

# 多类型气象科学数据在线汇交和共享平台建设与管理

■ 罗岚心 陈东辉 韩瑞 高静

数据汇交是促进数据资源共享、实现数据资源增值的重要手段。为提高数据汇交工作效率，降低数据管理与分析难度，避免数据遗失泄露风险，国家气象科学数据中心基于微服务架构实现从数据接入、处理、标识到服务的一站式气象数据在线汇交平台，提供多用户、多类型的数据资源一体化管理，保障数据汇交高效安全。

中图分类号：P4

文献标志码：A

DOI：10.3969/j.issn.2095-1973.2024.03.013

科学数据汇交能有效促进科学数据的流转、利用和增值，提高数据资源对科技创新和经济社会发展的支撑保障能力。2001年，科技部与中国气象局开展首次科学数据共享试点工程，建设首批国家科技基础条件平台之一——国家气象科学数据中心（简称数据中心），开启了我国科学数据开放共享新阶段。2015年，数据中心依托中国气象数据网面向社会开展气象相关领域数据汇交和科学数据共享服务。随着《科学数据管理办法》《国家科技资源共享服务平台管理办法》《中华人民共和国数据安全法》等国家层面的科学数据管理政策相继出台，明确了科学数据强制性汇交要求，针对农业、人口健康、生态环境等学科的科学数据汇交方案相继实现。气象领域数据资源普遍存在分类复杂、时间跨度范围大、时空分辨率不一致、数据格式多样、要素多样、加工方式复杂等特点，早期的数据汇交采用离线填写样表、数据刻盘等手段，数据维护、分析、更新与管理难度大，汇交工作周期长，且隐含数据遗失风险。为实现气象领域科学数据汇交的集约化、现代化管理，数据中心基于微服务架构研发了在线数据汇交平台（简称平台），实现元数据自动采集、数据加密传输、数据资源长期有序存储、汇交凭证自动生成等功能。平台的二级审核制度能保证汇交数据质量；分离式数据管存逻辑能防止已汇交数据被篡改或丢失。本文主要介绍平台的基本架构和运行维护，期望能为我国科技计划项目科学数据汇交管理的发展提供参考样本。

## 1 总体结构

平台总体架构分为五个层次（图1），自上而下分

为用户层、应用层、技术支撑层、数据层和基础设施与服务层。用户层分用户和管理者两种角色，分别通过数据网Web门户和管理站点登录应用；技术支撑层是业务应用拆分和调度的关键，包含流程自动化调度处理平台、凭证文档生成平台、Web门户集成技术和动态存储分配技术。

传统的单体应用架构和垂直切分架构都存在资源浪费、容错率低、响应较慢等问题，不适用于气象科学数据汇交情景。针对气象数据种类多、并发量大等特点开展技术选型，采用去中心化的Spring Cloud Alibaba的微服务架构部署应用和服务。服务独立部署在不同的Web容器中，在不改变全局业务应用逻辑的前提下新增服务，减少重构代码量。选用Nacos组件来实现服务注册发现功能，通过CoreDNS系统与分布式配置服务进行通信。

## 2 数据汇交流程与管理

### 2.1 数据汇交流程

气象科学数据汇交的对象为气象领域科技计划形成的科学数据，由科学计划主持单位提出汇交申请。平台支持气象及地球科学相关领域的结构化数据和非结构化数据汇交。在材料齐全、数据实体无错误的情况下，整个流程平均耗时在3~5个工作日内，规避早期离线汇交工作中因快递丢失、极端天气影响等意外因素造成的汇交延期风险。

