

# 地球系统科学数据共享标准规范体系研究与应用

王卷乐 孙九林

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

**摘 要:**地球系统科学是地球科学发展的一个新的历史阶段,其以地球系统及其整体行为作为研究对象,迫切需要大量多学科、多来源、多类型、综合性地学数据资源的支撑。针对这一需求,我国国家科技基础条件平台设立了“地球系统科学数据共享网”这一支撑条件。为了构建该支撑条件的标准规范环境,本文分析了地球系统科学数据的特征,剖析了“研究型”数据对标准规范的需求,指出了其面临的 3 个关键问题。研究了地球系统科学数据共享的概念模式,根据定义的 4 条基本原则,分析了地球系统科学数据共享标准规范体系的定位,构建其体系结构。该体系结构包括 4 个大类,即机制条例类、数据管理类、平台开发类、数据服务类,具体包括 18 项条例、办法、规范和技术标准。其中,地球系统科学数据共享联盟章程、核心元数据标准、数据质量管理办法、数据分类标准是该体系中的引领性、核心标准规范。经过近 6 年的研究和应用,目前该标准规范体系已经在地球系统科学数据共享网的总中心和 13 个分中心试用,取得了良好的运行服务效果。未来,地球系统科学数据共享标准规范将“向下”、“向上”两个方向继续发展。

**关 键 词:**地球系统科学;数据共享;标准规范;条件平台

## 1 引言

自 20 世纪 50 年代开始凸显的资源、环境、生态、灾害等问题直接涉及人类的生存与发展,渗透在社会生活的方方面面,使人类面临着来自赖以生存的地球的巨大挑战,也对科学研究提出了新的要求。为应对这些复杂问题,新的研究模式与研究思维开始快速发展,这促使地球科学的发展进入了一个新的历史时期。刘东生先生曾指出,地球科学在经历了“地球科学的系统”、“系统的地球科学”两个阶段后,从 20 世纪 80 年代中后期开始进入第三个阶段,即“地球系统的科学”时期<sup>[1]</sup>。地球系统科学把大气圈、水圈(含冰雪圈)、生物圈、岩石圈、地幔和地核、近地空间以及人类圈作为紧密联系的整体研究对象,采用复杂系统科学理论和方法以及现代高新技术手段研究其相互作用的过程和机理<sup>[2-3]</sup>。

地球系统科学是地球科学 21 世纪发展的前沿领域,具有一系列显著的特征,诸如全球系统观、全时空尺度、多学科交叉集成、高新技术应用体系化、高投入、高精度、信息数字化、强社会应用性、大科学计划推动、国际合作等<sup>[4]</sup>。地球系统科学的研究对

象是地球系统及其整体行为,研究方法是对全球环境变化进行观测、理解、模拟和预测。这二者决定了地球系统科学的研究对海量的、多样化的观测、探测、调查、试验数据的依赖,迫切需要大量多学科、多来源、多类型、综合性地学数据资源的支撑<sup>[3]</sup>。而这些数据主要来源于地学领域的科学研究项目,广泛分布于高校、科研院所以及科学家个人手中。

面对这一需求,我国自 2002 年启动国家科学数据共享工程首批 9 个试点时,就设置了“地球系统科学数据共享服务网”试点项目,并于 2005 年转入国家科技基础条件平台建设。该网的总体目标是整合集成分布在国内外数据中心群、高等院校、科研院所和野外监测台站及科学家个人手中历史的、现状的和未来的科学研究产生的数据资源,接收国家重大科研项目产生的数据成果及引进国际数据资源,加工、生产满足人地系统及地球系统各圈层相互关系研究的专题数据集,建立分布式地球系统科学前沿研究与全球变化研究数据支撑平台<sup>[5-7]</sup>。经过近 6 年的建设发展,目前地球系统科学数据共享网正在向长期运行服务阶段转变。

本文正是结合国家科技基础条件平台——地

收稿日期:2009-06; 修订日期:2009-09.

基金项目:国家科技基础条件平台建设项目(2005DKA32301);国家自然科学基金项目(40801180);资源与环境信息系统国家重点实验室自主研究项目。

作者简介:王卷乐(1976-),男,副研究员,主要研究方向是地学数据的集成与共享标准规范,遥感和 GIS 应用。

E-mail: wangjl@igsnr.ac.cn

球系统科学数据共享网的建设和发展过程,研究地球系统科学数据共享的标准规范体系的构建及其应用实践。

## 2 地球系统科学数据共享面临的标准规范问题分析

地球系统科学数据主要是来源于各类国家投入的地学领域的科研项目。此类数据不同于国家行业部门按照统一的制度、规范、标准长期采集和管理的科学数据,而是科学家根据研究需要,自主建立或整合的主要供本研究项目使用的综合性数据。

由于研究项目的数量众多,且每年都不断有新项目立项,持续积累的研究数据数量巨大,是重要的科技信息资源,受到国内外高度重视。2005年9月美国NSF发布了关于科学数据库的研究报告——《推动21世纪研究与教育的长期数字数据库》,该报告把科学数据分为三类,即研究型数据库、资源型数据库和参考型数据库<sup>[8]</sup>。

