

科学数据中心数据与知识集成

南卓铜¹ 王亮绪¹ 吴立宗¹ 祝忠明² 李新¹ 丁永建¹

(1. 中国科学院寒区旱区环境与工程研究所, 甘肃兰州 730000;

2. 中国科学院资源环境科学信息中心, 甘肃兰州 730000)

摘要: 文章回顾了国际上大型科学数据中心知识集成的现状, 认为越来越多的科学数据中心认识到数据中心需要集成更多的知识, 以提高科学数据的可用性以及满足长时期数据存储的要求。传统的元数据结构在应对知识集成时面临诸多挑战。中国西部环境与生态科学数据中心(西部数据中心)形成了从数据收集、规范化整理、数据集成挖掘到数据服务的完整体系, 集成了一批西部环境和生态研究的关键数据, 在数据中心层次上集成多元知识作了有益的开拓性的工作。文章还介绍了西部数据中心知识的表达形式, 描述了以元数据为核心, 数据文档和科学文献为补充的知识组织形式, 并介绍了西部数据中心开展的集成数据检索导航、模型数据集和在线模型服务等知识挖掘手段。

关键词: 数据中心; 知识集成; 知识挖掘; 元数据; 中国西部环境与生态科学数据中心

中图分类号: TP3

文献标识码: A

DOI: 10.3772/j.issn.1674-1544.2010.05.004

Experiences in Knowledge Integration of the Environmental and Ecological Science Data Center for West China

Nan Zhuotong¹, Wang Liangxu¹, Wu Lizong¹, Zhu Zhongming², Li Xin¹, Ding Yongjian¹

(1. Cold and Arid Regions Environmental and Engineering Research Institute, CAS, Lanzhou 730000;

2. Scientific Information Center for Resources and Environment, CAS, Lanzhou 730000)

Abstract: This paper reviews current status of knowledge integration made by the large international scientific data centers. With decades of development, scientific data centers have increasingly realized the demands of data-relevant knowledge integrated by data centers to enhance usability of data and meet challenges arising from long-term data stewardship. As traditional itemized metadata architecture is hard to include more unstructured knowledge, new approach or extension should be developed. Funded by the National Natural Science Foundation of China, Environmental and Ecological Science Data Center for West China (WestDC) was established to share valuable scientific data to support studies in West China. Besides its primary role on data sharing, WestDC has also carried out a number of tentative work to integrate more knowledge to extend data usability and applications. This paper describes possible forms of representing knowledge in the data center, followed by an introduction to the approach adopted in the WestDC to complement metadata with data documentation and scientific literature. Activities on data mining and knowledge discovery in the WestDC including comprehensive data navigation, model data sets, and online model sharing are also outlined.

Keywords: data center, knowledge integration, knowledge discovery, metadata, Environmental and Ecological Science Data Center for West China (WestDC)

第一作者简介: 南卓铜(1977-), 男, 副研究员, 研究方向: 地表过程建模环境、空间决策支持系统、科学数据中心的研究和开发。

基金项目: 国家自然科学基金委“中国西部环境与生态科学研究计划”重点项目(90502010)。

收稿日期: 2010年9月17日。

1 引言

随着观测仪器、测量手段和计算机技术的发展,科学数据的产生比历史上任何时候都要快且多。美国宇航局(NASA)主持的对地观测系统数据信息系统(EOSDIS)目前总存档地球科学数据量为4.2 PB,平均每天数据量增加1.8 TB。数据管理人员开始寻求更有效的数据管理办法,依托高速互联网建立物理上分布、逻辑上统一存取访问的科学数据中心。数据中心的基本功能是汇集、组织管理、存档和发布数据,为了实现这些功能,数据中心一般都通过元数据或者背景文档对数据加以描述,以辅助数据的存取访问和使用。随着数据中心的进一步发展,借助自身拥有的丰富数据资源,数据中心开始设计和实现深层次的知识挖掘和集成。美国NASA地球科学系统明确指出NASA数据信息系统在未来的6至10年内需要从数据管理跨越到知识管理^[1]。“中国西部生态与环境科学数据中心”(以下简称“西部数据中心”)在数据和知识集成方面作了一些开创性的尝试。2005年启动的西部数据中心是国家自然科学基金委资助的服务于中国西部区域的生态、环境和地表过程研究的科学数据中心,除了提供数据共享等基本功能,设置了文献知识平台、交流合作平台、数据工具等功能平台以促进数据使用和知识挖掘^[2]。本文首先介绍国内外数据中心数据和知识集成现状,然后结合西部数据中心介绍数据知识集成的方法和经验。