平台的汇交流程分为四个阶段。当数据体量较大（超过2 TB）或因特殊原因无法在互联网传输数据时，汇交人可联系平台提出离线汇交要求。离线汇交仅在

收稿日期：2022年7月27日；修回日期：2022年11月10日

第一作者：罗岚心（1992—），Email：luolx@cma.gov.cn

通信作者：陈东辉（1984—），Email：chendh@cma.gov.cn

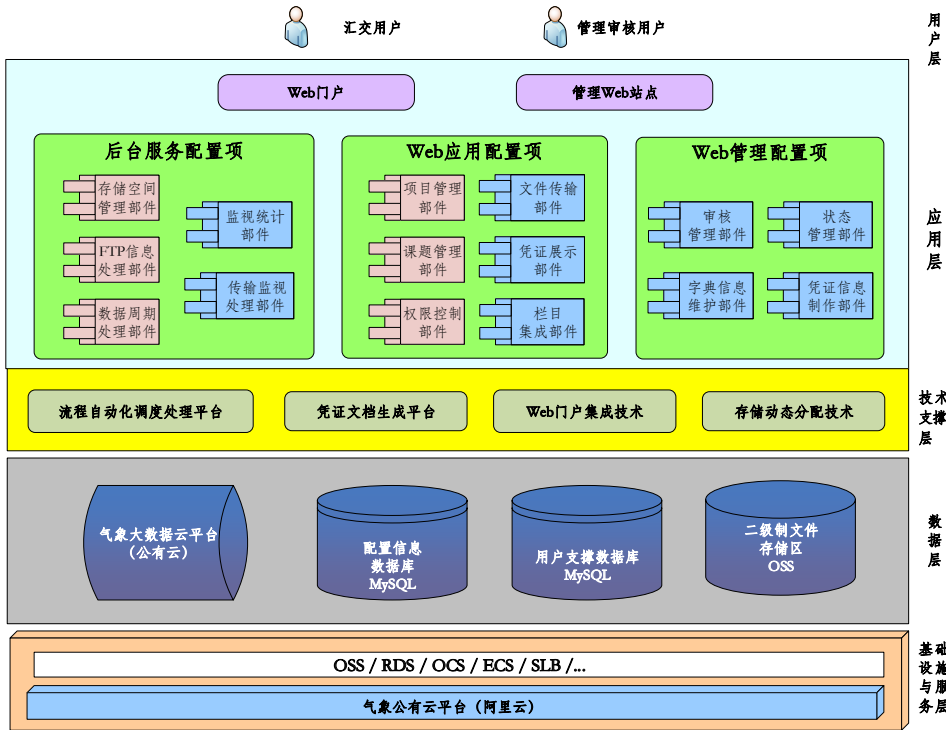


图1 气象数据汇交平台架构示意图

数据上传步骤与在线汇交不同，即使用移动存储介质（如移动硬盘、光盘）移交数据实体。

## 2.2 二级审核制度

为保证数据的科学性、有效性与可用性，平台组建一支拥有大气科学、地理信息和计算机科学等多元学术背景的高素质专家队伍，开展数据审核和数据质量评估评议，采用二级审核和交叉审核模式，确保汇交数据质量。

汇交审核采用日常性“即到即审”机制。系统一旦收到审核请求，会以邮件的形式将审核对象的关键内容告知审核专家组成员。参与审核的专家通过汇交管理系统的更新提示功能及时获取和追溯汇交项目的状态变更情况，提高审核响应速度，缩短审核等待时间。为保障信息安全，规避数据泄露风险，平台管理系统根据团队角色分组，分组管理各角色账户权限。

## 2.3 汇交数据综合管理

在环境建设方面：平台积极调研数据汇交需求和气象科学领域资源特点，基于阿里云架设数据仓库，现有在线数据存储空间1 PB，并基于在线数据仓库搭建虚拟Windows系统，实现远程登录审核功能，缩短专家审核数据实体时的传输耗时。

在支撑服务方面：平台根据用户需求灵活安排汇交方式，采取“在线申请填报，在线或离线数据传输”的工作模式，向客户提供标准化元数据填报系统、

高速数据传输接口和统一数据文档模板。在汇交专栏提供《汇交指南》，明确展示汇交申请入口、汇交工作流程、汇交凭证式样、相关政策说明和常见问题答疑，设立在线客服和热线答疑服务，最大程度提高汇交工作效率。

