·科学论坛·

# 国家自然科学基金委员会共享航次调查数据汇交体系构建

宋转玲[1,2*] 刘海行[2] 葛人峰[2] 李新放[2] 宋庆磊[2] 丁 明[2]

(1. 中国海洋大学，青岛 266100；2. 国家海洋局第一海洋研究所，青岛 266061)

[摘 要] 针对国家自然科学基金委员会共享航次调查数据汇交与共享的迫切需求，结合数据汇交工作中遇到的实际问题，提出构建面向科学研究服务的基金委共享航次调查数据汇交体系，包括数据汇交体系的总体框架，设计数据汇交内容、时间节点及以人工与网络结合汇交方式的汇交方案等。并从数据汇交的工作流程清晰化、明确数据生成者和数据管理者的责任、方便两者之间有效互动、提高数据汇交工作效率等诸方面阐述了构建的要素。

[关键词] 共享航次，汇交体系，元数据

国家自然科学基金委员会(简称“基金委”)于2009年开始试点推行共享航次计划(简称“共享航次”)，为必须进行海上考察的科学基金项目提供船舶运行时间，以确保自然科学基金资助项目海上考察任务的实施。并以此为契机，探索海上观测平台共享机制，加强海洋现场数据的长期积累，促进海洋科学研究多学科交叉与融合，以及科学家之间的交流与合作，这有助于推动和提升我国对海洋科学领域一些重大科学问题的研究。该项目启动以来至今已资助31个航次(截止到2014年底)，依托项目涉及基金委6个学部，几乎覆盖所有海洋科学基金项目类型。目前积累的TB级科学数据与资料大部分分散保存在于研究调查单位的相关各课题组。而海洋调查因成本大、难度高、耗时多，科研人员对这些散置数据的共享需求迫切。

数据是数据共享的主体，数据汇交是实现数据共享的前提和基础，构建完备的数据汇交体系是确保数据共享服务可持续深入拓展的根本。如何把这些弥足珍贵、纷繁杂乱的资料有效收集、管理，使调查数据得到有效的汇集与共享，并进一步实现科研价值最大化，是促使共享航次调查工作持续发展的基础工作。为了加强基金委共享航次调查数据的管理，规范数据汇交工作，基金委于2010年立项“船时共享航次调查资料管理模式研究”，由基金委青岛海洋科学数据共享服务中心(简称“数据服务中心”)负责实施开展共享航次科学调查数据汇交工作。共享航次数据的收集和管理对集成分散的数据资源、拓宽数据资源的应用范围、提高数据管理与共享航次管理水平具有重要的应用性价值和意义。

## 1 共享航次海洋科学数据共享汇交体系

共享航次海洋科学考察是由多航次、多区域、多单位协作参与的综合性、延续性的科研活动。所获数据跨时长、涉及学科广、类型杂、课题多、来源散，数据资源组织比较困难，而航次搭载的课题之间、学科之间、单位之间对数据的共享需求强烈，因此构建、改进、完备数据汇交体系是共享航次项目持续实施过程中亟待解决的问题。

在数据知识产权的所属关联不被更改及严格执行有关保密规定的前提下，搭载共享航次的科研项目负责人将科学调查过程中产生的所有数据资源汇交到数据服务中心，数据服务中心接收管理审核验收通过的数据，并整理生成数据集，形成共享航次调查数据汇交体系。

* Email：songzhuanling@fio.org.cn

本文于2014年7月16日收到。

共享航次数据汇交工作自 2010 年启动以来，了解到初期进展缓慢主要原因有两点：(1) 数据共享主观意识弱。一些数据持有者过高看重信息的价值，且因担心数据安全问题，将原本可以公开的信息封存自闭，形成信息资源“分割拥有、垄断使用”局面。(2) 数据汇交规则不明确。因为共享航次数据格式各异，科研人员不了解数据格式统一处理规则和数据汇交方法，主观臆想加大数据整理难度，阻碍了数据提交的积极性。

