

科学数据管理和数据汇交培训

科技计划项目数据汇交

国家天文科学数据中心

报告人：陶一寒 时间：2023-11-28

目录 CONTENTS

01

汇交政策背景及要求

02

汇交内容

03

汇交流程及具体操作

01

汇交政策背景及要求

01

项目数据汇交相关政策

- 2014年6月22日，科技部办公厅《**科技基础性工作专项项目科学数据汇交管理办法(试行)**》
- 2018年2月13日，科技部、财政部关于印发《**国家科技资源共享服务平台管理办法**》的通知（国科发基〔2018〕48号）
- 2018年3月17日，国务院办公厅关于印发《**科学数据管理办法**》的通知（国办发〔2018〕17号）
- 2019年12月26日，科技部办公厅关于印发《**科技计划项目科学数据汇交工作方案(试行)**》的通知（国科办基〔2019〕104号）
- 2019年2月19日，中国科学院印发《**中国科学院科学数据管理与开放共享办法(试行)**》
- 2022年8月19日，中国科学院印发《**中国科学院战略性先导科技专项科学数据汇交管理实施细则（试行）**》（科发规字〔2022〕47号）

数据汇交相关政策要求

《科学数据管理办法》明确提出科学数据“开放为常态，不开放为例外”的原则，**要求政府预算资金资助的各级科技计划（专项、基金等）项目所形成的科学数据，应由项目牵头单位汇交到相关科学数据中心。**

并要求各级科技计划（专项、基金等）管理部门应建立**先汇交科学数据、再验收科技计划（专项、基金等）项目**的机制；项目/课题验收后产生的科学数据也应进行汇交。同时鼓励社会资金资助形成的其他科学数据向相关科学数据中心汇交。

国科办基〔2019〕104号

科技部办公厅关于印发《科技计划 项目科学数据汇交工作方案 (试行)》的通知

各国家重点研发计划项目管理专业机构：

《科学数据管理办法》明确要求“政府预算资金资助的各级科技计划（专项、基金等）项目所形成的科学数据，应由项目牵头单位汇交到相关科学数据中心，各级计划（专项、基金等）项目管理部门应建立先汇交科学数据、再验收科技计划（专项、基金等）项目的机制”。

为落实《科学数据管理办法》和《国家重点研发计划综合绩效评价工作规范（试行）》的相关要求，规范科技计划项目科学数据汇交工作，科技部制订了《科技计划项目科学数据汇交工作方案（试行）》，现印发你们，请根据相关工作要求参照执行。

