

科技报告数据关联机制研究*

陈洁, 韩非, 武茜, 高尧

(陕西省科学技术信息研究所, 西安 710054)

摘要: 通过科技报告与其他科技信息资源关联机制研究, 力图解决目前科技信息资源存在的封闭、分散、碎片化问题, 以提高财政科技投入产出的质量和效率, 确保社会公众共享财政科技投入的成果。利用科技报告兼具文献和项目的双向特征, 提出通过科技报告数据关联科研成果信息、科技计划项目信息、机构知识库三种关联机制, 并对数据关联提出相关建议, 使科技报告资源更好地服务于科研管理。

关键词: 科技报告; 科技信息资源; 关联机制; 数据关联

中图分类号: G203

DOI: 10.3772/j.issn.1673-2286.2017.1.007

2012年以来, 我国相继颁布《关于深化科技体制改革加快国家创新体系建设的意见》(中发〔2012〕6号)、《关于改进加强中央财政科研项目和资金管理的若干意见》(国发〔2014〕11号)、《深化科技体制改革实施方案》等政策文件, 明确提出要建立国家科技报告制度。2013年, 科学技术部启动“国家重点基础研究发展计划”“国家高技术研究发展计划”“国家科技支撑计划”“国家科技重大专项”“国际科技合作专项”等国家科技计划项目科技报告的回溯工作。各地方积极响应, 相继出台地方科技报告实施意见、管理办法、建设方案等政策文件, 并陆续开展地方科技计划项目科技报告回溯工作。截至2016年10月, 国家科技报告服务系统已上线6万余份各类科技报告, 包括科学技术部、国家自然科学基金委、交通运输部的科技报告及浙江、辽宁、陕西、安徽、山东、四川、江苏、广东、吉林、甘肃、福建、河南等省份的科技报告。

科技报告作为国家重要的战略性资源, 随着其逐步收集积累及科技报告系统的逐步完善, 如何实现科技报告资源的最大化利用, 需要深入研究。本文通过科技报告数据与各类科技信息资源(包括科研产出信息、科技项目信息、机构知识库资源等)的关联机制, 力图

解决科技资源封闭、分散、碎片化的问题, 满足在科技管理体制改革的完善科技管理方式、提高资源配置效率的要求。

1 数据关联研究概述

1.1 数据关联基本理论

数据关联是实现信息资源内容整合的重要技术, 是基于内容的自然联系在不同信息资源间或同一信息资源内部不同子信息间建立对应关系, 从而实现从一种信息延伸至另一种信息, 属于信息资源整合方法中内容整合的范畴。从具体技术而言, 数据关联是一种数据描述技术, 通过指定映射关系或元数据定义等方式建立不同信息间的映射, 从而实现数据关联^[1], 是语义网技术的关键和核心。大多数学者认为数据关联的基石是RDF、URI和HTTP协议。Berners-Lee认为建立关联数据的4个原则包括(1) 使用统一资源标识符(Uniform Resource Identifier, URIs)命名事物, 每个数据库(如图像、视频、程序、HTML文档等)都要拥有唯一URI;(2) 通过HTTP URIs查询命名的事物, 用户检索资源

* 本研究得到陕西省科技专项“陕西省科技报告共享服务体系建设”(编号: 2015KZX-12)资助。

时,通过浏览器提交URI并在WEB环境下遵循HTTP协议,即可获取所需数据资源;(3)当用户查询相关URI时,使用RDF、SPARQL等标准为其提供有用信息;(4)提供其他URIs链接,以使用户发现更多相关信息,提高信息资源利用率^[2]。数据关联将数据集定义为针对某一特定主题,从某个数据源或业务过程产生,由某个单一提供者发布、维护或被某个管理者聚合和管理的一个资源描述框架三元组(主体、谓词、客体)集合,通过数据集内部关联和跨数据集关联等不同形式形成数据关联网^[3]。数据关联在数据层建立链接机制,很好地描述数据结构信息,并通过URI确保计算机能自动链接各种数据,为信息资源聚合的智能化和自动化提供基础。数据关联的最大优势是可提供多个分布式异构数据源整合的关联访问,将来自不同数据源的同一对象进行整合,减轻整合分布式异构数据源的复杂性,有助于实现信息资源聚合^[4]。