在资源管理方面：基于关系型数据库技术实现数据分级分类编码存储。专有数据库设为主备两套，分别架设于阿里云与中国气象局内部云，定期开展数据备份，确保科学数据安全。

在开放共享方面：数据中心贯彻落实“开放为常态，不开放为例外”原则，

严格根据《中华人民共和国数据安全法》《气象数据管理办法》和《气象数据分类分级方案（试行）》指导，参考各项目汇交计划填报的“共享方式”定制数据公开共享策略，分为“完全开放共享”“保护期限制（1a）”与“根据相关法规、政策暂不共享”三类。

## 3 安全管理机制

为打造更好的数据汇聚与应用环境，平台将气象科研数据按所属科学实（试）验、科研项目和科学考察产品的属性、影响程度和敏感程度分级为四级，级别越高受限越严，为管理科学计划项目汇交数据提供指导，保障科学数据资源安全。

平台建设参考《科技计划形成的科学数据汇通用数据元》（GB/T 39909—2021）、《科技计划形成的科学数据汇交 技术与管理规范》（GB/T 39912—2021）、《气象探测资料汇交规范》（QX/T 458—2018）和《气象数字对象标识符 总则》（QX/T 617—2021）。数据中心就数据资源管理和实现技术两个维度开展系列标准与规范研制工作，已立项的行业标准有《气象观测元数据》《公众气象观测规范》，在研项目有《气象数据资源分类分级方案（试行）》《气象数据开放共享清单》《气象科学项目数据汇交规范》，试行业务规范有《气象科学数据汇交工作流程》《气象探测资料汇交管理办法》。

气象数字对象标识符（MOID）通过标准化标识

位编码规则,实现气象数据产品的编码和确权。数据中心根据气象领域信息管理特点和共享服务需求,参考工业和信息化部数据资源唯一标识符体系建设相关要求,建设气象MOID节点管理系统和气象数据资源标识管理系统,应用于气象科学数据集汇交和发布服务,向汇交的数据集提供编码、查询、追溯等一体化服务,构建助力数据资源的确权、流通、保护,保障气象数据安全有序共享。

通过自研FIB(File Information Block)内容提取算法,平台实现了汇交资料文档的元数据自动提取,基于Mysql关系型数据库保存用户支撑数据和配置信息,设立加密接口实现按项目号、用户单位、时间段、数据名称等关键字的查询和调取。各项功能集成在Web门户网站,所有操作经服务层统一调度,实现一站式数据注册、传输、保存和管理工作。

#### 4 气象领域数据汇交共享成效

截至2022年3月,数据中心共协助北京大学、中国科学技术大学、南京大学、中国气象局等83家单位共49个项目完成数据汇交工作。汇交资源涉及六类国家重点专项(表1),共收到各类汇交数据集462个,文献类资料2800余篇,其中在线数据量4.1 TB,离线数据量超200 TB,涉及气象、气候、水文、自然灾害、大气成分等领域,极大程度丰富气象领域数据资源池。已汇交数据中部分开放共享的数据集已基于气象业务内网提供业务用户使用,气候模式预估产品已加入次季节到季节气候预测服务清单,面向48个国家的业务用户提供服务。

表1 汇交项目分类情况

专项名称	汇交项目情况		汇交数据集情况	
	个数	占比/%	个数	占比/%
重大自然灾害监测预警与防范	25	51.0	257	55.6
全球变化及应对专项	11	22.4	92	19.9
大气污染成因与控制技术研究	5	10.2	73	15.8
政府间国际科技创新合作重点专项	4	8.2	2	0.4
重大科学仪器设备开发	2	4.1	1	0.2
水资源高效开发利用	1	2.0	35	7.6
科技冬奥	1	2.0	2	0.4
合计	49	100	462	100

