

世界数据系统的体系框架及其数据门户初步应用

王卷乐 宋 佳

(资源与环境信息系统国家重点实验室, 中国科学院地理科学与资源研究所, 北京 100101)

摘要: 世界数据系统(WDS)是国际科学联合会(ICSU)的一个新的跨学科的科学团体,是在原世界数据中心(WDC)和国际天文和地球物理数据服务联合会(FAGS)的基础上,整合并纳入新的数据中心形成的。该系统于2008年经ICSU批准组建,即将于2011年正式启动。本文对WDS的体系框架进行分析,包括其组织架构、政策机制、功能体系。新的WDS数据门户已经开始设计和原型开发。基于OAI元数据收割协议,隶属于中国科学院的5个原WDC学科数据中心已经开展了与WDS数据门户的联网元数据检索,并初步实现了元数据互操作和资源共享应用。这在技术能力上,为我国加强与WDS的合作与交流创造了条件,为我国加强科学数据共享国际合作与交流提供了良好基础。

关键词: 世界数据系统, 世界数据中心, 数据共享, 中国中心

中图分类号: G351

文献标识码: A

DOI: 10.3772/j.issn.1674-1544.2011.04.010

System Framework Analysis of World Data System and Its Initial Application on Data Portal

Wang Juanle, Song Jia

(State Key Lab of Resources and Environment Information System, Institute of Geographic Sciences and Natural Resource Research, Chinese Academy of Sciences, Beijing 100101)

Abstract: World Data System (WDS) is new cross-discipline communities of International Council for Science Unions (ICSU). It is Body by incorporating the WDCs and FAGS as well as other 'state of the art' data centres and services. It was designed in 2008 and will be initiative in January, 2011. The paper analyzed WDS's general system framework, i.e., system architecture, data policy and operational mechanism, technical structure. WDS's data portal was developed before the new system launched. Five world data centers in Chinese Academy of Sciences joined the pilot portal project. Based on Open Archives Initiatives Protocol for Metadata Harvesting (OAI-PMH), metadata in different data centres are retrieved by one stop in the WDS data portal prototype. This is a good technology basis for more data sharing agencies and communities in China with WDS. It will promote China's scientific data sharing cooperation and exchange in the world.

Keywords: World Data System, World Data Centre, data sharing, China Centers

1 引言

为了推进全球科学数据和信息的共享, 国际科学联合会(ICSU)于1956年成立了国际天文和

地球物理数据服务联合会(FAGS)、1957成立了世界数据中心系统(WDC), 以此推进东西方、发达和落后地区的数据交换与共享^[1]。WDC和FAGS在过去的50年得到了长足的发展。WDC

第一作者简介: 王卷乐(1976—), 男, 博士, 中国科学院地理科学与资源研究所副研究员, 主要研究方向: 科学数据集成与共享关键技术标准研究、GIS与遥感应用。

基金与合作项目: 2010年中国科技资源导刊合作项目“基于实践的科学数据资源管理基本理论与方法问题探访”(DK2010-02010); 资源与环境信息系统国家重点实验室自主项目基金(2008)。

收稿日期: 2010年12月7日。

已在全球发展到 51 个数据中心。FAGS 也建立了长期存在的 12 个服务中心。但是,这两个系统还不能满足 ICSU 总体战略的需要,它们面临着在总体布局、支撑服务能力、系统互操作和可持续运行服务等方面的严峻挑战^[2]。在这种情况下,ICSU 在 2008 年的第 29 届大会上正式宣布,将原 WDC 与 FAGS 合并,并吸纳更多技术先进的数据中心,建立新的世界数据系统(World Data System, WDS)^[3]。WDS 于 2008 年开始组建以来,先后组建了 WDS 过渡工作组(WDS-TT)和 WDS 专家委员会(WDS-SC)筹划新的系统框架。经过 2 年多的努力,目前该体系架构已经基本建立,即将于 2011 年正式启动。本文将总体介绍 WDS 的组织架构、政策机制、功能体系,以及我国数据中心与 WDS 门户的初步元数据互操作和共享应用。

