

机构知识库在科研评价服务中的应用及实现*

姚晓娜 祝志明 刘巍 张旺强

(中国科学院西北生态环境资源研究院文献情报中心, 兰州 730000)

摘要: 机构知识库作为科研机构保存和管理知识成果的重要平台, 有效地融入科研流程和更好地支撑科研管理是未来的发展趋势, 也是当前国内外的研究热点。本文对目前机构知识库科研评价应用现状进行梳理, 针对数据来源、数据查重及关联、评价指标及方法、元数据标准及互操作等关键问题进行分析, 并基于中国科学院机构知识库CSpace系统进行若干功能的开发和实现。系统的科研评价功能已在多家科研院所及高校应用, 取得了较好的服务效果。

关键词: 机构知识库; 科研评价; CSpace

中图分类号: G250.76

DOI: 10.3772/j.issn.1673-2286.2020.06.004

科研评价具有完善科研资源分配、提升机构管理水平、促进科研机构可持续发展的意义。同行评议和文献计量是目前科研评价使用的主要方法。同行评议需要耗费较多的人力和时间, 结果可能存在一定的主观性; 而文献计量采用定量的方法对科研产出进行统计和评价, 结果较为客观, 被越来越多地用于科研评价^[1]。科研产出情况在各个世界著名大学排行中都是必要的核心数据: 在U.S. News Report全球最佳大学排行榜中, 文献计量学指标占65%; 泰晤士高等教育排名中, 研究产出量和论文引用率指标占总分比达60%^[2]。英国研究卓越框架 (Research Excellence Framework, REF) 科研评价体系^[3]中的科研产出占比为65%, 评价指标不仅包括期刊论文、会议论文及专利, 还包括产品、设备、新材料、软件及研究报告等。机构知识库作为科研机构保存和管理知识成果的主要平台和构建科研基础设施的重要组件, 具备了在科研评估中承担文献计量任务的基础条件。近年来, 机构知识库在传统的知识收集、存储、传播功能上, 扩展了知识管理及审计、科研规划及项目管理、科研评价等服务, 从而更好地嵌入科研流

程和支撑科研管理服务。机构知识库不再是孤立的系统, 而转变为整个科研管理系统的有机构成, 机构知识库已经“消失在科研流程中”^[4]。

中国科学院 (以下简称“中科院”) 将机构知识库的建设情况纳入全院年度科研信息化评估指标, 部分研究所将其作为年度科研成果统计奖励基础和依据, 机构知识库提供知识资产统计服务来支持有关科研成果的统计分析和辅助支持有关评估应用工作。中国农业科学院机构知识库^[5]也借助各院所科研处, 推行科研管理、项目申报、成果评价一体化服务, 将机构知识库发展与科研绩效评价挂钩。中国人民大学的机构知识库对每篇论文的收录及引用情况进行标识, 设置人气指数、评价、分享、收藏等指标用于统计机构知识库中的文献使用情况, 定期发布ESI学科动态^[6]。西安交通大学机构知识库在文献详细信息页面中嵌入Altmetrics指标聚合工具PlumX, 评价指标包括使用、引用、获取、社交媒体及讨论等多个方面; 提供年份和机构的成果类型统计、访问量统计、各类成果产出趋势统计、学院产出对比统计等科研评价功能^[7]。复旦大学上海医学院机构知识库通过

*本研究得到甘肃省青年科技基金计划项目“面向科研评价的机构知识库服务模式及关键技术研究” (编号: 18JR3RA386) 资助。

作者社交网络和语义分析技术,能够把海量的中外文数据准确地分配到对应的机构及学者个人,大大地提高了科研评价的正确性。此外还支持按院系、学科以及个人进行科研对比分析,可比较的评价指标包括成果数量、引用次数、H指数、TOP学科发文章、TOP期刊发文章、SCI分区发文章和中科院分区发文章等^[8]。

