

开放获取论文推送转发服务系统 iSwitch: 概念、功能与基本框架*

张晓林 梁娜 钱力 师洪波

(中国科学院文献情报中心 北京 100190)

摘要:【目的】支持开放获取论文从多个出版社向多个机构知识库自动推送转发。【方法】分析归纳论文自存储对作者的挑战和推送服务对出版社的挑战,分析多个出版社对多个机构知识库的推送转发服务需求。【结果】提出推送转发服务系统 iSwitch 的概念,提出该系统接收管理、转发管理、数据管理等功能模块及其具体任务,提出 iSwitch 设计的标准化要求和运行的合作管理要求。【结论】开放获取论文推送转发服务系统从出版社自动接收论文,识别论文作者、作者机构及资助机构,并自动推送到相应的机构知识库,有效保障机构知识成果的保存与传播。

关键词: 开放获取 作者自存储 推送服务 转发服务 iSwitch

分类号: G250.7

1 科研论文开放获取成为主流趋势

科研论文开放获取已经成为世界各国的共识^[1],例如,由全球主要科技国家组成的全球研究理事会 2013 年发布开放获取行动计划,要求全球公共科研资助项目的科研论文实行开放获取;美国白宫科技政策办公室发布行政指令^[2],要求所有科研资助金额达到一亿美元的联邦机构所资助项目的科研论文实行开放获取;英国研究理事会^[3]、欧盟地平线 2020 计划^[4]分别发布开放获取政策,要求所资助项目发表的论文实行开放获取。我国国家自然科学基金委员会于 2014 年 5 月发布开放获取政策声明^[5],要求所资助项目发表论文实行开放获取;中国科学院也于 2014 年 5 月发布开放获取政策声明,要求中国科学院作者获得公共资金资助的科研项目所发表论文实行开放获取^[6]。英国高等教育资助基金会甚至要求,在未来的高校研究绩效评估中,提交评估的

科研成果必须能通过高校的机构知识库开放获取^[7]。公共资助项目科研论文实行开放获取已经成为全球潮流。

科研论文开放获取的主要方式包括开放出版和开放存储,前者是论文在出版时即由期刊网站免费供公众获取,后者是在期刊出版后由作者将论文的经过同行评议审定修改后的录用版(称为 Final Peer-reviewed Manuscripts 或 Author Accepted Manuscripts, AAM)存储到作者所在机构或资助机构的知识库中,在一段时间(称之为开放时滞期——Embargo Period,例如 6 个月或 12 个月)后通过这些知识库免费供公众获取。如果论文本身是开放出版的,应存储出版时 PDF 版或网络版,并通过知识库立即开放;如果出版社允许,可提前开放,或存储 PDF 版。根据全球研究理事会 2014 年对全球各国资助机构的调查^①,几乎所有资助机构都同时支持开放出版和开放存储,多数资助机构要求强制实行开放存储。

收稿日期: 2014-08-13

收修改稿日期: 2014-08-22

*本文系中国科学院文献情报能力建设专项“国际开放论文国家交换服务中心示范系统”(项目编号:Y14008)的研究成果之一。

① Zhang Xiaolin, et al. Main Outcomes of the 2014 Survey of GRC Participating Organizations. The 3rd Annual Meeting of the Global Research Council, Beijing, China. 2014.

2 科研论文开放存储在存储操作上的挑战

2.1 开放存储对作者的挑战

通过开放存储,把公共资助项目科研论文及时、完全地存储到作者机构和资助机构的知识库中开放获取,是保存知识资产、传播科研成果、促进知识应用的重要保障。但是,对于作者,开放存储仍是一个比较复杂的过程,早在 2005 年 Alma Swan 和 Sheridan Brown 就研究了作者在自存储中遇到的挑战^[8],结合中国科学院机构知识库的要求和经验^[9],作者往往面对以下困难:

(1) 不熟悉开放存储细节规定,例如存储时间、存储版本、开放时滞期等;

(2) 不熟悉机构知识库存储操作流程,尤其是如果作者同时需要向所属机构和资助机构存储论文,或者作者同时隶属多个机构时,会带来更大困难;

