

基于元数据的计量科学数据汇交系统研究

陶毅¹ 苏爽² 赵正宜¹ 田锋¹ 韩德隆³ 杨一凡¹ 张加涛¹ 王亭亭¹ 刘宇东² 卢祝华²
(1. 中国计量科学研究院国家计量科学数据中心, 北京 100029; 2. 中国计量科学研究院,
北京 100029; 3. 山东省计算中心(国家超级计算济南中心), 山东济南 250014)

摘要: 随着大科学装置的不断建设以及大数据、人工智能等新一代信息技术的快速发展, 科学研究正在步入大数据时代, 生命、天文、生态等各领域陆续建立科学数据中心, 科学数据已成为支撑科技创新的重要资源。本文分析国内外科学数据管理情况, 明确计量领域科学数据特点, 基于元数据理念建立分层架构的计量科学数据汇交系统, 并在标准化、技术、数据共享生态、数据市场化配置等4个方面提出发展建议, 为计量领域科学数据汇交、管理水平的提升提供新的思考。

关键词: 科学数据; 数据汇交; 计量; 元数据; 数据共享

DOI: 10.3772/j.issn.1674-1544.2022.02.001

CSTR: 15994.14.issn.1674.1544.2022.02.001

中图分类号: G311

文献标识码: A

Research on Science Data Collection System in Metrology Field Based on Metadata

TAO Yi¹, SU Shuang², ZHAO Zhengyi¹, TIAN Feng¹, HAN Delong³, YANG Yifan¹, ZHANG Jiatao¹, WANG Tingting¹, LIU Yudong², LU Zhuhua²

(1. National Metrology Data Center, National Institute of Metrology, Beijing 100029; 2. National Institute of Metrology, Beijing 100029; 3. Shandong Computer Science Center(National Supercomputer Center in Jinan), Jinan 250014)

Abstract: With the construction of big scientific devices and the rapid development of the new generation information technologies such as big data and artificial intelligence, scientific research is stepping into the era of big data. Scientific data centers have been established in various fields such as life, astronomy and ecology, scientific data has become an important resource to support scientific and technological innovation. This paper defines the characteristics of scientific data in the field of metrology by analyzing the management of scientific data at home and abroad, and establishes a layered metrology scientific data collection system based on the

作者简介: 陶毅(1980—), 女, 中国计量科学研究院国家计量科学数据中心高级工程师, 研究方向为计量科学数据和数据分析; 苏爽(1984—), 女, 中国计量科学研究院科技管理部工程师, 研究方向为科研管理; 赵正宜(1985—), 女, 中国计量科学研究院国家计量科学数据中心工程师, 研究方向为大数据; 田锋(1973—), 男, 中国计量科学研究院国家计量科学数据中心高级工程师, 研究方向为计量科学数据(通信作者); 韩德隆(1993—), 男, 山东省计算中心(国家超级计算济南中心)助理研究员, 研究方向为大数据; 杨一凡(1992—), 女, 中国计量科学研究院国家计量科学数据中心助理工程师, 研究方向为自然语言处理和数据分析; 张加涛(1981—), 男, 中国计量科学研究院国家计量科学数据中心高级工程师, 研究方向为数字计量和软件测评; 王亭亭(1991—), 女, 中国计量科学研究院国家计量科学数据中心助理研究员, 研究方向为数字计量; 刘宇东(1987—), 男, 中国计量科学研究院工程师, 研究方向为计量管理; 卢祝华(1982—), 女, 中国计量科学研究院科技管理部副主任, 研究方向为科研管理。

基金项目: 国家科技基础条件平台项目“科学数据汇交管理研究及系统研究及应用”(APT2001)。

收稿时间: 2021年9月28日。

concept of metadata. Some suggestions for data collection are put forward in four aspects: standardization, technology, data sharing ecology and data market-oriented configuration, which provide a new thinking to the collection and delivery of metrology scientific data.

Keywords: scientific data, data collection, metrology, metadata, data sharing

0 引言

科学数据是科学研究的重要产物。随着信息技术的发展,在数据中心、物联网等信息基础设施的支持下,科学数据的管理共享、价值挖掘逐渐成为科学研究的重要环节。作为最基本、最活跃的一类科技资源,科学数据具有客观性、多结构性、分散性、时效性、共享性、易传递性和再创造性等特点,贯穿于科研活动的全过程。2009年,图灵奖获得者 Jim Gray 提出科学研究第四范式,即数据密集型科学发现,进一步突显了科学数据在科研活动中的重要性^[1]。目前,科学数据在科学研究过程中的重要作用和在知识创新中的战略地位已经得到越来越多的认同,正逐步成为支撑国家科技创新的重要资源^[2]。科学数据的汇交管理、长期保存和共享应用是解决科学数据分散重复问题,促进科学数据流转、利用和增值的有效路径,对于推动科学研究和科技成果产出,更好地发挥国家科研投入产出效益,提高我国科技创新水平具有重要意义^[3]。本文将在分析国内外科学数据管理现状的基础上,概述我国计量领域科学数据汇交和计量科学数据汇交系统,对今后科学数据汇交工作提出建议。

1 国内外科学数据管理现状

1.1 国外科学数据管理现状

随着科学数据价值的日益凸显,越来越多的国际组织开始关注科学数据可实现的价值、知识产权保护及数据共享参与各方的权利和责任等方面,以多种形式制定了不同领域的数据开放指南、标准规范等。如《北极数据管理原则与实践声明》鼓励声明范围内科学数据的开放,同时明确了数据共享的伦理要求^[4];《发展中国家数据共享原则》对数据的共享提出了持续性要求,有效提升了数据的可获取性,促进了发展中国家开展

或者参与更多科学研究^[5]。

作为全球科学数据共享工作的开拓者,美国于20世纪90年代开始提供全球变化研究数据的共享服务^[6];2013年,白宫出台《提高联邦资助科学研究成果获取》备忘录,要求政府资助科研项目所取得的成果对外免费开放^[7]。英国出台了《开放科研数据协议》,提出科研数据在被提供给其他人员使用分析前,其形式不仅要符合法律和道德的规范要求,还应符合所处学科和监管的规范要求^[8]。此外,法国、日本等国家均公布了促进科学数据开放的政策法规,可见科学数据的管理和开放已成为发达国家进一步提升科技水平的重要举措。

目前,国外已建设了大量成熟的科技平台,如 DataStaR、DataCite、GBIF 等,为科学数据的开放共享提供了有效保障。在科学数据仓储方面,美国、英国和加拿大等国家的数据仓储设施建设较多,且主要隶属于政府和机构,在学科分布上数量占比较高的学科为生物、环境、地球科学、多学科科学等^[9]。元数据是数据仓储的重要组成部分,其作用类似数据库管理系统的数据字典,保存了逻辑数据结构、文件、地址和索引等信息。以康奈尔大学的 DataStaR 为例,该系统主要由数据集存储库、语义元数据存储库、开源工具和内容转移协议组成,其必备元数据元素包括所有者、创建者、题名、识别符、元数据与数据获取许可、文件级元数据等;可选元素包括联系人员、与研究小组的关系、与其他资源的关系、时间范围等,DataStaR 元数据方案中需用户输入或选择的元素较少,支持信息资源的自动发现和互操作^[10]。在平台结构方面,以英美社会科学领域数据管理与共享服务平台为例,其结构在逻辑上主要由数据库、元数据库、平台网站、后台管理4个部分组成,如图1所示^[11]。

