·管理纵横·

面向数据中台的全周期数据安全管理研究与初步实践

——以国家自然科学基金数据管理为例

郝艳妮[1†] 李 东[1†] 韩陆超[1] 彭升辉[1] 刘西蒙[1,2*]

1. 国家自然科学基金委员会 信息中心，北京 100085
2. 福州大学 计算机与大数据学院，福州 350116

[摘 要] 国家自然科学基金委员会作为国家科研资助体系的重要组成部分，着力促进信息化与科研活动、科研管理体系的融合。科学基金数据作为新型生产要素，是数字化、网络化、智能化的基础，已快速融入科学基金服务管理等各个环节，用于持续提升科学基金资助效能，为推动基础研究高质量发展。本文介绍现有自然科学基金数据现状，分析目前数据管理中面临的挑战，设计了适用于现状的通用数据中台架构(以下简称“数据中台”)，构造了面向数据中台的安全系统结构，并给出基于数据中台的数据安全管理实践，该工作可以实现自然科学基金委数据中台中存储的数据在创建、存储、发布、访问、处理、重用过程中保证数据全生命周期安全性，有效促进自然科学基金委数据业务化数据长效优质管理的建设发展。

[关键词] 国家自然科学基金；系统性改革；服务架构；数据安全；数据管理；数据中台

中共中央、国务院印发《关于构建数据基础制度更好发挥数据要素作用的意见》(以下简称《数据二十条》)正式颁布，描绘了数据基础制度的“四梁八柱”，对于充分激发数据要素价值具有全局性、奠基性、引领性重要作用。数据作为新型生产要素，是数字化、网络化、智能化的基础，已快速融入生产、分配、流通、消费和社会服务管理等各环节，深刻改变着生产方式、生活方式和社会治理方式。数据来源也已从单一的业务数据向复杂的多源数据转变，数据格式也已经从以结构化为主向结构化与非结构化多种模式混合的方向转变，数据来源更加多元化，数据格式也更加多样化。近几年涌现出了大量新的数据应用技术，如非关系型的数据库(Non-relational Structured Query Language，NoSQL)、新的可扩展/高性能数据库(New Structured Query Language，NewSQL)和分布式数据库等，以及与数据采集、数据存储、数据建模和数据挖掘等大数据相关的技术。

现有信息系统已经从数据文件、数据仓库、数据平台向数据中台模式转变，如图 1 所示。数据中

刘西蒙 福州大学研究员，博士生导师，网络系统安全福建省高校重点实验室主任。主要研究领域为数据安全与密码学。致力于密态计算理论及相关安全理论研究。发表学术论文 100 余篇。主持国家自然科学基金面上项目、福建省自然科学基金“杰出青年”项目等多项研究课题。担任《通信学报》《网络与信息安全学报》等多个学术期刊的编委。

李东 国家自然科学基金委员会信息中心研究员。主要研究领域为科技政策与科研信息化。致力于科学基金信息化建设及相关政策理论研究，发表学术论文 20 余篇。

郝艳妮 国家自然科学基金委员会信息中心副研究员。主要研究方向为数据库管理与使用、计算机架构分析、计算机软件应用与维护。多年从事信息系统的建设与管理工作，发表学术论文 10 余篇。

收稿日期：2024-01-16；修回日期：2024-05-13

† 共同第一作者。

* 通信作者，Email：liuxm@nsfc.gov.cn

台[1]是指通过数据技术，对海量数据进行采集、计算、存储、加工，同时统一标准和口径。数据中台把数据统一之后，会形成标准数据，再进行存储，形成大数据资产层，进而为用户提供高效服务。数据中台的发展历程可以分为三个阶段：第一阶段是业务数据化，主要表现为数据采集、存储、处理的工具化；第二阶段是数据业务化，强调数据在业务场景中的应用，以数据驱动决策；第三阶段是数据智能化，通过引入人工智能等技术，实现数据的智能分析与预测。数据中台需求的出现，是企业数字化转型的一个标志性的转折，转型也正从流程优先走向数据优先，数据中台承担着赋予业务以数据和智能能力的职责，只有打通各业务系统的数据孤岛，将数据标准、口径、模型、存储统一，形成具备完整性、规范性、一致性、准确性和及时性的高质量数据，才能逐渐释放数据价值，数据中台是数据创新效率的保障[2]。