(3) 在存储论文时需要在知识库记载必要信息、为论文设置开放获取时的使用许可,或者要链接到期刊网站的 URL 等,作者对此往往更不熟悉;

(4) 存储操作必然会带来时间成本,这对科研时间紧张的作者也是挑战;

(5) 作者可能没有保留(或难以确认)正确的存储版本。

这些困难造成科研人员不进行存储、存储错误版本、存储时空缺重要信息等。即使作者委托机构的知识库管理人员进行存储,但这些管理人员也难以获得和确认正确版本,在机构作者规模较大时负担过重。

2.2 自动推送服务及其挑战

解决上述挑战的最佳方式是出版社直接将合适的论文版本及元数据推送到指定的机构知识库中(称为自动推送服务, Automatic Push),因为出版社准确知道哪些作者的论文被录用、哪些论文是开放出版、什么版本是 AAM 版。出版社可以批量推送论文、准确提供论文元数据、嵌入论文在期刊网站的详细 URL 等,有效避免作者自存储带来的成本、数据缺失或版本混乱等问题。

其实,许多资助机构已在积极推行出版社自动推送服务。例如,美国国立健康研究院(NIH)在 2008 年就强制实行所资助项目发表论文的开放存储。为有效支持开放存储,NIH 提供了 4 种存储方式^[10]。

(1) 由相关出版社自动把所有 NIH 资助论文的发表版推送到 PMC(主要是全刊入藏到 PMC 的期刊);

(2) 由作者告知出版社把自己的 NIH 资助论文的发表版推送到 PMC(主要是开放出版的论文);

(3) 由作者自行把论文 AAM 版存储到 PMC;

(4) 由出版社自动把所有 NIH 资助论文的 AAM 版推送到 PMC。

方法(4)包括几乎所有主要科技期刊出版社^[11],例如 Elsevier、Wiley、Springer、Taylor & Francis、NPG、LWW 等。

自动推送方法对于出版社满足资助机构政策要求、减轻作者负担、避免作者自存储错误版本、避免机构知识库信息缺失、共享机构知识库使用数据等,都带来明显好处。当然,出版社在自动推送中也会遇到一系列挑战,包括:

(1) 如何准确定位作者机构知识库并建立可靠的推送流程;

(2) 如何将期刊元数据转换为机构知识库元数据,包括开放政策元数据;

(3) 如何保证准确地推送审计;

(4) 如何为没有固定机构或所在机构暂时没有知识库的作者推送论文;

(5) 在科研交叉融汇、协同创新成为重要特色时,一篇论文可能由多个资助者资助、由多个机构的作者合作撰写,这使得上述挑战更加突出。

可通过多种方式应对这些挑战,例如建立一个超级知识库,各出版社可把一个国家内的开放论文都推送和存储到这里,只推送一次、只存储一处。但由于不同资助机构和科研机构都需要保存和传播自身资助或产生的成果,而且单一系统存在安全性和垄断性威胁,人们对超级知识库的可行性和长期可靠性存疑。

针对这个问题,英国 JISC 提出了科研论文转发服务系统(Repository Junction Broker, RJB)的概念^[12],现被称为 JISC Publication Router,即建立论文交换中心,所有出版社向该中心推送论文及其元数据,由该中心向论文作者所属机构和论文资助机构转发论文及其元数据。美国图书馆界提出的 SHARE 通知服务机制(SHARE Notification Service Architecture)也可以提供部分相应服务^[13]。

3 开放获取论文推送转发服务系统 iSwitch 框架

3.1 iSwitch 功能

中国科学院文献情报中心致力于推动开放获取^[14], 包括为科研论文的开放出版和开放存储提供强健的支撑服务体系。因此, 笔者提出建立开放论文推送转发服务系统 iSwitch, 支持把公共资助项目发表论文的规定版本及时完全地存储到所有中国科学院作者所在研究所的机构知识库和所有资助机构的知识库中, 提供流畅和可靠的存储审计, 积极扩展包括论文使用统计和政策履行度检查等服务。

iSwitch 应能可靠履行以下功能:

- (1) 接收相关出版社推送的论文及其元数据;
- (2) 解析论文元数据, 识别每个作者、作者机构和资助机构;
- (3) 将出版社提供的机构名转换为标准机构名, 并映射到其知识库接口;
- (4) 将出版社提供的论文元数据转换为目标知识库的元数据格式;
- (5) 将出版社推送的论文包分装为目标机构的论文包;
- (6) 向目标机构的知识库推送论文及其元数据;
- (7) 进行出版社到 iSwitch 和 iSwitch 到目标机构的推送审计及其数据汇集报送。

在此基础上, iSwitch 还可提供以下服务:

- (8) 为那些没有特定机构、无法确认目标机构、或者目标机构尚未建立知识库的作者保存论文(“孤儿论文”), 并提供“认领、转发”服务;
- (9) 利用推送审计汇集机制, 双向汇集报送在出版社网站和知识库上的论文使用数据;
- (10) 利用解析与推送审计等, 向资助机构提供开放存储政策履行检验服务;
- (11) 提供作者自存缴服务: 在出版社尚无自动推送、多个作者又需要向多个目标知识库存缴论文时, 可由第一存缴知识库(或某一作者)将论文数据载入 iSwitch、自动发现目标知识库并转发论文, 并可依据推送审计提供“存缴凭证”。

iSwitch 本身的可靠性可以通过双系统透明备份等予以保障。

3.2 iSwitch 系统框架

为了支撑上述服务功能, iSwitch 需要提供以下主要功能组件:

(1) 接收管理模块, 负责对出版社自动推送数据进行接收、解析、检验、审计和差错管理: 包括与来源出版社之间的推送接收流程管理、对接收到的论文数据包进行解析检验、对接收到的论文数据包中的论文元数据和论文格式进行解析检验; 对出版社推送数据中的错误进行统计、发送和保存; 对接收出版社数据推送的整体情况进行统计、发送和保存。

(2) 转发管理模块, 负责向目标知识库自动推送数据的转换、分装、发送、审计和差错管理: 包括针对不同目标知识库进行论文元数据转换、把分属于不同目标知识库的论文分装为相应的论文数据包、管理与目标知识库的推送接收流程、接收来自目标知识库的接收差错信息和审计数据、汇集并向出版社报告目标知识库的接收审计数据。

(3) 数据管理模块, 负责存储和管理开放论文元数据、推送与接收审计数据、“孤儿论文”数据、其他支持数据、以及相应的检索统计功能等。

(4) 支持数据, 包括来源出版社目录(包括其推送格式、元数据格式等定义信息)、目标知识库目录(包括其接收端口、可接收推送格式、可接收元数据格式等定义信息)、机构名称对照表、元数据转换对照表等。

iSwitch 系统结构如图 1 所示:

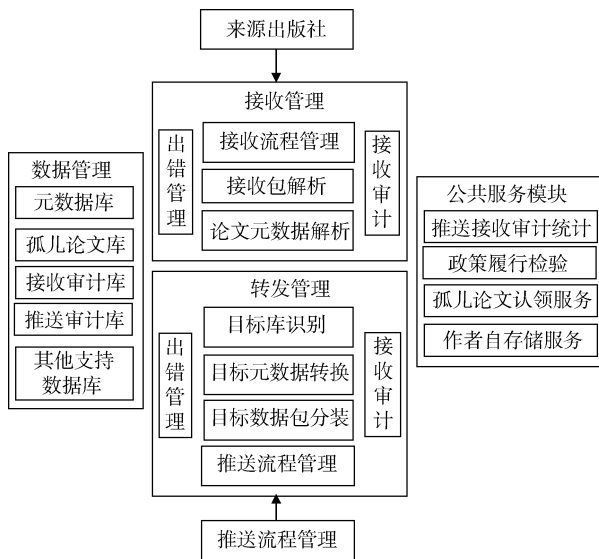


图 1 iSwitch 基本结构

4 iSwitch 建设机制要求

4.1 精细解析分发要求

iSwitch 系统必须能精细应对论文推送转发中的复杂性,包括:论文对象的复杂性,即“一篇论文”可能包括以 XML 格式表达和 PDF 格式表达的论文、图像图表以及用音视频表示的论文内容、论文的附属资料(数据、调查问卷、软件等)等;论文状态的复杂性,例如论文被撤销或出现更正后的处理;论文的作者与机构的复杂性,例如同一篇论文有多个分别隶属于不同机构、且受到不同资助机构资助的作者。iSwitch 对这些情况应能解析、验证、转换并可靠分发。