(1)研究型数据库是指某一个或者若干个固定的研究项目产生的数据集。这些数据集中的数据一般只经过有限的处理与管理,一般只为特定的研究群体服务。该类型数据库获得的资金资助较少,资助周期也较短,因而数据的标准化程度较低,数据的规模和覆盖的范围有限,数据可靠性稍差。

(2)参考型数据库旨在为大范围的科学与教育机构服务。这类数据库的典型特征是有一个大范围的、多样化的用户群体,包括来自于不同地域、不同学科、不同机构的科学家、学生、教育工作者。该类数据库遵照稳健和全面的数据标准为各类用户服务,经费预算通常很大,反映出其数据规模庞大和影响面广。它通常是由一个或多个机构提供长期的经费支持。典型的参考型数据库包括蛋白质数据库PDB、美国国立卫生研究院(NIH)的基因序列数据库GenBank等。

(3)资源型数据库是指那些服务于单一的科学与工程组织或者机构,其经费直接来源于相关的机构。通常该类数据库遵循一定的数据规范。地学、环境科学领域的资源型数据库隶属于某个数据中心(包括国家级的数据中心),能够对研究型、参考型数据库都提供支持。例如,美国国家航空航天局(NASA)的地球科学部的10个专业数据中心,多数

都具备了资源型数据库和参考型数据库的特征。

地球系统科学数据属于典型的研究型数据。这类数据普遍分散在各个研究项目中,主要目的是为项目本身提供服务,标准化程度普遍较低。如果没有体系化的标准规范环境,大量的科学研究项目数据很难得以交换、共享和再利用。这一难点主要体现在以下3点:①如何汇集这些分散在各个项目中的研究数据。这个问题是共享机制问题,即分散数据资源共享需要建立什么样的共享模式。②如何在统一的元数据标准框架下集成和共享多学科数据。③如何面对众多科学研究者的需求,提供便于用户使用的数据目录。

这也是本文分析的地球系统科学数据共享标准规范体系所要解决的核心问题。

## 3 地球系统科学数据共享标准规范体系的构建原则与定位

### 3.1 地球系统科学数据共享概念模式

地球系统科学数据共享的概念模式如图1、2所示。图1显示,基于项目A所产生的科学研究数据只提供给项目A的用户使用,并不直接服务于其他用户。这使得其他用户很难发现这些数据,即使其他用户能够找到并且访问这些数据,也因为缺乏数据共享环境,数据的分类体系、内容结构、概念语义、数据格式等方面信息不完整,很难理解和使用。同样,项目B所产生的数据也只提供给项目B的用户,其他用户也无法获得。这就导致大量的项目数据无法为更多人使用,没有发挥研究项目数据应有的科学价值。

反之,如果建立一个数据共享的标准规范环境,则可以按照统一的标准规范接纳项目A、项目B以及更多项目的数据。如图2所示。由于这些数据遵从同样的元数据标准、数据文档格式等标准规范,非常易于实现数据搜索、访问和获取,且便于用户理解。这些数据将随着用户的不断访问和使用,发挥国家投入科研项目数据的科学价值。

### 3.2 地球系统科学数据共享标准规范体系构建原则

地球系统科学数据共享标准规范体系的构建,遵从以下原则。

(1)标准规范体系内外协调原则。体系内部由于标准之间缺乏协调,会给数据的集成与综合造成很

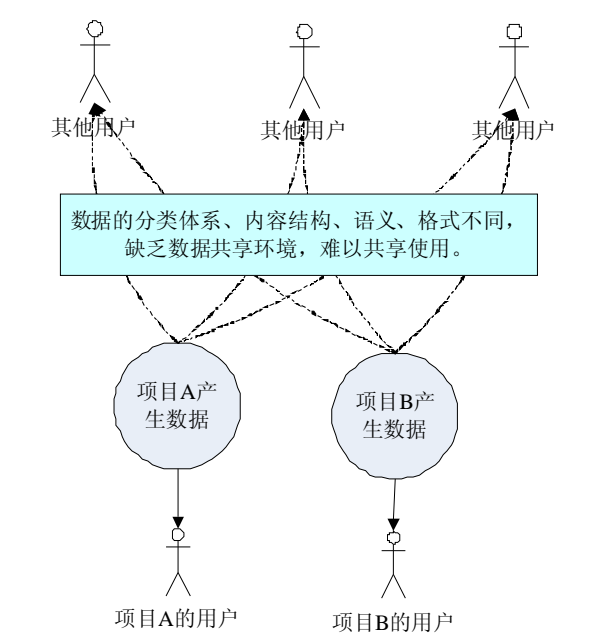


图 1 缺乏数据共享标准规范环境的项目数据

Fig.1 Research data without data sharing standards and specifications environment

大困难,有时甚至是不可能的<sup>[9-10]</sup>。因此,标准规范内部的各规范之间应该避免冲突,符合一致性规则。同时,标准规范体系的外部,应该与国际、国家现有的相关标准规范相协调。例如,地球系统科学数据共享平台作为国家科技基础条件平台的组成部分,其元数据标准要与国家科技基础条件平台的元数据标准(征求意见稿)、ISO TC211 的 ISO 19115 元数据标准相适应<sup>[11-12]</sup>。