尽管目前科学数据仓储、管理平台建设较

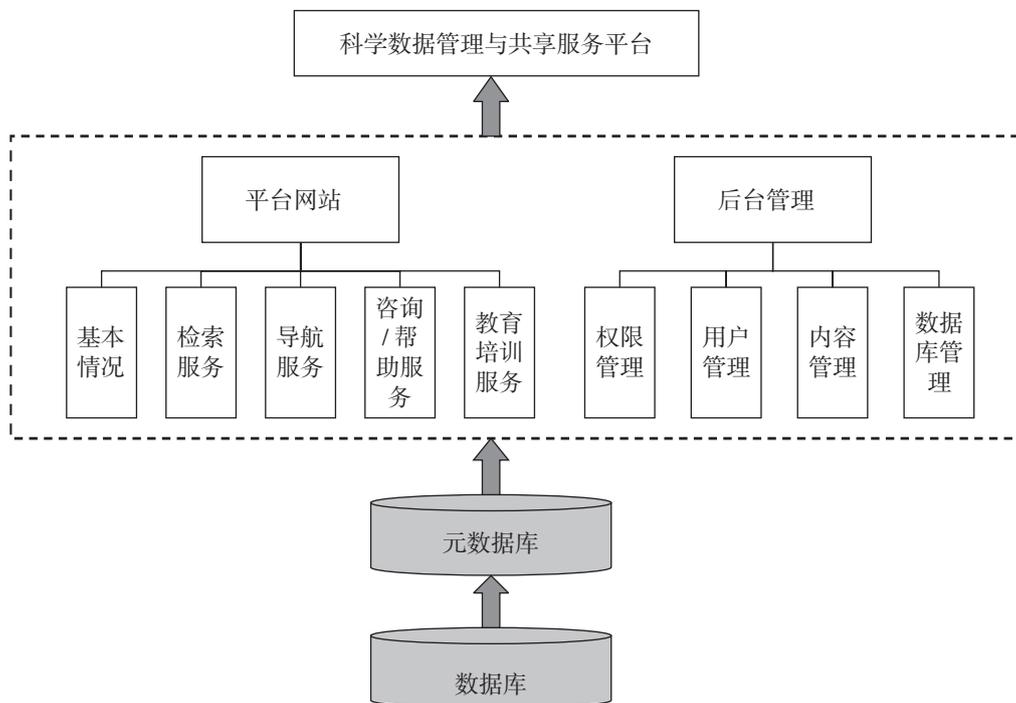


图1 英美科学数据管理与共享服务平台整体架构

为广泛，但是科研人员仍是科学数据汇交、共享的重要主体。有研究指出，目前科研人员对数据的共享积极性仍不高，如Federer等^[12]对2014—2016年间发表在PLOS ONE期刊上的47 593篇论文进行了分析，发现只有约20%的论文公开了相关数据。科研人员共享数据积极性较差的原因是多方面的，Pierce等^[13]认为共享数据的回报与期望值之间的差距是阻碍科研人员共享数据的主要因素之一。

1.2 国内科学数据管理现状

我国政府高度重视科学数据对科技发展的作用，相关政策环境不断完善。2007年修订的《中华人民共和国科学技术进步法》明确提出，利用财政性资金设立的科学技术研究开发机构，应当建立有利于科学技术资源共享的机制，促进科学技术资源的有效利用。在此之后，气象、海洋、地震科学等领域相继出台了指导本领域信息、数据管理的办法、标准。为进一步加强和规范科学数据管理，2018年出台的《科学数据管理办法》提出坚持“开放为常态，不开放为例外”的基本原则，明确要求政府预算资金资助的各级科技计划（专项、基金等）项目科学数据向科学数据中

心汇交。同年，科技部、财政部联合印发了《国家科技资源共享服务平台管理办法》，将承接科技计划项目实施所形成的科学数据的汇交、整理和保存作为国家科学数据中心重点任务之一。近年来，我国科研投入持续增长，大量科研项目的开展加速了科研数据的积累。据不完全统计，截至2017年年底，我国有效管理与保存的科学数据约为83.72 PB，覆盖地球与环境科学、天文与空间科学等领域^[14]。目前，我国已建设了包括国家计量科学数据中心在内的20个国家科学数据中心，具体如表1所示。同时，我国还建成了30个国家生物种质与实验材料资源库。这些数据中心和资源库已成为促进各领域科技创新的重要动力源泉。

我国科学数据汇交流程较为完备。针对科技计划项目形成的科学数据，国家科技基础条件平台中心制定了数据汇交的基本流程，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，以及科学数据使用与维护更新等，如图2所示^[15]。

为了进一步规范科学数据的汇交、共享等流程，提高科学数据管理成效，我国出台了一系列

表1 我国已建成的国家科学数据中心

序号	国家平台名称	依托单位
1	国家高能物理科学数据中心	中国科学院高能物理研究所
2	国家基因组科学数据中心	中国科学院北京基因组研究所
3	国家微生物科学数据中心	中国科学院微生物研究所
4	国家空间科学数据中心	中国科学院国家空间科学中心
5	国家天文科学数据中心	中国科学院国家天文台
6	国家对地观测科学数据中心	中国科学院遥感与数字地球研究所
7	国家极地科学数据中心	中国极地研究中心
8	国家青藏高原科学数据中心	中国科学院青藏高原研究所
9	国家生态科学数据中心	中国科学院地理科学与资源研究所
10	国家材料腐蚀与防护科学数据中心	北京科技大学
11	国家冰川冻土沙漠科学数据中心	中国科学院寒区旱区环境与工程研究所
12	国家计量科学数据中心	中国计量科学研究院
13	国家地球系统科学数据中心	中国科学院地理科学与资源研究所
14	国家人口健康科学数据中心	中国医学科学院
15	国家基础学科公共科学数据中心	中国科学院计算机网络信息中心
16	国家农业科学数据中心	中国农业科学院农业信息研究所
17	国家林业和草原科学数据中心	中国林业科学研究院资源信息研究所
18	国家气象科学数据中心	国家气象信息中心
19	国家地震科学数据中心	中国地震台网中心
20	国家海洋科学数据中心	国家海洋信息中心