1 机构知识库科研评价应用分类

科研评价的对象是各类科研实体,如机构、项目、人员、团队及成果等,而机构知识库作为机构保存和传播学术成果的数据库,本身对各类科研实体的信息表示及管理的能力有限。针对机构知识库科研信息管理的能力不足的问题,许多研究人员通过尝试引入科研信息管理的机制来优化或替代传统机构知识库,相关应用可以分为以下三类。

第一类是扩展机构知识库本身,支持与科研评价相关的科研实体进行存储、集成及管理,增加相关功能模块,开发面向用户需求的科研评价服务。如美国普渡大学(Purdue University)机构知识库可以把资助机构的数据管理规划转化为基于科研过程的数据管理模块,支持项目组围绕科研流程对各阶段的各类数据进行上传、共享、转移、保存、修改等,以及数据的采集、清理、审核、转换、提取、融汇、关联、可视化等,成为项目组的科研管理及统计分析平台^[9]。香港大学的机构知识库在2009年由DSpace扩展为DSpace-CRIS,从只提供开放存储全文索引的传统机构知识库提升为科研信息管理系统,不仅可以查询、保存、展示、分析香港大学相关的科研信息,还可以管理科研人员、院系、项目、专利、出版物等科研实体,支持不同层级的科研统计和评价分析^[10]。CNKI机构知识库依托于中国知网,保证了机构知识库数据的完整性和持续性,以及学术评价数据的全国性、全球性,在此基础上开发了一系列支持科研管理、评价和学科建设的扩展服务,如查查索引、学科评估、科研考评等,支持科研管理和学科建设方面的决策^[11]。

第二类是对已有的科研信息管理系统进行扩展,集成学术成果管理及共享功能,起到机构知识库的作用。如最初由丹麦奥尔堡大学(Aalborg University)开发的科研信息管理系统Pure,同时具有机构知识库的功能,目前已被爱思唯尔公司收购。其主要功能包括快速生成机构和科研人员简历,对管理目标进行实时分析和绩

效追踪,支持定制和共享各种报告^[12]。英国爱丁堡大学(the University of Edinburgh)于2011年采用Pure同时作为机构知识库和科研信息管理系统,收集了人员、成果产出、科研活动及奖励、文献计量数据、项目及资助信息、科研数据以及使用数据等,定期与人事系统、财务系统进行同步,经常应用于科研评价工作。

第三类是将机构知识库与科研信息管理系统有机链接,两个系统同时运行并进行信息交换,机构知识库从科研信息管理系统中获取科研实体信息,科研信息管理系统从机构知识库中获取学术成果及文献计量等信息,两者分别服务于不同的需求和领域,各自发挥不同的作用。科研信息管理系统以CRIS(Current Research Information Systems)系统为代表,CRIS系统可以为科研人员、科研管理部门、高校及科研机构、项目基金管理部门等收集、管理并提供该领域内的各类科研产出元数据,解决了在复杂的科研环境下对科研成果多角度、全方位信息化管理的问题^[13]。许多科研人员探索以CRIS+IR的组合服务构建语义环境下的研究信息综合管理框架,通过IR实现机构成果保存、传播与共享,通过CRIS实现成果登记注册、研究评估和自动报告。瑞典隆德大学(Lund University)将机构知识库与CRIS系统进行集成,支持在线学习、团队协作、成果可视化、文献计量、地图检索等,并可生成成果发布列表及其他科研信息^[14]。挪威CRIStin是为集成机构知识库和CRIS系统开发的,可以对科研机构的人员、项目、成果等进行统计分析,支持科研机构绩效评价及考核,从而作为政府拨款和项目资助的依据^[15]。

