].图书馆工作与研究,2014(3):72-75.
- [22] Library of Congress, "Update on the Twitter Archive At the Library of Congress" (January) [EB/OL]. [2015-11-15]. [http://www.loc.gov/today/pr/2013/files/twitter\\_report\\_2013jan.pdf](http://www.loc.gov/today/pr/2013/files/twitter_report_2013jan.pdf).
- [23] Zimmer M. The Twitter Archive at the Library of Congress: Challenges for information practice and information policy[J]. First Monday, 2015, 20(7).
- [24] Pulse U N G. Big data for development: Challenges & opportunities[M]. Naciones Unidas, Nueva York, mayo, 2012.
- [25] 中科院数据云. 科学数据与科技文献集成服务关键技术研究与应用 [EB/OL]. [2015-11-10]. <http://www.csdb.cn/linkdata/597.jhtml>.
- [26] 裴雷,孙建军.中国科技报告质量评价体系与推进策略[J].情报学报,2014(8):813-823.
- [27] 杨善林,罗贺,丁帅.基于云计算的多源信息服务系统研究综述[J].管理科学学报,2012,15(5):83-96.
- [28] 官思发,朝乐门.大数据时代信息分析的关键问题、挑战与对策[J].图书情报工作,2015,59(3):12-18,34.
- [29] 李广建,化柏林.大数据分析情报分析关系辨析[J].中国图书馆学报,2014(5):14-22.
- [30] Palmer C L. Weak information work and "doable" problems in interdisciplinary science[J]. Proceedings of the American Society for Information Science and Technology, 2006, 43(1): 1-16.
- [31] Renear A H, Palmer C L. Strategic reading, ontologies, and the future of scientific publishing [J]. Science, 2009, 325 (5942): 828-832.
- [32] 李广建,汪语宇,张丽.数字资源整合的实现机制及关键技术——对国外数字资源整合系统的实证研究[J].中国图书馆学报,2007,33(2):75-80.
- [33] 刘向,王伟军,李延晖.云计算环境下信息资源集成与服务系统的体系架构[J].情报科学,2014,32(6):128-133.
- [34] 王晴.论科学数据开放共享的运行模式、保障机制及优化策略[J].国家图书馆学刊,2014,23(1):3-9.
- [35] 李志芳,邓仲华.国内开放科学数据的分布及其特点分析[J].情报科学,2015(3):9.
- [36] 缪其浩.当情报遇到大数据——大数据热点情报跟踪的回顾与感悟[J].情报探索,2013(5):1-3.
- [37] Palmer C L, Cragin M H, Hogan T P. Weak information work in scientific discovery[J]. Information processing & management, 2007, 43(3): 808-820.
- [38] Malpas, Constance. Cloud-sourcing Research Collections: Managing Print in the Mass-digitized Library Environment [R/OL]. [2015-11-15]. <http://www.oclc.org/research/publications/library/2011/2011-01.pdf>.
- [39] Susan Kroll and Rick Forsman. A Slice of Research Life: Information Support for Research in the United States [R/OL]. [2015-11-15]. <http://www.oclc.org/research/publications/library/2010/2010-15.pdf>.
- [40] Lyon Liz. The Informatics Transform: Re-Engineering Libraries for the Data Decade[J]. The International Journal of Digital Curation, 2012, 7(1): 126-138.
- [41] 孙建军.大数据时代人文社会科学如何发展[N].光明日报,2014-07-07(11).
- [42] Chen Hsinchun, Chiang Roger-H.-L., Storey Veda-C.. BUSINESS INTELLIGENCE AND ANALYTICS: FROM BIG DATA TO BIG IMPACT[J]. MIS Quarterly, 2012, 36(4): 1165-1188.

**作者简介:**裴雷,男,南京大学信息管理学院副教授,研究方向:信息政策分析与信息资源管理;孙建军,男,南京大学信息管理学院教授,研究方向:网络信息计量与网络信息资源管理;肖璐,女,南京大学信息管理学院博士后,研究方向:数据挖掘与情报分析。