

农业立体污染防治重点领域信息门户建设实践*

赵瑞雪, 金慧敏, 鲜国建, 罗婷婷

(中国农业科学院农业信息研究所, 北京 100081)

摘要: 本文阐述NSTL重点领域信息门户农业立体污染防治专项的建设背景、门户构建策略与服务思路, 提出学科语义知识库构建思路和情报加工质量控制方法。结合专项门户实践, 总结学科门户资源建设的关键问题及采取的技术手段, 提出开放资源整合与门户建设是图书馆拓展服务渠道, 丰富知识服务内容, 提升服务能力的重要措施。

关键词: 农业污染防治; 信息门户; 本体构建; 情报产品; 服务模式

中图分类号: G250.76

DOI: 10.3772/j.issn.1673-2286.2017.12.008

1 引言

E-Science环境下, 图书情报机构依托自身资源和互联网开放资源构建特色学科知识服务平台, 面向学科领域提供专业信息资源聚合和深度知识服务, 成为当前深化服务的热点。Hemminger等通过调查发现越来越多的科学家使用在线科学数据库、虚拟研究实验室、数据共享中心、学科门户网站等开放电子资源, 这类资源的使用频率超过个人通信的使用频率, 接近电子期刊的使用频率^[1]。Vockner等提出学科门户必须是易用的, 不仅能向用户提供数据, 还可提供知识服务, 此外领域门户服务应该提供语义文本匹配与情报推荐^[2]。张晓林提出图书情报服务应该抓住知识服务的机遇, 满足新的支持知识服务的技术机制需求。该技术机制需充分支持数据挖掘和知识发现的技术应用, 支持个性化、专题化和智能化信息服务, 支持以用户为中心的信息交流、知识析取和知识应用, 充分融合用户信息资源和信息系统^[3]。

国务院发布的《“十三五”国家科技创新规划》(以下简称“《规划》”)中, 现代农业成为最新亮点。《规划》对农业污染防治与修复、农林资源环境可持续发展利用、节水农业、循环农业、盐碱地改良增效、农林生物质高效利用等农业前沿和关键技术研究进行部署, 以

期实现农业的绿色发展。力争到2020年, 建成信息化主导、生物技术引领、可持续发展的现代农业技术体系, 支撑农业资源与环境可持续发展。

2004年, 中国农业科学院的专家在多年农业污染防治研究基础上, 提出农业立体污染综合防治新概念。专家主张改变传统单一的农业污染防治思路与技术, 进行集成高新技术的多学科交叉研究, 加强对大气污染、水环境污染、土壤污染等环境问题进行治理。作为中国农业科学院国际合作重点领域的科研人员迫切需要一个综合政策法规、战略规划、研究热点、行业标准、科技文献等学科领域知识的一站式服务云平台。为此, 在国家科技图书文献中心(National Science and Technology Library, NSTL)的支持下, 中国农业科学院农业信息研究所开展了农业立体污染防治信息门户的建设与服务。

本文对农业立体污染防治门户的资源建设思路、领域本体构建方法和服务模式等问题进行分析总结, 以推动领域信息门户可持续发展, 拓展科研服务渠道, 深化学科知识服务。

2 门户建设思路

遵循“自动汇集领域信息资源, 提高网络资源利用

* 本研究得到国家科技图书文献中心“图书情报领域信息门户建设与服务”项目(编号: 2017XM37)和中国农业科学院科技创新工程项目(编号: CAAS-ASTIP-2017-AII)资助。

率,促进科技资源开放共享,为学科交叉研究提供开放性平台,为学科前沿热点研究提供优质高效土壤,为科技创新提供一站式信息支撑”的总体思路,基于互联网环境搭建学科门户服务平台。农业立体污染防治门户建设目标是在学科用户与知识间建立有效的链接通路。