国家自然科学基金委员会（以下简称“自然科学基金委”）作为国家科研资助体系的重要组成部分，着力促进信息化与科研活动、科研管理体系的融合，充分挖掘信息化对科研资助业务和科研资金管理的主动支撑作用[3]；在科学基金管理、开放共享、知识服务、办公自动化、基础设施等方面均取得了显著成效，建成了较为完善的科学基金信息化管理体系[4]。随着自然科学基金委全面深化改革工作的落实，以及科学基金全流程管理要求的变化，也出现了系统独立、互不连通，用户使用信息系统仍然存在高峰时段响应速度较慢、界面交互不够智能、缺少多语种支持等问题。数据中台的出现可以有效改变现有信息系统的现状。

本文以自然科学基金委为研究对象，介绍了自然科学基金委数据管理现状，分析出现有数据管理中遇到的瓶颈与挑战；给出了适用于自然科学基金委的数据中台框架，探索解决各个系统存在数据孤岛、高峰时段响应慢、数据访问安全等问题，破解现有数据中台的异构数据安全存储的难题；设计了全周期数据中台的隐私保护框架，包括：密码学算法模块、访问控制模块、身份认证模块、数据分类分级模块四大模块，实现了数据中台数据全周期过程中的安全与隐私性。

1　自然科学基金委数据管理现状与挑战

在数字新基建的大背景下，自然科学基金委将汇集多种模式下的数据，借助深度学习和人工智能等智能技术，优化业务流程，实现业务流程的智能化，通过用户行为分析提升用户体验，实现精准指派和安全风险管控，实现数字化和智能化的科学基金管理，提升数字智能化水平。这些技术解决实际业务问题的能力越来越强，但与此同时也增加了技术实现的复杂度。

目前，自然科学基金委主要管理的数据从不同的信息系统中产生，包含：科学基金项目基础数据、全流程项目管理数据、科学基金资助成果类数据、公文运转类数据、行政保障类数据、技术支撑类数据等。

科学基金项目基础数据与全流程项目管理数据主要来源于“科学基金网络信息系统”（以下简称“业务系统”），包含科学基金项目在申请、评审、立项、在研、结题、成果、资金、变更等业务过程中收集或产生的大量数据，有格式化数据 1 TB、文档类数据 56 TB，格式化数据从项目、人员、成果角度分不同的库存储。目前，业务系统注册用户超过 100 万，资助项目总数量超 73 万，申请项目总数量超过 360 万。

科学基金资助成果类数据主要来源于“国家自然科学基金大数据知识管理服务门户”[5]（以下简称“大数据服务”），该信息系统收集了科学基金资助项目研究过程中部分论文的元数据与全文数据，同时也收集了部分期刊的论文公开数据。目前，大数据服务已公开论文全文 112.1 万篇，涉及研究机构 28 万家，作者 101 万个；2023 年日均网页访问量 175.7 万次。

公文运转类数据与行政保障类数据主要来源于“国家自然科学基金委员会办公 OA 平台”（以下简称“办公服务”），该信息系统收集了自然科学基金委日常办公中的人事、考勤、公文运转、会议、资产等流

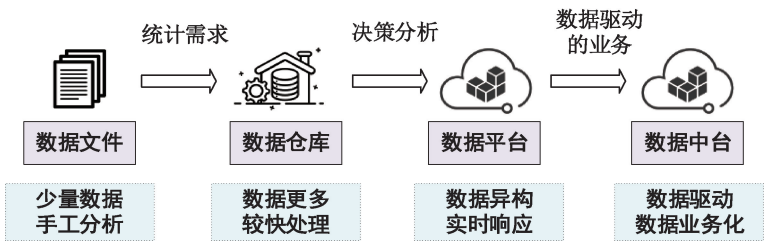

图 1　数据存储平台的演化

程数据以及文件类数据。目前，办公服务年登录量20.3万人次，公文数据6 800条，业务类数据3.03万条。

技术支撑类数据主要是由自然科学基金委数据中心（以下简称“数据中心”）负责管理。目前，数据中心总体架构包括生产中心服务、同城灾备中心服务、异城灾备中心服务三个部分组成。数据中心服务共存储了269.23 TB的数据，其中申请和成果数据约114 TB，Mongodb数据约41.2 TB，结构化数据库数据约6.03 TB，往年历史数据约108 TB。