2 数据和知识集成现状

科学数据中心的数据,可视为科学研究的一切输入,至少包括科学试验、观测和监测产生的数据、各种传感器数据、元数据、模型输出、建模的情景、定量行为数据、可视化数据、行政或者商业的各种统计数据等,而信息可被理解成科学研究的输出,通常也以数据为载体保存在数据中心里^[3]。科学数据中心的知识,除了数据和信息承载的知识,还包括在现有数据和信息基

础上挖掘而产生的有助于开展科研活动的新知识。数据、信息和知识通常没有明显的界线,比如当我们关注的对象是科学数据时,元数据通常作为一种知识提高我们对数据的理解,然而科学数据心里元数据也像其他数据一样被存储和描述。根据以上理解,在本文中涉及的数据和知识集成,是指数据中心除了单纯管理和发布科学数据,还应包括应用这些数据从事研究活动的知识。

世界数据中心成立于1957年(国际地球物理年)。数据中心的数据按学科分门别类,其任务主要是保管和发布原始科学数据。20世纪60年代后,欧美国家开始建立联邦、州、地区等各种级别的数据中心,主要任务是归档和发布数据,开始形成行业元数据标准。20世纪90年代,NASA建立了地球观测系统数据信息系统(EOSDIS)^[4],考虑到传统元数据与数据分离导致元数据容易丢失等问题,EOSDIS采用一种自描述的级联数据格式HDF,数据与数据背景信息保存在同一个文件中^[5]。然而,EOSDIS也存在各种不足,如使用了独特的元数据结构,同时各种卫星遥感数据产品的背景信息分散在地球观测系统网站上。为此,NASA先后实施了新数据信息系统和服务(NewDISS)^[1]和地球科学数据战略发展(SEEDS)计划,研究如何提供更好的数据和信息服务,并提出了从数据管理到知识管理的发展目标,包括为地球观测数据产品建立更好使用和更方便获取的分布式数据管理框架、集成同行科学家的知识和意见、提高系统效能和服务质量等^[6]。澳洲总理科学、工程和创新理事会下属工作组曾以“从数据到智慧”为题建议建立全国性的数据中心,通过信息基础设施保障数据共享,对数据管理、保存、使用、共享和合作、分析和挖掘等方面提出诸多有用意见^[7],指出一个理想数据系统需要在在一个数据系统层面上集成足够的知识,数据需要提升到知识层面进行管理。

在数据中心发展的过程中,元数据在数据管理上的作用得到肯定。元数据一般认为是描述数据的数据,用结构化的条目来描述尽可能详细的有用的数据信息。各学科形成有学科特色的元数据标准。国际标准组织(ISO)颁布了空间地理信

息元数据标准(ISO 19115)^[8]及其实现标准ISO 19139^[9]。其后,地学领域的元数据标准一般都采用在ISO标准基础上的扩展,比如美国和加拿大形成了北美ISO 19115专案^[10]。近些年,我国科学数据中心建设也取得长足进展,科技部组织了科学数据共享工程。在其支持下,中国气象局建立了中国气象科学数据共享网^[11],中国科学院启动了科学数据库及其应用项目^[12],国家自然科学基金委启动了中国西部环境与生态科学数据中心项目^[2]等,这些数据中心采用了不同的元数据标准^[13-17],强调原始数据的共享。

尽管基于元数据的数据中心结构体系被广泛应用,国际上大型数据中心实践表明^[18],条目式的元数据无法容纳更多的结构性不显著的知识,而这些存储于文档或者文献里的知识可能对于未来数据的使用至关重要,需要专人撰写易懂的文档以补充元数据的不足。大型野外试验的数据整理经验也表明^[19-20],对于野外观测数据的获取、处理流程、各种限制信息都需要仔细的文档撰写,以期在将来更容易地使用这些数据。元数据的局限也体现在为了包含各种知识,元数据条目越来越多,或者针对各自需求进行扩展,导致元数据标准之间不兼容,数据中心中的互访问变得困难甚至难以实现。