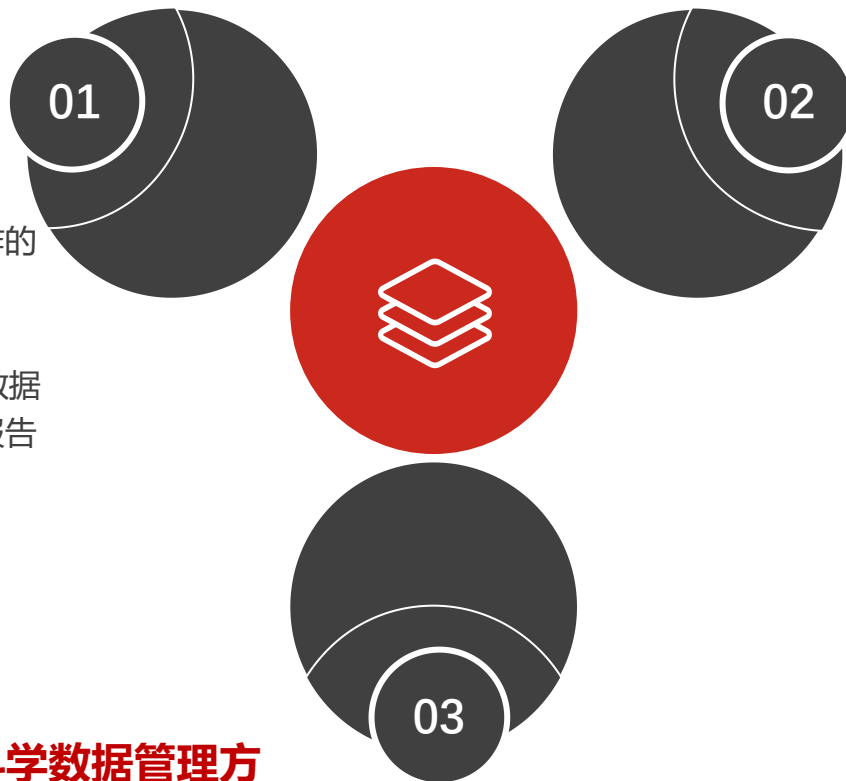
01

汇交的管理主体与职责

科技计划项目管理方

具体承担科技计划项目管理工作的机构，主要职责是：

1. 审批科学数据汇交计划；
2. 组织开展科技计划项目科学数据汇交，审批科学数据汇交审查报告（汇交凭证）；
3. 审批共享科学数据及目录。



科学数据管理方

指各类科学数据中心，是科学数据管理和开放共享的重要载体，主要职责是：

1. 制定科学数据的汇交操作规范并宣贯实施；
2. 承担汇交数据的接收、审查并保存与备份，出具审查报告（汇交凭证）；
3. 开展科学数据的分类编目与标识、加工整理和管理维护；
4. 发布科学数据及目录，评估科学数据应用情况，依法依规提供科学数据共享服务。

科学数据提交方

指科技计划项目的牵头法人单位，是科学数据汇交的责任人和义务人，主要职责是：

1. 组织制定科学数据汇交计划；
2. 按照相关标准规范组织开展科学数据采集生产和加工整理，形成便于使用的科学数据；
3. 建立科学数据质量控制体系，负责科学数据的质量控制，保证数据的真实性、准确性、可用性、完整性、一致性和安全性等；
4. 按照科学数据汇交计划向科学数据管理方提交科学数据。

01 国家天文科学数据中心

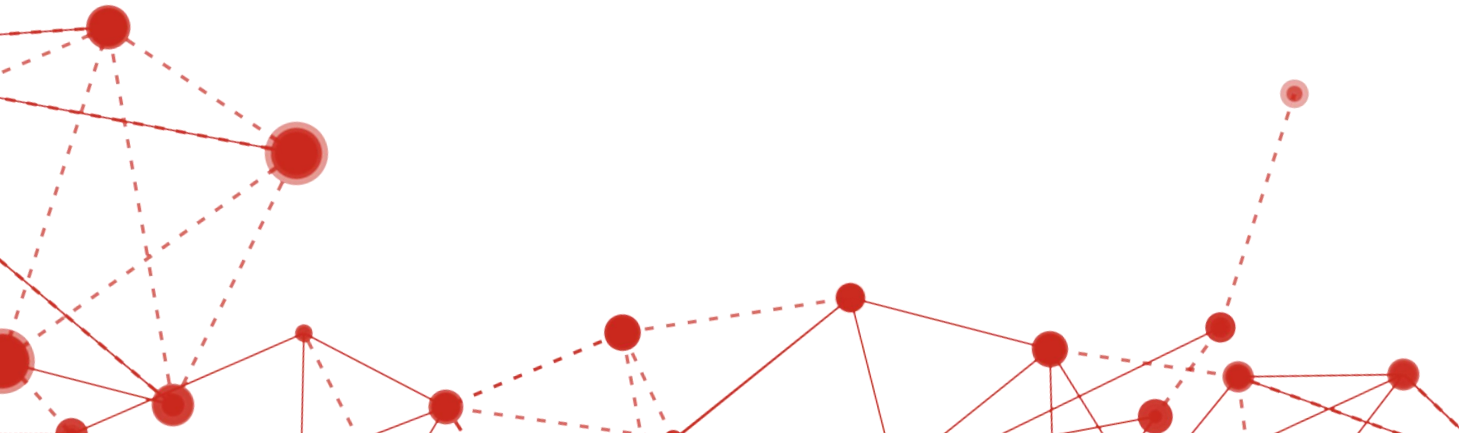
- 国家天文科学数据中心，英文名称为“National Astronomical Data Center”，简称NADC。
- 是由科技部、财政部认定的国家科技资源共享服务平台，属于基础支撑与条件保障类国家科技创新基地。
- 依托单位是中国科学院国家天文台。在中国科学院科学数据中心体系中同时为中国科学院天文科学数据中心。
- 负责汇交管理、整编、集成天文领域的科学数据，制定相关标准规范，建设天文科学数据资源体系。
- **承担天文领域科技计划项目汇交**
- 官网：<https://nadc.china-vo.org/>