张智雄等提出结构化监测的技术框架,在业务层次上将自动监测系统划分为目标资源层、监测信息采

集层、监测计算分析层和监测服务提供层^[4]。基于中国科学院文献情报中心的网络科技信息自动监测系统,农业立体污染门户建设团队与国家科学图书馆系统建设团队密切合作,构建农业立体污染门户。通过对互联网开放信息资源的监测采集、深度组织与挖掘分析、数据管理和情报服务,实现学科信息资源高效整合。农业立体污染门户的建设总体框架如图1所示。

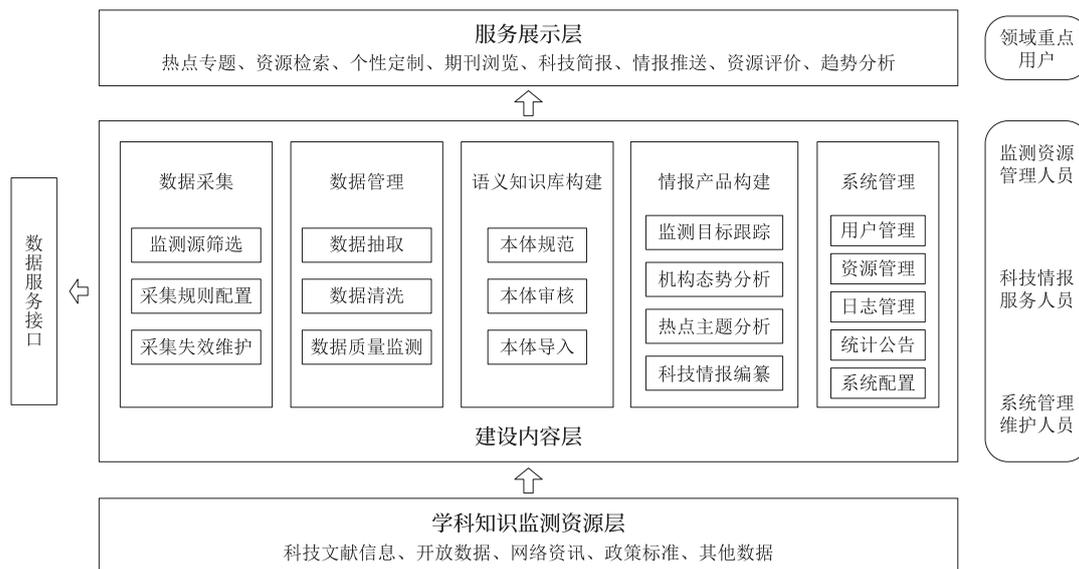


图1 农业立体污染门户建设总体框架

学科知识监测资源层应集成与整合科技文献信息、网站资讯、开放数据、国内外重大项目计划、政策标准、专利和科学数据等各种多源异构数据,构建学科信息资源的共建共享机制。门户资源的权威性、科学性、及时性、覆盖率均建立在学科知识监测资源管理基础上。门户建设内容层主要为数据采集、数据管理、语义知识库构建、情报产品构建和系统管理,这五项建设任务将实现学科开放资源的高效整合及再利用,并提供面向用户的增值服务。服务展示层主要包括为用户提供资源检索、资源评价、热点专题、情报推送等。

门户建设内容层的五项建设任务基础在于资源的发现与获取,所以对数据的采集与管理在门户建设期具有关键作用。数据采集服务主要采用多线程集群采集模式,以满足门户对采集广度与深度的要求。采集广度主要根据门户系统提供的采集配置,广泛采集国内外领域内包括期刊、新闻资讯、博客、行业标准、科技报告、科技成果等可开放获取的网络资源。监测资源管理人员在资源筛选评价基础上,根据门户监测源管理功能设置采集编码、采集周期、被采集目录的语种及采

集过滤规则等参数。采集正确率主要依赖采集配置的及时更新,采集时效性可以根据信息源更新频率自由配置。制定门户资源采集原则,构建资源评价指标体系,确立监测质量、监控机制,以及依据用户需求不断对监测资源进行更新维护,是门户持续发展的基石。