4.2 标准互操作要求

推送转发服务涉及到多个出版社与多个目标机构之间的 N:M 多对多转发关系,要建立一一对应的解析、转换将是少慢差费,应该尽可能采取业界通用的标准方式来把 N:M 关系转化为 N:1:M 关系。例如,可采用业界流行的 SWORD(Simple Web-service Offering Repository Deposit)协议^[15]进行出版社与转发系统和转发系统与目标知识库之间的数据传输,可要求所有出版社均以 JATS(Journal Article Tag Suite)格式^[16]标记论文元数据,应尽量采取规范方式表示论文的版本状态、开放获取状态和开放时滞期,尽量采用规范的唯一标识符(例如 DOI、ORCID^[17]、FundRef^[18]、事业单位证书号码等)来标识和确认论文、作者、资助机构和作者机构等,以及规范的格式和编码来标识差错状态和审计信息。文献^[19]曾对机构知识库互操作中的标准规范进行了系统的梳理,可供参考。

标准化要求还意味着开发标准化通用化的插件,例如将转发系统本身的接收管理模块封装为可方便嵌入到流行的多类机构知识库软件平台的插件,支持这些机构知识库采用与转发系统本身相同的传输协议、元数据格式、差错管理、接收审计方式来处理从转发系统推送的论文数据,方便各个机构接受推送转发服务。

4.3 运行管理要求

推送转发服务涉及多个出版社和多个目标机构,需要可靠高效的运行管理。

首先是与出版社的合作机制,包括与出版社的自动推送服务谈判、自动推送服务流程管理机制协商、自动推送服务测试、以及对自动推送服务的持续审计。

中国科学院文献情报中心已经与多家出版社(包括商业出版社和全开放出版社)达成合作意向。

其次是与目标机构的合作机制,包括与目标机构的自动推送服务安排、自动推送服务流程管理机制协商、自动推送服务测试、以及对自动推送服务的持续审计。中国科学院文献情报中心将利用中国科学院机构知识库服务网格,先行试验向中国科学院 100 余个研究所机构知识库进行推送服务。

在后续文章中,将专门讨论 iSwitch 的详细技术设计要求和试验系统开发,并讨论与出版社的自动推送服务合作机制。

参考文献:

- [1] Action Plan towards Open Access to Publications[OL]. [2014-08-10]. http://www.dfg.de/download/pdf/dfg_magazin/internationales/130528_grc_annual_meeting/grc_action_plan_open_access.pdf.
- [2] Increasing Access to the Results of Federally Funded Scientific Research [OL]. [2014-08-10]. http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf.
- [3] RCUK Policy on Open Access [EB/OL]. [2014-08-10]. <http://www.rcuk.ac.uk/research/openaccess/policy/>.
- [4] Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020 [OL]. [2014-08-12]. http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf.
- [5] 国家自然科学基金委员会关于受资助项目科研论文实行开放获取的政策声明[OL]. (2014-05-15). [2014-08-12]. <http://www.nsf.gov.cn/publish/portal0/tab38/info44471.htm>. (The National Natural Science Foundation of China Policy Statement on Open Access to Research Publications from Its Funded Projects[OL].(2014-05-15).[2014-08-12]. <http://www.nsf.gov.cn/publish/portal0/tab38/info44471.htm>.)
- [6] 中国科学院关于公共资助科研项目发表的论文实行开放获取的政策声明[OL]. (2014-05-15). [2014-08-12]. <http://www.cas.cn/xw/yxdt/201405/P020140516559414259606.pdf>. (Chinese Academy of Sciences Policy Statement on Open Access to Articles from Publicly Funded Scientific Research Projects [OL].(2014-05-15). [2014-08-12].<http://www.cas.cn/xw/yxdt/201405/P020140516559414259606.pdf>.)
- [7] HEFCE Policy Guide: Open Access Research [EB/OL]. [2014-08-12]. <http://www.hefce.ac.uk/whatwedo/rsrch/rinfra->

