

开放 共享 服务

政治工作数字资源开放平台建设研究

马海兵 著

世界图书出版公司

开放 共享 服务

政治工作数字资源开放平台
建设研究

马海兵 著

世界图书出版公司

上海·西安·北京·广州

图书在版编目(CIP)数据

开放 共享 服务:政治工作数字资源开放平台建设研究 / 马海兵著. —上海:上海世界图书出版公司, 2015.5
ISBN 978-7-5100-7924-5

I. ①开… II. ①马… III. ①数字技术—应用—军队
政治工作—研究—中国 IV. ①E22-39

中国版本图书馆 CIP 数据核字(2014)第 120866 号

出版人 陆琦
策划人 姜海涛
责任编辑 史旻
装帧设计 车皓楠
责任校对 石佳达

开放 共享 服务

政治工作数字资源开放平台建设研究

著 者 马海兵

出版发行 **上海世界图书出版公司** www.wpcsh.com.cn
地 址 上海市广中路 88 号 www.wpcsh.com
电 话 021-36357930
邮政编码 200083
经 销 各地新华书店
印 刷 上海市印刷七厂有限公司 如发现印装质量问题
开 本 890×1240 1/32 请与印刷厂联系 021-59110729
印 张 5.5
字 数 120 000
版 次 2015 年 5 月第 1 版
印 次 2015 年 5 月第 1 次印刷
书 号 ISBN 978-7-5100-7924-5/E·2
定 价 25.00 元

前 言

近年来,开放存取运动在全球广泛兴起。开放存取是指利用互联网进行学术信息和科研成果自由传播、免费利用的行动。其目的在于利用网络条件实现科学信息和科研成果广泛、快速地交流与传播,以提高信息资源的利用率,降低用户获取文献资料的成本,保障科学信息的长期保存。运用这一理念,充分利用全军政工网将全军院校、科研院所与基层部队互联互通的优势,建立军队政治工作数字资源开放平台,对于落实总部文件指示精神,推动政治工作信息资源建设意义重大。

所谓军队政治工作数字资源开放平台,是指在开放存取理念的指导下,以全军政工网作为分布式运行环境,通过机构知识库模式建设的一个实现政治工作信息资源,特别是原生数字资源的数据上传、分级审核、自动分类、集中整合,动态发布、长期保存和网络传播的智能化信息服务平台。平台基于OAI协议(开放文档倡议),能够在安全认证的保护机制下,最大限度地实现开放平台系统内部、系统之间的政治工作数字资源的共建共享。

本书以军队政治工作数字资源开放平台及其实现技术为

研究对象。在继承国内外学者对机构知识库理论研究和建设成果,特别是开源机构知识库系统 DSpace 的基础上,分析了军队政治工作原生数字资源开放平台的应用需求、体系结构,研究并实现了开放平台构建的一些相关技术,包括元数据交换技术、基于 XML 的系统界面和功能定制技术,利用 OAI-ORE 进行数据交换的技术等等。在此基础上,实现了军队政治工作原生数字资源开放平台原型系统,进行了系统功能测试和结果分析。

本书主要包括四部分内容。

第一,阐述了开放平台建设的基础理论和技术,综述了国内外机构知识库研究与建设现状。

第二,明确了开放平台的功能需求和总体设计,分析了利用 OAI-PMH 协议进行系统元数据获取功能的设计与实现,讨论了利用 DSpace 开源系统实现开放平台基本功能的原理。

第三,详细阐述了利用 Manakin 项目进行开放平台系统功能和界面定制的技术。明确了 Manakin 的设计思路,分析了其底层架构——Apache Cocoon 的基本原理,讨论了 Manakin 在 DSpace 集成开发中的应用实践。

第四,阐述了利用 OAI 协议进行开放平台之间互操作的原理。详细分析了利用 OAI-ORE 协议进行数据交换的基本原理,分析了 OAI-ORE 在 DSpace 中的具体实现。

本书的研究和探索有助于丰富我国机构知识库领域的理论研究,促进军队政治工作信息资源建设的发展实践。

目 录

0 绪论	(1)
0.1 研究的背景	(1)
0.1.1 开放存取运动在全球广泛兴起	(2)
0.1.2 数字信息资源长期保存面临威胁	(4)
0.1.3 政治工作信息资源建设迫切要求	(5)
0.2 研究的目的和意义	(8)
0.2.1 研究目的	(8)
0.2.2 研究意义	(10)
0.3 主要工作	(13)
0.3.1 研究内容	(13)
0.3.2 研究方法和技术路线	(14)
0.3.3 说明	(15)
1 基础理论和技术	(17)
1.1 开放存取理论	(17)
1.1.1 开放存取运动及其含义	(17)
1.1.2 开放存取的实现途径	(20)
1.1.3 开放存取运动在各国的进展	(22)
1.2 机构知识库理论	(26)