长时期数据存储是现有数据中心知识存储结构的另一挑战。长时期被定义为一个足够长的时间,在这个时间后技术出现变化,包括存储介质、数据格式甚至用户对象的变化,影响到现有信息的存储^[21]。目前的以元数据为核心的数据中心结构无法有效地存储和传递知识以回应长时期数据存储问题^[21]。长时期数据存储是NASA NewDISS和SEEDS建议的重点关注问题^[1,6]。

然而,在数据中心建设实践上,由于结构性的元数据适合计算机系统实现,国内外主要的数据中心多数是以元数据为核心而搭建的,存在上述各种问题。西部数据中心在数据和知识集成上开展了一些尝试性的工作,主要包括:(1)建立了元数据—文献文档关联体系,将数据与文档、文献集成到一个平台;(2)开展知识挖掘研究,提供集成数据检索导航、模型数据集和在线模型服务等特色服务。

3 西部数据中心数据和知识集成

3.1 总体结构

西部数据中心在功能结构上包括数据共享平台、知识积累平台、合作交流平台和数据科学平台^[2]。数据共享平台提供两种共享方式:在线共享和离线数据服务。前者通过Web页面提供数据发布、检索和下载,通过元数据库进行数据的组织和管理。针对国家政策规定的不能在线共享的特定数据,西部数据中心提供了专人负责离线数据服务。知识积累平台链接数据和文献、数据文档,是数据与知识集成的一个尝试。合作交流平台通过网络论坛、邮件列表等可能途径实现合作和交流,集成专家意见、汇集数据用户团体的知识。数据科学平台提供数据工具和模型数据集,开展数据挖掘研究。

3.2 知识表达

科学数据中心的知识集成体现在两个方面,一是集成辅助数据应用的有关知识,二是集成相关领域的专家知识进行建模和知识挖掘。这些知识不但要满足当前技术条件下的数据应用,还应当考虑未来技术发展时的数据应用。这些知识至少包括:(1)如何使用数据;(2)数据表示的科学意义以及数据如何获取和创建;(3)如何获取最原始的或者某一阶段的数据;(4)如何结合领域知识,使用数据工具和模型进行数据处理和应用,回答科学问题和满足社会需求。

西部数据中心5年的数据共享和服务实践表明,在环境生态领域,以下5方面知识是数据用户最为关心的:

(1)数据的科学基础和生成处理算法。这些知识部分可以包括在元数据的一些字段(如摘要)中,更多的知识则需借鉴数据文档和同行发表的科学论文。

(2)数据结构和格式。包括各参数和字段的含义、单位、格式等,在计算机上物理存储的格式,以及如何有效访问数据。

(3)数据质量评估信息。元数据结构可以描述部分数据质量评估信息,但必须注意到对于一个数据集,数据质量在不同的时间段、不同的空

间范围可能不一样。一些数据集对每一个数据都附加了数据质量标志,以描述数据采集、处理过程中的不确定性。有关数据在某些区域的适用性需要参考同行发表的科学论文。

(4)数据生产时的背景信息。包括生产和校验数据集时使用的其他数据;对于地学观测数据,至少应包括观测站点、仪器、周围环境等信息;对于遥感数据,还应包括传感器的各种特征、传感器标定的数据和方法等。背景信息有时甚至决定了数据的可用性。

(5)数据的应用情况。数据的已有应用情况可以帮助用户快速实现自己的研究目的。

上述非结构化的知识很难结构化到元数据条目上,为此,西部数据设计了以元数据为基础、文献和数据文档为补充的知识集成方案。在此方案下,知识被表现为:

(1)元数据。元数据提供了对数据的基本说明。建立了以ISO 19115地理信息元数据标准为基础、扩展了野外观测/监测数据描述的元数据专案,依托GeoNetwork作为元数据管理平台^[22]。包括元数据实体集信息、数据质量信息、数据内容信息、数据限制信息、标识信息、分发信息、参照系信息、联系人信息、范围信息,以及台站描述信息等扩展信息。