针对这些问题，数据服务中心根据共享航次调查科学数据的特点，采用基于元数据(metadata)的数据汇交模式。该模式在多学科、多类型、多格式的繁杂数据汇交管理上被普遍认同，并有广泛的应用与推广。

元数据也称中介数据、描述数据，是“关于数据的数据”(data about data)，是描述数据特征的数据，也是关于数据结构的数据。在数据仓库中，元数据被定义为描述数据特征及其环境的数据。它主要有以下 3 方面的作用：

(1) 描述功能：数据提供者利用元数据详细、全面的描述数据资源，阐明数据的数据量大小、分布空间、采集时间、采集手段、数据格式等基本特征，严格准确地区分数据资源。解决数据资源“是什么”的问题；

(2) 发现功能：数据使用者利用元数据确认和检索所需要的数据资源。解决数据对象“在哪里”的问题；

(3) 管理功能：数据管理者利用元数据组织数据信息对象，建立各数据信息对象之间的关系，提供数据资源的存储、管理和使用等方面的信息。解决数据对象“如何存储、如何发布、如何获取”的问题。

基于元数据的数据汇交模式已成为目前多学科数据汇交共享中的主要模式，有效化解数据汇交共享中各方利益冲突的矛盾。同时解决了多学科、多类型数据难以设计出统一标准格式，数据繁杂难以管理的问题。工作实践证明该模式在共享航次数据汇交工作中得到参与航次科学家的共同认可。

## 2 元数据标准格式及内容设计

对文件类型数据而言，数据集所包含的数据实体内容、数目、属性字段是因科研项目不同而相异，为能准确地说明数据实体便于检索，需要对于数据实体的产生条件、发展过程、属性字段等详细说明，即数据集的元数据。

虽然共享航次调查数据具有多学科、多类型、多格式等特点，但仍有共有属性，如数据所属项目、数据所属学科、数据采集时间、数据采集区域、数据采集单位、保管存放地点、数据联系人、数据内容描述、数据保护期、数据质量评价等。这些共有属性字段可设为数据的一级元数据，也称为数据的核心元数据。

考虑到数据服务中心需要反馈数据汇交回执，记录汇交事件。为简化汇交流程，提高汇交工作效率，在元数据部分设计 3 个表格包括共享航次搭载项目清单、拟采集数据信息表及数据汇交回执由航次首席和搭载项目负责人填写提交。

科研人员对数据共享的要求是便捷地获得所需的具有科研价值的调查数据或分析数据，不仅限于元数据。数据服务中心建立元数据库的同时，按照既定规则分类整合已汇交的数据形成实体调查数据集。

在共享航次数据汇交体系结构中，需要汇交的数据有元数据、实测数据、航次报告，构成数据的一个全面的描述体系。具体汇交内容如图 1 所示。

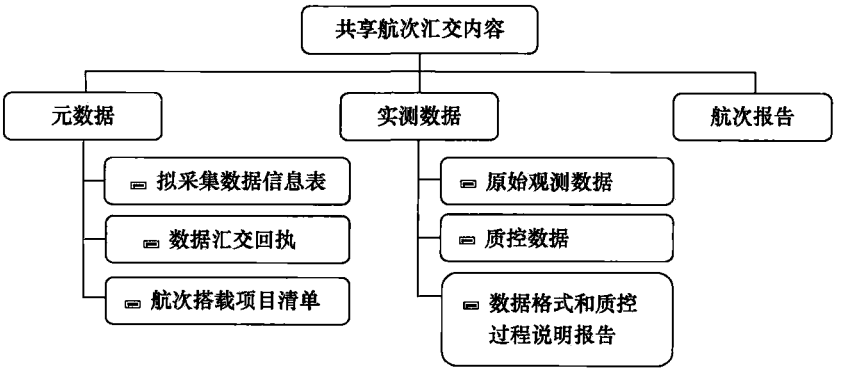


**图 1 数据汇交内容**

其中元数据包括的三个表格和航次报告按照 www. nsfcodc. cn 首页发布的文档格式编写提交。调查数据实体按照自由格式整理、提交。主要包括：

(1) 实测原始数据：如 GPS 数据、气象数据、ADCP 原始记录数据、CTD 原始记录的二进制码及仪器标定文件(hex、con、bl、hdr 等)；

(2) 质控数据：数据采集单位后处理的科学研究数据。如 CTD 依据规范程序处理得到的 1db 数据和采水层数据(cnv、ros)、采水层营养盐测试数据、柱状样测试数据等；