1.2 数据关联研究现状

黄冬春等从当前农业科技信息资源服务的需求出发,构建基于关联数据的农业科技信息资源服务平台架构模型^[5];白海燕等从技术角度介绍三种自动关联构建的方法:基于实体的文本映射、基于图相似度的映射和基于规则的关联构建^[6];贺德方等基于概念及概念关系、引证关系和科研本体等对馆藏资源进行语义聚合^[7];王涛研究关联数据在图书馆信息资源聚合中的应用,将其主要分为创建、发布、自动关联、浏览和链接维护5个环节^[4];夏立新等提出基于关联数据的科技报告语义组织与共享框架,构建科技报告语义描述模型和语义关联模型^[8]。目前已有研究主要集中在关联数据的技术实现及资源聚合方面,鲜有涉及科技信息资源间的关联机制研究,本文探讨的数据关联机制指科研成果信息、科技计划项目信息、科研机构知识库等不同类型的科技信息资源间的内容关联。

2 科技信息资源的存储利用现状

在科技项目研究执行过程中,会产生大量研究数据和成果(包括科研论文、科技报告、专利信息、科技成果、产品样本、标准文献、科技档案等),目前这些数据信息处于分散、割离、封闭、低效的存储状态,没有以科技计划项目为载体进行有效关联。受科技项目资

助产出的科研论文通常会在文中标注项目信息,但标注信息无法提取出来进行数据挖掘及统计分析,这也使社会公众无法直接了解科技计划项目产出情况,不能通过项目产出成果评价项目执行情况,从而增加检索成本,不利于科技项目的规范透明管理及科技资源的高效合理配置。

科技报告是科技人员为描述其从事的科研、设计、工程、试验和鉴定等活动的过程、进展和结果,按照规定标准格式编写而成的文献^[9]。国家科技报告服务系统于2014年3月正式开通运行,截至2016年10月,已有6万余份科技报告实现开放共享,通过实名注册的用户可在线浏览公开的科技报告全文。科技报告资源体系不同于其他文献资源体系,兼具文献和项目双向特征。随着国家科技管理体制改革不断深化及科技报告资源体系建设不断完善,利用科技报告资源的双向特征,将多种科技信息资源进行关联,可实现对科技资源的深度挖掘利用。

3 基于科技报告的数据关联机制研究

3.1 基于科技报告的数据关联框架

《关于加快建立国家科技报告制度的指导意见》明确规定,“财政性资金资助的科技项目必须呈交科技报告”^[10],即根据项目的研究性质和资助强度不同,每个项目至少须提交一份科技报告。科技报告依托科技项目而存在,每份科技报告中既包含大量科技计划项目相关信息,也体现项目的科研产出信息。利用科技报告知识要素及其相关成果间的关联关系可向公众提供基于科技报告的知识服务^[11]。本文提出通过科技报告数据关联科研成果信息、科技计划项目信息及机构知识库三种关联机制,在元数据间建立关系,深入揭示不同资源间的关联,帮助用户实现知识发现,使科技报告资源更好地服务于科研管理,提供决策支持(见图1)。