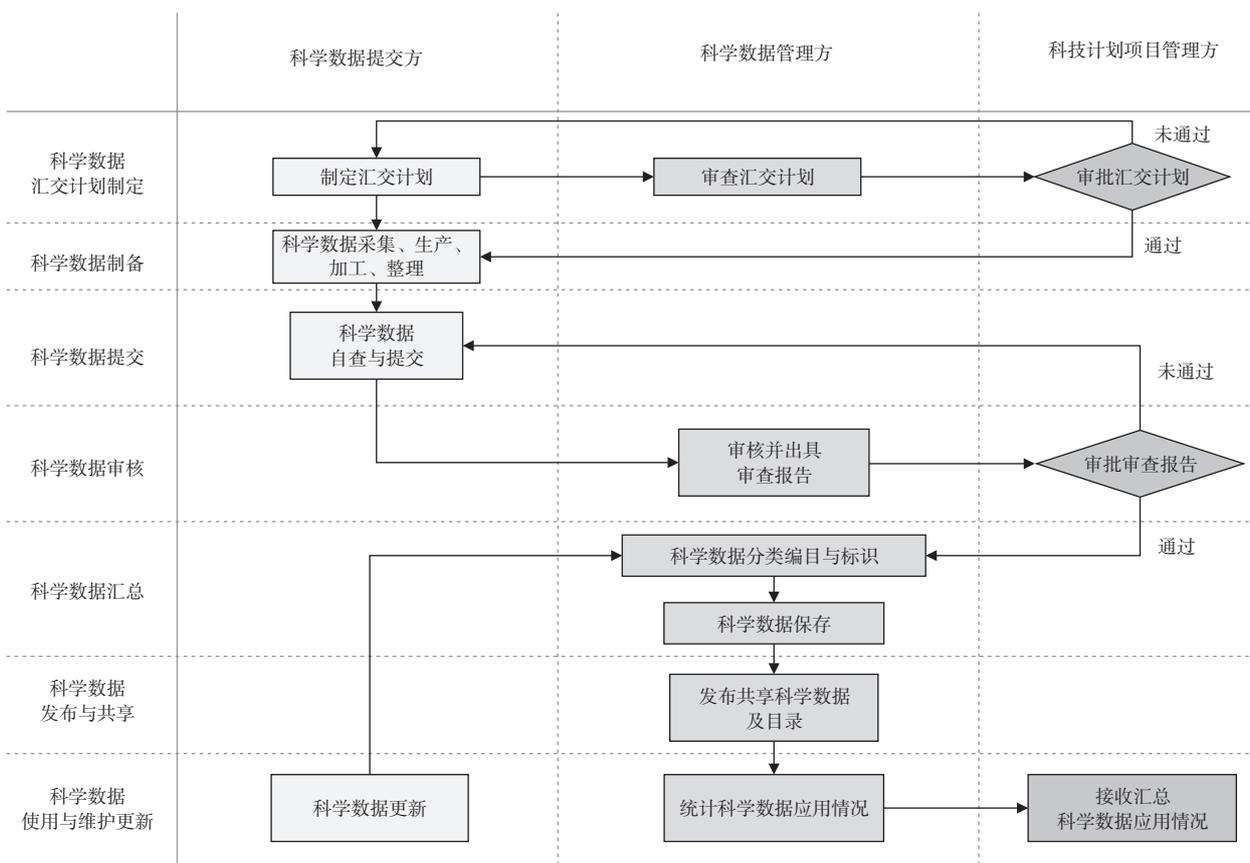


图2 我国科学数据汇交流程

国家标准，如表 2 所示，有效规范了科学数据管理工作。在数据汇交方面，根据《科技计划项目形成的科学数据汇交 技术与管理规范》要求，科学数据汇交内容包括科技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，其中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^[16]。科研数据的处理依赖于各种辅助软件工具，标准对数据汇交中应提供的辅助工具软件信息进行了详细要求，有效避免了软件差异所导致的数据分析误差。

我国在科学数据管理政策制度、平台建设、标准规范等方面已开展了大量工作，取得了一定成效。以国家基础科学数据共享服务平台为例，其数据累计下载总量已超过 2 290.42 TB，但是，与我国科学研究整体规模相比，当前科学数据的汇交、共享水平仍有较大提升空间。目前，国内

科研人员与同事或合作者私下共享数据的情况更为普遍，公开共享数据意愿较低^[17]。研究显示，数据共享态度、主观规范是影响科研人员共享数据的直接因素，感知行为控制、感知风险、感知有用性等为间接影响因素^[18]。科研人员之间信任程度的高低对科学数据共享的实现具有重要影响，多数研究者能够充分认识到科学数据对科研活动的重要性，但是在共享自身数据时存在较多顾虑，如对数据错用、滥用、保密等方面的担忧等^[19]。

为提升我国科学数据汇交、共享水平，除了进一步完善政策环境和管理机制，还应加强科学数据汇交、管理平台建设，通过功能完备、稳定可靠、方便易用的平台系统助力培育科学数据素养，引导科研人员开展数据共享，为我国科研水平的进一步提升提供坚实数据支撑。

2 计量科学数据的汇交

国家计量科学数据中心（以下简称“中心”）是承担计量领域科学数据汇集、管理、开放共享和保存的重要基础设施。中心负责管理的计量科学数据资源主要包含标准参考数据、计量科研数据、计量基标准数据、计量检测数据和计量信息数据等五类^[20]。与其他科研领域相比，计量领域

表 2 科学数据管理领域国家标准

序号	标准名称	标准号	发布时间/年
1	科技计划形成的科学数据汇交 通用代码集	GB/T 39908-2021	2021
2	科技计划形成的科学数据汇交 通用数据元	GB/T 39909-2021	2021
3	科技计划形成的科学数据汇交 技术与管理规范	GB/T 39912-2021	2021
4	科技平台 用户元数据	GB/T 39913-2021	2021
5	科技平台 元数据汇交业务流程	GB/T 32845-2016	2016
6	科技平台 元数据汇交报文格式的设计规则	GB/T 32846-2016	2016
7	土壤科学元数据	GB/T 32739-2016	2016
8	科技平台标准化工作指南	GB/Z 30525-2014	2014
9	科技平台 通用术语	GB/T 31075-2014	2014
10	科技平台 元数据标准化基本原则与方法	GB/T 30522-2014	2014
11	科技平台 资源核心元数据	GB/T 30523-2014	2014
12	科技平台 元数据注册与管理	GB/T 30524-2014	2014
13	机械 科学数据 第 1 部分：分级分类方法	GB/T 26499.1-2011	2011
14	机械 科学数据 第 2 部分：数据元目录	GB/T 26499.2-2011	2011
15	机械 科学数据 第 3 部分：元数据	GB/T 26499.3-2011	2011
16	机械 科学数据 第 4 部分：交换格式	GB/T 26499.4-2011	2011
17	生态科学数据元数据	GB/T 20533-2006	2006

的科研项目重点关注计量基标准装置建立、国家标准物质研制、国际比对实现等效及测量和校准能力实现国际互认等方面。计量领域科研项目的成果类型包括实验装置、标准物质、科技文献、新技术新方法、数据库及各类报告、标准、规程规范等，在成果形成过程中会产生大量科学数据，这些数据将成为计量科学数据实体的重要组成部分。

为落实《科学数据管理办法》和《国家重点研发计划项目综合绩效评价工作规范（试行）》相关要求，进一步规范计量领域国家科技计划项目科学数据汇交，提高科学数据共享开放水平，中心自主研发了计量科学数据汇交系统（以下简称“系统”）。作为计量科学数据向中心汇交的唯一端口，系统是为科研项目承担单位、项目负责人、课题负责人、专家、国家计量科学数据中心以及国家科技计划项目管理专业机构等各角色提供服务的统一信息化系统平台，可提供科学数据分类编目与标识、加工整理和管理维护等功能。系统所管理的国家科技计划项目（以下简称“项目”）是指以中央财政投入为主的国家自然科学基金、国家科技重大专项、国家重点研发计划、技术创新引导专项（基金）、基地和人才专项等科技计划（专项、基金）项目^[21]。系统所管理的科学数据是指项目实施过程中产生的计量科学数据。具体包括通过基础研究、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。在信息化不断向各行业领域渗透的背景下，系统实现了计量科学数据汇交的数字化，建立了集科学数据采集接入、交换汇聚、共享融合、管理运维等功能于一体的数据服务平台，为计量领域科学数据汇交、管理和共享提供了面向业务应用和共享服务支撑的底座能力。