传统关键词检索效率低,构建本学科领域的语义知识库有利于知识共享,实现基于本体的知识检索。系统对采集到的科技信息资源通过知识抽取、内容识别、监测对象,进行智能文本分析,将具有语义支撑的结构化数据存储到语义知识库。语义知识库构建在系统自动抽取的基础上,须结合国内外权威领域知识,汇总领域专家反馈的语料与建设意见,推进本体规范、本体审核与本体导入的流程规范,完善语义知识库建设。

门户不仅是对学科领域信息的聚合,更应该提供深层次知识服务。科技情报分析人员通过门户可以进行机构态势分析、热点主题分析和监测目标跟踪等,助力从事农业立体污染防治研究工作的用户群体及时了解和掌握领域研究热点及科研发展态势。科技情报分析人员根据系统提供的监测服务进行情报分析,深度加工情

报产品；情报分析人员可将编译的产品通过系统自动推送功能发送到用户邮箱，也可以通过系统自动推荐功能分享到微博、微信等各大社交媒体平台。利用门户资源为用户提供领域监测报告，这种情报产品的电子化、自动化、网络化有利于学科领域知识快速传播。

门户知识需多样化聚类展示，应用元数据、XML、OAI的通用数据规范，基于学科叙词表的本体构建、文本挖掘等信息技术，为门户建设提供有效支撑。系统管理人员通过网页配置、访问控制、用户管理、资源管理、日志管理、访问统计分析等功能对整个门户进行管理。

通过数据接口服务可以实现与外部系统直接的数据交互、数据同步、数据共享等操作，减少信息系统间的数据冗余，解决信息孤岛问题。

3 门户建设重点

Tolle在科学第四范式报告中指明自然科学研究在大数据时代利用互联网技术建立面向交叉学科的科学社区，进行数据分析挖掘会促进研究人员科学研究的新火花^[5]。

探索如何建设符合学科实际需求的服务门户，如何从海量互联网资源中遴选更好的学科领域研究热点，如何将多元动态的科学研究数据进行融合并个性化展示，使研究人员获得信息的途径不再是单向灌输而是共享的、动态的。交互方式是门户建设服务需要研究的重点，所以在门户建设构建过程中，需要着重关注以下关键点：（1）用户需求分析关乎门户建设内容的领域相关度，门户服务要支持用户研究、学习以及学术交流的知识需求；（2）信息监测源管理要持续更新维护，监测源质量提升需要制定监测原则、数据源筛选标准以及著录规范；（3）系统精准检索、个性化推荐、资源评价等更深层次的知识关联揭示依赖于领域语义知识库建设；（4）服务于专项的情报产品需要可行的工作流程。

3.1 用户需求分析

随着科研环境数字化、网络化，农业污染防治研究呈现精细化、集成化、及时性等特征。门户建设要满足重点领域科研团队的多样性、动态性、复杂性的知识创新需求，需要深入调研用户个性需求，将用户需求与反馈机制贯穿研究团队的整个科研生命周期。以有研究通过定性研究方法以电子邮件的形式发放调查问卷

来收集用户需求。该方法收集反馈周期长，服务响应时效差。在建设初期，门户采用情报分析工具综合定量分析方法获取国家重点研发专家队伍与研究热点高频词辅助了解用户研究背景、研究轨迹，并通过专家会谈、邮件交流、微信、微博等多种沟通机制及时了解用户当前研究热点、及时收集评价监测源、及时进行语料添加和审核，从而详细分析本领域主体分类、创新团队、科研项目等领域重点对象的知识需求。

门户用户个性需求获取融入环节：（1）利用ESI、Citespace、TDA等情报分析工具综合定量分析方法，对本领域的学科概况与发展趋势进行调研分析，了解学科不同发展阶段研究热点、科技成果、主要面临的技术问题等，主动与用户沟通，在调研基础上进一步挖掘领域科技创新团队的理论研究和科技发展需求；（2）建立责任与反馈机制，建立与科研团队的联络人机制，建立多种沟通渠道，达到快速获取用户反馈和实时收集用户需求的目的；（3）深入潜在用户群体，宣传门户服务内容并获取持续新增的深层次服务需求；（4）根据专家、学生、企业等不同用户类型制定个性化培训方案，组织系统用户交流会，解疑答惑，收集门户建设意见；（5）根据系统提供的个性化定制服务结合用户的游览行为、检索行为及系统日志进一步动态分析用户需求。