上述三类应用各有利弊,科研机构要根据自身需求和现有条件选择相应的方式。第一类应用比较适合评价需求主要为成果统计的情况,对机构知识库本身的改动较少,如果评价需求涉及项目、基金等更多科研信息的时候,对机构知识库的改动较多,成本较高;第二类应用要求科研管理信息系统具有机构知识库的功能,而大多数机构的科研管理信息系统主要在机构内部使用,与机构知识库开放获取的理念相悖,实施起来难度较高;第三类应用对机构已有的科研管理信息系统和机构知识库改动少,但对系统间的集成和互操作要求较高。euroCRIS于2015年启动了一项关于CRIS/IR的调查,结论是CRIS和IR相辅相成,不能互相取代,IR作为管理出版物及论文等成果的首选,CRIS通常用于管理机构科研信息,其中包括成果的元数据^[16]。目前国内的相关应用大多属于第一类,科研评价服务主要

还是对成果收录、引用和使用情况的简单统计。而国外在第三类的相关应用较多,更关注机构知识库和科研信息管理系统的集成及互操作。

2 机构知识库科研评价关键问题分析

2.1 数据来源

数据来源的完整和准确性决定评价分析结果是否可信。张晓林^[17]提出首先要评价和核实机构的知识产出在多大程度上已经收集进机构知识库,为此可以与综合文摘索引系统进行比对分析,或者与论文引证系统进行对比分析,或者与科研管理/学位管理系统申报数据进行比对分析。通过人工方式对成果数据进行检查和核对费时费力,Scopus、Web of Science、arXiv等综合文摘索引系统提供了机器可读的接口,可以自动采集已发表成果的元数据(包括收录类别信息)。一些论文引证系统可以提供引用次数、期刊影响因子及分区(JCR分区、中科院分区)、学科排名等重要评价指标数据。不过这些机器可读的接口一般是需要购买相应的服务才可以使用。

学术机构在基于机构知识库开展科研评价服务时,应充分利用已购买的文摘索引服务接口。此外,还可以将机构知识库与机构已有的科研管理/学位管理系统进行系统对接和数据集成,如西安交通大学的机构知识库集成了本校的国际论文库和学位论文库,香港大学学术库集成了来自研究成果申报系统的数据。

2.2 数据查重及关联

在统计分析之前,需要对数据进行清洗,识别重复的成果、科研人员以及项目等科研实体,匹配并建立正确的关联关系。在对评价对象进行唯一标识的基础上,还要通过各种匹配策略,实现人员、项目与研究成果之间的多向关联和统计。对于从文摘索引系统集成的成果,可以通过文摘索引系统分配的唯一标识符(如DOI、WOS ID、PubMed ID等)进行查重。由于人员名称存在同名同姓、汉字拼音多元化、错字、非拼音的英文名称等问题,仅凭成果元数据中的作者名称,很难识别并匹配到正确的科研人员,造成论文和引用无法正确统计的问题,从而影响科研人员评价的结果。马雨萌等^[18]指出科研人员是科研机构的主体,以“人”为中心

计量科研影响力,作为机构评价的重要参考。可基于ORCID(Open Researcher and Contributor Identifier,开放研究者和贡献者标识符)围绕科研人员生成成果清单,并集成科研成果在各个系统的利用数据。ORCID作为研究者标识符方面的通用国际规范,通过为研究者提供16位唯一标识符的注册来解决作者名称歧义问题。

建议各学术机构在具体应用时,结合外部公共服务(如文摘索引系统、ORCID)和机构内部数据库(如人员数据库)为各类科研实体分配唯一标识符,在此基础上进行识别管理,需要的时候可加以人工辅助,通过半自动的方式进行查重和关联。

2.3 评价指标及方法

科研评价从某种程度上来说可以看作采用文献计量学的方法对知识成果及相关实体进行统计分析。赵洁洁等^[19]对机构知识库学术评价体系进行总结,把评价指标分为传统计量学指标(收录、引文、影响因子、核心期刊等)和替代计量的指标(使用、获取、提及、社交媒体等)。

