

973 计划资源环境领域项目数据汇交实践与思考

王卷乐 孙九林 杨雅萍 宋佳 诸云强 朱华忠 乐夏芳 张金区

(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: 随着国内科学数据共享的推进, 我国科技计划项目的数据汇交逐渐得到重视和发展。本文基于实践, 对已经实施的 973 计划资源环境领域项目数据汇交的进展进行阶段回顾, 介绍了数据汇交在政策制定、汇交中心组建、组织实施策略、汇交环境建设、数据接收管理及共享服务方面的具体做法和进展。并在此基础上, 针对 973 计划资源环境领域项目数据汇交未来在其他领域的推广与汇交数据的深度开发利用进行了初步思考。

关键词: 数据汇交; 科技计划项目; 国家重点基础研究发展计划; 数据共享

中图分类号: G352

文献标识码: A

DOI: 10.3772/j.issn.1674-1544.2011.03.001

Data Archiving Practice and Consideration of National Basic Research Program of China (973 Program) in Resource and Environment Field

Wang Juanle, Sun Jiulin, Yang Yaping, Song Jia, Zhu Yunqiang, Zhu Huazhong, Le Xiafang, Zhang Jinqu

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

Abstract: With the development of scientific data sharing in china, research data archiving derived from national funding scientific research program are valued and developed. Data archiving in national basic research program of china (973 Program) in resource and environment field was launched in 2008. Based on its practice, the paper summarized its main progress in 6 aspects, i.e., data archiving policy, data archiving center building, data archiving strategy, data archiving environment construction, data collection and management, and data sharing. For future development of scientific data archiving, 2 aspects are stressed, i.e., data archiving model, data sharing and using.

Keywords: data archiving, scientific research program, national basic research program of china, data sharing

1 引言

2008年, 科技部在国家重点基础研究发展计划(973计划)资源环境领域开展数据共享试点, 并依托于资源与环境信息系统国家重点实验室, 建立973计划资源环境领域项目数据汇交管理中心, 开展我国科技计划项目数据汇交的实践^[1-2]。

973计划资源环境领域项目数据汇交的内容是项目在研究过程中新产生的各类数据, 具体包括新增原始数据、研究分析数据以及应用软件等^[3]。新增原始数据指项目产生的观测数据、监测数据、探测数据、试验数据、实验数据、调查数据、考察数据等。研究分析数据指对原始数据进行处理和加工后形成的数据。应用软件指项目支持开发

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研究项目“基于实践的科学数据资源管理基本理论与方法问题探讨”(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。

的数据处理、加工和分析软件及其使用说明。973计划资源环境领域项目数据汇交工作自2008年3月启动以来,已经有2年多的进展。本文将基于这一数据汇交试点的实践,对其进展进行回顾和分析,以期形成共识,为我国科技计划项目数据汇交的推广提供借鉴,促进科技计划项目产生数据资源的共享效益。

2 数据汇交实践

973计划资源环境领域项目数据汇交在实施过程中,主要包括6个环节,即数据汇交管理办法的制定、数据汇交中心的组建、数据汇交实施策略、数据汇交环境建设、数据接收和管理、数据共享服务。

2.1 数据汇交管理办法与数据汇交中心

数据汇交管理办法主要回答3个方面的疑问:(1)数据交到什么地方去?(2)交什么?怎么交?(3)数据汇交后的权益如何保护?2008年初科技部制定了《973计划资源环境领域项目数据汇交暂行办法》(以下简称《办法》),内容包括数据汇交的组织管理、汇交内容、数据汇交计划、数据汇交流程、数据管理、权益保护、监督与信用管理等,系统回答了需要解决的这三大疑问。

针对疑问一,数据交到什么地方去?《办法》规定要建立数据汇交管理机构,即数据汇交中心。该中心将在科技部的领导下,具体负责数据的接收、归档、管理和共享服务。

针对疑问二,交什么?如何交?《办法》规定汇交的数据内容包括项目新增原始数据、研究分析数据以及应用软件等,汇交的数据集应有元数据和数据说明。在汇交方式上,要求汇交的项目数据以数字化形式提交,数据标准按数据汇交中心规定的标准执行。《办法》同时确立了数据汇交的工作流程,具体包括4个阶段,即数据汇交计划制定阶段、数据汇交准备阶段、数据汇交阶段和数据管理与共享服务阶段。