02

汇交内容



根据《科学数据管理办法》，**科学数据**主要包括在自然科学、工程技术等领域，通过基础研究、应用研究、试验开发等产生的数据，以及通过观测监测、考察调查、检验检测等方式取得并可用于科学研究活动的**原始数据及其衍生数据**。

项目科学数据汇交内容包括科技计划项目执行过程中产生的科学数据实体、科学数据描述信息和科学数据辅助工具软件。



科学数据实体

原始数据及基于原始数据或研究分析数据所形成的完整数据库或数据文件



科学数据描述信息

描述信息、数据说明文档及过程资料



科学数据辅助工具软件

项目形成的用于科学数据 处理、加工和分析的专门辅助软件工具等

汇交的主要内容

科学数据实体

- 科学数据实体是在科技计划执行过程中形成的**原始数据**及基于原始数据或研究分析数据所形成的完整**数据库或数据文件**。
- 数据库：**结构化**的数字对象的表达，通用或专用的数据库格式（如星表数据库）。
- 数据文件：**非结构化**的一个或多个数字对象的集合（如图像、光谱、光变曲线等文件或多个文件基于逻辑关系组合形成的数据集）。

汇交的主要内容

科学数据描述信息

- 科学数据必须提供相应的**描述信息**，各领域可根据科技计划项目的实际需要进行扩展，同时，各领域根据实际需要确定相应的**数据说明文档及过程资料**。
- 统一的元数据**
 - 通用元数据
 - 领域专有元数据
- 数据说明文档**：对数据文件组织形式、数据格式、处理过程和使用方式进行说明

1987-2011年北京怀柔太阳磁场观测数据 发布时间: 2022-09-01

数据资源1987年-2011年怀柔多通道磁场望远镜常规观测的磁场数据,包括太阳光球和色球的纵向磁场,太阳色球矢量磁场。数据处理过程中对数据信息进行了完善和规范化,保存成国际通用的fits格式。资源中包括软件程序,可以对数据进行读取和处理。其中文件名带有4061的数据是色球纵向磁场数据;文件名带有5324的数据是光球磁场数据。数据文件头内包含观测数据信息。

· 数据访问 ·

数据下载

· 数据简介 ·

英文名称	Observation data of the solar magnetic field in Huairou, Beijing from 1987 to 2011
关键字	太阳磁场, 太阳地球, 太阳物理
数据量	386048.0000 MB 185362 文件
共享途径	线上共享
共享范围	完全共享
申请流程	不需要申请
DOI	10.12149/100764

· 联系数据作者 ·

作者 刘颖
邮件 liusl@nao.cas.cn

· 联系数据发布者 ·

邮件 support@atchina-vo.org
电话 +86-10-64807973
机构 国家天文科学数据中心

· 数据标签 ·

观测波段 光学波段
观测装置和计划 怀柔磁场望远镜
子学科 太阳物理
数据类型 图像数据
用户对象 科研

元数据

- 名称
- 数据集介绍
- 访问方式 (检索/下载)
- 关键字
- 数据量
- 共享途径和范围
- 唯一标识符 (DOI/CSTR/IVOID)
- 子数据集和星表
- 数据作者、发布者
- 数据分类标签

分类标签

- 观测波段
- 观测装置和计划
- 数据类型
- 子学科
- 生产年代
- 用户对象

汇交的主要内容

科学数据辅助工具软件

- 科学数据辅助工具软件是指科技计划形成的**用于科学数据处理、加工和分析的专门辅助软件工具等**。
- 项目实施中采购的工具软件可不必汇交，但需提供与汇交数据处理相关的工具软件**使用说明**。
- 提交的科学数据辅助工具软件包括**软件工具本身或网络调用接口和属性信息**，应符合相关软件安全规定。
- 属性信息主要包括软件名称、用途、开发工具、运行环境、开发单位、版本号、使用手册等。

03

汇交流程及具体操作

03

汇交流程

“科技计划项目形成的科学数据的通用汇交流程包括**科学数据汇交计划制定**、科学数据制备、科学数据提交、科学数据审核、科学数据汇总、科学数据发布与共享和科学数据使用与维护更新等流程。”