在实际应用中,建议学术机构可以根据自身的需求,对上述指标体系进行灵活扩展或综合计算。在发文量方面,除了统计院系/部门、人员、项目等不同实体的发文量,论文、专利及奖励的发文量可作为单独的评价指标进行统计;通过统计各个年份的发文量可以评价目标统计对象的发文趋势;通过统计国际合作发文量及国际合作发文占比来评估国际合作的水平。在引文评价方面,除了基本的引用次数,还可以计算篇均被引次数、人均被引次数、高被引论文数、高被引论文百分比以及H指数等扩展指标,进行更加有效的评价。替代计量中的使用统计一般是指成果在机构知识库内部的浏览量和下载量,这些数据由系统自动记录。某些文摘索引系统和学术网站也会提供成果在自身系统里的使用情况,可以通过相应的接口集成到机构知识库中。成果的人气指数、评价、分享以及收藏等行为,可以通过PlumX等Altmetric聚合指标工具^[20]获取。

2.4 元数据标准及互操作

为更好地了解科研活动的整体形势,优化科研资源配置,高层次的管理者需要对多个机构的统计结果

进行集成和比较。统一的元数据标准可以在整合数据的过程中减少重复劳动,以及烦琐的跨系统元数据对应工作。欧盟科研元数据标准CERIF (the Common European Research Information Format, 通用科研信息格式)^[21]从科研管理流程出发,分析和明确了研究计划、项目、成果、出版、专利、研究者等各个环节所涉及的主要科研实体及其科研关系,从而支持科研信息的交换和共享。CERIF定义了评价实体(measurement)和指标实体(indicator)来扩展定量评价能力,通过与机构实体、成果实体的关联,描述科学研究活动的评价结果,支持跨机构间进行标准化的评估。CERIF模型目前广泛地应用于欧盟组织的科研信息管理系统及机构知识库中,香港大学学术库DSpace-CRIS也是基于CERIF模型构建的。CERIF模型没有对具体的评价指标和方法作出定义,因此对于某些机构知识库内部的统计,如使用情况的统计,仍然缺乏统一的标准来互操作。英国PIRUS项目(Publisher and Institutional Repository Usage Statistics Project)研究利用COUNTER标准^[22]统计机构知识库成果的使用情

况,从英国各个成员知识库中获取原始使用数据(由知识库推送给IRUS-UK),然后统一进行分析和处理,生成符合COUNTER标准的统计报告,通过SUSHI协议提供给各个成员知识库或者其他集成服务系统,实现了使用统计数据的共享和互操作。

建议学术机构在制定元数据标准和互操作方案时,尽量参考和借鉴国外已有标准和协议,如CERIF模型和COUNTER标准,如果当前系统内部已有数据模型改动难度较大,可通过接口的方式,将已有数据模型部分映射到CERIF模型描述并发布,便于实现与第三方系统的互操作。