针对疑问三,汇交后的数据权益如何保护?《办法》规定对汇交数据进行分类、分级存储和管理,确保项目数据的物理安全,不得擅自修改和删除汇交的项目数据。项目数据可设置保护期,

其中保护期内的项目数据仅供项目和课题承担单位及其授权范围内的用户访问和使用;保护期结束后,数据汇交中心以在线、离线等方式分期、分批向全社会提供数据共享服务。用户利用汇交数据产生的研究成果须注明数据源。

2009年科技部依托于资源与环境信息系统国家重点实验室成立“973计划资源环境领域项目数据汇交管理中心”。中国科学院地理科学与资源研究所则承担数据汇交的具体工作。数据汇交中心采用理事会领导下的主任负责制。理事会由973计划资源环境领域的咨询和项目专家组成员组成。理事会下的协调办公室设在科技部基础司基础性工作与综合处。数据汇交中心设置综合办公室、标准规范研究组、数据接收管理组、数据平台开发组、数据共享服务组,分工开展汇交管理工作。

2.2 数据汇交实施策略

在开展数据汇交工作之初,数据汇交中心确立了两条基本的数据汇交实施策略。

(1)“分阶段、分类型”的数据汇交实施策略

分阶段是指按照《办法》规定,数据汇交工作将按数据汇交计划制定、数据汇交准备、数据汇交和数据共享服务4个阶段有序开展。分类型是指在2008年开展数据汇交工作之初,已经实施的973计划资源环境领域项目处于不同的进展状态,可分为三大类型,即新启动项目、在研项目(包括中期进展和即将结题项目)、已结题项目。针对不同类型的项目,根据其当前状态对应不同的阶段,直接开展相应的汇交工作。不同类型的项目与数据汇交阶段的对应关系如图1所示。

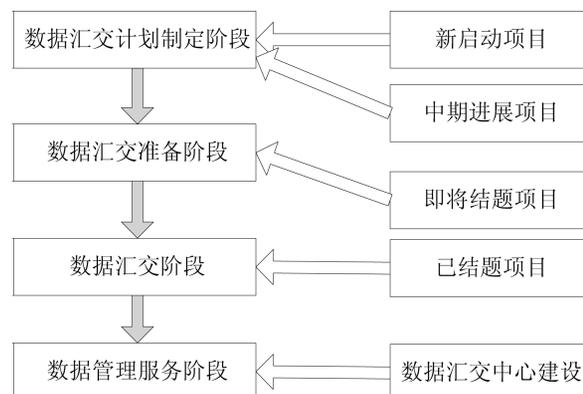


图1 分阶段、分类型数据汇交对应关系图

(2)“先服务、后汇交”的数据汇交实施策略

数据汇交中心秉承为各 973 计划资源环境领域项目提供服务的核心理念。在汇交工作开展之前，数据汇交中心首先征求各项目对于项目研究和执行中的数据需求，基于中心此前承担的国家科技基础条件平台——地球系统科学数据共享网的已有数据基础为这些项目提供数据共享服务。这包括两个方式：一是如果本中心有相关的数据则直接提供数据共享服务或数据资源加工定制服务；二是如果本中心没有相关数据，则尽量为其提供资料的来源渠道或资源导航信息。为了具体落实数据汇交的协调和共享服务工作，数据汇交中心还专门设立 4 个工作联络组，分别有专人对应项目进行沟通和联系，保障汇交工作的有序开展。

2.3 数据汇交环境建设

数据汇交工作伊始，数据汇交中心就通过自身建设落实《办法》规定的各类措施的执行，尽可能地为各项目数据汇交提供便利的条件。这包括标准规范制定、汇交软件工具研发、汇交管理与共享服务平台研发、数据汇交存储环境建设 4 个部分。

(1)数据汇交标准规范：973 计划资源环境领域产生的数据资源内容复杂多样，类型不一。为了保证汇交数据的一致性，《办法》规定了数据汇交要遵从一致的标准规范。为此，数据汇交中心制定了若干具体的规范，并提前下发到各项目开展应用。已经制定的标准规范包括数据汇交计划格式、元数据标准、数据文档格式、数据质量

检查规范、数据光盘刻录规范、首席科学家审查报告格式、数据汇交工作方案格式等。这些已制定的规范与国家已经实施的科学数据共享工程和国家科技基础条件平台中制定的相关标准规范保持一致。