3 计量科学数据汇交系统的构建

3.1 元数据理念

元数据（Metadata）是描述数据及其环境的数据，可用于数据的发现、互操作、质量控制

等^[22]。科学数据的元数据是以科学数据为描述对象的元数据，是对科学数据开展描述、组织、出版等工作的重要工具，可以帮助用户快速搜索发现并了解数据集基本情况，对于促进科学数据共享有着重要的意义^[23]。根据《科技平台资源核心元数据》，元数据元素（Metadata Element）是元数据的基本单元，如科研项目名称是定义科研项目的一个元数据元素。元数据实体（Metadata Entity）则是一组说明数据相同特性的元数据元素或实体的集合，如科研项目是由科研项目名称、科研项目编码、项目负责人等元素或实体的集合^[24]。

元数据的应用和管理贯穿科学数据汇交业务、运行管理、标准规范、安全和数据质量等方面。作为系统的基础之一，元数据实体设计的质量直接影响到整个系统的质量，必须最大限度地规避冗余、递归、重复等问题。以科研项目为例，首先对科研项目实体从定义、描述、中文名等进行定义，然后对元数据实体所包含的元数据元素或实体通过标识、中文名、英文名等进行定义。对于需要使用专用软件查看的科学数据，系统设计了软件元数据实体，包含软件所属科学数据、软件名称、开发单位、运行环境、开发工具、用途等要素。表3所示为系统元数据实体设计的要素，其中项目负责人、项目承担单位等元数据实体都是独立定义的单位。

3.2 系统结构框架

系统基于计量科学数据运行管理体系、标准规范体系、安全保障体系和质量保证体系构建，系统逻辑架构可分为云基础服务层（包含基础设施、数据系统和商用基础系统软件）、服务层、应用层和展示层，如图3所示。其中，系统最底层为基础设施层，其次是数据系统层和商用基础系统软件层。这3层充分利用了中心建设的私有云设施构建，也可统称为云基础服务层。

服务层是系统的关键部分，其所包含的数据管理平台不仅是服务层的核心，也是系统的核心。数据管理平台贯穿科学数据标准管理、数据治理管理、元数据管理、数据安全管理和数据生

表 3 科研项目元数据实体

标识	中文名	英文名	定义	数据类型	值域	约束/条件
project_Name	项目名称	project name	项目的名称	字符串	不限长文本	必选
project_Number	项目编号	project number	项目的编号	字符串	按照一定规则的编号	必选
government_funds	财政经费	government funds	国家财政所拨经费	数值型	数值范围	可选
horizontal_self_funds	横向自筹经费	horizontal self-funds	通过横向合作自筹经费	数值型	数值范围	可选
unit_self_funds	单位自筹经费	unit self-funds	承担单位自筹经费	数值型	数值范围	可选
government_direct_funds	财政直接经费	government direct_funds	国家财政所拨经费中用于直接经费的部分	数值型	数值范围	可选
government_indirect_funds	财政间接经费	government indirect_funds	国家财政所拨经费中用于间接经费的部分	数值型	数值范围	可选
self_direct_funds	自筹直接经费	direct_self-funds	自筹经费中用于直接经费的部分	数值型	数值范围	可选
self_indirect_funds	自筹间接经费	indirectself-funds	自筹经费中用于间接经费的部分	数值型	数值范围	可选
project_type	所属专项	project type	项目所属专项类别	字符串	不限长文本	必选
start_date	开始日期	start date	项目的开始日期	日期型	日期和时间	必选
end_date	结束日期	end date	项目的结束日期	日期型	日期和时间	必选
actual_end_date	实际结束日期	actual end date	项目的实际结束日期	日期型	日期和时间	必选
project_leader	项目负责人	project leader	项目的负责人	复合型		必选
undertaking_unit	项目承担单位	undertaking unit	项目承担单位	复合型		必选
project_contact	项目联系人	project contact	项目联系人	复合型		必选
secrecy_level	密级	secrecy level	密级	字符串		
project_type	项目类型	project type	项目类型	字符串		

命周期管理，具有计量科学数据汇聚、数据融合、数据组织、数据服务等功能和服务能力。智能服务平台是使用数据挖掘技术、传统AI或创新的学习型模型，通过标准化的统一接口提供服务。数据即服务（Data as a Service, DaaS）是指在运行管理、标准规范、安全保障以及质量保证等体系规范下，对科研数据、统计数据等实现统一的数据即服务，数据挖掘、分析以及人工智能等应用产生的数据成果，同样以DaaS的业务模式提供服务。工作流引擎同样位于服务层，为科学数据汇交各流程提供工作流技术支撑。

在运行管理、标准规范、安全保障以及质量保证体系下，针对计量科学数据的特点，构建以数据管理平台为核心、分层架构的系统框架，可以有效支撑计量科学数据汇交业务的在线运行，在提升数据汇交、管理水平的基础上，进一步实现计量科学数据的增值和安全共享。

3.3 系统角色及用例

根据科学数据汇交要求，系统的主要角色有系统管理员、科学数据管理员、科学数据主管、数据中心主任、项目承担单位等 10 个。各角色的简要说明如表 4 所示。

系统各角色之间的关系及其相关用例如图 4 所示。以下主要是 6 个用例的具体功能。

(1) 汇交计划制定和评审。由项目负责人制定汇交计划，然后由专家评审计划，汇交计划及专家评审结果交由项目承担单位、科学数据管理员进行形式审查，审查通过后报送项目管理机构审核。