2 WDS 体系框架

2.1 目标

WDS 的核心任务是针对 ICSU 的使命和目标,面向国际科学界和其他利益相关者,建立一个国际框架,在该框架内提供科学驱动的数据和产品的长期服务,对这些数据的访问、分发以及对其发展和设计的质量评价,促进和支持多学科集成和科学发展。因此,WDS 的总体目标可以概括为:确保对科学数据、数据服务、产品和信息的普遍和平等访问;保证服务的长期性、便捷性;促进数据标准和规范的一致性;确保数据、

产品和信息的质量保证;缩小数字鸿沟,最终为科学进步提供更好的支持^[4]。

2.2 组织结构

WDS 的组织结构与科学和技术数据委员会(CODATA)等其他合作伙伴并列归属于 ICSU,如图 1 所示^[4]。在国际项目办公室(IPO)的协调下,整合原有的 WDC 和 FAGS 成员机构以及更多符合条件的、技术先进的中心。值得说明的是,原有 WDC 和 FAGS 成员不能自动成为新的 WDS 的成员,如果希望加入,则需要重新申请和接受评估。

2.3 政策机制

WDS 的政策机制可以通过其章程体现^[5]。WDS 已经制定了章程(草案),将在确立国际项目办公室(IPO)后,提交 ICSU 批准和执行。其中明确规定,WDS 的成员应该满足 WDS-SC 通过的认定条件。WDS-SC 初步制定的认定条件见表 1 所示^[6]。

WDS 成员的义务包括科学数据和信息的采集、分析、编辑、分发、归档和网络支持,面向 ICSU 服务的科学团体和广大科学界提供数据服务。ICSU 不为 WDS 成员的发展提供资金支持。作为科学基础设施的核心部分,这些机构应该得到本国和国际的资金支持。

2.4 功能体系

WDS 是一个“数据系统的系统”,其融合了多种类型的机构。例如,以数据存储、管理和服务为核心的原世界数据中心,以面向科学计划提

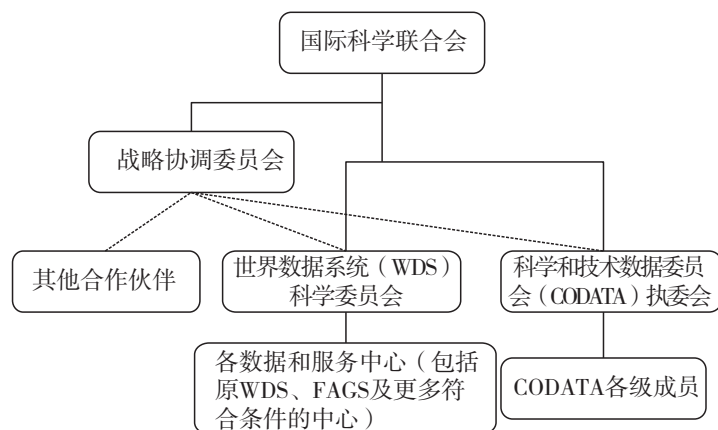


图 1 ICSU 的数据和信息体系战略结构图

表 1 WDS 的认定条件(草案)

评估内容	编号	条件要求
WDS总体需求和政策	1.1	签署协议信, 同意加强与ICSU的合作
	1.2	具有相关的国际专家, 并能够向WDS的节点提供建议和指导
	1.3	参加WDS每两年一次的会议
	1.4	推动科学家和其他用户的交流
	1.5	提供全面、开放、及时、非歧视和不限定的元数据、数据和服务访问, 不收取费用或者不超过用户获得数据的费用
组织框架	2.1	机构已经确定具有以下明确定义: (1) 所拥有数据、数据产品、服务的范围。(2) 长期保存这些数据、产品和服务。(3) 服务对象和用户群体。(4) 用户访问和使用数据的权利。(5) 针对科学需求变化的技术革新能力。
	2.2	组织机构具有足够的资金、质量合格的员工、组织架构和长期计划
	2.3	主管机构的专家提供本地的监督
	2.4	具有长期计划, 如果依托机构转变其工作重点或不再继续支持数据中心建设仍然能够运行
	2.5	同意接受正式的、周期性的检查和评价, 以确保能够对科学技术和发展的需求变化做出反应
数据、产品和服务的管理	3.1	确保数据集在获取、归档存储、数据质量评价和分析、产品生产、访问和分发过程中的集成和认证
	3.2	整合、集成按照规定标准采集、整合和生产出的数据集
	3.3	按照定义的规范对数据集进行归档和存储
	3.4	允许高效地使用归档数据集、数据产品和服务, 鼓励采用开放的标准(可搜索、可访问、可使用和服务)
技术设施	4.1	设施功能基于支持良好的操作系统和其他核心软件基础
	4.2	能够使用软件和硬件技术支撑其设计的服务功能
	4.3	具有安全管理技术能力, 保护设施内的用户、数据、产品和服务安全