项目立项阶段

科学数据汇交计划制定

科学数据提交方

- 按照科技计划项目管理方的相关规定，**编制科学数据汇交计划**，提出质量控制的手段和方法，以及科学数据开放时间及共享方式等共享条件。
- 数据汇交计划编制完成后需将其提交至科技计划项目管理方。

科学数据管理方

- 根据需要制定本领域科学数据汇交计划格式
- **审查汇交计划**

科技计划项目管理方

- **审批汇交计划**

03

汇交流程

“科技计划项目形成的科学数据的通用汇交流程包括科学数据汇交计划制定、**科学数据制备**、科学数据提交、科学数据审核、科学数据汇总、科学数据发布与共享和科学数据使用与维护更新等流程。”

	科学数据提交方	科学数据管理方	科技计划项目管理方
项目执行阶段	科学数据制备 <ul style="list-style-type: none">遵照科学数据汇交计划和相关标准规范，进行规范化的科学数据实体采集生产与加工整理，按规定格式形成科学数据的数据元信息。建立科学数据质量控制体系，完成科学数据的质量控制	<ul style="list-style-type: none">对科学数据的制备工作提供指导	—

03

汇交流程

“科技计划项目形成的科学数据的通用汇交流程包括科学数据汇交计划制定、科学数据制备、**科学数据提交**、**科学数据审核**、科学数据汇总、科学数据发布与共享和科学数据使用与维护更新等流程。”

项目结题阶段

	科学数据提交方	科学数据管理方	科技计划项目管理方
科学数据提交	<ul style="list-style-type: none"> 按照科学数据汇交计划，对计划汇交的科学数据进行汇总整理，对科学数据质量进行自查，编制科学数据质量信息报告，并将科学数据提交至科学数据管理方。 	—	—
科学数据审核	<ul style="list-style-type: none"> 若提交的科学数据存在问题，科学数据提交方应及时进行修改并重新提交。 	<ul style="list-style-type: none"> 按照科学数据汇交计划和科学数据质量控制体系要求进行形式审查，通过开展科学数据质量同行评议等方式对科学数据质量进行评估。 当审核通过时，出具审查报告作为汇交凭证。 	<ul style="list-style-type: none"> 审批汇交凭证

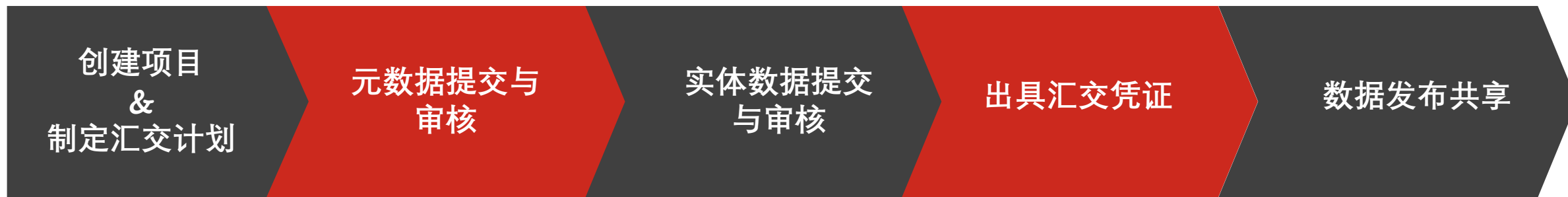
03

汇交流程

“科技计划项目形成的科学数据的通用汇交流程包括科学数据汇交计划制定、科学数据制备、科学数据提交、科学数据审核、**科学数据汇总**、**科学数据发布与共享**和**科学数据使用与维护更新**等流程。”

	科学数据提交方	科学数据管理方	科技计划项目管理方	
项目结题后	科学数据汇总	—	—	
	科学数据发布与共享	<ul style="list-style-type: none"> 对于涉密数据，负责按照相应程序定密，经科技计划项目管理方审批后，按照相应的保密要求进行管理 	<ul style="list-style-type: none"> 保障科学数据安全，根据科学数据的开放条件、开放对象和审核程序等，对外公布科学数据及目录 建设运营科学数据共享发布系统 	<ul style="list-style-type: none"> 审批项目数据对外发布共享
	科学数据使用与维护更新	<ul style="list-style-type: none"> 鼓励对所提交的科学数据进行持续更新与归档 	<ul style="list-style-type: none"> 做好科学数据的管理与维护工作，对科学数据共享应用情况进行统计，并将统计情况反馈至科技计划项目管理方 	—