数据中心按照国家科学基础条件平台关于科学数据汇交相关要求,基于气象及相关领域数据资源特征,利用微服务架构和结构化数据库搭建了气象数据在线汇交平台,提供气象相关科学数据在线汇交、传输、存储管理和共享一站式服务。平台在实名注册的前提下,采用二级审核制度,保障科学数据安全与汇

交数据质量;通过在线填报系统和在线加密数据传输通道,促进数据汇交高效进行,降低数据管理分析难度;完善汇交资料元数据管理系统,建立气象数字资源唯一标识符机制,促进气象学科数据产权保护,推动气象类科技计划项目数据安全有序流动和依法依规使用,提高气象数据规范性、权威性。相较于早期纸质填报、离线邮寄、人工转储的科学数据汇交方式,现有在线数据汇交具备流程简单、存储有序、检索高效等优点,合理规避数据安全隐患,有效减少人工重复性劳动,缩短数据汇交周期。

气象领域数据汇交开展数年,现行运维稳定,但仍存在数据价值挖掘不深入、填报与审核不够智能、数据确权和激励机制无力等客观情况存在。下一步将继续调研国际先进科学数据管理和汇交工作经验,完善系统功能,改善用户体验,提高数据资源开放共享水平,加强和规范科技资源积累和开放共享,释放数据价值,更好支撑国家科技创新、经济社会发展和国家安全。

#### 深入阅读

- 白燕,杨雅萍,2020.科技基础性工作专项数据汇交实践与启示[J].中国科技资源导刊,52(4):70-79.
- 国务院,2015.国务院关于印发促进大数据发展行动纲要的通知[EB/OL].(2015-09-05)[2022-07-25]http://www.gov.cn/zhengce/content/2015-09/05/content\_10137.htm.
- 黄铭瑞,李国庆,李静,等,2019.国家科学数据中心管理模式国际对比研究[J].农业大数据学报,1(4):14-29.
- 李集明,熊安元,2004.气象科学数据共享系统研究综述[J].应用气象学报,15(增刊1):1-9.
- 刘润达,诸云强,2007.科学数据共享关键问题探索——以地球系统科学数据共享网为例[J].地理科学进展,26(5):118-126.
- 秦顺,邢文明,2019.开放·共享·安全:我国科学数据共享进入新时代——对《科学数据管理办法》的解读[J].图书馆(6):36-42.
- 吴思竹,钱庆,周伟,等,2020.面向人口健康领域科研项目数据汇交的数据仓储设计与实现[J].数据分析与知识发现,4(12):2-13.
- 熊安元,冯爱霞,高梅,等,2021.青藏高原地气系统气象科学数据集成和共享[J].高原气象,40(4):724-736.
- 杨杰,宋佳,诸云强,等,2017.科技基础性工作专项数据汇交共享平台建设[J].中国科技资源导刊,49(5):52-59,67.
- 张晶,王琰洁,黄小锋,2017.一种微服务框架的实现[J].计算机系统应用,26(4):82-86.
- 张明锋,林广发,2020.生态环境科学数据汇交平台设计与实现[J].福建电脑,36(8):122-124.
- 张志强,张强,陈东辉,等,2018.基于混合云架构的气象数据服务平台研发及应用[J].中国科技成果,19(17):55-56.
- 赵红伟,崔运鹏,2019.农业科技工作数据汇交系统建设[J].内蒙古农业大学学报(自然科学版),40(4):85-93.
- Balalaie A, Heydarnoori A, Jamshidi P, 2016. Microservices architecture enables DevOps: migration to a cloud-native architecture[J]. IEEE Software, 33(3): 42-52.
- Flower M, 2003. Design -who needs an architect?[J]. IEEE Software, 20(5): 11-13.

(作者单位:国家气象信息中心)

(编辑:卢冰)