1.2.1	机构知识库的内涵	(27)
1.2.2	机构知识库的特点	(29)
1.2.3	机构知识库的功能和作用	(30)
1.3	数字对象管理技术	(32)
1.3.1	OAIS 参考模型	(32)
1.3.2	元数据标准	(37)
1.4	互操作技术	(42)
1.4.1	OAI-PMH 协议	(42)
1.4.2	OAI-ORE 协议	(44)
1.4.3	OpenURL 协议	(46)
2	国内外机构知识库的发展现状与分析	(48)
2.1	国内外机构知识库建设现状	(48)
2.1.1	国外机构知识库建设现状	(49)
2.1.2	中国机构知识库建设现状	(51)
2.1.3	国内外机构知识库建设比较	(55)
2.2	构建机构知识库的软件介绍	(56)
2.2.1	机构知识库系统软件分类	(57)
2.2.2	机构知识库主要开源软件	(58)
2.3	机构知识库发展面临的问题	(61)
2.3.1	建立机构知识库联盟	(61)
2.3.2	机构知识库范围扩展	(62)
2.3.3	机构、出版商和 OAI 服务商之间的广泛合作	(62)
2.3.4	机构知识库与其他系统和环境的无缝连接	(63)
3	开放平台的功能需求 and 设计	(64)

5.2	Manakin 的设计思路	(108)
5.2.1	Manakin 的设计目标	(108)
5.2.2	选择 apache cocoon 作为底层架构的原因	(110)
5.3	Apache Cocoon 基本原理	(111)
5.3.1	Cocoon 的基本构成	(111)
5.3.2	Cocoon 的结构体系	(112)
5.4	Manakin 的架构分析	(113)
5.4.1	Manakin 的基本构成	(113)
5.4.2	Manakin 处理一个用户请求的过程	(114)
5.4.3	生成内容的国际化	(115)
5.5	Manakin 在集成开发中的应用实践	(116)
5.5.1	Eclipse 集成开发环境的建立	(116)
5.5.2	界面汉化、中文检索和全文检索	(117)
5.5.3	对 Manakin 的调试	(118)
5.6	结语	(120)
6	利用 OAI-ORE 进行数据交换的原理与实现	(122)
6.1	引言	(122)
6.2	应用背景介绍	(124)
6.2.1	OAI-ORE 国内外应用情况	(124)
6.2.2	数字资源共享中心的建设思路	(126)
6.3	应用 OAI-ORE 进行数据交换的基本原理	(127)
6.3.1	利用 OAI-PMH 实现元数据交换	(127)
6.3.2	OAI-ORE 的基础——RDF 模型	(130)
6.3.3	OAI-ORE 完整的数据模型	(132)
6.4	OAI-ORE 在 DSpace 平台中的具体实现	(135)

6.4.1	OAI-ORE 和 DSpace 之间的映射	(135)
6.4.2	资源图的发布	(137)
6.4.3	数字对象的获取	(138)
6.4.4	收割调度系统	(140)
6.5	结语	(140)
7	系统功能测试与演示	(142)
7.1	系统功能测试	(142)
7.2	系统功能演示	(147)
7.2.1	创建用户和组织机构	(147)
7.2.2	创建专题	(150)
7.2.3	在专题中进行资源项的提交	(152)
7.2.4	搜索和浏览	(154)
7.2.5	系统功能的进一步完善	(156)
	参考文献	(157)

0 绪 论

“军队政治工作数字资源开放平台”是应用开放存取理论和机构知识库模式建设的一个实现政治工作信息资源,特别是原生数字资源的数据上传、分级审核、分类、检索、发布、长期保存和网络传播的智能化信息服务平台,是军队政治工作信息资源建设与研究的重要组成部分。论述本书研究背景、目的和意义,介绍研究内容、方法和技术路线。

0.1 研究的背景

随着以计算机和现代通信技术为核心的信息技术发展,特别是网络环境的形成,信息的生产、存储和传递的方式发生了革命性的变化。数字信息资源以传统信息资源难以比拟的优势逐渐成为信息资源的主体。数字信息资源是一个国家的智力资源,是学术研究信息的数字存档,一个国家的科技创新能力,以及与此相关的国际竞争力,都依赖于其快速、有效的开发与利用数字信息资源的能力。以数字信息资