3 基于CSpace的机构知识库科研评价功能实现

3.1 系统功能框架

本文基于中科院机构知识库CSpace系统进行若干科研评价功能的开发和实现,系统功能框架如图1所示。

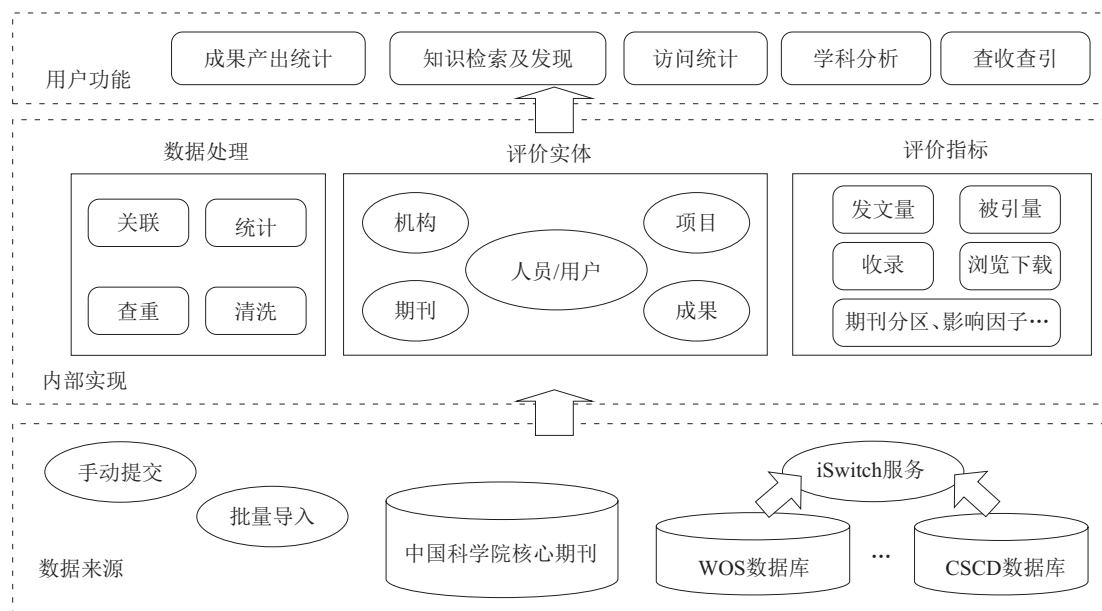


图1 基于CSpace的科研评价功能框架

系统支持多种途径的数据来源,除了科研人员或图书馆馆员可以手动提交和系统管理员批量导入之外,还支持通过iSwitch服务自动集成本机构已发表的学术成果,并定期更新收录情况和引用次数;此外,还可以基于中科院核心期刊名称规范服务获取期刊的收录类

别、影响因子、5年影响因子、JCR分区以及中科院分区等信息,从技术角度保证了数据来源的完整和准确。

数据查重方面,系统为成果定义了系统内部唯一的标识符,存缴流程中也有一定的查重机制。对于从文摘索引系统集成的成果,可以通过文摘索引系统分配

的唯一标识符(如DOI、WOS ID、PubMed ID等)进行查重。对于科研人员或研究者,支持采用ResearcherID或ORCID作为作者的唯一标识符。系统借助用户别名库,给用户分配本地化的唯一标识符,自动匹配作品和用户间可能的关联,并发送邮件给对应用户认领,保存认领结果,解决了因作者姓名变体形式导致的难以建立作者与作品之间准确关联关系的问题。

评价指标方面,系统支持大部分传统计量学指标的统计,可以从不同角度针对不同对象进行发文量及被引量的统计,并借助中科院核心期刊名称规范服务完成收录、影响因子、核心期刊、分区等指标的计量。此外还自动记录了成果的浏览及下载数据,并排除了机器访问,提供了较为客观和准确的使用指标数据。

3.2 用户功能实现

科研评价的服务场景多种多样,机构的不同部门有着不同的评价需求,CSpace系统在具体应用中应对用户需求开发了相应的功能,如成果产出统计、访问统计、学科分析、知识检索及发现、查收查引等。成果的产出统计是机构知识库科研评价的基本功能,科研管理部门通过不同维度的成果产出统计可以了解机构整体、团队或科研人员的成果产出情况,包括发文量、被引量、收录分布、成果类型分布及趋势分析。访问统计自动记录成果的浏览及下载情况,记录的信息包括访问时间、访问对象、来访IP、来访用户等数据,并在此基础上开发了访问统计功能,采用自定义统计条件的方式,使得用户可以从各种角度进行统计。在成果、项目、人员、学科等数据都比较完善的情况下,CSpace系统的学科分析功能使得科研管理部门可以对机构各个学科的学科队伍、教学成果、科研成果、科研获奖及科研项目等方面进行分析及统计,如在学科队伍方面可以统计科研人员数量、职称、学历及年龄段分布;在教学成果方面可以按自定义的成果及奖励类型进行统计;在科研成果方面,可以统计成果的类型、论文收录和引用分布以及ESI高被引论文;在科研项目方面,可以按照项目类型或者年度统计项目数量及经费数据。系统的知识检索及发现模块支持对产出成果进行关键词提取及聚类分析,通过研究指纹、主题树、关键词云等多种评价分析功能,使得科研管理及决策层可以了解机构的研究主题分布以及知识能力布局。学术论文的收录引证报告是课题申报、职称评定的重要参考依据,查收查引