自然科学基金委目前的数据已分别实现了业务类和办公类的集成整合。通过上述特点，发现自然科学基金委几类数据仍面临下述挑战：（1）业务系统数据与大数据服务数据依赖性强，两系统存储的小文件定期需要共享、大量的格式化业务数据需要同步，成果等信息需要实时共享，业务系统的并发访问量大，现有系统建设架构已经无法满足当前文件与数据的访问需求；（2）业务系统数据和办公数据仍相对独立，在业务系统和办公服务之间尚未做到充分的资源共享和互联互通，如：人员信息变动、相关规定限制等系统组件无法做到相互复用；（3）基础设施需重复建设，未来新建系统需要重复构建，自然科学基金委各系统间的协同管理和互联互通需要整体重新规划。面临上述挑战，以下将探索适用于自然科学基金委的通用数据中台架构，可有效解决存储文件琐碎、系统相对独立、设施重复建设三大挑战。

2 面向自然科学基金委通用数据中台架构设计

自然科学基金委可通过重新设计，减少功能冗余和提高功能复用为原则，体系架构综合考虑了数据中台的各种要素，实现一次规划、分步实施，可以有效提升数据资产价值。新的架构设计充分考虑了自然科学基金委在数据管理中一些实践基础，新设计的通用数据中台架构主体如图2所示。数据中台可以解耦为6个可以分别独立建设、演进的功能子框架，包含数据存储框架、数据采集框架、数据处理框架、数据治理框架、数据安全框架及数据运营框架六大部分[6]。

2.1 数据采集框架

数据中台的采集框架应对纳入数据中台的各种源数据进行统一采集管理。数据采集框架中应提供多种数据采集方式，如：文件上传采集、数据库采集、接口应用程序接入采集、流式采集及网络爬虫采集。同时，数据采集框架应按照数据采集规范对源数据进行预处理，从而去除明显的多余数据，并对采集过程进行管理。自然科学基金委的项目基础数据、全流程项目管理、公文运转类数据、行政保障类数据主要用到数据库与文件上传采集，科学基金资助成果类数据用到了流式、网络爬虫与接口接入采集。因此，数据采集框架可以涵盖现有自然科学基金委数据管理形式。