源为基础的信息服务也逐渐成为信息服务的主导。自 20 世纪 90 年代以来,各国政府、企业、科研部门、信息机构等对数字信息资源建设和开发利用给予了高度关注。美国政府把数字信息资源的产生、传播、获取和利用,作为国家信息化建设的关键和重点。加拿大在 2002 年提出的国家创新体系中,将建立国家数字科技信息网作为其重要组成部分。欧洲也从 1998 年开始启动了面向数字信息资源整合开发的项目。在我国,中共中央办公厅、国务院办公厅印发的《2006—2020 年国家信息化发展战略》将“大力发展以数字化、网络化为主要特征的现代信息服务业,促进信息资源的开发利用”作为我国信息化发展的战略之一,充分说明国家对数字信息资源建设和以数字信息资源为基础的信息服务的高度重视^①。

军队政治工作数字资源开放平台建设是在开放存取理念在全球广泛兴起的大背景下,面向军队政治工作科研院所,对如何有效进行政治工作数字资源,特别是对原生数字资源开发利用及其关键实现技术的研究,是军队政治工作信息资源建设与研究的重要组成部分。

0.1.1 开放存取运动在全球广泛兴起

随着互联网的普及和电子出版技术的完善,一方面,网络技术的发展和成熟使科研人员的学术交流活动日益频繁、形式日益多样化,为全球的知识共享活动创造了良好的条件;另

^① 肖希明等:《数字信息资源建设与服务研究》,武汉:武汉大学出版社,2008 年,第 1 页。

一方面,传统的学术期刊交流模式相对封闭滞后,难以适应科研活动的新要求,在一定程度上阻碍了知识的交流和共享。开放存取作为一种新的学术信息共享的理念和出版机制应运而生。开放存取是利用互联网进行学术信息和科研成果自由传播、免费利用的行动,是一种新型的学术出版和交流模式。其目的在于利用网络条件实现科学信息和科研成果广泛、快速地交流与传播,以提高信息资源的利用率,降低用户获取文献资料的成本,保障科学信息的长期保存,提高科学研究的效率。开放存取近年来引起了各国政府及国际组织的广泛关注。

开放存取运动的历史并不久远。2001年12月1日至2日,美国开放社会研究所(Open Society institute, OSI)在匈牙利的布达佩斯召开了小型但意义深远的会议,这是一次旨在推进全世界各学科领域研究论文在因特网上免费传播进程的会议,这次会议的结果便产生了“布达佩斯开放存取先导计划”(Budapest Open Access Initiative, BOAI)。该倡议首次给出了开放存取的完整定义。

2003年10月22日,德国、法国、意大利等国的科研机构根据BOAI的精神,起草了《对科学与人文科学知识开放存取的柏林宣言》,通常简称《柏林宣言》,旨在BOAI的基础上,将因特网作为全球科学知识的基地和人类思考的重要设施,提供给研究政策制定者、研究机构、资金代理商、图书馆、档案馆和博物馆。《柏林宣言》进一步明晰了开放存取必须满足的两个方面的条件。2004年5月,中国科学院院长路甬祥和中国国家自然科学基金委员会签署了《柏林宣言》,表明中国科学界和科研资助机构支持开放存取的原则立场。

0.1.2 数字信息资源长期保存面临威胁

随着数字技术和数字出版的发展,越来越多的有价值的著述内容已经只能以数字形式来管理、保存和使用。根据美国加州大学伯克利分校住处管理与系统学院的统计,仅 2002 年产生的纸张、胶片、磁盘和光盘存储的信息大约有 5 艾字节 (5EB; 1EB=1024PB, 1PB=1024TB, 1TB=1024GB, 1GB=1024MB, 1MB=1024KB, 1KB=1024Byte)。可见,数字信息资源的数量很大。美国“国家数字信息基础设施和保藏计划”(national digital information infrastructure and preservation program)网站明确指出:“在许多情况下,数字资源比物理资源更脆弱。这些文档本身更容易被毁坏,或者它们的载体很容易被淘汰。”

根据数字信息资源产生的形态,数字资源可以分为数字再造资源和原生数字资源。数字再造资源是对以前存在的资源进行数字化再造而形成的“数字拷贝”,是原物的一种表现形式,而原生数字资源则是仅仅以数字形式存在的数据。例如,万维网网站、电子出版物、多媒体产品、文化和科技数据库等^①。“原生数字资源”没有其他的存储形式,一旦破坏,会永远丢失。“因为通常没有单独制造出来数字格式资料的模拟(物理)材料版本,这些作为历史资源的所谓原生数字资源即面临着更大的消失和不可获得的风险,或阻止未来研究人员利用它们原始的形式研究它们的风险。数以万计的数字资

^① Linda Ashcroft, *Long-term preservation of our digital asset is a step closer: report on first year's activity by Digital Coalition published*, *New Library World*, 2004, Vol. 105, Iss. 5/6:227.