(2)数据方便访问原则。地球系统科学数据共享的根本目的是为分散的科学研究项目数据提供一个数据汇集、检索、访问、浏览、获取的标准规范环境。因此,便于用户访问和共享使用是一条根本原则。在这一原则指导下,地球系统科学数据应呈现出良好的数据分类体系和完整的元数据描述信息。

(3)参考模型指导原则。地球系统科学数据共享涉及数据的汇集、集成、管理、分发等一系列过程。相应的,地球系统科学数据共享标准规范体系应该涵盖和体现这些过程的标准化技术要求。为此,要参照 ISO 标准参考模型的指导思想,建立起整体标准规范框架。

(4)标准与软件相结合原则。地球系统科学数据共享活动主要依赖于分布式的网络平台实施。尽可能地将标准的概念模型、逻辑模型转化成为基于分布式网络平台的物理模型,在网络平台功能中实现

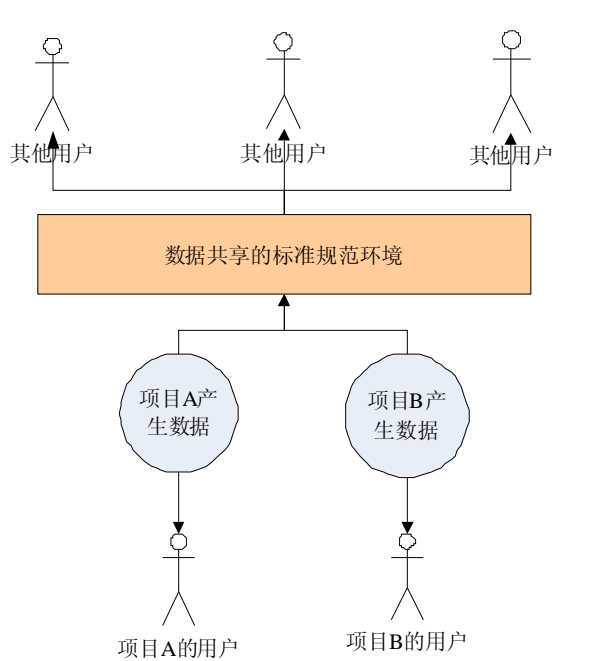


图 2 进入数据共享标准规范环境的项目数据

Fig.2 Research data within data sharing standards and specification environment

标准规范的有关规定,是标准规范应用、实践和推广的最好方法。

3.3 地球系统科学数据共享标准规范体系定位

尽管目前还没有专门为地球系统科学研究活动制定标准规范的专门机构,但是从数据共享的角度,国内外已有部分与此相关的标准化基础。相关的标准主要包括 ISO 19100 地理信息系列标准、OGC 相关地理信息互操作标准、国家科学数据共享工程标准规范体系、国家科技基础条件平台基础标准、国家电子政务相关标准、我国地理信息相关基础标准等。

由于地球系统科学数据共享网是国家科技基础条件平台中数据共享子平台的组成部分,因此地球系统科学数据共享标准规范体系将首先立足于国家科技基础条件平台的基础标准和国家科学数据共享工程的标准规范框架上。按照标准规范内外协调的基本原则,地球系统科学数据共享标准规范将充分参考、采用、改造国际和国家标准。其体系定位如图 3 所示。

需要说明的是:①由于地球系统科学数据共享网目前是以项目形式存在的,因此本网所制定的标准规范多数还是首先定位于满足项目数据管理和数据服务的需要。②本项目产生的章程、条例、规范和技术标准(草案)等均纳入地球系统科学数据共享



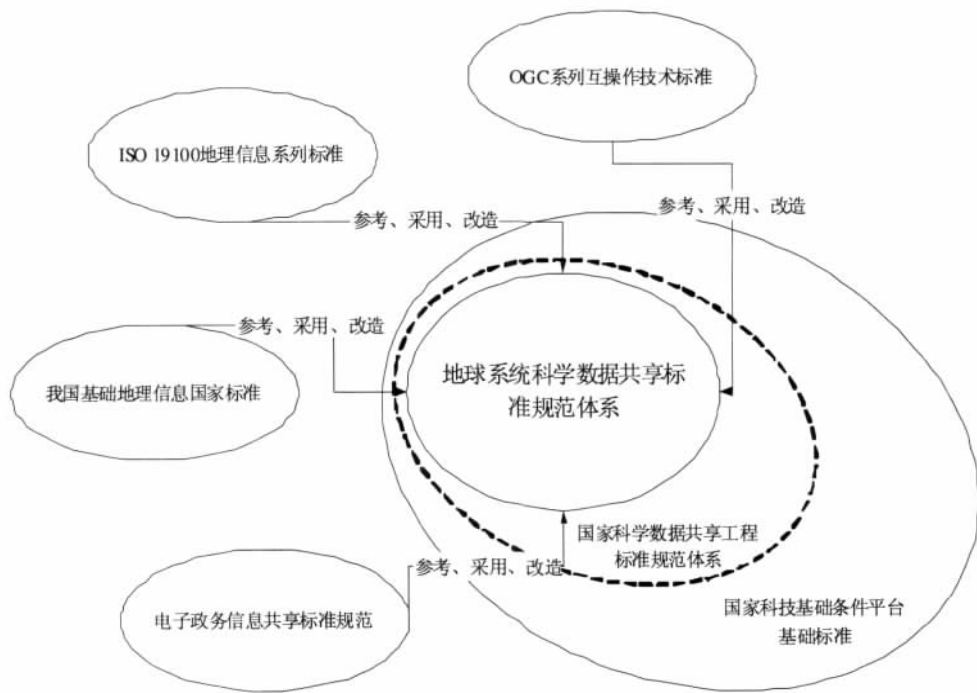


图3 地球系统科学数据共享标准规范体系定位

Fig.3 Orientation of standards and specifications system for earth system science data sharing