(2)数据文档^[23]。作为元数据的补充,对数据相关的背景知识、数据处理过程、数据属性、数据的使用方法和数据来源及引用信息进行描述。

(3)关联文献。依托数字文献资源库,提取中国西部生态和环境研究文献子集,汇集“中国西部环境与生态科学研究计划”历年来所支持的研究项目和课题的知识产出、建立有关西部环境与生态研究的专题数据库以及对国内外环境与生态科学领域的各种有价值的开放性学术资源的集成利用。通过关键词与数据进行模糊关联。

(4)模型和数据再分析。形成若干种可被广泛使用的模型数据集和具有可靠质量的再分析资料等标准数据集,并且开发一系列用于数据处理和分析的共享工具,开展数据融合、数据集成、数据同化等前沿研究。

(5)专家和数据用户知识集成。一方面,数据生产者或领域专家直接参与或者指导元数据和

数据文档的编写,数据用户进行数据的评价和反馈;另一方面,西部数据中心提供数据用户之间、数据用户与数据提供者之间的沟通机制,并可以通过自由讨论产生新想法。

3.3 知识组织

越来越多的数据中心开始认识到元数据的不足,逐渐强调数据文档对于补充数据信息、提高数据可用性的作用。西部数据实现了元数据和数据文档、科学文献的关联(图1)。

数据文档内容^[23]包括:(1)数据制备、项目支持等背景信息;(2)数据处理过程,包括资料准备、数据采集方法和取样方法,仪器的相关说明,数据处理的方法和详细过程;(3)数据属性信息,包括元信息、时空分辨率、投影信息和数据格式等信息;(4)数据读取和使用说明等;(5)数据来源及引用信息,数据文档中提供明确的数据来源和引用信息。为了保证数据文档的准确和权威性,数据生产者和领域专家参与或指导数据文档的撰写和编辑,在技术上采用基于网络的协作式写作Wiki系统。

图1关联的文献来自于中国科学院资源环境科学信息中心的文献库。在开源数字知识库软件DSpace的基础上,遵循OAIS(开放档案信息系统)参考模型,实现对知识内容进行采集、描述、存储、发布、浏览和利用等多方面的功能和服务。在技术上包括:(1)权威学术资源系统专题知识资源采集模块,从中文权威学术资源系统中筛选有关(西部)环境与生态研究的学术文献,并进行自动采集,集成到知识库;(2)网上开放专题知识内容采集模块,基于OAI-PMH元数据收割协议,从学科专题知识库或机构知识库等开放数字知识库中自动发现并采集与环境生态专题相关的知识内容。

文献与数据中心的集成通过开放互操作接口实现,包括RSS聚合、OAI-PMH开放式档案存取协定、基于网络服务的SRW/U检索集成接口。通过关键词模糊检索实现数据和文献的关联。

对数据用户来讲,技术上的关联是透明的。当用户在检索和访问数据条目时,如果数据文档可用,将以链接的方式给出;关联的文献以“建议的参考文献”列出,如果文献版权许可,在各