源,诸如安装在早期互联网上的网站,已经消失了”^①。对于一些特殊的信息资源而言,其长期保存的难度更大。如数字地理空间数据的保存涉及数据转换、文件大小、适当的数据格式、版权以及文档格式的复杂性等。事实上,由于计算机界面的变化,有的数据已经无法再读取,如1960年美国人口调查统计数据存储只有用UNIVAC Type II-A Tape Drive才能读取,全世界只有两台机器可以读,一台在日本,另一台在美国Smithsonian博物馆。

数字化信息的长期可获得性是从事知识、加工、管理专业学者研究的一个专业化问题,是为了确保信息能够被下一代获得,保障我们创造的知识能够进一步得到发展。数字化信息的基础性和普遍性使得长期保存的负担甚至已经远远超出单个和某些机构的能力。许多国家已经认识到保存数字信息资源的紧迫性,为此进行了一系列的研究和实验。2003年8月,美国国家自然科学基金(NSF)和美国国会图书馆联合发布了《现在是行动起来的时候了:数字存档与长期保存的研究挑战》报告^②,提出这一问题的紧迫性和解决对策。这一问题在我国同样存在,军队政治工作数字信息资源面临的问题更加突出。

0.1.3 政治工作信息资源建设迫切要求

军队政治工作信息化建设的实质,是在政治工作领域中

① Library of Congress announces awards of \$13.9 million to begin building a network of partners for digital preservation. [EB/OL] [2014-01-05] <http://www.loc.gov/today/pr/2004/04-171.html>.

② It's about time: research challenges in digital archiving and long-term preservation. [EB/OL] [2014-01-08] <http://digitalpreservation.gov/about/NSF.pdf>.

对现代信息技术的全面运用和对信息资源的广泛开发利用^①。军队政治工作信息化建设与社会信息技术的发展密切相关。在军队政治工作领域,信息资源同样经历了从传统资源为主体到数字资源为主体的巨大变化。政治工作信息资源建设是军队政治工作信息化建设的核心任务,是体现我军政治工作特色的精彩华章。

军队政治工作信息资源,指的就是军队政治工作活动中产生的、经过有序化处理的、能对部队政治工作效能产生影响的有用信息的集合。从内容上讲,其范围包括党、国家和军队领导人讲话、中共中央、中央军委文献、解放军四总部和武警总部指示、政治工作法规制度、政治工作经验和学术研究成果、部队官兵思想行为数据,社情、民情等等。军队政治工作信息资源建设,就是对处于无序状态的各类原始政治工作信息进行选择、采集、组织和开发等活动,使之形成可资利用的信息资源体系的全过程。在军队政治工作信息化建设发展布局中,基础设施建设与信息资源建设无疑是两个重要组成部分。人们形象地将其喻为“路”和“车”的关系。但从军队政治工作信息化发展现状来看,政治工作信息资源建设与服务无论是实践还是理论研究都相对滞后。因此,研究如何有效借鉴、充分吸引现代信息技术和理念进行军队政治工作数字资源的开发和利用,是军队政治工作信息化建设的紧迫课题。

军队政治工作数字资源同样面临着长期保存和利用这一

^① 沈国权、罗剑明:《军队政治工作信息化建设研究》,北京:解放军出版社,2007年,第32页。

问题的挑战。随着数字学术的繁荣,军队政治工作信息资源同其他学科的信息资源一样,正以指数级速度急剧增长。海量的信息与数据的生成和积累为军队思想政治工作提供了取之不尽的信息资源。但现实情况是,这些宝贵的数字资源要么处于无序或分散的状态,要么以一种垄断方式被管理和存储,不利于共享利用。同时,没有可以信赖的机构能够做到及时永久保存这些数字资源,使其随时面临因技术过时和载体变质而损失的危险。以军队政治类院校为例,许多集众多领导和学者集体智慧,由专家教授、机关干部呕心沥血才形成的政治工作相关特色资源,如领导讲话、教学简报、报告、课件、讲义、调研报告、录音、录像等等数字资源散落于各个机关、教研室和教员手中。其中绝大多数的原生数字化智力资源无法、也不能在单位以外出版,即使是单位内,也无法有效实现交流与共享。这不利于政治工作学科建设,也使许多重要的学术与文献资源失去了传承的机会,并导致大量的重复性智力劳动,使政治工作信息资源的建设和使用效益不高。为此,许多政治工作者通过计算机、个人网页、博客、E-mail 等方式来存储自己的教学与科研成果,但是由于个人无法保障系统安全性、网站周期性地重组网页以及容量限制等原因,使得学者科学研究的心血在任何时候都处于极大的变数之中。由此可见,从事军队政治工作研究和教育的机构,迫切需要运用开放理念,建设一个平台,去记载、保存、组织、发布、评价并与同行交流这些政治工作特色的数字资源,充分利用全军政工网将全军院校、科研院所与基层部队连接互通的优势,建立开放的军队政治工作数字资源交流与共享平台。

以上背景,构成本书的研究动机。