也是科研评价的重要手段之一。系统的查收查引功能支持生成科研成果列表,包括收录类别、引用次数、他引次数统计、施引文献等详细信息,支持收录引证报告的自动生成及信息导出。

3.3 应用效果

CSpace系统目前已在120多家科研院所及高校进行部署和应用,其中的成果产出统计、访问统计、学科分析、知识检索及发现等功能在一定程度上满足了许多机构的科研评价需求,得到了用户的认可。以上海科技大学为例,在CSpace系统基础上构建的知识管理系统,作为唯一的机构科研成果数据源,支持与校内相关应用系统的定制集成和应用,无缝支持学校教学科研管理应用。在科研评价方面支持面向机构各层级部门提供灵活的知识分析功能,如为学院的科研人员年终考核、科研成果统计报告提供分析数据,为教务处学位点业务的报表填制提供分析数据,为图书馆情报分析服务的个人竞争力分析、机构对标分析提供分析数据等,从而有机嵌入学校智能决策过程,服务科研发展^[23]。

4 结语

目前国内机构知识库的科研评价功能及服务尚在起步阶段,主要还是对成果收录、引用和使用情况的简单统计,相关技术也不够成熟,很多关键问题缺乏相应的对策和解决方案。某些文献计量学的指标获取受到限制,如Altmetrics工具都是基于国外流行的社交网络(如Youtube、Slideshare等)计算和衡量,而国内的用户很少使用这些平台,因此存在“水土不服”的问题,这些需要科研管理部门、出版商数据库、社交网站等多方协作,建立良好的数据开放和共享的机制;随着DOI、ORCID等唯一标识符的广泛使用,以及机构知识库存缴和认领机制的完善,数据的查重和匹配的正确率将会大幅提高;评价指标和方法的选取要因地制宜,根据不同机构不同需求定制不同的评价功能,这就要求机构知识库的科研评价功能可扩展可定制,在设计开发阶段就要充分考虑到这一点;目前国内的机构知识库在科研评价方面缺乏统一的元数据标准和互操作协议,建议在面对数据集成和共享的应用需求时,尽量参考和借鉴国外已有标准和协议,对评价结果的元数据进行描述和发布。