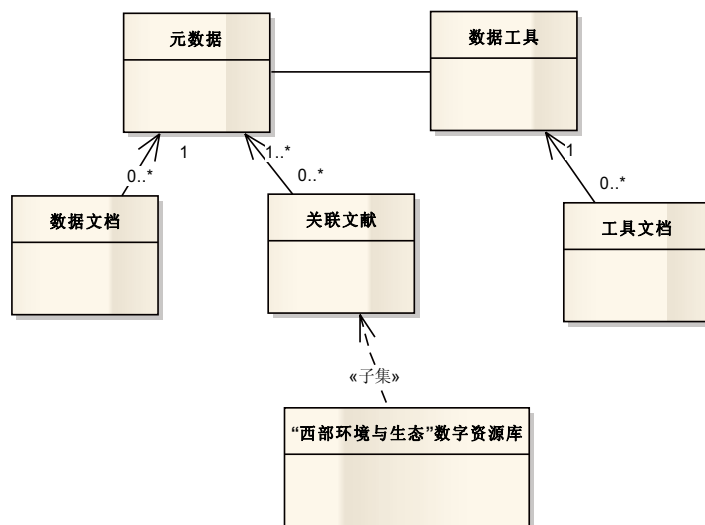


图1 元数据—文献文档关联示意图

项参考文献右侧给出下载链接。

3.4 知识挖掘

借鉴数据库知识挖掘的概念^[24]，知识挖掘是应用数据解决问题的过程，包括5个步骤：检索数据、数据的前处理、数据转换、数据挖掘以及结果的解译和评价。西部数据中心针对地学科学研究的特点，设计和实现了一系列有特色的知识集成和挖掘工具，如数据导航、模型数据集、在线模型服务等。

(1) 集成数据检索和数据导航。西部数据中心是一个逻辑上有统一入口、物理上分布式存储数据的系统。多源数据集成不仅是同一数据中心内多源数据的集成，也包括不同数据中心基于元数据互操作接口进行多源数据的集成。西部数据中心提供了统一的数据浏览、检索和导航系统，能够快速定位所需数据。基于元数据信息实现的数据快速导航模块包括基于数据集序列的、基于数据集类别的、基于数据时间的、基于数据空间位置的和基于关键词的多种导航模式。在数据检索的设计上，根据用户需求设计了快速搜索和高级搜索，其中高级搜索提供了更多的选项，包括关键词和时空信息，以便更准确地快速定位数据。

(2) 数据工具及其网络服务。尽管科学数据各种各样，但一些数据处理的需求是比较集中的，根据服务实践，西部数据中心提供一些常用的工具，包括元数据工具、数据格式转换工具、

空间数据处理工具、可视化工具、建模工具。多数工具仍然是在单机上运行，或者依靠IDL等商用软件运算。随着Web服务技术的成熟，一些数据工具通过网络运行，一方面可以使用服务器相对强大的计算资源，另一方面也可以解决商用软件的版权问题。

(3) 模型数据集^[2]。模型是环境和生态领域的重要研究工具，针对环境、生态、水文领域的常用模型，西部数据中心与模型应用者合作，形成可以用于模型运行的数据集，并发布到网络共享。模型数据集具有很强的专业性、地域性，往往需要长时间制备。在西部数据中心共享网站上公布的几个黑河流域模型数据集是下载次数最多的数据之一，表明模型数据集是可行的科学数据整合和集成方法之一。

(4) 在线模型服务^①。通过Web服务，模型可以被在线运行。西部数据中心做了一些尝试性的工作，建立了一种灵活的、结合在线数据服务和在线模型服务的集成在线服务(图2)。当一个客户程序在客户端调用一个位于西部数据中心服务器的远程模型，在线模型服务将调用西部数据中心的在线数据服务，对于西部数据中心不具备的数据，由客户端提供，模型在服务器上运行结束后，结果返回到客户端机器，客户可进行进一步分析和处理。

(5) 数据整合、同化和再分析。西部数据中

① 参见朱仕杰、南卓铜、陈昊等《基于Web Service的在线水文模型服务研究》文。

心开展了一些数据融合、数据集成、数据同化等数据科学的前沿研究。目前西部数据中心网站公布的中国陆域2000年1公里土地覆盖数据^[25]即为多源数据融合和同化的产出。

4 总结与展望

中国西部环境与生态科学数据中心经过多年(2006-2010)的运行,形成了从数据收集、规范化整理、数据集成挖掘到数据服务的完整体系。当国际科学数据中心考虑从数据管理到知识管理的角色转变,西部数据中心就知识集成作了有益的开拓性的工作,主要表现在:

(1)多种知识表达途径,包括元数据、数据文档、科学文献、模型和再分析数据,以及专家和用户的知识集成,在一定程度上解决了传统元数据手段无法表达更多数据知识的问题。

(2)依托中国科学院资源环境科学信息中心数字知识资源库,西部数据中心实现了与西部环境生态研究有关的科学文献的集成,通过数据可以导航到相关专业文献。

(3)面向西部环境和生态科学问题,通过与领域专家的紧密配合,西部数据中心在知识挖掘领域做了一些有益尝试,包括建立了集成数据检索和数据导航、模型数据集、在线模型服务等一系列有特色的服务。