(2) 汇交方案制定和评审。其相关流程与汇交计划制定和评审用例基本相同。

(3) 科学数据汇交和审核。包含项目目标/成果/考核指标的信息管理，以及课题信息管理、指标映射管理、数据集管理、数据提交和各级

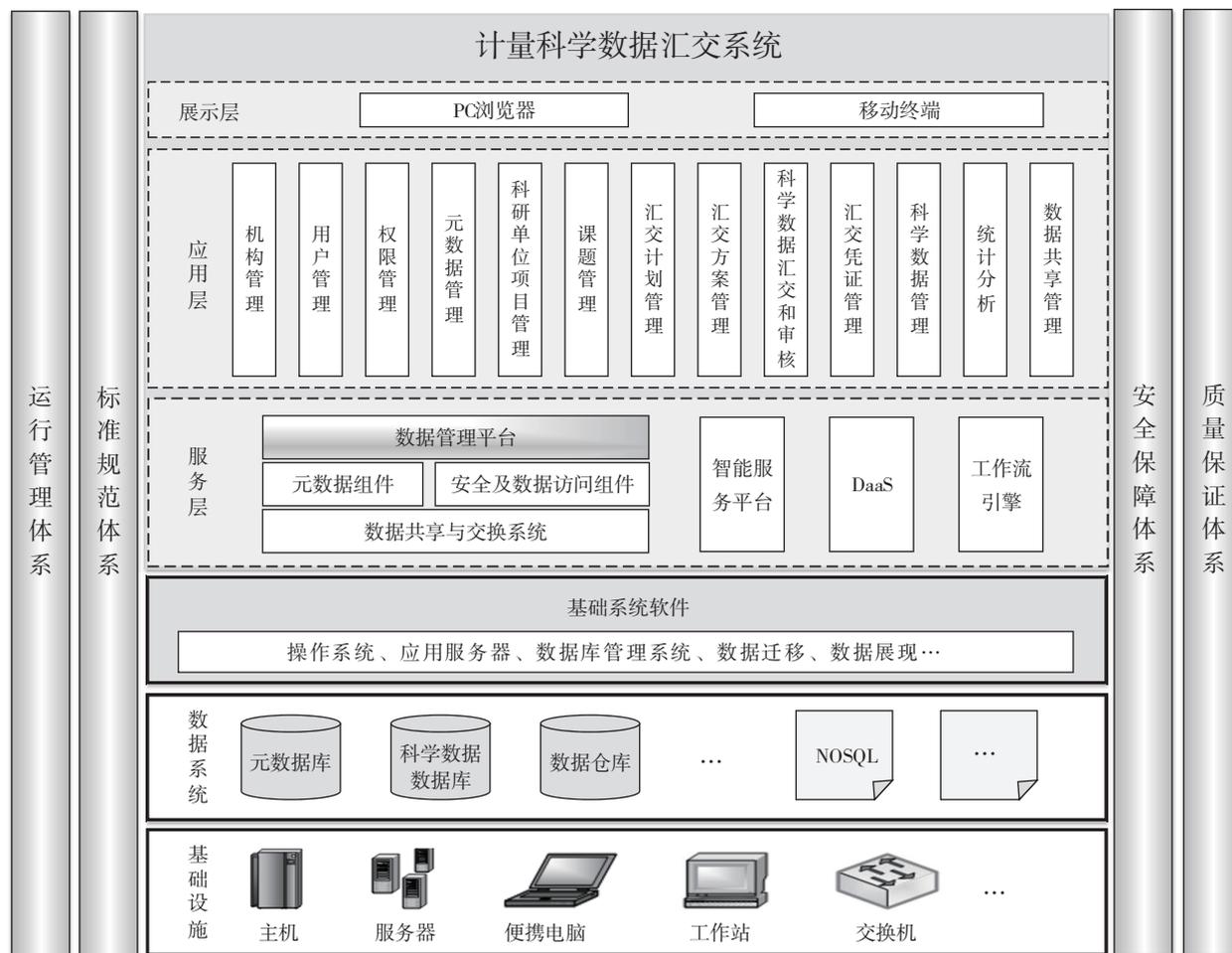


图3 系统架构

表4 系统角色及其说明

角色	简要说明
系统管理员	完成系统的设置、备份等管理的工作
科学数据管理员	负责科学数据提交方维护、科学数据管理等相关具体的工作，可以完成权限范围内的查询统计分析工作
科学数据主管	科学数据管理的负责人，主要负责对科学数据管理工作的检查，相关流程的审批，具有科学数据查询统计分析权限
数据中心主任	国家计量科学数据中心主管科学数据的主任，主要完成相关流程的审批，以及科学数据的查询统计分析工作
项目承担单位	又称科学数据提交方，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
项目承担单位管理员	又称科学数据提交方管理员，每个单位只有一个管理员账号，负责维护本单位项目负责人和项目信息
项目负责人	课题管理、项目科学数据汇交及相关审核工作
课题负责人	制备并提交课题的科学数据
专家	对汇交计划、汇交方案等进行评审
项目管理专业机构	系统产生的审查计划和报告等，需要以电子或纸质形式报送专业机构

审核。

(4) 科学数据管理。主要包含元数据管理、科学数据架构和模型管理、科学数据标准管理、科学数据质量管理、科学数据安全等功能。

(5) 统计分析。面向主题，通过数据选取、

数据清洗、数据加载、数据加工计算处理等过程，使用报表工具展示统计分析结果。该用例可引入人工智能、数据挖掘等技术，一方面通过智能化模型挖掘计量科学数据的隐含价值，另一方面通过知识积累和智能场景化学习的结合感知客

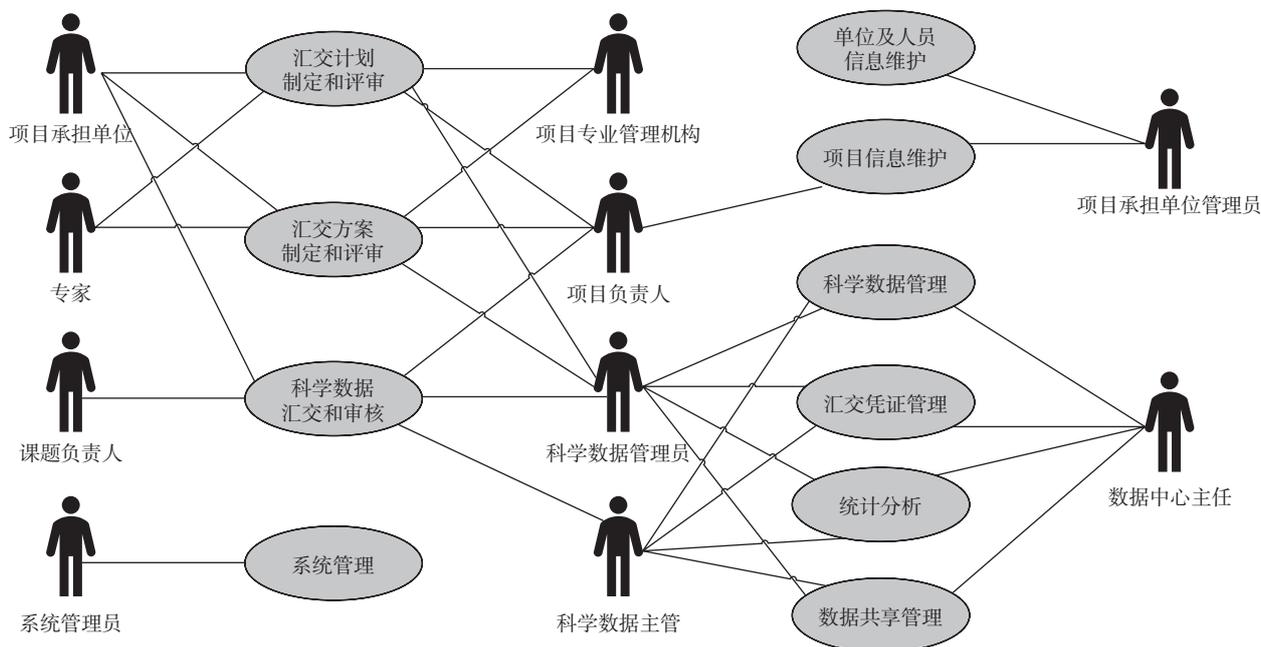


图4 系统用例图

户需求，提供智能化统计分析服务。

(6) 数据共享管理。根据数据的安全级别提供科学数据共享服务，部分统计分析结果将共享到门户网站。

3.4 科学数据汇交技术流程

在计量科学数据汇交业务流程中，参与角色较多的主要流程为汇交计划审核、汇交方案审核、汇交数据审核及汇交凭证审批。所有流程都是基于工作流引擎实现的，可满足定制化需求，系统在提供以上功能的同时，重点关注、优化科学数据在系统内的流转、管理。在业务流程的基础上，技术流程可分为数据接口、数据安全、数据标准管理、科学数据管理、计量科学数据汇交门户5个部分，如图5所示。