标准规范体系。待试用和完善后,再逐步梳理和上升到推荐性国家或行业标准层面。

## 4 地球系统科学数据共享标准规范体系构建

经过地球系统科学数据共享网项目近6年的探索,初步建立了地球系统科学数据共享标准规范体系框架。如图4所示。该标准规范体系目前包括18项标准规范,分属于机制条例类、数据管理类、平台开发类、用户服务类4大类。

### 4.1 机制条例类

机制条例类规范具体包括《地球系统科学数据共享联盟章程》、《地球系统科学数据共享平台章程》、《地球系统科学数据共享平台暂行管理办法》、《地球系统科学数据共享平台运行管理规范》、《地球系统科学数据共享平台数据共享条例》5项。本部分规范在外部联系上,与《国家重点基础研究发展计划资源环境领域项目数据汇交暂行办法》(本文中简称973数据汇交暂行办法)对数据资源的格式和内容要求上相一致。

机制条例类规定了标准规范体系的总体环境,

以《地球系统科学数据共享联盟章程》为核心,建立地球系统科学数据汇集和共享机制。这对应于开篇提出的第1个问题。章程中规定了联盟的宗旨是积极吸引从事地球系统科学相关研究的研究院所、高等院校、数据组织和科学家个人加入,倡导地球系统科学数据共建共享环境建设,促进数据的流通、使用和增值;加强面向国民经济建设和国家创新需求研究,为地球系统科学等基础与前沿科学研究和科技创新提供数据支撑;确立共享联盟的组织机构、业务范围、联盟成员的权力和义务等核心内容。

### 4.2 数据管理类

数据管理类标准规范具体包括《地球系统科学数据共享平台元数据标准》、《地球系统科学数据共享平台元数据编写规范》、《地球系统科学数据共享平台数据文档规范》、《地球系统科学数据共享平台数据备份规范》、《地球系统科学数据共享平台国际数据资源引进与服务规定》、《地球系统科学数据共享平台数据质量管理办法》、《地球系统科学数据共享平台矢量数据库建设规范》、《地球系统科学数据共享平台栅格数据库建库规范》、《地球系统科学数据共享平台属性数据库建库规范》等9项。本部分规范在外部联系上,与国家科技基础条件平台元数据标准(征求意见稿)相一致。其中元数据标准和数