然而,西部数据中心的数据和知识集成程度仍然不高。本文介绍的数据与知识的集成和挖掘更多的是示范和研究性质的,还没有得到大范围实施。

从数据到知识的转变,首先是理念上的提高,将来的数据中心不是数据的简单存档和管理,而是对与本学科研究有关知识的管理。在技术上,科学数据中心将来的发展也要考虑大规模的数据应用服务,比如在线模型服务,基于云的数据和模型服务等。考虑到知识挖掘是面向科学问题的,西部数据中心要加强与其他数据中心、环境和生态领域专家的合作,在数据和知识集成的广度和深度上均需努力,进一步促进数据共享和提高知识挖掘水平。一个理想的数据中心应当如国际科学委员会ICSU^[3]所归纳的,除了开放和完全的数据访问,数据本身应当权威,用户通过数据中心能得到足够的信息,鼓励数据应用和学科交叉研究,并有足够的弹性满足将来用户对数据的需求。

参考文献

- [1] NewDISS Team. NewDISS: A 6-to 10-Year Approach to Data Systems and Services for NASA's Earth Science Enterprise[R]. Washington DC, US: NASA, 2002:1-51.
- [2] Li Xin, Nan Zhuotong, Wu Lizong, et al. Environmental and Ecological Science Data Center for West China: In-

