



面向科学数据全生命周期的动态安全评估机制

聂晓伟^{1,2,3}, 潘小多^{1,3*}, 李新¹, 汪寿阳^{4*}, 金婧^{1,5}, 杨洋^{2,3}

1. 中国科学院青藏高原研究所, 国家青藏高原科学数据中心, 北京 100101;

2. 西藏大学生态环境学院, 拉萨 850000;

3. 广东省灵山论坛科学中心, 广州 511466;

4. 中国科学院数学与系统科学研究院, 北京 100190;

5. 中国人民大学生态环境学院, 北京 100872

* 联系人, E-mail: panxd@itpcas.ac.cn; sywang@amss.ac.cn

科学数据是指通过基础研究、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用于科学研究活动的原始数据及其衍生数据。科学数据不仅是科学研究创新和社会经济发展的战略性、基础性资源^[1], 也是政府部门制定政策、进行科学决策的重要依据, 是科技进步、国家安全的重要保障, 在国家科技创新格局中占据重要地位。加快建设科技强国, 实现高水平科技自立自强离不开科学数据的引擎作用。科学数据的生命周期包括: 数据采集、存储、传输、处理、销毁。

科学数据安全性是采取措施保障科学数据能在持续安全状态下处于有效保护和合法利用的状态。基于数据安全保密性、完整性、可用性三大特性, 科学数据安全更具动态性, 伴随数据处理活动及数据安全检测, 数据处理者、用户可信度、数据保密级等处于动态变化之中, 相应等数据安全管理工作也应动态调整以达到保护目的。

随着数据密集型科学研究范式的兴起, 世界各国都在积极推进科学数据开放共享, 我国新成立的国家数据局出台了《“数据要素×”三年行动计划(2024~2026年)》, 强调科学数据在推动产业发展和区域发展中的作用。如何在促进开放共享的同时保障科学数据安全这一新挑战随之而来。国务院办公厅于2018年3月17日印发《科学数据管理办法》, 专门提出建立我国科学数据管理的基础制度, 推进落实科学数据安全保障。但是在实施过程中我国科学数据管理仍存在制度体系不健全^[2,3]、政策工具使用不够均衡、生命周期管理有待完善^[4]等问题。

本文总结和分析了科学数据安全评估研究现状及存在的不足, 提出了面向科学数据全生命周期的动态安全评估模型、安全指标体系, 以及评估方法, 形成了科学数据安全评估机制。并将上述成果应用到青藏高原科学数据的安全访问中, 验证了安全评估机制的科学性和可行性。



潘小多 博士, 中国科学院青藏高原研究所特聘骨干研究员、博士研究生导师。主要从事区域气候变化、数据同化、数据集成和大数据分析等方向的研究工作。



汪寿阳 博士, 中国科学院特聘研究员, 中国科学院预测科学研究中心主任, 上海科技大学创业与管理学院院长, 发展中国家科学院院士、国际系统与控制科学院院士。主要研究方向为决策分析、最优化与供应链管理、金融创新风险管理、经济分析与预测、能源环境经济学与政策分析。

1 科学数据的研究趋势和进展

2000年至今, Web of Science(WOS)数据库中以“data security”为主题文献总数达566722篇, CNKI数据库中以“数据安全”为主题文献总数达43337篇, 可见大数据背景下的数据安全问题已引起学界广泛关注, 就年度发文趋势可见, 数据安全相关研究热度呈逐年上升趋势。通过全时段数据安全相关文章词云分析发现数据安全相关论文主要聚焦在数据隐私、区块链、物联网等领域(图1)。同时, 数据安全相关研究



在密码学、数据模型、医疗保健、数据共享等方面也具有广泛的研究潜力.

ISO 27001是国际上最广泛接受的信息安全标准之一^[5]。该标准提供了一套全面的信息安全管理系统(ISMS)框架,涵盖了安全策略、组织安全、人员安全、资产管理、访问控制等多个方面。ISO 27001的优点是全面性和灵活性,适用于各种规模和类型的组织。但它也存在一定的劣势,如实施成本高,需要专门的资源和专业知识,以及对持续改进的长期承诺。中国已经建立了相对完整的数据安全评价标准和体系,包括《信息安全技术 信息系统安全等级保护基本要求》(GB/T22239-2008)、《信息安全技术 网络安全等级保护基本要求》(GB/T 22239-2019)等。这些标准强调了防御深度、防御全面、持续改进和应对未知威胁等原则,对指导我国的数据安全工作具有重要参考作用。然而,这些标准的缺点在于固定性较强,对新的威胁和技术挑战适应性不强。综合国内外的数据安全评价标准和体系,普遍存在灵活性和通用性不足的问题,没有一套完全适合所有情况的标准或体系。因此,对于科学数据安全,应综合考虑特定需求、资源和风险,基于动