在技术流程中，数据接口、数据安全、数据标准管理和科学数据管理在系统运行中起到了重要作用。

(1) 数据接口。通过数据接口可获取元数据查询服务；科研人员在用户界面录入的数据、上传的文件通过数据接口进入系统；线下汇交的科研数据通过数据接口的“其他来源”进入系统；建有科研项目管理系统的单位可直接调用本系统接口，实现科研数据的在线汇交。

(2) 数据安全。由元数据算法管理、密钥版本管理、保密策略管理等子模块组成，根据密级对数据进行分级管理。

(3) 数据标准管理。在数据标准方面，经过标准发布、执行、监控和集合等过程形成数据标准，最终由该模块统一管理。

(4) 科学数据管理。在逻辑类型方面，经过数据注册、发布、运营和评估等过程形成逻辑模型，最终由该模块统一管理。

4 计量科学数据汇交系统的应用及成效

4.1 科学数据共享

系统汇交数据共享技术架构如图6所示。对科学数据进行分类、编目、标识、保存、加工、整理及管理，形成科学数据及其目录，通过门户网站对外公布，推动科学数据的共享、再利用。在数据共享方式方面，目前数据汇交系统提供了完全开放共享、协议共享、不予共享3种数据共享方式。完全开放共享数据指提供给国家计量科学数据中心且无附加共享利用条件约束的科学数据资源；协议共享数据指按约定的协议条件共享利用的科学数据资源；不予共享数据指不宜共享利用的科学数据资源。此外，对于涉密数

据，项目承担单位负责按照相应程序定密，经国家科技计划专业机构审批后，按照相应的保密要求进行管理。对于项目承担单位申请保护的科学数据，在保护期满后，系统将根据其开放条件、

开放对象和审核程序等对外公开并提供共享与服务。

从数据共享的角度来看，计量科学数据汇交入库后，在逻辑上处于数据层，可进一步分为非

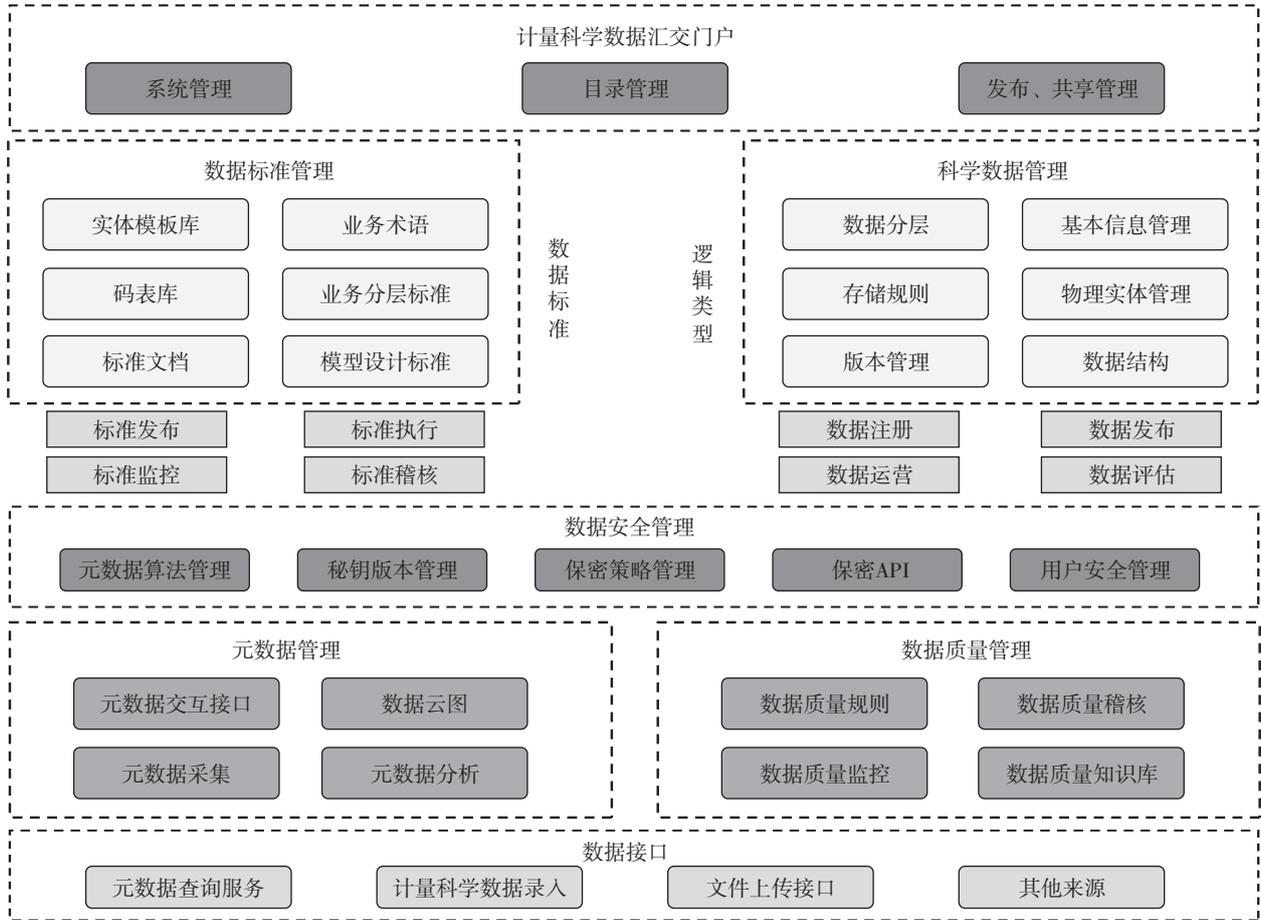


图5 计量科学数据汇交技术流程

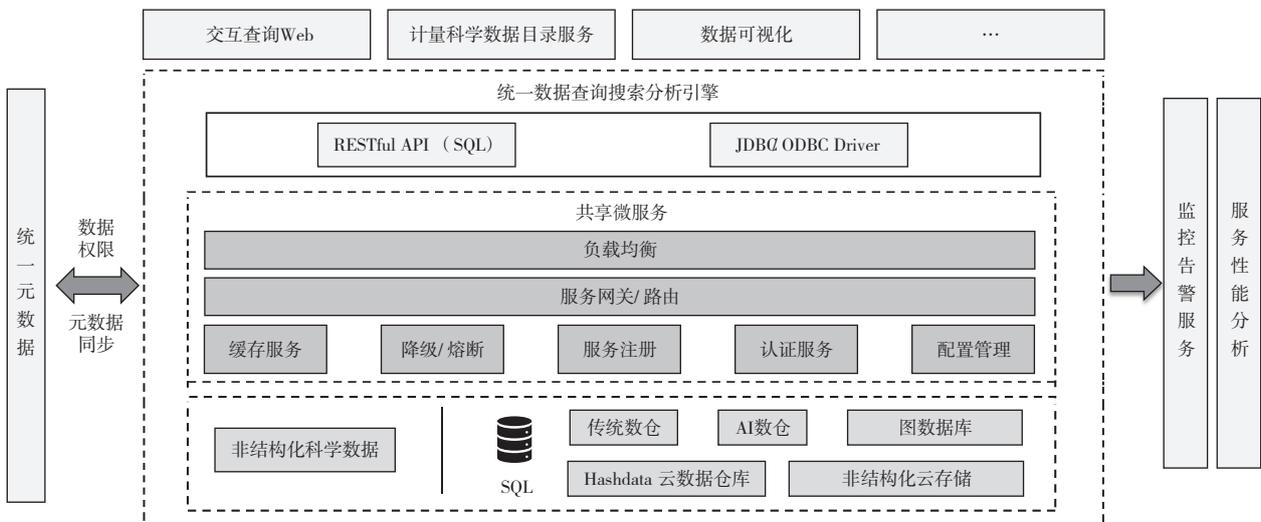


图6 计量科学数据共享模式技术架构

结构化的科学数据、非结构化的云存储、传统数仓、人工智能数仓以及云数据仓库等。为了提供高质量的开放共享服务，系统构建了共享微服务层，更高层的Web交互查询、目录服务、数据可视化等，可以通过Restful API，也可以直接通过JDBC等驱动直接访问数据，为科学数据的进一步开放共享提供技术保障。