供服务的世界数据服务机构, 提供数据的分析和服务的世​​界数据分析服务中心以及已有的兼有数据管理和分发职能的其他中心^[7]。这些机构将通过统一的功能体系, 融合在新的 WDS 功能体系内, 其结构如图 2 所示。

由图 2 可见, 深色背景框内显示的是核心功能, 无背景色框内显示的是这些功能与其他相关者的关系。其中, 核心功能包括元数据和数据服务, 数据归档、存储和发布功能, 数据采集和处理功能等。除了核心功能, 在整个 WDS 的功能体系中, 还包括数据的可视化功能、数据分析功能、数字图书馆的数据归档、编目功能以及通用的数据检索、目录服务、获取及分析服务等功能。

3 WDS 数据门户初步应用

为了加强未来 WDS 对 ICSU 的相关科学团体的数据交换与服务, 原 WDC 于 2008 年尝试建

立了一站式的数据服务系统, 即数据门户(Data Portal)^[8]。科学数据门户是科学数据共享设施发展到较高层面的一个产物, 它是专门供科研人员使用的用于获取科学数据资源及相关服务的一个门户系统, 是众多的科学数据共享网络、科学数据中心的延伸, 是充分地利用了 Web 技术构建的特殊门户网站^[9]。

WDS 数据门户设想采用 OAI-PMH 标准协议, 实现统一的元数据访问服务。数据门户的前期开发由位于德国不莱梅大学的世界数据中心海洋学科中心负责, 作为试点研究的一部分, 隶属于中国科学院的 5 个 WDC 学科中心参加了联网试验。这 5 个中心分别是 WDC 天文学科中心、空间学科中心、地球物理学科中心、冰川冻土学科中心和可再生资源与环境学科中心。由于我国的 5 个中心在数据资源体系上已经被国家科技基础条件平台——地球系统科学数据共享网(<http://>