杰安全评估制定一个最佳标准或体系。

数据分级分类目前是国际上数据安全管理的的重要手段之一,近年来逐渐成为学术界和产业界的热点。Alonge等人^[6]构建了一个网络安全风险评估的数据分类模型,为网络安全管理提供了有力工具。Ahmadi等人^[7]对医疗领域的数据分类进行了系统性的回顾,总结了该领域的研究现状和挑战。然而,尽管各种数据分级分类方法层出不穷,但往往缺乏对实际应用环境的适应性和灵活性及对于分级分类的长期效果和影响的深入探讨。

综上, 目前针对科学数据在隐私保护, 存储、访问和传输安全, 合规性等方面都进行了大量研究和实践, 但仍存在如下问题: (1) 科学数据的访问需要根据数据的全生命周期开展动态调节, 因此已有的静态、单一安全机制难以实现科学数据的安全贡献; (2) 尚未建立有针对性的科学数据评价标准, 难以保证科学数据评估科学性; (3) 缺乏有针对性的评估方法, 难以保障科学数据评估有效性。

因此,本文立足于科学数据安全管理的切实需求,给出了安全评估指标体系,提出了基于动态信任度的科学数据安全

全评估模型, 设计了科学数据评估方法, 构建了科学数据评估体系, 通过科学数据中心的实践, 验证了所提评估体系的科学性和有效性.

2 科学数据安全评估体系

2.1 评估指标

科学数据安全评估应该根据数据安全标准制定相应的指标体系. 按照机密性、完整性和可用性三大核心原则. 机密性确保敏感信息仅对授权用户可见, 防止数据泄露. 完整性保障数据在存储、传输过程中维持其准确、未被非法修

改. 可用性指数据和资源在需要时可被授权用户访问, 确保业务连续性. 这三层指标共同构成了数据安全的基石, 帮助识别并缓解潜在的安全风险. 数据安全评价指标树状结构如图2所示. 评估指标应该包括基础设施安全、数据传输安全、数据存储安全、数据备份和恢复、安全管理制度等多方面内容, 指标如表1所示.

2.2 动态科学数据安全模型

数据的安全等级在数据全生命周期中是动态变化的, 因此本文给出了动态科学数据安全模型(dynamic security model, DSM), 如图3所示. 该模型结合国家青藏高原科学数据中心