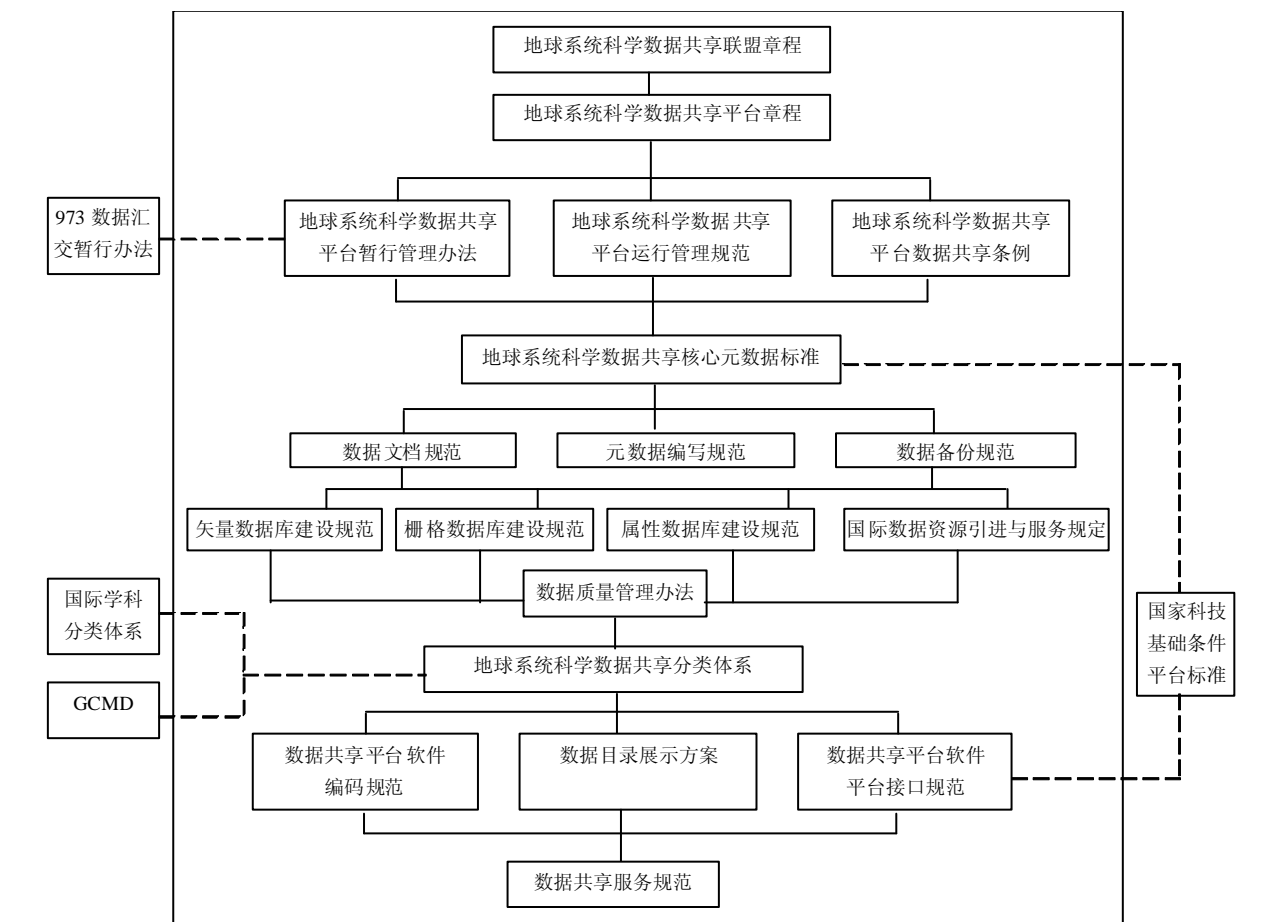


图 4 地球系统科学数据共享标准规范体系结构图

Fig.4 Standards and specification system structure for earth system science data sharing

据质量管理规范是数据管理的核心标准规范。

(1)地球系统科学数据共享核心元数据标准

数据共享中,元数据为各种形态的数字化信息单元和资源集合提供规范、普遍的描述方法和检索工具;元数据为分布的、由多种数字化资源有机构成的信息体系提供整合的工具与纽带<sup>[13]</sup>。离开元数据的各种数据信息将是一盘散沙,将无法提供有效的检索和处理。地球系统科学数据共享网正是以元数据为核心,规范化汇聚、集成和共享来自于联盟成员的多学科、多类、异构数据资源。《地球系统科学数据共享核心元数据》是数据管理活动的核心标准。当前的地球系统科学数据共享核心元数据标准包括 188 个元数据项,其中核心元数据项为 22 个。

从技术角度,地球系统科学数据共享核心元数据的作用体现在以下 5 点。①保证数据汇交的统一元数据格式。从入口上,定义了数据汇交的元数据标准,保证所有联盟成员汇集的数据遵从相同的核心元数据,但允许不同学科数据基于核心元数据进

行扩展。②保证基于元数据的全局搜索。基于核心元数据信息,可以实现分布式网络体系的全局数据搜索。例如以数据集名称、关键词、摘要、数据集作者等核心元数据项作为搜索条件,可以搜索到全网的所有相关数据。③数据描述和版权维护作用。元数据信息清晰地描述了数据集的产生背景、产生过程、主要内容、质量情况等说明信息,同时数据集的生产者信息、使用限制等也反映了数据的知识产权信息。④数据管理作用。通过元数据信息中的一些公共接口,例如元数据 ID、数据分类标识、数据保护期等信息,可以灵活实现对数据分类管理、数据保护管理、用户权限管理等。⑤促进数据交换和互操作。核心元数据保证所有数据描述信息在格式上的统一性和语义上的一致性,便于系统内部及与外部的交换与互操作。例如,地球系统科学数据共享网中的所有元数据都可以与国家科技基础条件平台门户进行互操作。