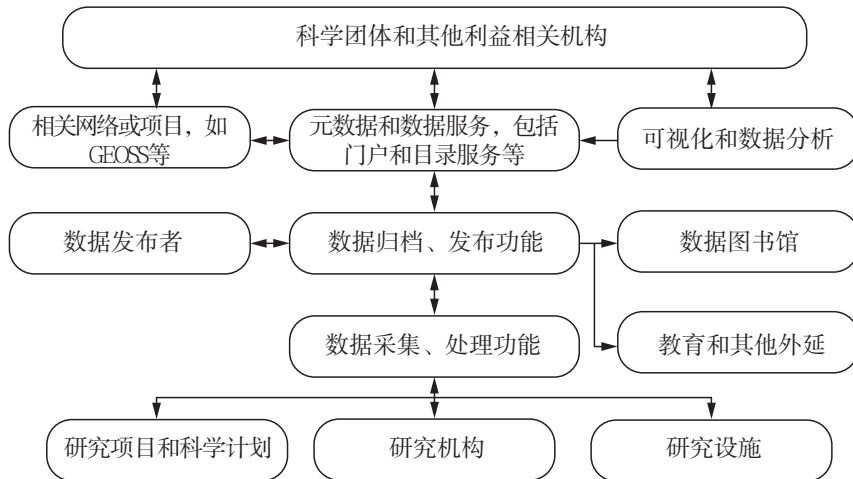


图 2 WDS 的体系功能及其与外部的关系

www.geodata.cn) 进行整合, 因此统一由该网对外建立元数据服务接口。

3.1 元数据互操作方法基础

OAI 意为开放文档先导。该协议得到了数字图书馆联盟、美国网络信息联合会、美国国家科学基金会等机构支持。OAI 元数据采集协议框架把系统的参与者分成数据提供者 (DP) 和服务提供者 (SP) 两类, 前者是指支持 OAI 元数据采集协议的元数据发布者, 后者通过 OAI 元数据采集协议从数据提供者那里采集元数据, 并以此为基础提供一站式服务。协议规定数字资源库中的每一条元数据记录要按统一标准制定唯一标识符, 并与数字对象进行链接^[10]。

实现 OAI 协议, 对服务提供者 (即数据采集方) 而言, 就是按照协议规定的格式和参数发送指令; 对数据提供者而言, 就是能对由采集方发出的请求指令按协议规定的格式返回对方想要的数。OAI-PMH 协议定义了 6 个谓词 (verb),

这些谓词完成 DP 与 SP 之间的交互 (表 2)^[11]。这 6 个谓词中, 有 3 个谓词用于提供描述存档本身的元数据, 它们分别是: Identify、ListSets 与 ListMetadataFormats, 它们提供了对存档的描述、联系方式、用户协议以及其他有关规定。元数据收割的主要实质性工作是由 ListIdentifiers、GetRecord 以及 ListRecords 这 3 个谓词完成的。

OAI 协议的实现, 对数据提供方而言, 要能通过 Web 服务器接收采集方发送的 6 条指令, 并用 XML 格式返回相应的结果; 对采集方而言, 能发送 HTTP 请示到指定的 BaseURL, 并对接收到的 HTTP 相应进行解析。

3.2 元数据互操作接口

对于参加联网的数据中心, 实现元数据互操作需要 3 个步骤。(1) 定义一个合理的数据集粒度, 并且进行描述。(2) 选择合适的标准建立元数据目录, 例如 ISO19115、DIF 或者 FGDC 等开放标准。(3) 建立一个符合 OAI-PMH 的目录服

表 2 OAI-PMH 中的 6 个谓词列表

谓词	说明
Identify	返回该存档的详细描述 (名称、联系方式等)
ListSets	返回库中可用的资源分组的名称
ListMetadataFormats	返回存档所用的元数据格式
ListIdentifiers	返回满足查询条件的记录的标识
GetRecord	返回与指定标识对应的记录
ListRecords	返回满足查询条件的记录

务, 提供接口服务。其中与 WDS 数据门户最终联接的一步即为开发元数据互操作接口。

为此, 在经过数据集整合、描述、编目的基础工作后, 地球系统科学数据共享网在其英文网站平台 (<http://eng.geodata.cn:8080>) 上开发了 OAI-PMH 的调用接口, 访问接口信息如下:

<http://eng.geodata.cn:8080/geonetwork/srv/en/main.home/oaipmh?verb=Identify>

<http://eng.geodata.cn:8080/geonetwork/srv/en/main.home/oaipmh?verb=ListMetadataFormats>

<http://eng.geodata.cn:8080/geonetwork/srv/en/main.home/oaipmh?verb=ListSets>

<http://eng.geodata.cn:8080/geonetwork/srv/en/main.home/oaipmh?verb=ListIdentifiers&metadataPrefix=iso19115>

<http://eng.geodata.cn:8080/geonetwork/srv/en/main.home/oaipmh?verb=ListRecords&metadataPrefix=iso19115>

<http://eng.geodata.cn:8080/geonetwork/srv/en/main.home/oaipmh?verb=GetRecord&identifier=84841320-8105-11dc-9843-00188b49aadb&metadataPrefix=iso19115>

该协议接口为位于德国不来梅的 WDS 门户与位于我国的地球系统科学数据共享平台 (英文网站) 之间建立了元数据标准和收割的接口。WDS 门户能够通过该接口收割中国的数据资源。根据协议参数, WDS 门户将按照 ISO19115 的元数据标准进行元数据收割, 实现元数据互操作。