数据关联框架的基本结构包括数据层、描述层、关联层和服务层四个部分,如图2所示。

(1)数据层。处于关联数据基本架构的底层,是进行数据关联的基础;汇集科技报告数据、科研论文数据、科技项目数据、专利数据、成果数据等原始数据库资源。

(2)描述层。URI为技术关键,同时利用RDF对资源标识进行组织与交换。基本过程是利用通用资源

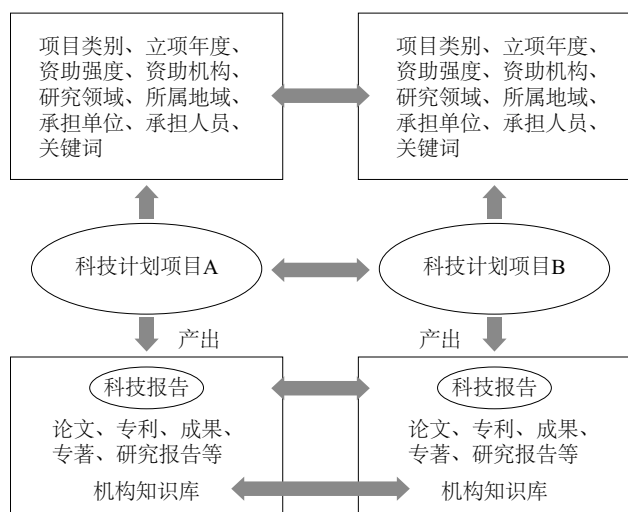


图1 基于科技报告的数据关联框架

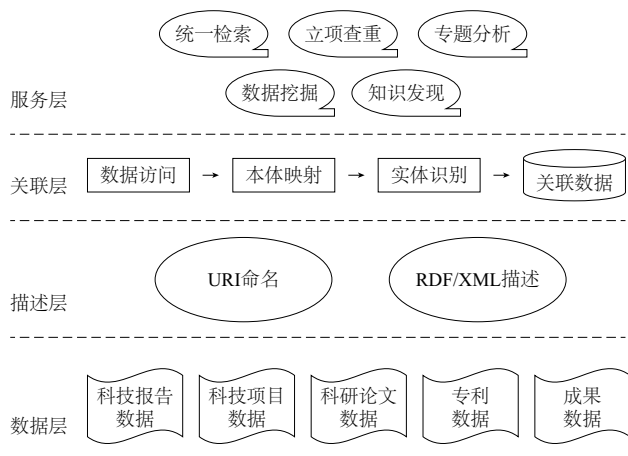


图2 数据关联的基本架构

标识符对数据、图像、视频等资源进行有效标识，再通过HTTP协议对所标识资源进行存取与传输。

(3) 关联层。包括数据访问、本体映射、实体识别和关联数据。实现对结构化数据资源的相互关联，从而支持用户对不同资源的检索。

(4) 服务层。利用关联数据源中的关联关系，实现有序地组织、集成和关联信息资源，进行信息内容的关联聚合和深层展示，实现多类型信息资源的整合和集成，为用户提供包括统一检索、立项查重、专题分析、数据挖掘、知识发现等多层次信息资源聚合服务。

3.2 关联科研成果信息

财政资金资助的科技项目必须呈交科技报告才能

申请结题验收，除科技报告外，项目执行过程中还会有专著、论文、专利、成果鉴定、新工艺、新方法、获奖、人才等多种类型的科研产出。科研产出的完成质量反映项目的创新水平和项目成员的学术作风，体现科研单位及人员的科研实力和科学态度，有利于在科技界倡导诚信务实的科学精神，有助于深化科研诚信管理、完善科技评价机制^[12]。通过科技报告系统实现对同一科技计划项目科研产出的关联，为科技计划项目管理提供过程监控、成果评估等统计分析服务，统计分析科研项目的执行情况、阶段进展、项目产出、成果质量等信息，为科技管理部门监督科技计划项目实施进程、考评实施效果、考核科研机构和科研人员绩效提供参考依据，有助于实现对科研项目的过程监控^[13]（见图3）。科技界同行和科研管理部门可通过科技报告系统中科技报告及其他科研产出，对政府立项的科研项目进行项目评估和成果评价。所以，科技报告与其他项目相关成果的关联使社会公众对科研项目执行情况有全面的了解，促进科研工作的公开化、透明化，杜绝科研工作中的虚假行为，防止学术腐败，间接提升科研工作水平^[14]，使从事科技活动的人员共同受益。

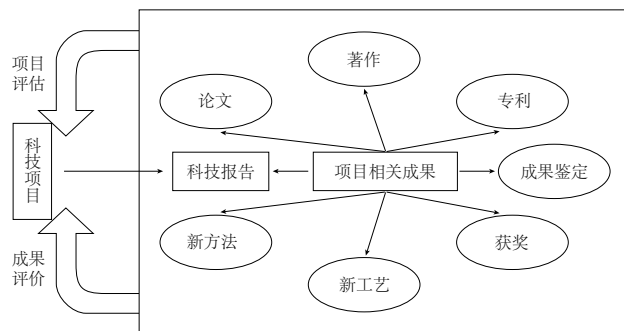


图3 通过科技报告数据关联科研成果信息

3.3 关联科技计划项目信息

胡红亮等通过对我国科技计划项目管理的调研，提出科技计划项目管理中缺乏以科技报告为核心的科技档案和科技资料系统，影响科技成果资源的推广应用^[15]。通过科技报告数据与科技计划项目信息的关联，实现科技报告与科研信息的同步公开，在接受社会监督的同时，实现科技计划项目信息的公开化和透明化，推进科技管理体制变革。

从科技报告基本信息表可得到项目名称、项目编号、计划名称、承担单位、合作单位、主管部门、科技领

域、经费信息、项目负责人等信息。根据这些信息,可进行计划类别、研究时间、资助机构及强度、热点领域、领军人物、地区分布、承担单位等不同类别的统计分析(见图4);还可进一步挖掘数据间内在关系和规律,为技术评估和技术预见提供数据支持,了解我国科研活动的主要特征和发展趋势。通过科技报告与科技计划项目信息的有效关联,开展立项查重、科技查新、定题服务、情报咨询等增值服务,以促进科技资源交流共享,抬高后续研究起点,提高科技投入产出效率。