随着系统中科学数据总量的不断增加，数据量将超过用户可以接受的程度，导致用户难以在海量数据中定位所需信息，出现“信息过载”的问题，阻碍数据的共享、再利用。针对这一可能出现的问题，中心正在基于知识图谱、人工智能开发数据推荐技术，通过对用户研究领域、浏览记录等进行分析，为用户推荐可能感兴趣的科学数据、科研项目等信息，从而促进科学数据的共享、再利用。

4.2 平台应用效果

计量科学数据汇交系统已于2021年6月投入运行，系统截图如图7所示。截至2021年9月，系统已登记注册机构350家、机构管理员113名、用户107名，实现了对计量领域192个项目/课题科学数据的汇交，极大地提升了计量领域科学数据管理水平。

通过对系统已有数据的分析可以发现，目前系统汇交的计量科学数据中60%以上为仪器制

造类数据，包括测试数据、比对数据、论文、专利、测试报告等。在科学数据共享服务方面，目前系统服务对象主要为项目组成员、计量领域/跨领域科研人员以及计量领域第三方实验室，并根据三类服务对象的不同需求，系统提供了不同的共享服务方案。

(1) 项目组成员。主要需求为项目数据的系统存储和整理加工，以便后续开展成果转化和推广应用。针对这一需求，中心将基于汇交系统将深入分析每个汇交项目，与项目组建立密切联系，根据需求为其提供定制化的信息推送服务。

(2) 计量领域/跨领域科研人员。主要需求为特定科学数据的获取及分析。这一需求可由汇交系统自身的数据检索功能实现，此外中心还将基于汇交数据定期编制计量领域各细分方向的研究报告，为用户提供数据共享、分析服务。

(3) 计量领域第三方实验室。主要需求为计量行业发展趋势、细分领域的技术指标及市场需求分析。针对这一需求，中心将通过汇交系统开展数据分析，为用户提供定制化的分析报告。

5 对科学数据汇交工作的建议

5.1 坚持标准先行

基于数据生命周期管理角度，围绕科学数据产生、汇交、管理、应用等各环节，在现有科学

项目/课题名称	项目/课题编号	所属专项	承担单位	项目负责人	开始日期	结束日期	状态	操作
项目 荧光数字基因测序分子检测...	2018YFF01012100	重大科学仪器设备开发	西安天隆科技有限公司	彭年才	2018-10	2021-09	上传项目档案	查看项目信息 全部查看 相关附件
项目 原子分辨率无耦合的高光谱...	2017YFE0130200	政府国际科技创新合作重点专项	中北大学	李艳君	2019-08	2021-12	上传项目档案	查看项目信息 全部查看 相关附件
项目 基于薄片层叠介质的系...	2016YFE0202500	战略性新兴产业重点专项	华中科技大学	朱刚	2017-09	2021-12	邮寄中	查看项目信息 全部查看 相关附件 退回原因
项目 应对国际贸易食品安全法规...	2017YFC1601600	食品安全关键技术研发	中国检验检疫科学研究院	张健	2018-01	2021-12	邮寄中	查看项目信息 全部查看 相关附件 退回原因
项目 海洋石油海洋管桩研制	2018YFC0310300	深海关键技术装备	中国石油集团工程材料研究院有限公司	肖长康	2018-07	2021-12	邮寄中	查看项目信息 全部查看 相关附件 退回原因
项目 金属增材制造在线监测...	2017YFB1103900	增材制造与激光制造	武汉大学	刘烈	2017-07	2021-06	数据汇交中	查看项目信息 全部查看 相关附件
项目 原子分辨率无耦合的高光谱...	2017YFE0130200	政府国际科技创新合作重点专项	中北大学	马宗敬	2019-08	2021-12	上传项目档案	查看项目信息 全部查看 相关附件
项目 开发突发食品溯源溯源化学...	2018YFC1602600	食品安全关键技术研发	中国人民解放军军事科学院军事医学研究院	Jianwei Xie	2018-12	2021-12	邮寄中	查看项目信息 全部查看 相关附件 退回原因
项目 CT核心部件高速磨环研...	2019YFC0117400	数字诊疗装备研发	北京凯星机器制造有限公司	徐鹏飞	2019-12	2021-12	邮寄中	查看项目信息 全部查看 相关附件 退回原因
项目 冬奥会口罩快速通关智能监...	2019YFC0302500	科技冬奥	中国海关科学技术研究中心	汪万春	2019-10	2022-06	数据汇交中	查看项目信息 全部查看 相关附件
项目 利用建筑“四节一环保”火...	2018YFC0704300	绿色建筑及建筑工业化	住房和城乡建设部科技与产业化发展中心	为敬雄	2018-07	2021-12	成果提交录入中	查看项目信息 全部查看 相关附件

图7 计量科学数据汇交系统截图

数据相关标准体系基础上,细化制定适合计量领域科学数据管理的标准,实现对计量科学数据更加规范、统一、高效的管理。

5.2 强化技术支撑

提高科学数据共享平台的管理水平,通过数据标签等手段对科学数据实行分级分类管理,运用数据认证技术加强对数据完整性的核验,在实现对科学数据合理管控的同时保证数据的可用性。强化平台数据处理能力,为科研人员提供在线数据分析服务。尝试引入区块链技术,凭借其不可篡改、可追溯的特点,强化科学数据保护能力,减少学术不端、侵犯知识产权等事件的发生,提升科研人员共享科研数据的意愿。

5.3 培育数据共享生态

围绕科学数据的共享和应用,举办科学数据大会、科学数据应用大赛等活动,为科研人员提供交流科学数据共享、应用经验的平台,营造良好的科学数据共享、应用氛围。此外,以科学数据为中心,通过知识图谱、人工智能等技术进一步分析科研机构、科研人员之间的联系,发现潜在合作机会,推动同一研究领域的科研人员、机构之间形成高度互信的合作关系,在此基础上逐步培育数据提供者、数据管理者、数据使用者密切协作的科学数据共享生态。

5.4 推动科学数据市场化配置

强调需求导向,完善科学数据供需对接相关功能,畅通科研人员寻求、获取科学数据的渠道,通过需求带动数据共享。研究根据数据性质完善产权性质,探索建立规范化数据交易平台,如对于政府资金支持下产生的科学数据,其交易收入可用于支持平台运作;对于非政府资金支持科研项目所取得的数据,其收益归数据生产者,以此进一步激发科研人员 and 科研机构创造、共享科研数据的积极性。