3.2 信息监测源管理

监测资源库建设是门户服务的根基，门户建设需要广泛收集农业污染防治领域内可开放获取的数字资源，并对资源内容、资源形式进行评价。从信息资源精度、广度、深度三个维度，提高门户专业信息资源质量。

为达到门户监测信息的精准高效，信息源监测需要遵循四个原则。（1）采集范围具有学科针对性。建设初期重点采集领域核心网站资源，由具备专业知识背景的学科馆员筛选出国内外权威学科领域门户网站；建设完善阶段主要根据用户需求定时、定量更新资源，保证资源的学科针对性。（2）采集信息类别具有全面性。采集的资源包括国内外权威政策法规与行业标准、领域SCI核心期刊以及中文核心期刊、领域重要科技报告、国内外项目计划网站公开的数据、学协会科研动态、高等院校与科研机构的学术动态、试验器械设备研发市场动态等类别。（3）保证采集信息持续性。采集监测源必须稳定性高，经审核发布的内容具有权威性，然后根据监测信息源更新频率制定不同的采集频率，持续采

集。(4) 信息可维护性。建立质量监控机制, 采集的数据须定期维护, 对网站真实性、科学性进行有效的动态监控。对于失效的监测目录, 采集规则要及时修改。高质量信息可进行情报加工后直接推荐重点用户, 而对低质量信息及时删除。

根据以上原则, 在门户建设过程中使用多线程集群采集监测方法, 并通过以下手段支撑监测资源管理: 数据源质量控制需要制定明确的数据源筛选标准和著录规范; 领域本体质量控制有利于采集管理员提高效率、保证质量; 完善采集器自动采集的采全率、采准率, 更好地实现人机结合, 提高资源研判的准确率。

国外图书馆已开展大量网站影响力评价和电子资源采集标准研究, 而国内研究主要集中在理论研究方面。由于学科领域标准不同, 需要制定符合各学科领域的评价指标体系。通常需要采用定性分析方法与定量分析方法结合的综合评价方法^[6]。其中, 定性分析方法指根据专家组的反馈意见对监测目标进行评分, 由15位领域资深专家对已筛选资源进行评审, 对需要量化打分的资源给出分数及意见, 综合专家意见评定资源的合理分值; 定量分析方法根据资源类型、网站更新频率、网站数据资源量以及学科影响力(应用文献定量分析方法计算在领域发文量排名情况), 计算得出综合分值。通过定性和定量方法结合的综合评价方法会更贴合目前学科发展态势。

依据监测源筛选标准实行严格的质量控制对要采集的监测源进行采集配置, 保证信息源的稳定性和采集内容的权威性, 保障领域学科范围的全覆盖。

通过文献计量学方法, 科学、客观、定量地描述领域科学发展脉络。总结农业立体污染防治重点领域近二十年发展历程和研究进展, 分析污染因素及防治技术, 为平台建设提供依据, 对筛选领域重点监测机构、基金项目等监测源与语料建设起到重要支撑作用。

通过专家回访、问卷调查、项目组走访等途径收集专家建设意见, 加大对信息源丰富学科、新兴交叉学科未来发展方向把握, 扩大监测资源池。

通过自主开发的工具采集美国及欧盟的项目信息, 收集国家基金项目信息以及国内外学科相关科技报告和政策规划等开放资源, 充盈监测资源池。

3.3 领域语义知识库建设

由于领域交叉学科范围庞大, 领域本体库构建需

要获得专家组支持, 由领域专家参与审核。同时由于科技词汇的动态性, 决定了本体库需要不断更新, 增加新概念以及与此概念相关的各种关系。本体构建将增强对内容的语义级理解^[7]。语义库建设是有效实现基于本体个性化服务的重要技术支撑。目前, 领域知识库建设多采用机器自动学习方法来自动抽取本体, 但只采用自动抽取方式, 不辅以人工审核修正, 将会影响本体质量, 导致学科相关性差。在机器自动学习, 自动抽取术语、人物、项目、机构、科技规划等本体的功能基础上, 由门户监测管理员对抽取的本体进行筛选、导入本体库、依据本体类型进行管理, 丰富本体库建设。