(2)地球系统科学数据质量管理规范

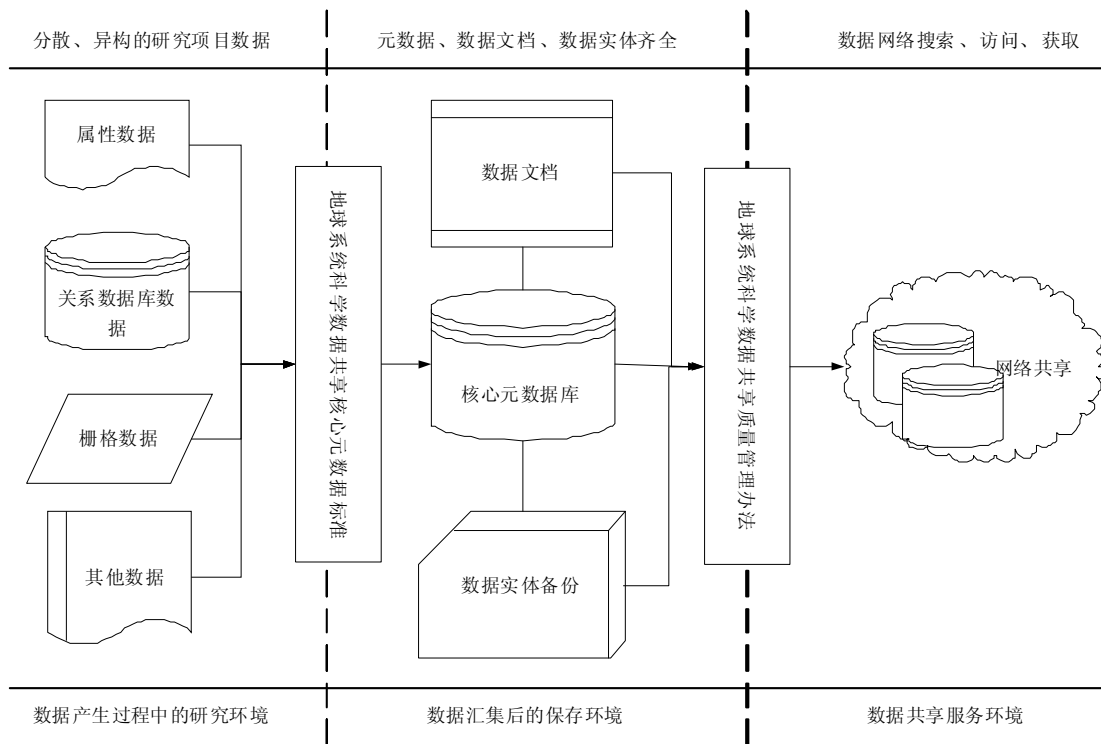


图 5 数据资源进入不同环境的准入条件

Fig.5 Admittance condition for different data sharing environment

以上分析可以看出,遵从统一的元数据标准是数据资源进入地球系统科学数据共享网存储环境的基本准入条件。但这并不意味着这些数据能够进入数据共享服务环境。与元数据标准作为数据汇集的准入条件相对应,《地球系统科学数据共享质量管理办法》则是所有数据能否进入数据共享服务环境的准入条件。二者的联系和关系如图 5 所示。

数据质量管理的重点在于 3 个方面。①数据完整性。所有的数据都必须有元数据、数据文档和数据集实体备份;②格式规范性。所有的元数据、数据文档和数据实体都必须遵循地球系统科学数据共享网规定的相关标准要求;③数据质量审核。所有的数据都必须经过两级审核,第一级为联盟核心成员机构对本机构提供的数据资源进行审核;第二级为地球系统科学数据共享网总中心数据管理员质量审核。通过以上 3 方面检查的数据,才能最终进入数据共享服务环境。

#### 4.3 平台开发类和用户服务类

平台开发类和用户服务类标准规范具体包括

《地球系统科学数据共享平台数据分类体系》、《地球系统科学数据共享平台软件平台编码规范》、《地球系统科学数据共享平台软件平台接口规范》、《地球系统科学数据共享平台数据共享服务规范》4 项。本部分规范在外部联系上,与我国学科分类与代码(GB/T13745-92)及国际全球变化主目录(GCMD, Global Change Mater Directory)\*相一致。

平台开发、用户服务类标准规范以《地球系统科学数据共享平台数据分类体系》及相应的数据目录展示技术方案为核心,指导平台软件的开发和便捷的用户服务。这对应于开篇提出的第 3 个问题。

地球系统科学数据分类体系规定了地球系统科学数据的时间尺度分类、空间尺度分类、数据目录分类,及相应的编码<sup>[14]</sup>。该分类体系是数据资源后台组织和前台展示的基础,也是软件平台开发界面方案的主要依据。地球系统科学数据分类体系贵在应用,其结构体系的优劣要通过数据共享应用进行评价。结构良好的数据分类体系便于用户检索、查询、访问和获取数据,提高数据共享服务的效率。