03 具体操作



- 在汇交系统上创建项目
- 编制项目科学数据汇交计划

- 在汇交系统上提交拟汇交数据集的元数据

- 在汇交系统上提交相应的实体数据
- 邀请同行专家审核数据质量，填写评议表

- 完成以上步骤并收到签字盖章的汇交计划后，NADC出具汇交凭证

- 依据汇交计划填写的数据公开时间及开放条件，NADC对数据进行发布共享

项目数据汇交系统 <https://nadc.china-vo.org/res/>

创建项目

- 项目组数据汇交工作对接人在汇交系统上创建项目，填写项目基本信息，包括项目名称、编号、项目类型、牵头单位、项目简介和数据介绍等

MMDCS

项目

点击“创建数据汇交项目”，进入项目创建页面

+ 创建数据汇交项目

项目编号	项目名称	项目类型	牵头单位	状态	操作
nafc123	检测	国家任务/国家重大科技专项	国家天文台	项目已审核通过	查看项目信息 汇交计划 数据汇交

上一页 1 下一页

国家科技资源共享服务平台 yihan tao 操作台 退出 English 全站 请输入检索关键字 搜索

NADC National Astronomical Data Center 国家天文科学数据中心 申请访问 科学数据 专题服务 云资源 公众服务

MMDCS 汇交说明

项目创建: [提交](#) [返回项目列表](#)

项目名称: 项目编号:

关键词: 编号分号, 隔开 推荐应用领域: 编号分号, 隔开

项目类型:

项目牵头单位:

[添加参与单位](#)

参与单位: [删除](#)

[添加项目成员](#)

用户名: 邮箱: [删除](#)

项目简介: 数据介绍:

[提交](#)

制定汇交计划

- NADC建立项目数据汇交微信群沟通指导汇交计划编写，根据项目任务书审查汇交计划
- 项目组根据汇交计划模板及填写说明，编写汇交计划，定稿盖章后将最终版上传系统

目录

一、科学数据概述	3
二、科学数据资源清单	3
表1 计划汇交的科学数据清单	4
三、科学数据的质量控制说明	5
四、科学数据的软件工具说明	5
表2 科技计划项目工具软件描述信息	6
五、科学数据的衍生数据的使用原则	7
六、科学数据的使用期限与长期保存	7
七、科学数据汇交技术方案	7
八、其他补充说明	7

汇交计划是项目开展科学数据汇交的重要依据

制定汇交计划

一、科学数据概述

简要描述科技计划项目拟产生的科学数据情况，包括但不限于数据内容、采集方案、采集地点、采集时间、设备情况等基本信息。

描述项目（拟）产生的科学数据总体情况，建议包括以下内容：

- 项目背景及重点任务简介；
- 关于项目（拟）产生的数据内容、类型、数据量和共享方式等方面的综述；
- 项目承担单位的数据生产条件、数据主要来源、生产方式和生产时间等方面的综述。

制定汇交计划

二、科学数据资源清单

表1 计划汇交的科学数据清单

序号	科学数据集名称	数据类型	预估数据量 /记录数	数据格式	共享方式	公开时间	对应考核指标

- **应包含任务书考核指标对应的数据资源。**
- **共享方式**可填写完全共享、有条件共享、暂不共享，如没有特别需要，均应选择完全共享，其余情况均按要求提供情况说明，经科技计划项目管理方审批。
- **数据公开时间**原则上为项目完成验收后即开放。对于数据集在汇交时因汇交方相关研究仍在进行当中，为保护其首发研究成果的权益，可以设置不超过24个月的保护期，按月填写。主张保护期需提交证明材料，并经科技计划项目管理方审批。

制定汇交计划

三、科学数据的质量控制说明

描述科学数据生产所采用的相关数据质量控制情况，包括对汇交科学数据的来源、采集、加工、处理等各环节的质量控制措施等内容。

- 主要用于说明数据产生的方式、流程以及各流程所采用的质量控制措施。
- 具体描述数据采集加工的时间、地点、所使用的仪器设备、采用的标准规范、处理的模型方法、质量保障手段等。