3.3 元数据互操作应用

基于 OAI-PMH 的互操作, 实现了 WDS 数据门户与我国 5 个 WDC 学科中心 (依托地球系统科学数据共享平台) 的元数据互访。其应用界面如图 3、图 4 所示。

如图 3、图 4 所示, 当输入搜索关键词 “land use” (土地利用) 后, 可以搜索到 599 个数据集结果, 其中多数是来自于中国科学院的数据内容。例如, 第一条数据集就是由中国科学院地理科学与资源研究所生产的中国土地利用数据库。这些数据的详细元数据信息可以通过数据门户网站直接浏览, 点击数据链接可以访问到数据实体对象。

4 结语

WDS 即将重装待发, 但面向未来的发展仍然存着诸多机遇与挑战。在机遇方面, 其具有的优势主要有: 原有 WDC 和 FAGS 等机构具有长期数据服务的基础; 对科学和数据有深刻的理解; 可开放访问所有已有数据资源; 具有较强的全球数据管理能力及学科交叉应用能力, 等等。在挑战方面, 其发展劣势在于: 缺乏专项的资金支持; 数据服务的组织和数据质量管理不够连续; IT 技术的支持还不足; 系统整体性不强、资源分散, 等等。然而, 对于我国来说, WDS 的发展将为我国科学数据共享提供更多机遇, 这主要体现在以下几点。(1) 新的 WDS 将允许除原有的 WDC 和 FAGS 成员单位外的更多数据共享机构加盟, 这为我国更多数据共享机构和团体提供了参与国际数据共享合作交流的机会;(2) 我国已经在技术体系上与 WDS 门户建立了合作关系, 在技术系统的国际接轨方面具有一定的优势;(3) 我国近年来在国家科技基础条件平台等数据共享基础设施的建设上取得可喜的进展, 通过与 WDS 的合作与交流可以提高我国在科学数据共享领域的国际影响;(4) 促进我国与 WDS 体系内其他国家数据中心的联系和沟通, 引入新的技术经验、政策规范及数据资源, 进而推动我国科学数据共享的整体发展。

致谢: 感谢德国不来梅大学 WDC 海洋学科中心的 Michael Diepenbroek、Robert Huber 在技术上的支持与协调。

参考文献

- [1] Wang Juanle, Sun Jiulin, Development of China WDC Systems for Data Sharing[J]. China Basic Science Research, 2007(2): 36-40. (in Chinese)
[王卷乐, 孙九林. 世界数据中心(WDC)中国学科中心数据共享进展[J]. 中国基础科学, 2007(2): 36-40.]
- [2] Wang Juanle, Sun Jiulin. World Data Center's Review, Reform and Prospect Analysis[J]. Advances in Earth Science, 2009, 24(6): 612-620. (in Chinese)
[王卷乐, 孙九林. 世界数据中心(WDC)回顾、变革与展望[J]. 地球科学进展, 2009, 24(6): 612-620.]
- [3] ICSU. International Science Community Agrees on First