(2)汇交软件工具：《办法》规定了所有汇交数据都必须同时提供数据的元数据和数据文档信息。为了保证各项目都按统一的标准采集元数据信息，数据汇交中心开发了元数据汇交工具。这包括离线采集工具和在线采集工具两类。离线采集工具是单机版的录入系统，可在 windows 操作系统下方便使用，具有数据的录入、修改、模板管理等功能，其界面如图 2 所示。在线采集工具则是基于 B/S 的网站系统，具有数据在线录入、修改和管理功能，其界面如图 3 所示。

(3)数据汇交管理与共享平台：面向数据汇交管理与共享服务两方面的功能需求，开发了数据汇交管理与共享平台。其数据管理功能包括汇交数据的进度管理、内容管理、统计管理等功能，界面如图 4 所示；针对数据用户的需求，开发了数据的元数据查询和目录服务功能，界面如图 5 所示。

(4)建立数据汇交的备份存储环境：为了确保数据的物理安全，数据汇交中心构建了 100TB 存储容量的磁盘阵列存储系统，建立统一编目的光盘存档系统，实现了双备份的独立存储环境。

2.4 数据接收、管理与共享服务

截至目前，973 计划资源环境领域有 66 个项目（1998-2010）参加数据汇交，其中已结题



图 2 离线元数据软件采集工具界面



图 3 在线元数据采集工具界面



图4 数据汇交进度管理界面



图5 汇交数据查询界面

项目 39 个 (1998-2006), 在研项目 27 个 (2007-2010)。已结题项目已全部完成数据汇交, 汇交数据集 1800 多个, 数据量约 2TB。汇交的数据包括属性数据、矢量数据、栅格数据、文本数据等多种数据类型。在研项目全部完成数据汇交计划的编制。数据共享服务工作按照《办法》的规定, 与数据管理同步展开。当前提供的数据服务包括 4 种类型: (1) 数据汇交共享服务网站的数据查询、元数据浏览和信息服务; (2) 数据实体的离线申请服务; (3) 部分整编汇交数据的内容访问及再分析服务; (4) 提供数据汇交简报、标准规范及工作资料下载服务。

3 数据汇交的初步思考

973 计划资源环境领域项目数据汇交工作还只是开始, 面向未来发展, 还有很多方面需要不断完善和提高。结合实践, 提出两个方面的思考: 一是如何总结和形成科技计划项目数据汇交的模式, 供相关领域推广和开展数据汇交工作时借鉴参考; 二是如何对已汇交的数据资源进行加工分析和再利用, 实现数据汇交工作的科技支撑效益最大化。这两个方面的问题需要在实践的基础上深入分析。笔者在此提出几点看法。

3.1 数据汇交工作模式

数据汇交的工作模式包括政策的制定、政策的执行和政策的效果落实 3 个部分。这其中涉及 3 个利益主体的关系, 即科技计划项目管理部门、项目负责人和数据生产者以及数据汇交中心。只有在其工作模式上, 协调好 3 个方面的关系, 才能保证汇交工作的顺利开展。基于 973 计

划资源环境领域项目数据汇交的经验, 初步提出的工作模式如图 6 所示。即在数据汇交工作开展之前, 首先要在政策和制度层面制定数据汇交管理的办法和条例, 并由权威机构发布。在此基础上, 通过遴选已有机构或组建新机构的方式建立数据汇交中心。在管理办法 (条例) 的指导下, 同步开展数据汇交中心建设和科技计划项目的数据准备和汇交工作。最终要在数据接收、管理的基础上开展存档和共享服务, 实现科技计划项目产出数据的价值和效益。在数据共享的过程中, 数据生产者、数据用户以及存储数据资源的数据汇交中心之间会产生很多关联, 同时也会发现许多问题。这将进一步反馈到数据汇交中心的建设和数据汇交工作的流程, 进而反馈到数据汇交管理办法 (条例), 以确定其是否合理, 是否需要完善。这种正向的工作流程与反向的反馈流程将不断交互、螺旋式发展, 逐渐实现科技计划项目数据汇交效益的最大化。

3.2 数据资源利用

科技计划项目汇交数据的利用是汇交工作的最终目的之一, 包括数据的重复使用和加工分析应用。其一, 重复使用是共享的基本方式, 其关键在于如何确保数据的质量和可靠性, 这是数据共享的基本问题。其二, 数据的加工分析应用, 则需要有明确的科学目标导向。由于科技计划项目汇交数据本身是分散的且不完全以科学目标为导向, 因此如何组织和开发利用这些数据资源有很大挑战性。就目前的认识来看, 将科技计划项目数据汇交与某些有导向性的国家重大战略、重大科技计划相结合, 有益于数据的开发利用。