基于统计算法的文本挖掘技术包括文本分类、文本聚类、文本摘要、重要性分析等, 都需要领域知识语料库的支撑, 经门户三年的实践发现, 通过自动抽取语料、专家审核方式建设和积累语料库, 可以有效提升数据资源权值的准确率。农业立体污染防治门户采取五种构建模式: 领域专家推荐平台管理员批量添加; 学科叙词表导入, 如STKOS超级词表^[8]、GEMET词表; 核心期刊高频关键词抽取; 系统根据采集内容自动抽取本体后, 由专业人员规范; 重点用户自主添加。

实践发现, 建立有效的语料甄别机制将有助于领域语料库的建设; 新闻资讯类的主题词与科技文献中的关键词侧重点不同; 语料评分机制将影响资源重要性算法的精准性; 如何权衡各语料来源的权重将影响门户资源的重要度。

3.4 情报服务产品加工

门户情报服务人员利用资源管理和资源推送功能, 通过检索筛选、编辑专业资源, 主动推送给领域重点用户。情报加工需要清晰流畅的工作流程作为支撑, 及时报道领域热点。遵循快报生产过程中基本信息流的运行规律, 农业立体污染防治科研监测快报流程具有可操作性和一定的操作弹性。门户情报产品工作流程分为需求调查、监测选题、编辑加工、审校定稿、发布推送与沟通反馈, 其工作流程如图2所示。

(1) 需求调查。根据各子专题的研究动向, 以用户需求目标为驱动力, 为用户提供实用的科技情报服务。

(2) 监测选题。在情报源获取、情报源筛选、情报源优化的基础上, 专项情报人员对资源进行选题组稿。目前门户主要的情报源为系统每天定时自动采集的有价值的情报资源以及系统外的资源, 如专家推荐的资