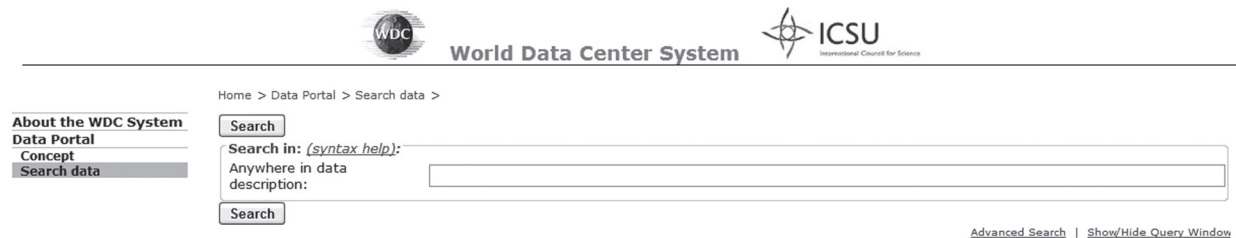


图3 WDS数据门户搜索界面

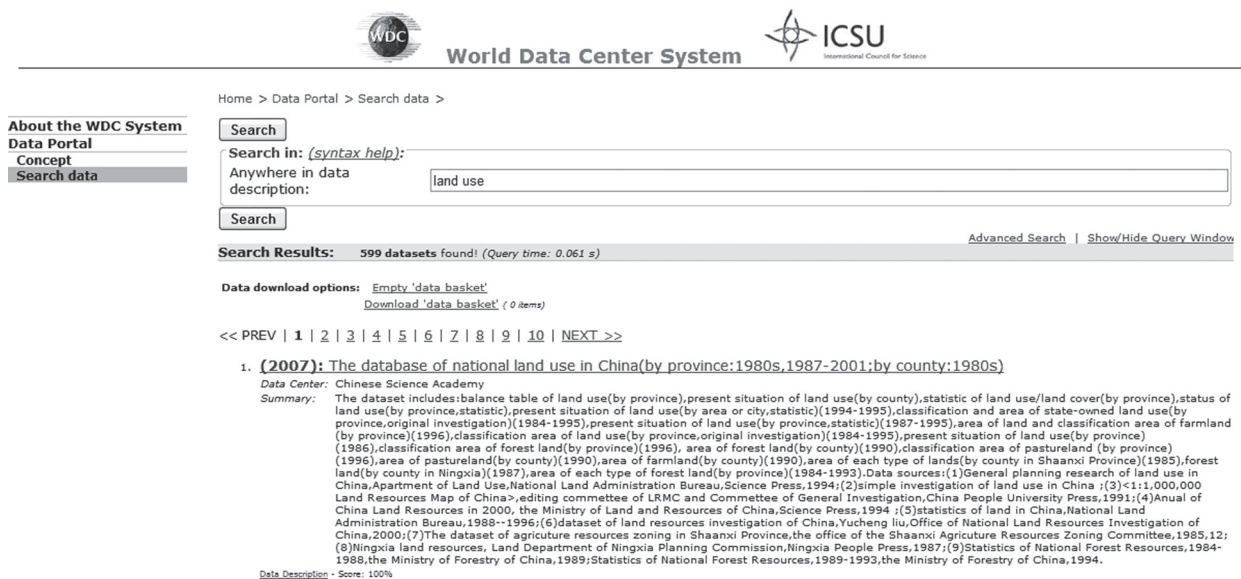


图4 WDS数据门户的跨数据中心搜索结果

- Steps to Establish a Global Virtual Library for Scientific Data[EB/OL].[2008-09-29]. http://www.icsu.org/3_mediacentre/GA_29.html.
- [4] Ad hoc Strategic Committee on Information and Data Final Report to the ICSU Committee on Scientific Planning And Review[EB/OL]. [2008-03-30]. <http://www.icsu.org>.
- [5] World Date System.Constitution of the International Council for Science World Data System (draft)[EB/OL]. [2010-10-30]. <http://www.world-data-system.org/>.
- [6] Certification of World Data System Facilities and Components (draft)[EB/OL].[2010-10-23]. <http://www.icsu-wds.org/>.
- [7] Jean Bernard Minster. Presentation of the ICSU World Data System: 22nd International CODATA Conference, South Africa, 24-27 Otc 2010.
- [8] Data Portal of WDC System 2.0[EB/OL].[2010-12-30]. <http://www.world-data-system.org/>.
- [9] Liu Runda. Scientific Data Portal & Its Building Practices—Earth System Science Data Sharing Network as an Example[D]. Beijing: Chinese Academy of Sciences, 2009.(in Chinese)
〔刘润达.科学数据门户与构建实践研究——以地球系统科学数据共享网为例[D].北京:中国科学院博士论文,2009.〕
- [10] Zheng Xiaohui, Zhang Haitao. Study on Open Archives Initiatives Protocol[J]. Library, 2004,23(1):55-58.(in Chinese)
〔郑小惠, 张海涛. 开放资源互操作协议OAI研究[J].图书馆杂志, 2004,23(1):55-58.〕
- [11] Dong Hui, Ding Botao. Study on OAI-PMH[J]. Library Information and Knowledge, 2004(6): 70-73.(in Chinese)
〔董慧,丁波涛. OAI-PMH协议初探[J].图书情报知识,2004(6):70-73.〕