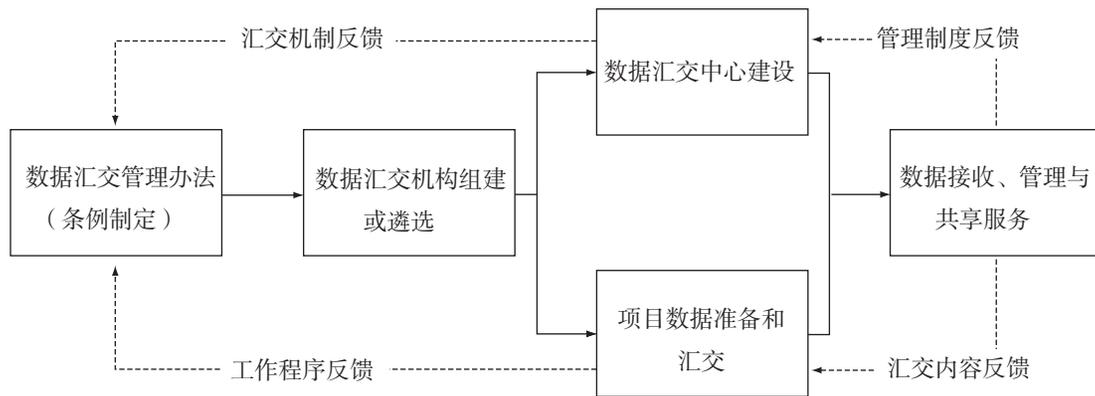


图6 科技计划项目数据汇交管理概念模式

针对数据资源的利用, 973 计划资源环境领域项目数据汇交管理中心已经开展了扎实的数据共享工作, 包括: (1) 基于数据汇交管理中心承担的“国家科技基础条件平台——地球系统科学数据共享网”为 973 项目 / 课题提供无偿数据共享服务; (2) 按照《办法》要求, 所有汇交数据的元数据信息将在接收数据 1 个月内, 对外开放, 目前接收到的 1833 条元数据全部上网 (www.973geodata.cn), 对外开放, 提供检索和离线数据申请; (3) 按照数据保护期要求, 对于可公开的数据资料, 以离线方式提供数据实体服务, 这要求申请数据的用户提供遵守数据知识产权的数据使用声明; (4) 已经开始并将逐步增加开放资源的在线下载服务, 进一步加强数据资源共享的力度。以上这些工作为汇交数据资源的开发利用创造了基本的共享环境, 随着越来越多数据资源的共享, 必将在资源开发利用模式上会有更多的认识。

4 结语

科技计划项目数据汇交工作是科技资源管理中的一个热点和难点。本文结合实践, 对 973 计划资源环境领域项目数据汇交在政策制定、汇交中心组建、组织实施策略、汇交环境建设、数据接收管理及共享服务方面的具体做法和进展进行了初步分析。这些工作已经在 973 计划资源环境领域的的数据资源积累、项目研究数据的规范化、

数据共享利用等方面发挥了初步效益。希望随着各方面工作的推进, 不断完善提高, 尤其在数据汇交模式提炼、数据资源深度开发利用等方面进一步提高认识, 充分发挥数据汇交的价值和效益。

参考文献

- [1] Lin Hai, Wang Juanle. Data Archiving Work Was Launched in National Basic Research Program (973) in Resource and Environment Field[J]. *Advances in Earth Science*, 2008, 23(8): 895-896. (in Chinese)
〔林海, 王卷乐. 国家重点基础研究发展计划 (973) 资源环境领域项目数据汇交工作正式启动 [J]. *地球科学进展*, 2008, 23(8): 895-896.〕
- [2] Wang Juanle, Yang Yaping, Zhu Yunqiang, et al. Data Archiving Progress and Data Types Analysis of National Basic Research Program of China (973 Program) on in Resource and Environment Field[J]. *Advances in Earth Science*, 2009, 24(8): 947-953. (in Chinese)
〔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 [J]. *地球科学进展*, 2009, 24(8): 947-953.〕
- [3] Ministry of Science and Technology of the People's Republic of China. Notification of National Basic Research Program Data Archiving in Resource and Environment Field[S], 2008, 3. (in Chinese)
〔科技部. 关于开展国家重点基础研究发展计划资源环境领域项目数据汇交工作的通知 (国科发基 [2008]142 号) [S]. 2008, 3.〕