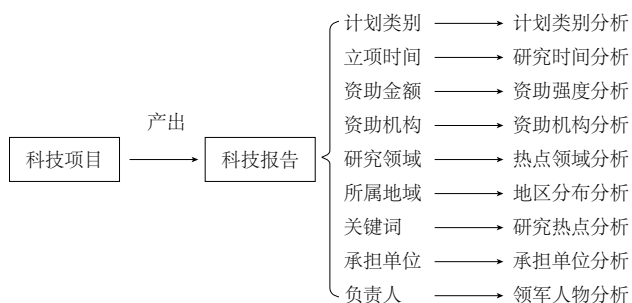


图4 通过科技报告数据关联科技计划项目信息

十八届五中全会提出要推动政府职能从研发管理向创新服务转变,建立系统思维、清单管理、链条部署。随着科技报告资源的持续积累及科技报告管理体系的不断完善,通过科技报告将实现科技计划项目信息、科技人才信息、科研成果信息、科技经费投入等不同信息间的关联分析,及对科技活动产出、分布、规模、趋势等的分析,为科技管理部门提供重要的决策情报。

3.4 关联机构知识库信息

机构知识库是知识开放获取的产物,是高校以及科研机构对其知识资产进行有效管理的工具,也是机构知识能力建设和服务能力提升的重要机制^[16]。机构知识库指利用网络及相关技术,依附于特定机构而建立的数字化学术数据库,其收集、整理并长期保存该机构及其社区成员所产生的学术成果,并将这些资源进行规范、分类、标引后,按照开放标准与相应互操作协议,允许机构及其社区内外的成员通过互联网免费获取使用^[17]。目前,科技活动产生大量科技信息并以档案形式分散保存在各单位,很难得到充分利用^[18],通过对科技报告与其他科研成果的关联,将零散的信息聚合,实现对知识资源的积累和沉淀,从而完善本单位的科

研绩效管理体系及知识共享服务体系(见图5)。以中国科学院机构知识库为例,其资源收录类型包括期刊论文、会议论文、学位论文、专利、出版专著、图像、学术报告、成果、文集、标准、年报、传媒扫描等^[19]。科技报告因其内容详尽、新颖,剪剪性强,是对机构知识库内容的丰富与重要补充,通过关联科技报告与机构知识库资源,分析科研机构间开展的合作情况,有利于科研机构提升知识管理能力。

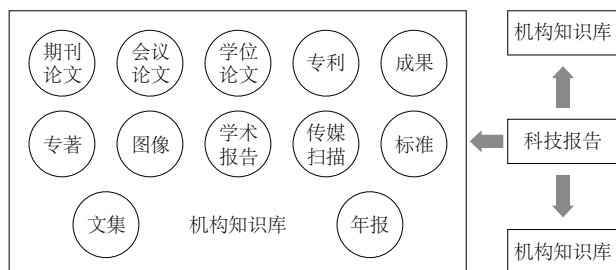


图5 通过科技报告数据关联机构知识库信息

4 基于数据关联建设的建议

要实现科技报告资源体系与科研成果信息、科技计划项目信息及机构知识库信息等的有效关联,在科技报告资源体系建设过程中,需要注意5个问题。(1)元数据格式统一。无论是国家部委还是地方的科技报告资源系统,其元数据均应采用统一格式,这为不同科技报告间的调用和关联创造便利条件。(2)确保不同系统中元数据的准确性及一致性。准确的数据源是不同系统数据资源间建立关联的基础。(3)规范科技报告附带的参考文献、论文、专利、获奖等成果信息格式。目前科研项目产出的成果在科技报告中以附录的形式出现,但并未对成果种类进行区分,建议明晰分类条目,便于后期数据分类抓取及数据关联。(4)科技报告延期公开权限的界定。为加强科技报告资源的开发利用,需明确科技报告是否符合延期公开的条件及延期公开的时限,避免个别人利用延期公开权限隐藏科技报告资源。在保护科技报告著者知识产权的前提下尽量实现科技报告全文公开共享,确保资源时效性。加强科技报告公开权限的审核,控制延期公开科技报告的数量,如果系统大部分科技报告均是延期公开状态,科技报告系统就违背其建设初衷,也不利于与其他系统间的数据交互。(5)由于科技报告涉及大量技术数据,需通过颁布相关政策法规及提供必要的技术支持,确保