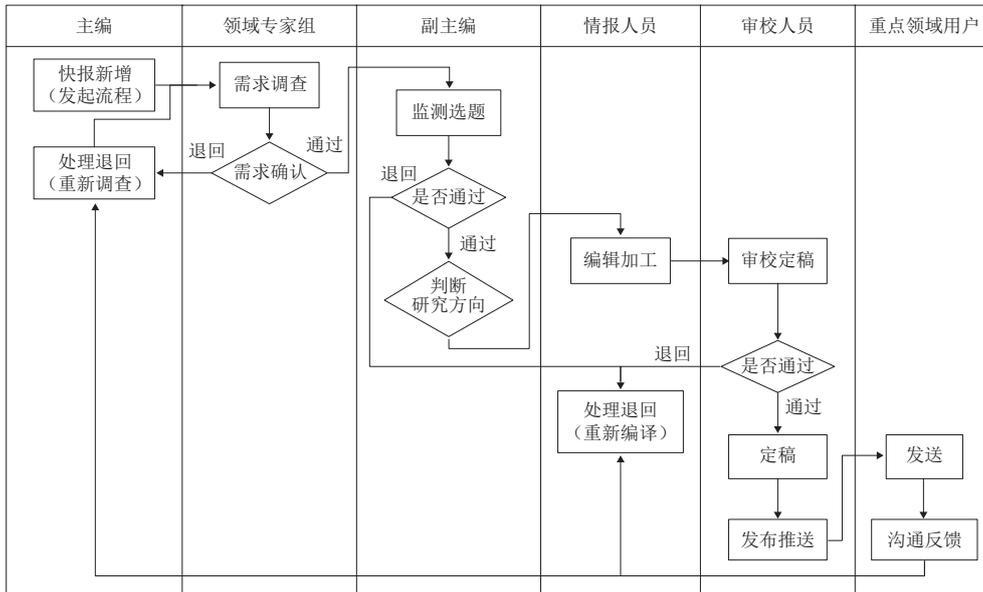


图2 情报产品工作流程

源、农业科技热点网络监测平台内资源、科学数据，以及专利数据库等外部资源。

(3) 编辑加工。农业立体污染防治情报产品的编辑加工主要分为内容加工与文字加工。情报人员在内容加工方面需要通读全文，确定主题是否明确，逻辑顺序是否层次分明，文章标题是否贴切，及文章科学数据的准确性、科学性；在文字加工方面，情报人员需要着重注意科技术语的规范化。

(4) 审校定稿。审查初稿中误译、错译等编译质量问题，审校初稿的文法、语法、错别字与图表排版等编辑排版的质量。对文章主旨阐述是否精确，翻译是否正确，分析报告是否合理，审校人员可咨询领域专家。

(5) 发布推送。利用平台情报产品发送功能，将新发布的情报产品推送至领域重点服务用户的邮箱。

(6) 沟通反馈。通过有效的沟通机制，与用户保持充分互动；通过多次循环迭代的过程，实现快报服务质量的持续提升。

4 门户服务模式与成效

随着科学研究信息化的飞速发展，国内外科研团队研究趋向协同化、实验设备网格化，促使海量知识资源成为科研工作创新发展的核心动力。阳广元等对2002—2013年的国外E-Science研究进行趋势分析，发现以英国为首的欧洲国家科研实力最强，其次是以美国为首的美洲国家，亚洲和大洋洲发文量只占总发文

量的4%—5%；提出国外研究聚焦在网格、Web服务、云计算等方面^[9]。英国曼彻斯特大学、牛津大学等国外高校都构建了知识资源服务平台，提供面向学科的知识服务。如英国印第安纳大学的Fox等论述了Web 2.0的服务方式以及应用Tagging与书签等技术发现科学家重叠的研究兴趣领域，实现科学社区建设与服务^[10]。

除技术外，平台服务也需要创新数字图书馆服务模式。面向科学家、创新团队、重点实验室等科研群组，围绕科研流程，组建跨学科的研究机构网络。服务主导型模式、嵌入式学科服务模式^[11]及门户网站式服务模式^[12]的有机结合，不仅为用户提供了便捷获取网络开放学科知识的平台，还实现了以用户个性化需求为导向的增值服务。门户提供的专题服务、定制资源推送等主动式服务，将网络信息资源进行整合，提供的一站式定制服务模式增强了服务的可见度和用户黏性。

门户根据专项进展情况制定年度培训方案。根据与领域内重点科研机构沟通情况，完成用户培训资料发放、用户培训、答疑等。迄今为止，门户组织了面向中国农业科学院、华中农业大学等机构共计10次用户培训会。门户通过确定领域内的目标用户群，制定年度详细推广方案及传播策略。具体方案包括针对学术论坛，通过编写软文进行宣传推广；面向用户，直接向其发放门户宣传册；通过网站链接，宣传推广门户；通过微博发布门户最新推荐的热点资源；通过微信传递最新情报产品；通过邮件与电话，进行用户反馈跟踪；通过接口调用实现数据交互等。截至目前，门户筛选收录614

个监测目录,基本实现专项权威网络资源覆盖的建设目标;通过系统收录的2 556个监测本体,增强网络资源的语义级理解能力;通过门户整理出有价值的信息资源进行编译,共计400条国内外领域新闻报道,推荐1 093条热点专题;编写了反映最新农业立体污染防治学术水平和科技进展及最新科技成果的37期科研监测快报;定期向重点领域服务对象发送专业情报信息,通过微博、微信与微信公众号等多种途径推送情报产品。根据2016年访问统计结果,门户全年用户访问数超过6万人次,具有良好的服务效果,成为支撑和保障本领域科研需求可靠的科技信息门户。