6 结语

本文对我国计量领域科学数据汇交和计量科学数据汇交系统及其数据共享与应用进行阐释,

并且计量科学数据汇交系统的总体框架和基于元数据的开发理念在系统建设过程中得到了充分的验证,系统技术思路和架构满足了计量科学数据汇交需求。目前,系统已部署上线,成功完成了多个计量科研项目的科学数据汇交工作。随着计量科学数据汇交工作的进一步深入,国家计量科学数据中心将继续探索区块链、人工智能等新一代信息技术在科学数据汇交领域的应用,提升计量领域科学数据管理水平,为计量科学数据的开放共享和价值挖掘提供有力支撑。

参考文献

- [1] TANSLEY S, TOLLE K M. The fourth paradigm: data-intensive scientific discovery[M]. Redmond, WA: Microsoft research, 2009.
- [2] 沈梦轩. 国内科学数据资源建设与共享研究[J]. 图书馆工作与研究, 2019(S1): 137-141.
- [3] 黎建辉, 沈志宏, 孟小峰. 科学大数据管理: 概念、技术与系统[J]. 计算机研究与发展, 2017, 54(2): 235-247.
- [4] IASC. Statement of principles and practices for arctic data management[EB/OL]. (2013-04-16) [2021-07-14]. https://iasc.info/images/data/IASC_data_statement.pdf.
- [5] Participants of the Workshop on Open Data for Science and Sustainability in Developing Countries. Data sharing principles in developing countries[EB/OL]. (2015-07-30) [2021-07-14]. <http://doi.org/10.5281/zenodo.22117>.
- [6] ALLAN B D. Policy statements on data management for global change research[J]. Informatics Studies, 2018, 4(4): 57-61.
- [7] Executive Office of the President. Increasing access to the results of federally funded scientific research[EB/OL]. (2013-02-22) [2021-07-14]. http://.obam-awhitehouse.archives.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf.
- [8] 王丹丹. 英国《开放科研数据协议》的解读与启示[J]. 数字图书馆论坛, 2018(8): 25-30.
- [9] 胡芳. 国外典型科学数据仓储实施的元数据方案及启示[J]. 图书与情报, 2015(1): 117-121.

(下转第92页)

- 业为例[J]. 科技进步与对策, 2021, 38(1): 1-9.
- [11] 王超, 许海云, 方曙. 颠覆性技术识别与预测方法研究进展[J]. 科技进步与对策, 2018, 35(9): 152-160.
- [12] 石慧, 潘云涛, 苏成. 颠覆性技术及其识别预测方法研究综述[J]. 情报工程, 2019, 5(3): 33-48.
- [13] 赵格. 基于多源异构数据的颠覆性技术识别[D]. 武汉: 华中科技大学, 2017.
- [14] 蒋林杉. 萌芽颠覆性技术的识别与成长性判断[D]. 北京: 北京工业大学, 2018.
- [15] 孙永福, 王礼恒, 孙棕檀, 等. 引发产业变革的颠覆性技术内涵与遴选研究[J]. 中国工程科学, 2017, 19(5): 9-16.
- [16] KELLER A, HUESIG S. Ex ante identification of disruptive innovations in the software industry applied to web applications: the case of Microsoft's vs. Google's office applications[J]. Technological forecasting & social change, 2009, 76(8): 1044-1054.
- [17] 邓雪, 李家铭, 曾浩健, 等. 层次分析法权重计算方法分析及其应用研究[J]. 数学的实践与认识, 2012, 42(7): 93-100.
- [18] 傅耀威, 相红, 宋阳, 等. 支撑智能社会的颠覆性技术评价体系构建[J]. 中国基础科学, 2020, 22(5): 59-66.
- [19] 张守明, 张斌, 张笔峰, 等. 颠覆性技术的特征与预见方法[J]. 科技导报, 2019, 37(19): 19-25.
- [20] 张佳维, 董瑜. 颠覆性技术识别指标的研究进展[J]. 情报理论与实践, 2020, 43(6): 194-199.
- [21] 郑国雄, 李伟, 刘溉, 等. 基于德尔菲法和层次分析法的“卡脖子”关键技术甄选研究: 以生物医药领域为例[J]. 世界科技研究与发展, 2021, 43(3): 331-343.
- [22] 刘梦明, 何文英, 李珂, 等. 联合使用德尔菲法与层次分析法构建某附属医院学科建设评价指标体系的研究[J]. 中国医药科学, 2019, 9(17): 18-22.
- [23] 阎凌宇, 程丽云, 李威君, 等. 基于AHP-联系熵的应急预案实施过程风险评估[J]. 中国安全生产科学技术, 2020, 16(2): 130-135.
- [24] 郭海燕. 工业机器人国内外行业发展现状概述[J]. 机电产品开发与创新, 2021, 34(3): 143-144, 154.
- [25] 刘媛, 姚缘, 钱琳. 工业机器人产业竞争情报分析[J]. 中国科技资源导刊, 2017, 49(6): 36-44.
- [26] 黄鲁成, 蒋林杉, 吴菲菲. 萌芽期颠覆性技术识别研究[J]. 科技进步与对策, 2019, 36(1): 10-17.

(上接第12页)

- [10] 杨鹤林. 康奈尔大学: 让研究者共享科学数据[J]. 中国教育网络, 2014(4): 32-34.
- [11] 覃丹. 英美社会科学数据管理与共享服务平台调查分析[J]. 图书情报工作, 2014, 58(16): 67-75.
- [12] FEDERER L M, BELTER C W, JOUBERT D J, et al. Data sharing in PloS one: an analysis of data availability statements[J]. PloS one, 2018, 13(5): e0194768.
- [13] PIERCE H H, ANURUPA D, EMILY S, et al. Credit data generators for data reuse[J]. Nature, 2019, 570(7759): 30-32.
- [14] 国家科技基础条件平台中心. 国家科学数据资源发展报告2018[M]. 北京: 科学技术文献出版社, 2019: 8.
- [15] 陈军, 李大博, 杨智君, 等. 计量科学领域科技项目数据汇交管理研究[J]. 中国科技资源导刊, 2021, 53(4): 83-89.
- [16] 全国科技平台标准化技术委员会. 科技计划形成的科学数据汇交 技术与管理规范: GB/T 39912-2021 [S]. 北京: 中国标准出版社, 2021.
- [17] LI X, CHENG G, WANG L, et al. Boosting geoscience data sharing in China[J]. Nature geoscience, 2021, 14(8): 541-542.
- [18] 庄倩, 何琳. 科学数据共享中科研人员共享行为的演化博弈分析[J]. 情报杂志, 2015, 34(8): 152-157.
- [19] 何琳, 常颖聪. 科研人员数据共享意愿研究[J]. 图书与情报, 2014(5): 125-131.
- [20] 国家计量科学数据中心. 平台介绍[EB/OL]. (2021-09-23) [2021-09-23]. <http://www.nmdc.ac.cn/ptjs/index.jhtml>.
- [21] 中国计量科学研究院. 国家计量科学数据中心项目数据汇交管理办法[Z]. 2020.
- [22] 赵华, 周国民, 王健. 基于元数据的科学数据属性特征分析[J]. 情报杂志, 2015, 34(7): 173-178.
- [23] 王维佳. 基于SWOT分析的数据科研环境下图书馆数据素养教育策略研究[J]. 图书与情报, 2017(5): 104-113.
- [24] 全国科技平台标准化技术委员会. 科技平台 资源核心元数据: GB/T 30523-2014 [S]. 北京: 中国标准出版社, 2014.