\* <http://gcmd.nasa.gov>

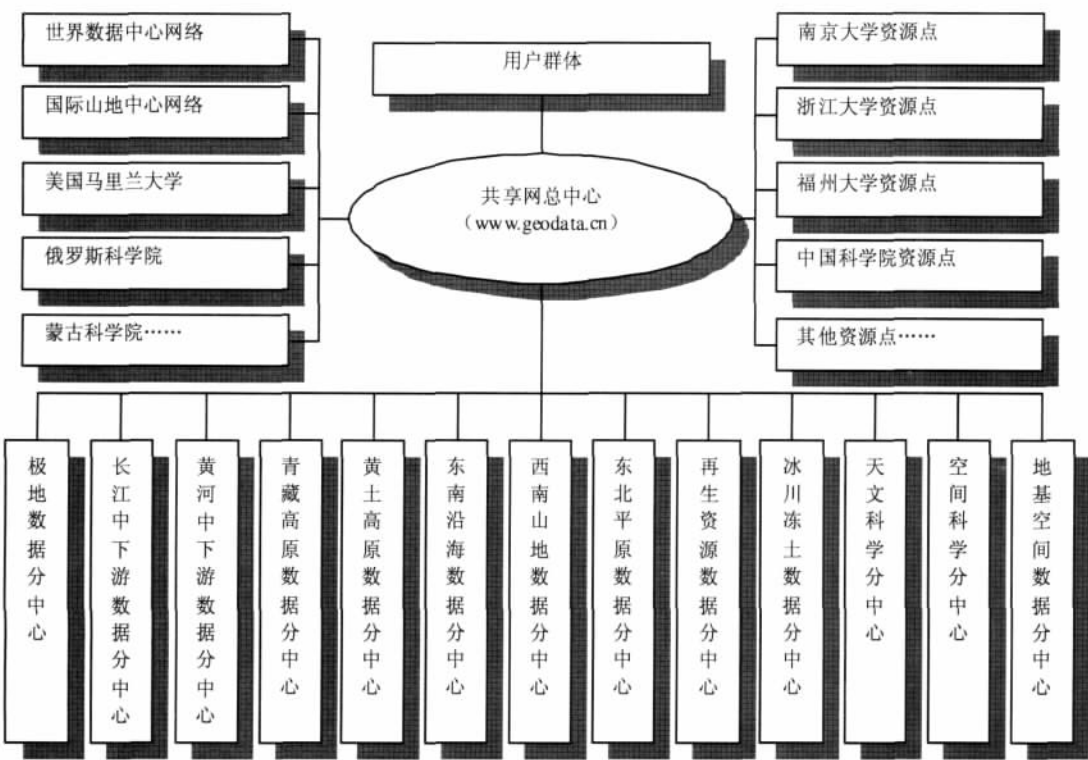


图 6 地球系统科学数据共享网分布式结构体系

Fig.6 Distributed architecture for data sharing network of earth system science

5 地球系统科学数据共享标准规范的贯彻与应用

按照本文提出的标准与软件相结合原则,地球系统科学数据共享网的 18 项有关条例、办法、规范和标准已经与地球系统科学数据共享网的分布式物理平台整合,实现了全网应用。地球系统科学数据共享网的分布式平台体系如图 6 所示。具体包括 1 个总中心和 13 个学科和区域分中心。

地球系统科学数据共享标准规范体系中的各项标准规范均发挥着自身的作用。下面仅就本文重点提到的 4 个核心标准规范,简述其应用效果如下。

(1)地球系统科学数据共享联盟章程。在《地球系统科学数据共享联盟章程》的总领下,配合多种分散数据资源整合集成的具体模式<sup>[15-16]</sup>,截止 2008 年底先后有 40 余家数据共享联盟成员加入,除了国内的研究机构和高校外,还包括美国马里兰大学、世界数据中心(WDC)<sup>[17]</sup>、国际山地中心(ICMOD)等国际机构。其中,核心联盟成员 22 个,主要包括中国科学院地学领域研究所、国内地学领域的一些研究性高校、研究机构和数据组织。

(2)地球系统科学数据共享核心元数据标准。基于 XML Schema 扩展技术和 Java 开发语言,地球系统科学数据共享元数据标准已经嵌入分布式的网络平台。目前参与共享的各联盟成员均按照统一的核心元数据标准(或者扩展后的元数据标准)汇集数据<sup>[18]</sup>。这一机制,保证了所有数据元数据信息在全网 13 个分中心内部可互访,并且与总中心实现同步收割。截止到目前,该网已经汇集了 1436 个数据集/元数据,数据总量超过 18TB。

(3)地球系统科学数据共享数据质量管理办法。数据质量管理办法对数据的完整性、规范性和数据质量审核要求均已嵌入软件平台。元数据格式不规范、缺少数据文档、缺少数据实体备份的数据无法通过数据审查,因而也无法进入数据共享服务环境。经过这一控制,目前地球系统科学数据共享网发布的 1436 个数据集,均具有数据集实体和完备的说明文档,便于用户使用。

(4)地球系统科学数据分类体系。按照地球系统科学数据分类体系及相应的数据目录展示方案,地球系统科学数据共享网开发了网络服务界面(<http://www.geodata.cn>),可为用户提供便捷的数据目录服务<sup>[19]</sup>。目前,网上展示的主分类体系包括:地表过程



与人地关系、典型区域、日地系统与空间环境、地球系统系统集成、国外数据资源、对地观测数据及产品 6 个大类,具体包括 28 个二级类。

除了以上应用以外,地球系统科学数据共享标准规范已经被国家科技计划项目数据汇交工作借鉴。当前正在实施的“国家 973 计划资源环境领域项目数据汇交”<sup>[20]</sup>工作参考了本网站的标准规范体系和多项具体标准规范。

## 6 讨论

地球系统科学数据共享是一项长期的事业,支撑其发展的标准规范环境也需不断完善。展望未来,地球系统科学数据共享标准规范还需要“向上”、“向下”两个方面发展。“向下”发展要深入挖掘地球系统科学数据共享标准规范的概念模型,要建立健全整个标准规范体系的参考模型,通过更多的抽象和概括总结地球系统科学数据共享的基本规律。“向上”发展则要在现有应用体系的基础上,深入分析地学数据在其生命周期过程中不同阶段的特点,加强数据产品类和服务类标准规范的研究和应用。例如,当前只是侧重在数据整合阶段的标准规范,对于数据集成、融合以及面向科学研究问题的数据同化等方面还缺少标准规范的支持。

另外,如何使侧重于科学研究数据集成与共享的标准规范成为国家标准还有很长的路要走。还需要与相关行业的标准化技术委员会(TC)相结合,与国家现有标准体系相融合。例如,加强与全国地理信息标准化技术委员会、全国科技平台标准化技术委员会的合作等。

致谢:感谢国家科技基础条件平台——地球系统科学数据共享网项目组同仁对本项工作的指导。感谢科技部基础司滕绵震先生对本项研究的指导。

## 参考文献

[1] 刘东生. 全球变化和可持续发展科学. 地学前缘, 2002, 9(1):1-9.