(3) 数据报告：包括数据质控方法介绍、数据分布区域、数据可视化图等。

## 3　数据管理模式与质量控制

### 3.1　数据管理模式

数据管理是共享航次数据汇交体系重要组成部分。目前采用集中式管理的方式完成数据存储的管理。集中式管理要求数据生产者和数据提交者把元数据和实测数据都提交到数据服务中心，数据服务中心负责元数据对外发布与数据共享服务。这种管理模式有利于提高数据分发效率、保障数据安全存储、整体规划。

为确保数据安全，防止数据零丢失。根据实际需要，本体系的数据容灾方案暂时采用级别较低的本地硬盘和光盘库冷备份方式。硬盘备份方法包括

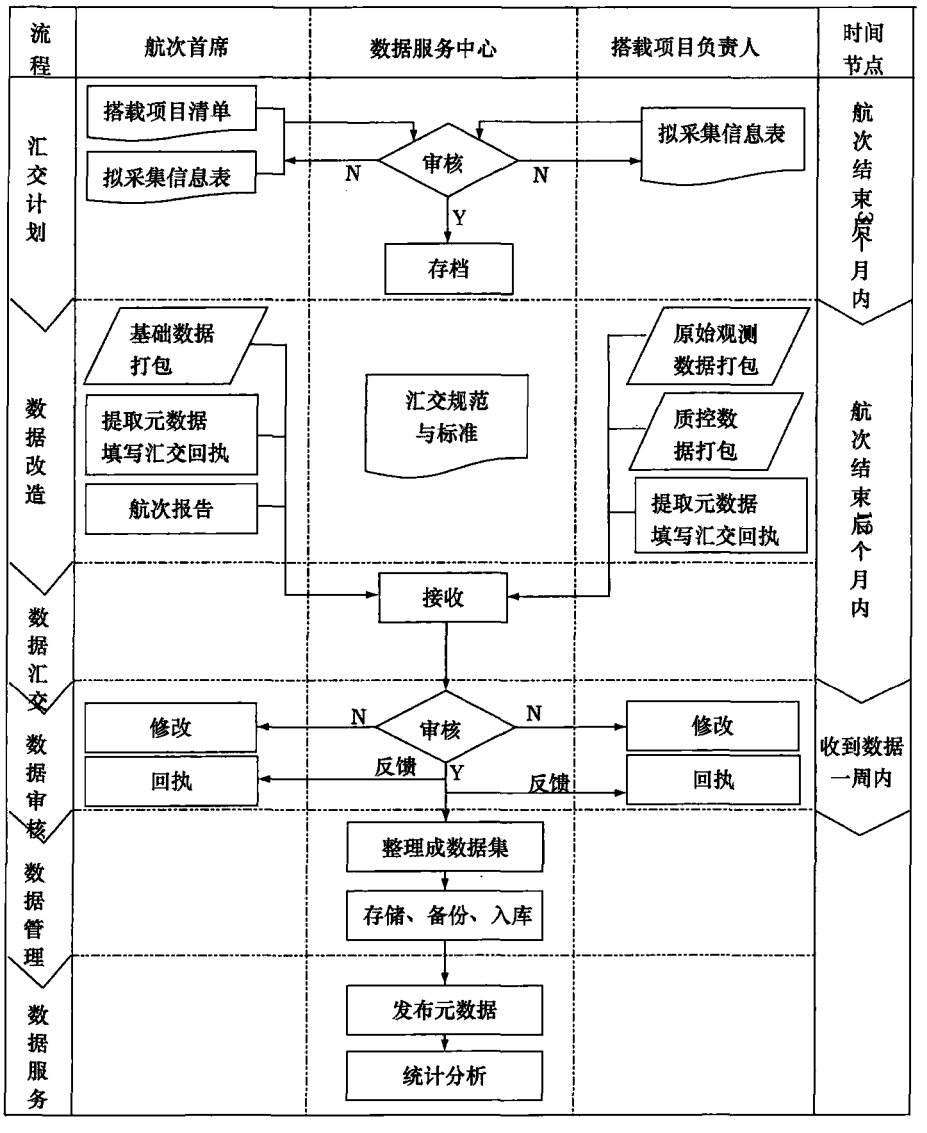


图 2　数据汇交流程

定期增量备份与差异备份，光盘增量备份后参考《管理规范和电子文件归档与电子档案管理办法 GB/T18894—2002》与《青岛市电子文件归档》实行严格的电子档案管理。

由于共享航次数据共享模式处于调研阶段，数据服务中心暂不提供数据共享服务。但航次内部可以共享数据，数据提供者可以充分优先使用数据，数据需求方如有数据分发需求，可通过元数据了解到数据存放信息联系数据提供者获取，以保护数据生产者的知识产权。

### 3.2 数据质量控制

数据质量是关系到数据能否有效应用的根本。在共享航次汇交的数据整理过程以及用户共享申请的数据内容中发现提交基础性最强、共享需求度最旺盛的必备环境数据质量良莠不齐。如 CTD 数据因各航次使用的仪器型号各异，采样频率和仪器精度也有很大的差别，且提交的 CTD 数据内容不一，有的仅提交了机器码格式的原始数据，无法看到可读数据；也有的仅提交了成果数据，没有原始数据，不能溯源查询；当然也有提交比较全面的，既有原始数据，也有处理过的成果数据，并附加了现场航次报告，便于了解数据内容及其生产过程。