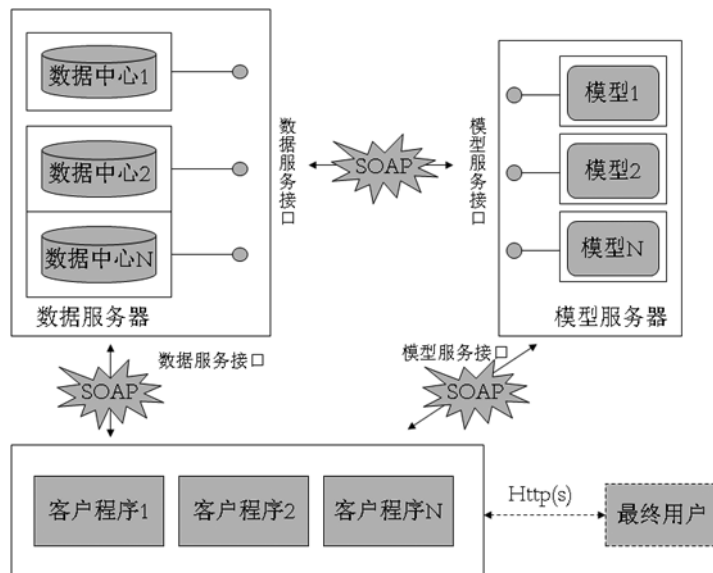


图2 西部数据中心在线模型服务

- tegration and Sharing of Environmental and Ecological Data[J]. *Advances in Earth Science*, 2008, 23(6): 628–637.(in Chinese)
〔李新, 南卓铜, 吴立宗, 等. 中国西部环境与生态科学数据中心：面向西部环境与生态科学的数据集成与共享 [J]. *地球科学进展*, 2008, 23(6): 628–637. 〕
- [3] Ad Hoc Strategic Committee on Scientific Information and Data. Draft Final Report to the ICSU Committee on Scientific Planning and Review[R]. Paris, France: ICSU, 2008:1–39.
- [4] Kobler B, Berbert J. NASA Earth Observing System Data Information System (EOSDIS)[C]// 11th IEEE Symposium on Mass Storage Systems. Los Alamitos, CA: IEEE Computer Society Press, 1991:18–19.
- [5] Meyer T, Suresh R, Ilg D, et al. Mosaic, HDF and EOSDIS: Providing Access to Earth Sciences Data[J]. *Computer Networks and ISDN Systems*, 1999, 28(1/2): 221–229.
- [6] SEEDS Formulation Team. Strategic Evolution of Earth Science Enterprise Data System (Seeds) Formulation Team Final Recommendations Report[R]. USA: NASA, 2003:1–80.
- [7] Prime Minister's Science Engineering and Innovation Council (PMSEIC) Working Group. From Data to Wisdom: Pathways to Successful Data Management for Australian Science[R]. Australia: Prime Minister's Science Engineering and Innovation Council, 2006:1–94.
- [8] International Standard Organization. Geographic Information—Metadata[S].ISO 19115:2003. ISO TC211, 2003:1–140.
- [9] International Standard Organization. Metadata – XML Schema Implementation[S]. ISO/TS 19139:2007. TC/SC, 2007:1–111.
- [10] NAP Metadata Working Group. North American Profile of ISO19115:2003 – Geographic Information – Metadata[S].Reston, Virginia, US: North American Profile Committee, 2007:1–282.
- [11] Qin Dahe. Meteorology Data Sharing Is Essential to Effective Utilization of National Resources[J]. *China Territory Today*, 2003(2): 27–29.(in Chinese)
〔秦大河. 气象科学数据共享是国家资源有效利用的必然选择 [J]. *今日国土*, 2003(2): 27–29. 〕
- [12] Gui Wenzhuang. The 20-year Development of CAS Scientific Database [J]. *Bulletin of the Chinese Academy of Sciences*,2007(1): 87–89, 91.(in Chinese)
〔桂文庄. 迎接科学数据库发展的新阶段——中国科学院科学数据库发展 20 年的回顾与思考 [J]. *中国科学院院刊*,2007(1): 87–89, 91. 〕
- [13] State General Administration of the People's Republic of China for Quality Supervision and Inspection. Metadata for Ecological Data[S].GB/T 20533 – 2006. Beijing: China Standard,2007:1–169.(in Chinese)
〔国家质量监督检验检疫总局. 中华人民共和国国家标准—生态科学数据元数据 [S].GB/T 20533 – 2006. 北京: 中国标准, 2007:1–169. 〕
- [14] State General Administration of the People's Republic of China for Quality Supervision and Inspection. Geographic Information – Metadata[S].GB/T 19333.15—200X/ISO 19115:2003. Beijing: China Standard,2004:1–45.(in Chinese)
〔国家质量监督检验检疫总局. 中华人民共和国国家标准—地理信息元数据 [S].GB/T 19333.15—200X/ISO 19115:2003. 北京: 中国标准, 2004:1–45. 〕
- [15] Scientific Database Center of CNIC. Chinese Academy of Sciences Scientific Database Core Metadata (Version 2.0)[R]. Beijing: Computer Network Information Center, Chinese Academy of Sciences, 2004:1–178.(in Chinese)
〔中国科学院计算机网络信息中心科学数据库中心. 中国科学院科学数据库核心元数据标准 V2.0[R]. 北京: 中国科学院计算机网络信息中心, 2004:1–178. 〕
- [16] Li Jun, Zhou Chenghu. Overview on Metadata Standards of GEO Spatial Data[J]. *Progress in Geography*, 1998, 17(4): 55–63.(in Chinese)
〔李军, 周成虎. 地球空间数据元数据标准初探 [J]. *地理科学进展*, 1998, 17(4): 55–63. 〕
- [17] CMA. Core Metadata Content of Meteorological Dataset[S].QX/T 39–2005, 2005:1–13.(in Chinese)
〔中国气象局. 气象数据集核心元数据 [S].QX/T 39–2005, 2005:1–13. 〕
- [18] Parsons M A, Duerr R. Designating User Communities for Scientific Data: Challenges and Solutions[J]. *Data Science Journal*, 2005(4): 31–38.
- [19] Parsons M A, Brodzik M J, Rutter N J. Data Management for the Cold Land Processes Experiment: Improving Hydrological Science[J]. *Hydrological Processes*, 2004, 18(18): 3637–3653. DOI:10.1002/hyp.5801.
- [20] Newcomer J A, Huemrich K F, Landis D, et al. Managing and Supporting Large Integrated and Interdisciplinary Field Studies: The Boreas Example[J]. *Journal of Geophysical Research—Atmospheres*, 2001, 106(D24): 33517–33528.
- [21] Duerr R, Parsons M A, Marquis M, et al. Challenges in Long-Term Data Stewardship[C]//21st IEEE Conference on Mass Storage Systems and Technologies. Adelphi MD, US: IEEE Computer Society,2004: 47–67.