科技报告资源在关联过程中的安全合理使用。在扩大资源传播范围的同时,也对各种资源的集成揭示、集中管理和使用控制提出更高要求。

5 结语

科技报告是国家战略性信息资源,只有充分共享利用才能真正发挥其价值。随着科技报告系统资源的不断丰富及科技报告体系的不断完善,科技报告在我国科技体制改革中将发挥越来越大的作用,通过科技报告资源与科研成果信息、科技计划项目信息及机构知识库信息等的关联,实现科技报告资源更好地为科技管理服务,提高财政科技投入产出质量和效率,确保社会公众共享财政科技投入成果。

参考文献

- [1] 杨勤.公安信息资源关联系统研究与实现[D].成都:电子科技大学,2010.
- [2] BERNERS-LEE T. Linked Data[EB/OL]. [2016-12-06]. <https://www.w3.org/DesignIssues/LinkedData.html>.
- [3] 赵龙文,莫荔媛,潘卓齐.基于关联数据的政府数据开放实现方法研究[J].情报资料工作,2016(6):55-62.
- [4] 王涛.基于关联数据的馆藏信息资源聚合研究[J].图书馆学刊,2012(8):44-46.
- [5] 黄冬春,姜昊,张学福.基于关联数据的农业科技信息资源服务平台研究[J].中国农机化学报,2015,36(5):248-250,270.
- [6] 白海燕,朱礼军.关联数据的自动关联构建研究[J].现代图书情报技术,2010(2):44-49.
- [7] 贺德方,曾建勋.基于语义的馆藏资源深度聚合研究[J].中国图书馆学报,2012(4):79-87.
- [8] 夏立新,李成龙.基于关联数据的科技报告语义共享框架设计与实现[J].数字图书馆论坛,2015(9):2-9.
- [9] 贺德方.科技报告资源体系研究[J].信息资源管理学报,2013(1):4-9,31.
- [10] 国务院办公厅转发科技部关于加快建立国家科技报告制度指导意见的通知[EB/OL]. (2014-09-10)[2016-09-12]. http://www.gov.cn/jzhengce/content/2014-09/10/content_9071.htm.
- [11] 王星,赵捷.国家科技报告服务系统构建研究[J].中国科技资源导刊,2015(5):26-34.
- [12] 曾建勋.关于构建国家科技报告管理条例的思考[J].中国科技资源导刊,2014(1):14-20.
- [13] 侯人华.科技报告政策体系及服务方式研究[J].情报学报,2013,32(5):472-477.
- [14] 冯长根,饶子和,王隽德,等.建立国家科技报告体系势在必行[J].科技导报,2011(21):15-16.
- [15] 胡红亮,周萍,龚春红.中国科技计划项目管理现状与对策[J].科技管理研究,2006,26(8):1-5.
- [16] 王思丽,祝忠明.利用关联数据实现机构知识库的语义扩展研究[J].现代图书情报技术,2011(11):17-23.
- [17] 柯平,王颖洁.机构知识库的发展研究[J].图书馆论坛,2006,26(6):243-248.
- [18] 邹大挺,沈玉兰,张爱霞.关于建设中国科技报告体系的思考[J].情报学报,2005,24(2):131-135.
- [19] 中国科学院机构知识库服务网格(CAS IR GRID)[EB/OL]. [2016-09-12]. http://www.irgrid.ac.cn/search-filter?field=dc.type_filter&advanced=false&&offset=0&author_sort_type=count.

作者简介

陈洁,女,1980年生,硕士,助理研究员,研究方向:科技报告、科技评估及情报研究, E-mail: echo_cargill@hotmail.com。
韩非,男,1979年生,本科,馆员,研究方向:科技评估、科技报告。
武茜,女,1984年生,硕士,助理研究员,研究方向:科技查新、科技报告。
高尧,男,1985年生,硕士,工程师,研究方向:经费监管、科技报告。

Research on the Association Mechanism of Science and Technology Information Resources Based on the Science and Technology Report

CHEN Jie, HAN Fei, WU Xi, GAO Yao
(Institute of Scientific and Technical Information of Shaanxi, Xi'an 710054, China)

Abstract: Study on the association mechanism through the science and technical reports and other information resources of science and technology, in order to solve the current science and technology information resources are closed, scattered and fragmented, improve the financial science and technology input and output quality and efficiency, ensure the public sharing of financial investment in science and technology achievements. The use of science and technology report both two-way characteristics of literature and projects, put forward three kinds of association mechanism through science and technology report and science and technology achievement information, science and technology project information, institutional repository, then give the relevant suggestions of data association, make the science and technology report resources to better serve the scientific research management.

Keywords: Science and Technology Report; Science and Technology Information Resources; Association Mechanism; Data Association

(收稿日期: 2016-12-19)