5 结束语

在门户建设过程中,收集用户信息进行服务的同时应注意隐私保护,文章编译报道时应注意知识产权保护。无论是个人空间体验,还是资源使用、检索、情报产品服务的体验,都应着重考虑用户的感受,以及用户使用过程中的便捷程度。门户工作需要善于结合各类信息分析工具,汇聚图书馆各类服务资源,实现知识管理。只有注重数据质量、增强服务深度、改善运行机制以及完善工作流程,才能提升整个重点领域信息门户的服务能力,满足用户持续增强的服务需求。

参考文献

[1] HEMMINGER B M,LU D,VAUGHAN K T L.Information seeking

behavior of academic scientists[J].Journal of the American Society for Information Science and Technology,2007,58(14):2205-2225.

[2] VOCKNER B,RICHTER A,MITTLBÖCK M.From geoportals to geographic knowledge portals[J].ISPRS International Journal of Geo-Information,2013,2(2):256-275.

[3] 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.

[4] 张智雄,张晓林,刘建华,等.网络科技信息结构化监测的思路和技术方法实现[J].中国图书馆学报,2014,40(4):4-15.

[5] HEY T,TANSLEY S,TOOLE K.The Fourth Paradigm:Data-Intensive Scientific Discovery[M].The Science Press,2009.

[6] 周琼琼,冯楚建.我国科技信息资源影响力评价体系设计与研究[J].中国科技资源导刊,2013(5):1-5.

[7] 袁静.基于本体的数字图书馆个性化服务研究[J].图书馆建设,2009(1):66-69.

[8] 孙坦,刘峥.面向外科技文献信息的知识组织体系建设思路[J].图书与情报,2013(1):2-7.

[9] 阳广元,邓进.国外E-Science研究论文的计量研究[J].西南民族大学学报(人文社科版),2015(3):234-240.

[10] FOX G C,PIERCE M E,MUSTACOGU A F,et al.Web 2.0 for e-science environments[C]//Conference on Semantics,Knowledge and Grid(SKG 2007).Shaanxi,2007:1-6.

[11] 田芳.高校图书馆嵌入式学科服务发展理念与服务模式研究[J].农业图书情报学刊,2012,24(7):205-208,225.

[12] 王丽敏.大数据环境下图书馆信息服务模式探析[J].情报工程,2015(2):91-95.

作者简介

赵瑞雪,女,1968年生,博士,研究员,博士生导师,研究方向:信息管理与信息系统、信息资源管理、知识组织及数字图书馆, E-mail: zhaoruiXue@caas.cn。

金慧敏,女,1978年生,助理研究员,研究方向:信息管理与信息系统、知识服务, E-mail: jinhuimin@caas.cn。

鲜国建,男,1982年生,博士,副研究员,研究方向:知识组织、关联数据、语义出版、信息系统开发, E-mail: xianguojian@caas.cn。

罗婷婷,女,1985年生,硕士,助理研究员,研究方向:数字资源加工、元数据规范、知识组织、关联数据, E-mail: luotingting@caas.cn。

The Information Portal Construction Practice in Key Fields of Agricultural Multi-Dimensional Pollution

ZHAO RuiXue, JIN HuiMin, XIAN GuoJian, LUO TingTing
(Agricultural Information Institute of CAAS, Beijing 100081, China)

Abstract: This paper described the construction background, strategies and service thoughts of NSTL agricultural multi-dimensional pollution portal, and put forward the develop idea of semantic knowledge database and quality control method of information processing. Combined with the practice of special portals, this paper summarized the key issues and technical measures of building the subject portal resources. Open resource integration and portal construction were the important measures to expand the service channel, enrich the contents of the knowledge service and enhance service capabilities in the libraries.

Keywords: Agricultural Multi-Dimensional Pollution; Information Portal; Ontology Construction; Intelligence Product; Service Mode

(收稿日期: 2017-11-17)