制定汇交计划

四、科学数据的软件工具说明

描述科技计划形成或使用的用于科学数据处理、加工和分析的专门辅助软件工具等。其中自主研发的软件工具应汇交，默认共享方式与相关数据集相同。

表 2 科技计划项目工具软件描述信息

序号	软件名称	用途	开发工具	开发单位	运行环境	相关数据集序号	是否自主研发	备注

- 既包括项目**使用**的第三方的数据处理工具，也包括项目**自行研发**的专门用于其产生的数据成果的相关加工处理的软件工作。
- 项目实施中采购的第三方工具软件可不必汇交，但需提供与汇交数据处理相关的工具软件使用说明
- 项目自主研发的工具软件应纳入汇交范围，操作手册等资料应一并汇交。

制定汇交计划

五、科学数据的衍生数据的使用原则 描述衍生数据的利用、再加工的政策。

- 如果科学数据提交方对共享用户衍生使用数据有任何方面的要求，可提出相关的使用政策。
- 对于**已经完全开放共享的科学数据**，衍生数据的使用原则主要关于知识产权，如是否允许再发布、是否允许演绎、所产生的演绎作品是否需要署名、是否允许用于商业目的等。比较常见的声明方式如Creative Commons知识共享许可协议。
- 对于**尚未完全开放共享的科学数据**，如有条件共享，不公开共享等可以起草一个协议模板，通过与用户签署使用协议的方式，提出具体的衍生数据使用原则。

制定汇交计划

六、科学数据的使用期限与长期保存

说明科学数据的保护期限及保护原因，并指明拟进行科学数据存档的数据中心。

- 针对《表1 计划汇交的科学数据清单》中“共享方式”一项填写为非完全共享的，以及在“公开时间”一项列出保护期的，说明其保护原因。

七、科学数据汇交技术方案

说明科学数据汇交时拟采用的技术方案，包括但不限于数据目录及文件命名规则、安全策略等，以保证科学数据高质量且快速高效的汇交到科学数据管理方。

- 项目方进行数据整理的方案，包括目录结构、数据集结构（例如包括数据文件+支撑文件，表格数据可列出各字段）、命名规则、安全策略等。

元数据提交与审核

元数据提交： 对应汇交计划中的资源清单，列出的每一个数据集都需要提交相应的元数据，数据集名称也应保持一致。

- “添加元数据”：填写表单逐个数据集添加元数据
- “下载元数据模板” + “上传元数据”：数据集较多时，可用模板编辑批量上传添加多个数据集的元数据

元数据审核： 汇交专员将对照汇交计划对元数据进行审核，如不符合要求将退回修改。

测试 项目元数据列表

添加元数据
下载元数据模板
上传元数据
留言反馈
返回项目列表

编号	标题	修改时间	状态	操作

实体数据提交与审核

上传实体数据:

- **线上提交:**
 - 5GB以内, 网页直接上传
 - 5GB以上, 可使用rsync上传
- **线下提交:** 也可以直接提交硬盘 (硬盘不归还, 将直接存档)。
- 对于数据量很大, 且有能力面向用户提供数据下载服务, 无条件支持长期稳定运行, 可签署承诺书, 不上传实体数据, 提供实体数据访问链接。

The screenshot displays the '测试 项目元数据列表' (Test Project Metadata List) interface. At the top, there are buttons for '添加元数据', '下载元数据模板', '上传元数据', '留言反馈', and '返回项'. Below this is a table with columns: 编号 (ID), 标题 (Title), 修改时间 (Modification Time), 状态 (Status), and 操作 (Action). The table contains one entry with ID 100563, title '测试数据集2 test', and status '元数据审核通过'. The '操作' column for this entry has three buttons: '查看元数据', '上传/管理数据' (highlighted with a red box), and '提交数据'. Below the table is a 'Data Upload' section with two options: (1) '通过网页直接上传' (Upload directly via website) with a '选择文件' button and an '上传' button; (2) '如数据量较大, 可使用rsync上传' (If data volume is large, use rsync upload) with a 'Get rsync account' button and a form for 'username' (yfwang@bao.ac.cn) and 'secret' (secret). Below the upload options is a 'test data set' section showing '已上传文件 (0 文件 0B)' and navigation arrows.