(下转第 36 页)

享网站系统,集成了一批西部环境与生态乃至整个中国大陆地球表层科学方面的关键数据集,为西部计划等项目及科研团体与个人提供了持续的数据服务。西部数据中心将继续在数据集成的广度和深度方面努力,并与其他数据中心合作,争取建立以环境与生态领域的科学问题为导向,紧密服务在寒区旱区开展的各类科学计划的特色科学数据中心。

参考文献

- [1] Leng Shuying, Li Xiubin, Cheng Guodong, et al. The Progress of Studies on the Environmental Change and Ecological Issues in Western China[J]. Science Foundation in China, 2005(5):262-267. (in Chinese)
〔冷疏影,李秀彬,程国栋,等.中国西部环境和生态科学重大研究计划阶段性进展及深入研究的问题[J].中国科学基金,2005(5):262-267.〕
- [2] Li Xin, Nan Zhuotong, Wu Lizong, et al. Environmental and Ecological Science Data Center for West China: Integration and Sharing of Environmental and Ecological Data[J]. Advances in Earth Science, 2008, 23(6):628-637. (in Chinese)
〔李新,南卓铜,吴立宗,等.中国西部环境与生态科学数据中心:面向西部环境与生态科学的数据集成与共享[J].地球科学进展,2008,23(6):628-637.〕
- [3] Wang Liangxu, Wu Lizong, Nan Zhuotong, et al. Application of Open Source Technologies in Geoscientific Data Centers[J]. China Science & Technology Resources Review, 2010, 42(3):17-23, 35. (in Chinese)
〔王亮绪,吴立宗,南卓铜,等.开源技术在地球科学数据中心中的应用[J].中国科技资源导刊,2010,42(3):17-23,35.〕
- [4] Zhu Zhongming, Ma Jianxia, Chang Ning, et al. A DSpace-based Sharing Environmental and Ecological Knowledge Space[J]. Library and Information Service, 2007, 51(4): 71-74, 108. (in Chinese)
〔祝忠明,马建霞,常宁,等. SEEKSpace——基于DSpace的环境与生态科学知识积累平台[J].图书情报工作,2007,51(4):71-74,108.〕
- [5] Li X, Li X, Li Z, et al. Watershed Allied Telemetry Experimental Research[J]. Journal of Geophysical Research, 2009, 114: 19.
- [6] Pan Xiaoduo, Li Xin, Nan Zhuotong, et al. Research on Data Description Document [J]. China Science & Technology Resources Review, 2010, 42(3):30-35. (in Chinese)
〔潘小多,李新,南卓铜,等.科学数据文档的研究[J].中国科技资源导刊,2010,42(3):30-35.〕
- [7] Li Hongxing, Wang Jian, Nan Zhuotong, et al. Data Service of Environmental and Ecological Science Data Center for Western China [J]. China Science & Technology Resources Review, 2010, 42(3):24-29. (in Chinese)
〔李红星,王建,南卓铜,等.西部数据中心的数据服务实践[J].中国科技资源导刊,2010,42(3):24-29.〕
- [22] Wang Liangxu, Wu Lizong, Nan Zhuotong, et al. Application of Open Source Technologies in Geoscientific Data Centers[J]. China Science & Technology Resources Review, 2010, 42(3): 17-23, 35.(in Chinese)
〔王亮绪,吴立宗,南卓铜,等.开源技术在地球科学数据中心中的应用[J].中国科技资源导刊,2010,42(3):17-23,35.〕
- [23] Pan Xiaoduo, Li Xin, Nan Zhuotong, et al. Research on Data Description Document[J]. China Science & Technology Resources Review, 2010, 42(3): 30-35.(in Chinese)
〔潘小多,李新,南卓铜,等.科学数据文档的研究[J].中国科技资源导刊,2010,42(3):30-35.〕
- [24] Fayyad U, Piatetskyshapiro G, Smyth P. The KDD Process for Extracting Useful Knowledge From Volumes of Data[J]. Communications of the ACM, 1996, 39(11): 27-34.
- [25] Ran Y H, Li X, Lu L. Evaluation of Four Remote Sensing Based Land Cover Products Over China[J]. International Journal of Remote Sensing, 2010, 31(2): 391-401.

(上接第21页)