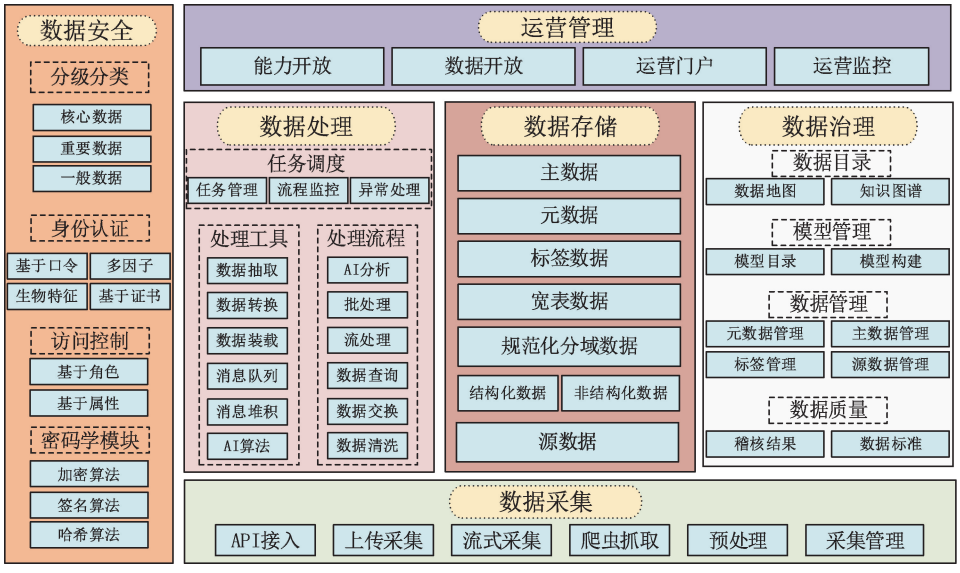

图2 面向通用数据中台的整体架构

2.2　数据处理框架

数据中台的核心是数据，数据通过采集系统获取，框架应对不同类型的数据进行存储。数据存储框架中，无论数据采用对象存储、块存储还是数据库存储技术，各种中台数据可按照图 2 所示分类管理。源数据主要由采集框架进行管理，数据治理框架按照数据特征把数据简单分为结构化和非结构化数据两大类，而规范化分域数据则是数据治理框架对全量数据的规范化分域整理。宽表数据是数据关联的结果，利用宽表数据可以对人、事、地、物、组等对象进行完整的数据画像，宽表数据也可以作为上层模型数据的中间层数据。元数据和标签数据都是对数据的描述，其中元数据用来对数据的客观属性进行表示，标签数据更倾向于管理者对数据的主观表述及等级划分，比如：质量等级标签、安全标签、属性标签等。主数据需要在各系统间频繁更新、交换，且需要独立的存储空间进行维护管理。数据存储框架将对自然科学基金委正在进行的数据分级分类工作起到指导性的作用。

2.3　数据存储框架

数据处理是每个数据应用的基本环节之一，经典的数据抽取、转换和加处理流程在数据采集预处理、数据整合、数据建模等多个地方均要使用。单独建设数据处理框架有利于数据处理工具组件的集中开发与管理，也有利于数据中台数据处理任务的协调与调度。数据处理框架专门负责数据处理相关的任务，包括：批处理、流处理、人工智能分析、数据清洗、数据交换及查询，此外数据处理的相关工具组件可在处理框架中配置。任务调度模块在数据处理框架中处于居中指挥的作用，并对运行的数据处理任务进行监控及异常处理等操作。目前，自然科学基金委的大数据服务中已经引入了流水线的数据处理方式，可针对不同来源的论文数据进行清洗与处理。在新数据中台架构中，可将目前已经在用的流水线做为一种数据处理的成熟手段加以推广应用。

2.4　数据治理框架

数据治理框架包含数据目录、数据管理、模型管理和数据质量四个模块：数据目录主要作用是展示数据的属性及相互关系，包括：数据地图、数据资产目录、知识图谱及数据血缘。数据管理模块用作对各个数据的管理，可以细分为元数据管理、主数据管理、标签数据管理及源数据管理。模型管理能提高数据中台对外部应用需求的反应能力，固化的中间模型数据需要专门管理。数据质量管理模块按照制定的数据标准及数据稽核规则对数据中台中的数据进行质量管理。数据治理相当于对基础数据通过一些技术进行初步的整理与规划，为上层的数据运营提供基本支撑。自然科学基金委目前已在一些小的场景中进行探索应用，比如构建学科知识图谱。

2.5　数据安全框架

数据安全框架是数据中台必不可少的组成部分，叠加在数据中台其他功能框架之上，数据中台在数据采集、处理、交换、共享等每个环节均必须实施安全控制策略。安全框架可以分为分级分类、用户认证、权限访问控制管理及加解密等功能模块，框架也可以对外提供安全能力封装，展示数据中台的安全态势及安全视图。近几年，自然科学基金委面临的最大信息安全风险就是数据安全，如何建立一套有效的数据安全控制框架已迫在眉睫。

2.6　数据运营框架

数据中台的核心功能是综合众多数据应用的数据处理及数据治理功能，如：集中建设、集中管理、减少冗余、增加复用。数据运营框架一方面有利于针对外部用户提供针对性功能，另一方面作为用户与数据中台核心数据服务之间的中间层，可有效隔离外部用户直接控制、接触核心数据及应用，并保护数据中台的安全性及内部功能的稳定性。现有自然科学基金委的数据总线，简单实现了上层不同信息系统之间的数据交换，但是由于没有一套完整的数据生产、使用、管理的规范，还远远没达到数据运营的程度。