- [2] 周秀骥. 对地球系统科学的几点认识. 地球科学进展, 2004, 19(4):513-515.
- [3] 黄鼎成, 林海, 张志强. 地球系统科学发展战略研究. 北京: 气象出版社, 2005.
- [4] 中国科学院地学部地球科学发展战略研究组. 21 世纪中国地球科学发展战略报告. 北京: 科学出版社, 2009.
- [5] 孙九林, 施慧中. 科学数据管理与共享. 北京: 中国科学技术出版社, 2002.
- [6] 孙九林, 林海. 地球系统研究与科学数据. 北京: 科学出版社, 2009.
- [7] 孙九林, 施慧中. 中国地球系统科学数据共享服务网的构建. 中国基础科学, 2003, (1):76-82.
- [8] National Science Foundation. Long-lived Digital Data Collections: Enabling Research and Education in 21st Century, September, 2005.
- [9] 姜作勤, 马智民, 李东来, 等. 地质信息服务体系框架研究. 中国地质, 2007, 34(1): 173-178.
- [10] 姜作勤, 刘若梅, 姚艳敏, 等. 地理信息标准参考模型综述. 国土资源信息化, 2003, (3):11-18.
- [11] 李莉, 曾澜, 朱秀丽, 等. 电子政务—自然资源 and 地理空间信息库标准体系研究. 地理信息世界, 2006, (6):6-20.
- [12] 蒋景瞳, 刘若梅, 贾云鹏, 等. 国内外地理信息标准化现状与思考. 国土资源信息化, 2002, (4): 8-13.
- [13] 王卷乐. 科学数据交换中心元数据关键问题研究: 以地球系统科学数据共享网为例. 中国科学院地理科学与资源研究所博士学位论文, 2005.
- [14] 廖顺宝, 蒋林. 地球系统科学数据分类体系研究. 地球科学进展, 2005, 24(6):93-98.
- [15] 孙九林. 分散数据资源整合策略和模式研究. 中国科技资源导刊, 2008, (3): 6-11.
- [16] 刘润达, 诸云强. 科学数据共享关键问题探索. 地理科学进展, 2007, 26(5):118-126.
- [17] 王卷乐, 孙九林. 世界数据中心(WDC)回顾、变革与展望. 地球科学进展, 2009, 24(6): 612-620.
- [18] 冯敏, 诸云强, 王卷乐, 等. 分布式多标准地学元数据共享研究与实践. 地理与地理信息科学, 2007, 23(6):8-13.
- [19] 诸云强, 冯敏, 宋佳, 等. 基于 SOA 的地球系统科学数据共享平台架构设计与实现. 地球信息科学学报, 2009, 11(1):1-9.
- [20] 王卷乐, 杨雅萍, 诸云强, 等. “973”计划资源环境领域数据汇交进展与数据分析. 地球科学进展, 2009, 24(8): 947-953.



# Study on Scientific Data Sharing Standards and Specifications System for Earth System Science and its Application

WANG Juanle, SUN Jiulin

(Institute of Geographic Sciences and Natural Resources Research,  
State Key Lab of Resources and Environment Information System, CAS, Beijing 100101, China)

**Abstract:** Earth System Science (ESS) has developed into a new historic stage. Because it takes the whole earth system and its interactive actions among spheres, its development needs lots of multi-disciplinary, multi-sources, multi-type, integrated geosciences data resources. According to this requirement, Data Sharing Network of Earth System Science (DSNESS) was established at the beginning of China Scientific Data Sharing Program (SDSP) launched in 2003. It has been one of the long-term data sharing platforms in National Science & Technology Infrastructure (NSTI) since 2005. Data sharing in DSNESS needs standards and specifications environment. Based on the ESS data characteristics' analysis, 3 main problems needed to be studied at present, i.e., how to establish the ESS data sharing mechanism? How to integrate and share multi-disciplinary data? How to make data easily find and access for use? Basic data sharing concept models are studied firstly for these questions. According to four principles designed, ESS data sharing standards and specifications system and its relationship among related international and domestic standards systems are studied. Designed standards and specifications system includes four main classes, i.e., mechanism and rules class, data management standards and specification class, platform development specification class and data service specification class. A total of 18 standards, rules and specifications have been drawn up. Faced to the problems mentioned above, there are four key standards and specifications in the architecture, i.e., ESS data sharing union constitution, core metadata standard, data quality control measurement specification and data classification system. Through almost 6 years research and application, nowadays all of these 18 standards and specifications have been used successfully in the distributed data sharing network of ESS which include one general center and 13 sub-centers. In the near future, standards and specification environment of DSNESS will be further developed towards two directions, one is basic concept reference model, and the other is data fusion and assimilation standards and specifications.

**Key words:** earth system science; data sharing; standards and specifications; facility