具体操作

实体数据提交与审核

实体数据审核：

- 形式审查：**汇交专员对提交的实体数据进行形式审查，查看文件数、数据量、文件类型等是否与汇交计划和元数据相符。
- 质量评估：**项目组邀请2名以上同行专家从规范性、标准性、完整性和可用性四个方面审核数据并填写评议表，可采取会评或函评的形式。如能提供证明数据质量的材料可以免审，如第三方检测报告等。

科技计划项目科学数据汇交 数据质量同行专家评议表

专家个人信息

姓名		单位	
职称/职务		电子邮件	
通讯地址		手机号	

请在阅读科学数据描述信息，访问并实际试用数据实体样本和相关辅助工具软件的基础上获得关于数据集质量情况的感受，并在下表中勾选适合的评价，填写评议意见。

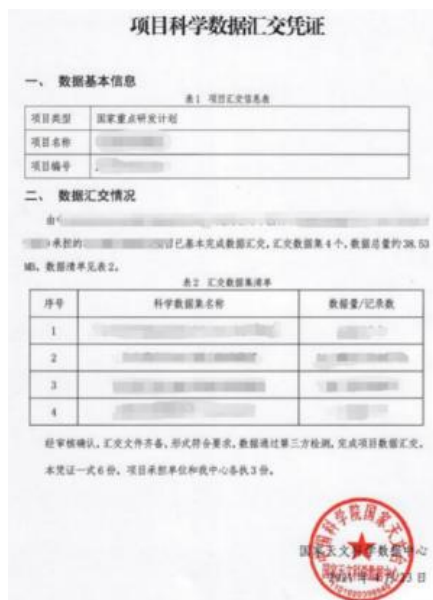
质量要素	评价因素	质量要素评价
规范性	数据符合数据标准、数据模型或权威参考数据的程度。包括但不限于以下因素： <ul style="list-style-type: none"> 数据集组织形式和文件格式符合研究领域常用数据格式和标准。数据集内同类数据文件格式统一； 《数据说明文档》对于数据集内容组织、类型说明、含义描述、使用方式及所需的工具软件等进行充分说明。 	优秀 <input type="checkbox"/> 良好 <input type="checkbox"/> 合格 <input type="checkbox"/> 不合格 <input type="checkbox"/>
准确性	数据内容对数据所研究内容的反映、表现是否准确及其准确程度；数据形式对数据内容的表述、表达是否准确及其程度。包括但不限于以下因素： <ul style="list-style-type: none"> 数据实体的精度达到《科学数据汇交计划》及项目任务书中所叙述的相关流程、方法、装置（设备）应该达到的指标； 数据集的误差达到同类研究的平均水平。不影响科技项目研究使用。可以支持大多数情况下的二次分析使用。 	优秀 <input type="checkbox"/> 良好 <input type="checkbox"/> 合格 <input type="checkbox"/> 不合格 <input type="checkbox"/>
完整性	数据集结构和内容是否完整及与《数据汇交计划》描述一致的程度。包括但不限于以下因素： <ul style="list-style-type: none"> 数据集体结构完整，包含科技项目研究论证所需的全部信息。也不存在因为缺少关键潜在信息而导致无法进行二次分析的情况； 	优秀 <input type="checkbox"/> 良好 <input type="checkbox"/> 合格 <input type="checkbox"/> 不合格 <input type="checkbox"/>

	<ul style="list-style-type: none"> 数据实体在时间、空间等要素上具有良好的连续性，不存在大量断档； 数据实体抽查未发现大量内容空缺，对缺项、异常值及错误值的处理在《数据说明文档》中进行了说明。 	
可用性	数据集内容满足用户使用需求的程度。可以但不限于从以下角度衡量： <ul style="list-style-type: none"> 数据集内容可用于验证项目研究的原始过程； 数据集内容可以支撑项目主要结论性成果； 数据集内容具有一定的共享再利用价值； 数据集内容具有一定的独特性、新颖性。 	优秀 <input type="checkbox"/> 良好 <input type="checkbox"/> 合格 <input type="checkbox"/> 不合格 <input type="checkbox"/>
评审意见	(请对数据集进行综合评价)	
评审结果	经过对数据集及相关描述信息的阅读与试用，我认为本项目（编号：_____；名称：_____）提交实体数据、相关说明文件及相关辅助工具软件的质量水平总体上_____（达到/未达到）_____（优秀/良好/合格/不合格）水平。_____（建议/不建议）正式发布，并开具汇交凭证。 本人承诺对该数据集所进行的专家评议工作按照《科技计划项目科学数据汇交工作方案（试行）》的要求进行，保证所做出的评议结果是基于真实的检查工作和学术水平得出的科学结论。本审核结果可用于《科学数据汇交审查报告》的编制和《科学数据汇交凭证》的出具。 评议专家（签字）：_____ 评议完成时间：____年__月__日	