3　面向自然科学基金通用数据中台架构设计

数据安全架构作为数据中台其他框架的基础，贯穿数据中台全流程管理。目前，自然科学基金委在数据管理中存在一些的安全与隐私问题[7]：(1) 异构设备在采集不同类型数据在上传过程中加密标准都不一致性。现有的服务建设框架异构，数据采集标准不一致，运用的加密手段也不一致，导致数据安全框架无法为其他数据中台基础模块提供基础加密服务。(2) 现有多元异构业务数据安全存储格式不一致。现有不同业务类型的数据利用不同数据库单独进行存储，数据在不同业务服务中利用的安全存储算法与规则也不经相同，这就导致多元异

构数据在不同业务服务中的数据难以安全融合存储。(3) 存储加密数据在密态数据下实现数据处理。现有加密技术算法在对数据进行加密过后,加密数据的密文信息无法反应出明文的任何信息。(4) 数据输出过程中的访问控制,进行严格的访问授权,以确保数据无法被非法访问。

自然科学基金委建设面向数据中台安全架构可以有效解决数据中台的安全难题,为数据提供数据安全[8]、访问控制[9]、身份认证[10]、数据分级分类[11]等服务。当数据中台为上层服务提供数据接口服务,数据中台可以在隐私保护的情况下,利用多种的数据安全模型,对数据进行相对应的计算,并将计算结果依据不同的业务类型根据不同的访问控制结构进行加密。当上层数据服务需要数据中台提供数据时,中台将数据传输给服务业务请求方,请求方根据自身的访问控制权限,利用密钥对数据进行解密,获取最后的业务数据。在这种业务模型下,数据安全框架应包括以下组成部分:密码学算法模块、访问控制模块、身份认证模块、数据分类分级模块。

3.1 密码学算法模块

密码学算法基础模块用于支撑数据加解密对数据基础进行保护,实现数据的可验证,包括:对称加密算法、公钥加密算法、哈希算法、数字签名算法以及其他可扩展的密码算法。对称加密算法用于加密数据块,实现对数据、文件的保护;公钥加密算法用于实现密钥传输与对称密钥保护;哈希算法能将目标文本转换成具有相同长度的、不可逆的杂凑字符串,用于形成消息摘要对整个消息的完整性进行校验;数字签名算法用于数据的完整性保护;可扩展的密码算法包括现有可证明安全的主流其他加解密算法,用于支撑数据访问控制与数据可信流通,如:同态加密算法用于实现加密数据的计算、属性基加密算法实现密文存储过程中的细粒度访问控制等。

3.2 访问控制模块

访问控制模块可以防止对任何资源进行未授权的访问,从而使计算机系统在合法的范围内使用。目前被广泛采用的两种权限模型为:基于角色的访问控制和基于属性的访问控制。

(1) 基于角色的访问控制(Role-based Access Control,RBAC),指的是通过用户的角色(Role)授权其相关权限,这实现了更灵活的访问控制,相比直接授予用户权限,要更加简单、高效、可扩展。只需要为该角色制定好权限后,给不同的用户分配不同的角色,后续只需要修改角色的权限,就能自动修改角色内所有用户的权限。

(2) 基于属性的访问控制(Attribute-based Access Control,ABAC)是一种非常灵活的授权模型,通过各种属性来动态判断一个操作是否可以被允许。在 ABAC 的执行过程中,决策引擎会根据定义好的决策语句,结合对象、资源、操作、环境等因素动态计算出决策结果。每当发生访问请求时,ABAC 决策系统都会分析属性值是否与已建立的策略匹配,如果有匹配的策略,访问请求就会被通过。

3.3 身份认证模块

数字中台中身份认证模块是在计算机网络中确认操作者身份的过程而产生的有效解决方法。作为防护网络资产的第一道关口,身份认证有着举足轻重的作用。

基于口令的身份验证依赖于用户名和口令或个人身份识别码(Personal Identification Number,PIN)。口令是最常见的身份验证方法,但更容易受到网络钓鱼和暴力攻击。

双因素身份验证(Two-factor Authentication,2FA)要求用户提供除密码之外的至少一个附加身份验证因素,附加因素可以是其他身份验证类型或通过文本或电子邮件发送给用户的一次性密码,其中附加因素与原始设备位于不同通道上,以减轻中间人攻击。