为了保障数据的安全性、准确性、完备性，数据服务中心对需求程度高的基础数据依据数据质量控制方法给予质量审核与评价管理，并反馈给数据提供方数据质量评价报告。随着汇交工作的进展，质量控制也将根据不同专业建立数据质量准则，逐步扩大质控覆盖面。

### 3.3 数据汇交流程

为了确保数据汇交工作的顺利进展，数据服务中心引入流程管理思想用来规范数据汇交与管理。明确汇交步骤与双方责任，方便数据提交者清晰地了解汇交环节、内容和过程，保质保量汇交到数据服务中心。数据汇交流程包括数据的改造、数据的审核、数据的管理，具体工作流程如图 2 所示。

数据提交负责人按照汇交方案和数据汇交的格式标准进行各自专业领域的数据加工、整理，根据时间节点进行汇交。数据服务中心根据数据验收质量控制标准及相关管理办法验收上交的数据，对于合格的数据和数据集提供数据汇交证明，以确保数据的知识产权和使用范围。

## 4 结语

共享航次调查数据汇交工作实施过程中从最初的基于元数据的资料汇交模式逐步扩展到元数据和实体数据相结合的汇交管理模式，拟定了数据汇交体系的总体框架，设计了数据汇交内容、时间节点及以人工与网络结合汇交方式的汇交方案。在数据汇交工作初期，多采用人工汇交。至今数据服务中心已收集到 284GB 共享航次调查数据，搭载共享航次的科学家逐步认可数据汇交工作。

为了保证数据传输便捷，本课题组目前正在研发基于网络的共享航次数据汇交系统，科学规划数据汇交流程，以保证汇交数据的统一性、全面性和便捷性，为数据汇交和共享提供自动化技术平台。另外共享航次调查数据汇交体系尚处于试运行阶段，汇交方案在具体实施过程中根据专家意见和实际情况不断调整，力争切实可行，建议建立科学、有效的数据汇交管理机制，为后期推进共享航次科学数据共享打下基础。

**致谢** 本文工作得到国家自然科学基金资助（资助号：41306094）。

## 参 考 文 献

[1] 王亮绪，吴立宗，李红星，等. 面向黑河流域生态水文过程集成研究的科学数据汇交与管理. 遥感技术与应用，2013，28(3)：362—369.