具体操作

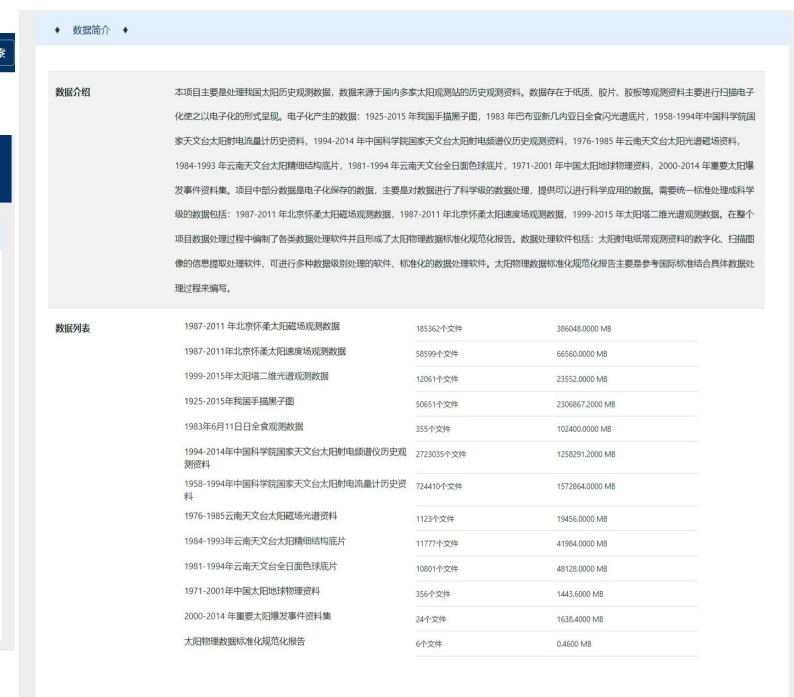
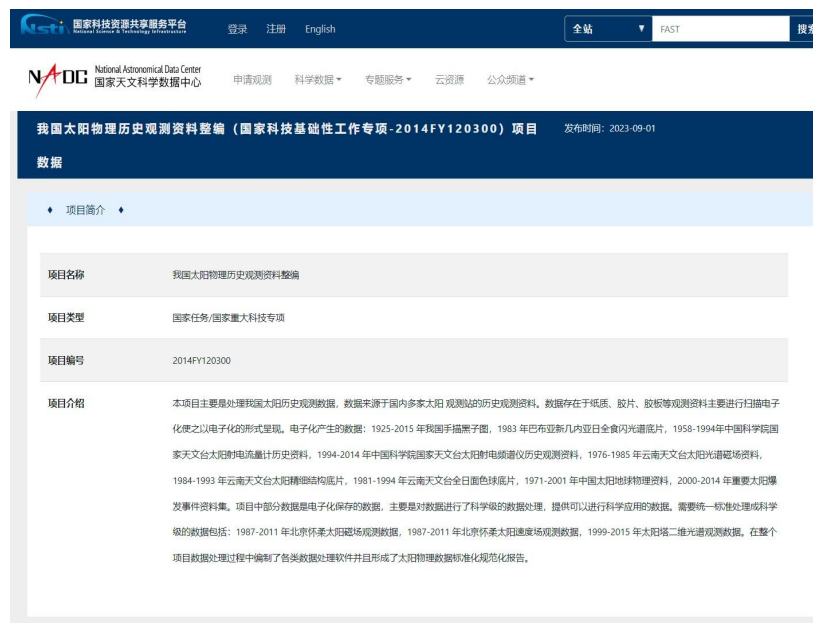
出具汇交凭证

- 项目组提交盖章签字的汇交计划及专家签字的数据质量评议表
- NADC出具汇交凭证



数据发布共享

- 当数据过保护期后，经项目管理方审批，NADC对数据进行整编并发布共享。



THANK YOU

谢谢

数据汇交事宜及问题可联系：

陶一寒 y.tao@nao.cas.cn 15810537347

王有芬 yfwang@cas.ac.cn 15010880359