生物识别技术使用用户自身生物特征进行验证,由于生物识别身份也是唯一的,这使得破解帐户变得更加困难。常见的生物识别类型包括:指纹扫描、面部识别、虹膜识别、行为生物识别等。

证书身份验证技术使用证书颁发机构颁发的数字证书和公钥加密来验证用户身份。证书存储身份信息和公钥,而用户拥有虚拟存储的私钥。这种身份验证类型适用于雇用临时需要网络访问的用户。

3.4 数据分级分类模块

数据分类分级是数据安全保护最重要的基础性工作,针对不同类型、不同级别的数据制定和采取不同的安全保护措施,从而实现数据安全保护与数据流通利用的平衡。数据分类分级模块旨在对数据资产盘点梳理并进行标准化、专业化管理,将常见、稳定的属性或特征作为数据分类的依据,从而便于依

照类别建立完善有序的数据架构，以实现高效准确的数据管理与使用。

数据分类分级模块基于数据在经济社会发展中的重要程度，以及一旦遭到泄露、篡改、破坏或者非法获取、非法利用，对国家安全、公共利益或者个人、组织合法权益造成的危害程度，通过定量与定性相结合，根据数据分级要素开展数据影响分析，从而为数据划分不同级别。数据处理者应针对不同级别的数据，制定对应的安全策略、确定适当的对外开放程度并在全生命周期采取不同的安全保护措施。根据数据分级与影响对象、影响程度的关联关系，将数据按照级别从高到低分为核心数据、重要数据、一般数据三类。可以按照重要数据、核心数据、一般数据的顺序确定数据级别，在参考重要数据识别相关标准识别出重要数据后，再根据是否可能直接影响政治安全、国家安全重点领域、国民经济命脉、重要民生、重大公共利益来确定相关数据是否为核心数据。如果不属于重要数据或者核心数据，则相关数据可确定为一般数据，具体见表 1。

4　基于数据中台数据安全的管理建议与展望

建立起统一的数据中台后，自然科学基金委业务数据在输出的过程中需要建立严格的数据访问控制机制，在进行访问控制与授权情况下，对中台数据进行访问，以确保数据进行越界或者非法访问。除了利用技术手段确保数据中台的安全性，自然科学基金委数据中台中存储的数据在创建、存储、发布、访问、处理、重用过程中，采取一系列管理措施保证数据全生命周期安全性[12]。在数据创建阶段，一方面针对数据内容本身进行安全性评估，根据涉密程度利用数据分类分级模块进行分级分类；另一方面要定期审查新兴、颠覆性技术对数据中台的安全问题带来新的挑战，分析现有人工智能等技术[13, 14]，对数据中台中存储的数据带来的关联、挖掘、融合风险。在数据存储阶段，《中华人民共和国数据安全法》《中华人民共和国个人信息保护法》也明确要求数据安全、个人信息保护的制度建设，也对于自然科学基金委处理数据（尤其是个人信息）提出了更为明确、清晰的要求，密码学算法模块、访问控制模块、身份认证模块用于应对数据流失带来的数据主权和安全问题。在发布访问阶段，根据数据分级分类的等级，通过访问控制模块与数据加密传输模块，保证自然科学基金委的数据受控、合法的被访问和传输。在重用阶段，经过识别和评估长期保存的数据，应当进行全生命周期的监控，以保证数据资源的质量、完整性、机密性和安全性不受到时间推移受到损害。

表 1　数据分级分类方法

影响对象	影响程度		
	特别严重危害	严重危害	一般危害
国家安全	核心数据	核心数据	重要数据
经济运行	核心数据	重要数据	重要数据
社会稳定	核心数据	重要数据	一般数据
公众利益	核心数据	重要数据	一般数据
组织/个人权益	一般数据	一般数据	一般数据

当前，《中华人民共和国网络安全法》《中华人民共和国数据安全法》《中华人民共和国个人信息保护法》共同构建了我国信息安全的法律框架，数据中台安全性除了符合上述法律法规之外，还需要：（1）制定自然科学基金委相关安全的规章制度。制定网络与信息技术安全总体规划，加强数据安全管理与技术研究，并在实际信息系统建设中予以落实。（2）开展系统性实证研究，在实现总体安全下，适应自然科学基金委数据中台的特点、兼顾新技术和社会发展需求的数据中台安全治理方法研究。（3）进一步明确数据规模等对数据中台的具体界限，指定既能保证数据安全，又能保证数据在各系统开放贡献的解决方案。

参　考　文　献

[1] 吴信东，应泽宇，盛绍静，等. 数据中台框架与实践. 大数据，2023，9(6)：137—159.