参考文献

- [1] 万昊, 谭宗颖, 朱相丽. 同行评议与文献计量在科研评价中的作用分析比较 [J]. 图书情报工作, 2017 (1): 134-152.
- [2] 杨鹤林. 科研评估环境下图书馆利用机构库实施文献计量服务研究 [J]. 大学图书馆学报, 2017, 35 (4): 92-97.
- [3] REF 2021 [EB/OL]. [2020-03-20]. <https://www.ref.ac.uk>.
- [4] HORSTMANN W. Invisible repositories, re-use and reproducible research [C] // Proceedings of the Open Repositories, 2013: 2013-07.
- [5] 中国农业科学院机构知识库 [EB/OL]. [2020-03-20]. <http://caas.agriir.cn>.
- [6] 中国人民大学机构知识库 [EB/OL]. [2020-03-20]. <http://ir.lib.ruc.edu.cn>.
- [7] 西安交通大学机构知识库 [EB/OL]. [2020-03-20]. <http://ir.xjtu.edu.cn>.
- [8] 复旦大学上海医学院机构知识库 [EB/OL]. [2020-03-20]. <http://shmuir.fudan.edu.cn>.
- [9] Purdue University Research Repository [EB/OL]. [2020-03-20]. <http://purr.purdue.edu>.
- [10] ELIOT Z H, DAVID T. 香港大学学术库——机构知识库的应用扩展 [J]. 大学图书馆学报, 2015, 33 (4): 68-75.
- [11] CNKI机构知识库 [EB/OL]. [2020-03-20]. <http://oversea.cnki.net/marketing/ir/gb/intro.html>.
- [12] Pure [EB/OL]. [2020-03-20]. <https://www.elsevier.com/solutions/pure>.
- [13] 龚晓阳, 刘啸, 邵波. CRIS技术发展及应用对中国高校机构知识库建设启示 [J]. 图书馆学研究, 2019 (10): 38-44, 67.
- [14] RABOW I. Research information systems in the nordic countries—infrastructure, concepts, and organization [J]. Hprints Org, 2009 (6): 64.
- [15] 石长慧. 挪威科研信息管理系统介绍及其对我国的启示 [J]. 全球科技经济瞭望, 2018 (2): 40-43.
- [16] Report on EUNIS-euro CRIS joint survey on CRIS and IR [EB/OL]. [2020-05-05]. <https://investigarlainvestigacion.wordpress.com/2016/05/31/report-on-eunis-eurocris-joint-survey-on-cris-and-ir>.
- [17] 张晓林. 机构知识库的发展趋势与挑战 [J]. 现代图书情报技术, 2014, 30 (2): 1-7.
- [18] 马雨萌, 黄金霞. iAuthor面向机构科研评价的应用实现 [J]. 中国图书馆学报, 2016, 42 (2): 97-108.
- [19] 赵洁洁, 詹华清, 介凤. 高校机构知识库学术评价功能研究 [J]. 图书馆杂志, 2017, 36 (9): 20-25.
- [20] 刘春丽. altmetrics工具与机构知识库的整合与效果——以PlumX为例 [J]. 图书情报工作, 2015, 59 (24): 39-46.
- [21] Main features of CERIF [EB/OL]. [2020-03-20]. <http://www.eurocris.org/cerif/main-features-cerif>.
- [22] Project COUNTER [EB/OL]. [2020-03-20]. <http://www.projectcounter.org>.
- [23] 孙小影. 上海科技大学知识管理系统服务学校科研发展和智慧决策的实践 [EB/OL]. [2020-03-20]. <http://ir.las.ac.cn/handle/12502/10906>.

作者简介

姚晓娜, 女, 1985年生, 硕士, 馆员, 研究方向: 知识管理系统, E-mail: yaoxn@llas.ac.cn。

祝忠明, 男, 1968年生, 博士生导师, 研究馆员, 研究方向: 知识管理与知识系统。

刘巍, 男, 1980年生, 硕士, 副研究馆员, 研究方向: 知识仓储。

张旺强, 男, 1985年生, 硕士, 馆员, 研究方向: 知识仓储。

The Application and Practice of Institutional Repository in Scientific Research Evaluation

YAO XiaoNa ZHU ZhongMing LIU Wei Zhang WangQiang
(Lanzhou Information Center, Chinese Academy of Sciences, Lanzhou 730000, China)

Abstract: Institutional Repository is an important platform for institutions to preserve and manage knowledge achievements, of which development trend in the future is integration of research processes and better support for research management services. It is also a research focus domestic and overseas. This paper investigates the current research status of scientific research evaluation of institutional repository, analyzes key issues such as data sources, data deduplication and correlation, evaluation indicators and methods, metadata standards and interoperability, and develops and implements several functions based on the CAS institutional repository CSpace system. The scientific research evaluation function of the system has been applied in many scientific research institutes and universities, and has achieved good service effect.

Keywords: Institutional Repository; Scientific Research Evaluation; CSpace

(收稿日期: 2020-04-12)