[2] 范玉. 成果地质资料电子文档的汇交与保管. 云南地质，2005，24(2)：207—211.

[3] 刘青荣. 谈高校档案管理分类编号工作. 合肥教育学院学报，2000，17(4)：69—70.

[4] 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.

[5] 张文君. 澳大利亚地质资料汇交管理制度及其启示. 兰台世界，2010，(4)：14.

[6] 姜竹君. 成果地质资料汇交中存在的问题及建议. 甘肃地质，2013，22(3)：85—87.

[7] 张志娜. 对成果地质资料汇交的几点看法. 西部探矿工程，2013，(3)：188—190.

[8] 耿庆斋，安波，朱星明. 基于元数据的水利科学数据汇交体系研究. 水利水电技术，2009，40(5)：81—85.

[9] 赵瑞雪. 农业科学数据共享中数据汇交与管理研究. 科技管理研究，2009，(8)：284—286.

[10] 林丹红，钟伶. 中医药实验性研究课题科学数据汇交探讨. 科技管理研究，2006，(11)：125—127.

## Constructing Survey Data Collection System for NSFC Sharing Cruise

Song Zhuanling[1,2]　Liu Haixing[2]　Ge Refeng[2]　Li Xinfang[2]　Song Qinglei[2]　Ding Ming[2]
(1. *Ocean University of China*, *Qingdao* 266100; 2. *The First Institute of Oceanography*, *SOA*, *Qingdao* 266061)

**Abstract**　Based on the urgent need for data collection, we propose to establish a data collection system of survey data from the NSFC sharing cruises. We think the overall system framework should include the content and timeline of data collection, manual and network collection modes. In order to improve the efficiency of data collection, we think it is important to make the working process of data collection clear, and to get data producer and administrator understanding their duties and to brings effective interaction between them.

**Key words**　Sharing cruise; Data collection; Metadata

·资料信息·

# 《2015 年度国家自然科学基金项目指南》征订通知

国家自然科学基金委员会编制的《2015 年国家自然科学基金项目指南》(以下简称《项目指南》)将于 2014 年 12 月中旬出版发行。《项目指南》中的部分学科代码等内容有了新的变化，为了更好地了解国家自然科学基金的资助政策，学科资助范围，正确选择资助类别、研究领域及研究方向，准确选择申请代码，请广大基金申请人和管理者踊跃订购《项目指南》。

《项目指南》针对 2015 年度集中接收的各类项目进行介绍，充分体现 2015 年科学基金资助工作的指导思想、最新资助政策和管理办法，是指导申请国家自然科学基金的重要依据，是广大科学基金申请人、管理者和评审者必读的参考文献。

1. 为了便于订购和邮寄，请各单位详细填写《征订单》(http://www.nsfc.gov.cn/publish/portal0/tab38/info45207.htm)，用挂号信或传真发至国家自然科学基金委员会机关服务中心办公室。《项目指南》定价 38 元，2014 年 12 月 15 日后发行《项目指南》。

通信地址：北京市海淀区双清路 83 号 邮政编码：100085

联系人：国家自然科学基金委员会机关服务中心　张艳东

联系电话：010-62326973/62327218；

传真：010-62327220

2. 请购书单位在银行或邮局的汇款单上务必详细注明单位名称和姓名及数量，否则无法开具发票和按时寄书。(财务室电话：010-62327020)

**银行汇款**

开户银行：中国工商银行北京北太平庄支行

单位名称：国家自然科学基金委员会机关服务中心

银行帐号：0200010009014450296

**邮局汇款**

单位名称：国家自然科学基金委员会机关服务中心财务室

单位地址：北京市海淀区双清路 83 号

邮政编码：100085

附件：《2015 年度国家自然科学基金项目指南》征订单

国家自然科学基金委员会 机关服务中心

2014 年 9 月 15 日