[2] 苏萌，贾喜顺，杜晓梦，等. 数据中台技术相关进展及发展趋势. 数据与计算发展前沿，2019，1(5)：116—126.

[3] 李东，郝艳妮，彭升辉，等. 国家自然科学基金委员会信息化建设现状及智能化发展展望. 中国科学基金，2023，37(2)：307—312.

[4] 国家自然科学基金委员会. 国家自然科学基金“十四五”发展计划. (2022-11-22)/[2024-06-28]. https://www.nsfc.gov.cn/publish/portal0/tab1392/.

[5] 姚畅，王晓帆，杜一，等. 国家自然科学基金大数据知识管理服务平台总体方案及关键技术研究. 中国科学基金，2019，33(1)：55—61.

[6] 李东，郝艳妮，彭升辉，等. 国家自然科学基金委员会网络安全现状与展望. 网络与信息安全学报，2022，8(6)：92—101.

[7] 杨世旺，赵萍，黄剑，等. 中台数据库信息共享适用性和安全性的增强. 云南师范大学学报(自然科学版)，2023，43(5)：54—58.
[8] 冯登国，张敏，李昊. 大数据安全与隐私保护. 计算机学报，2014，37(1)：246—258.
[9] 王静宇，张伟. 结合属性和RBAC的访问控制模型及算法研究. 小型微型计算机系统，2022，43(7)：1523—1528.
[10] 张淑娥，田成伟，李保罡. 基于区块链技术的身份认证研究综述. 计算机科学，2023，50(5)：329—347.
[11] 胡康，郭金成，李健. 数据分类分级标准化探索——以某研究院为例. 中国信息化，2023(9)：54—56.
[12] 李宜展，刘细文，李泽霞，等. 科学数据安全边界概念模型研究——基于利益相关者视角. 中国科学基金，2022，36(2)：339—347.
[13] 刘西蒙，谢乐辉，王耀鹏，等. 深度学习中的对抗攻击与防御. 网络与信息安全学报，2020，6(5)：36—53.
[14] 熊金波，毕仁万，田有亮，等. 移动群智感知安全与隐私：模型、进展与趋势. 计算机学报，2021，44(9)：1949—1966.

Full-cycle Data Security Management Research and Consulting Practice for Data Middleware System—Take National Natural Science Foundation of China Data Management as An Example

Yanni Hao[1†] Dong Li[1†] Luchao Han[1] Shenghui Peng[1] Ximeng Liu[1,2*]

1. *Information Center, National Natural Science Foundation of China, Beijing* 100085
2. *College of Computer and Data Science, Fuzhou University, Fuzhou* 350108

Abstract As an important part of the national scientific research system, the National Natural Science Foundation of China (NSFC) strives to promote the integration of the committee's informatization with scientific research activities and scientific research management systems. As a new production factor, science fund data is the basis for digitization, networking, and standardization. It has been integrated into all aspects of the bladder science fund service and management to continuously improve the goals and tasks of the science fund and promote the high-quality development of basic research. This article introduces the current status of the existing Natural Science Foundation data, analyzes the current challenges faced in data management, designs a data middle platform construction architecture suitable for the current natural science fund data status, builds a security system structure for the data middle platform, and provides It has developed management practices based on data security in the data center. This work can ensure the security of the entire life cycle of data in the data creation, storage, release, access, processing, and reuse processes stored in the data center of NSFC, and effectively promote the natural science and technology development of natural science and technology. The construction and development of data businessization and long-term high-quality data management of NSFC.

Keywords National Natural Science Fund of China; systemic reform; service architecture; data security; data management; data middle platform

(责任编辑 陈 磊 张 强)

† Contributed equally as co-first authors.
* Corresponding Author, Email: liuxm@nsfc.gov.cn