- struct/oa/policy/.
- [8] Swan A, Brown S. Open Access Self-archiving: An Author Study [OL]. [2014-08-12]. <http://eprints.soton.ac.uk/260999/1/jisc2.pdf>.
- [9] 关于中国科学院科研论文开放获取政策的问答[EB/OL]. [2014-08-12]. <http://ir.las.ac.cn/handle/12502/7186>. (Questions and Answers about Open Access Policy of the Chinese Academy of Sciences Research Papers[EB/OL]. [2014-08-12]. <http://ir.las.ac.cn/handle/12502/7186>.)
- [10] NIH Public Access Policy: Submission Methods [EB/OL]. [2014-08-12]. http://publicaccess.nih.gov/submit_process.htm.
- [11] NIH Public Access Policy. Method D Publishers[EB/OL]. [2014-08-12]. <https://publicaccess.nih.gov/Method%20D%20Publishers>.
- [12] Jisc Publication Router[EB/OL]. [2014-08-12]. <http://broker.edina.ac.uk/>.
- [13] SHARE Notification Service Architectural Overview [OL]. [2014-08-12]. <http://www.arl.org/storage/documents/publications/SHARE-notification-service-architectural-overview-14apr2014.pdf>.
- [14] 张晓林, 刘细文, 李麟, 等. 研究图书馆推进开放获取的战略与实践——以国家科学图书馆为例[J]. 图书情报工作, 2013, 57 (1): 15-19, 48. (Zhang Xiaolin, Liu Xiwen, Li Lin, et al. The Strategies and Practices of Research Library to Support Open Access——Taking National Science Library as an Example [J]. Library and Information Service, 2013, 57 (1): 15-19, 48.)
- [15] Simple Web-service Offering Repository Deposite [EB/OL]. [2014-08-12]. <http://swordapp.org/about/>.
- [16] JATS: Journal Article Tag Suite[OL]. [2014-08-12]. http://www.niso.org/apps/group_public/download.php/10904/z39.96-2012.pdf.
- [17] ORCID[EB/OL]. [2014-08-12]. <http://orcid.org/>.
- [18] FundRef [EB/OL]. [2014-08-12]. <http://www.crossref.org/fundref/>.
- [19] 梁娜, 张晓林. 机构知识库的互操作需求和互操作规范框架[J]. 现代图书情报技术, 2013(9): 1-7. (Liang Na, Zhang Xiaolin. The Interoperability Needs and Standards Framework for Institutional Repositories [J]. New Technology of Library and Information Service, 2013(9): 1-7.)

作者贡献声明:

张晓林: 提出 iSwitch 服务概念、基本框架和功能流程, 撰写和审核论文;

梁娜: 参与提出 iSwitch 的功能流程和标准要求, 参与撰写论文; 钱力, 师洪波: 承担 iSwitch 试验系统技术开发。

(通讯作者: 张晓林 E-mail: zhangxl@mail.las.ac.cn)

Router Service Engine iSwitch for Open Access Articles: The Concept, Strategy, and Framework

Zhang Xiaolin Liang Na Qian Li Shi Hongbo
(National Science Library, Chinese Academy of Sciences, Beijing 100190, China)

Abstract: [Objective] Support automatic push and routing of open access articles from multiple publishers to multiple repositories of funders or research organizations. [Methods] Summarize the challenges of author self archiving for authors, challenges of push services to publishers, and analyze the needs of route services. [Results] Propose the concept of the iSwitch, a router service system, suggest the functional modules for ingest, affiliation/funder identification, routing, and data management, set out the standardization requirements for the system design and the collaboration requirements for its operation. [Conclusions] iSwitch can automatically receive open paper from presses, identify the authors of the paper, the author agencies and funding agencies, then automatically pushed to the corresponding institutional repositories, effectively guarantee preservation and spread of institutional achievements.

Keywords: Open access Author self archiving Automatic push Automatic routing iSwitch