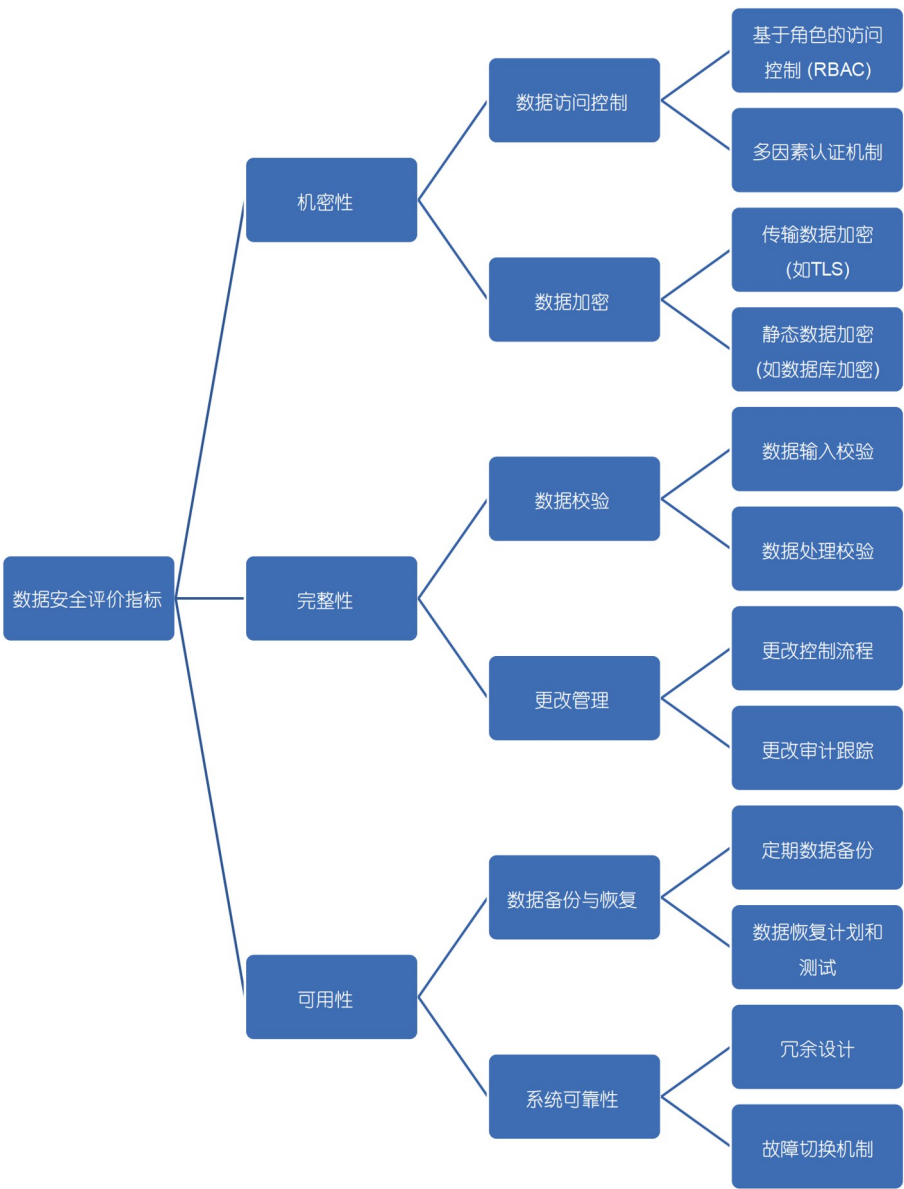


图 2 (网络版彩色)数据安全评级指标树状结构图

Figure 2 (Color online) Tree structure diagram of data security rating metrics

表 1 科学数据安全指标列表

Table 1 List of scientific data security indicators

数据安全环节	注意事项	释义
科学数据的生成	数据源验证	科学数据的来源是否可信、其采集过程是否合法
	数据完整性	数据在存储、传输和处理过程中是否准确、完整, 没有冗余
	数据一致性	科学数据在格式、机制等方面的一致性
	元数据	与科学数据相关的元数据信息, 考察其采集时间、地点等
科学数据的处理	数据格式标准化	是否采用通用的数据格式标准, 确保数据的互操作性和可用性
	数据压缩存储	是否采用较优的数据压缩算法和存储策略, 确保较高的数据存取效率, 同时防止压缩、解压缩过程中数据的破坏
	数据可用性	数据是否具备相应的访问权限并且可用
	加密技术	是否有强鲁棒性的加密技术对数据进行保护
科学数据的传输	传输安全	数据在传输过程中, 是否采取措施防止被拦截或窃取
	日志和监控	记录数据传输的日志信息是否及时、完整、准确
	数据存储	数据的存放位置是否合理, 存储方式是否适当
	数据隐私	数据中的个人或单位信息(如姓名、地址、电子邮件地址、电话号码等)是否有访问权限保护
科学数据的存储	数据备份	是否有定时将数据复制到其他安全的位置以防止数据损坏或丢失的措施
	审计日志	记录所有与数据使用相关事件以便检查和追踪
	身份验证	是否通过身份验证来确保只有授权人员才能访问数据
	防火墙	数据载体是否建立防火墙来保护数据系统的安全
科学数据的管理	病毒扫描	是否定期通过病毒扫描来检测和清除可能存在的计算机病毒
	弱密码检测	是否存在弱密码检验和扫描机制
	存储扩展	确保有足够存储空间, 并具备充分的可扩展性以适应可能的数据增长
	安全培训	是否对数据管理员进行定期的数据安全培训
	规范审查	是否根据实际情况制定相应的数据管理规范
	访问控制	是否有对数据进行安全授权和管理的措施
	数据分类分级	是否根据数据类型进行分类分级采取不同管理措施

将数据类型分为: 常规数据、科技项目产出数据、投稿论文关联数据, 对其安全评估并确定安全状态为: 可共享、不能共享和条件共享, 分别采取: 开放共享、严格访问、安全访问。该模型可以实现在数据全生命周期中, 根据数据类型的动态变化, 根据安全评估动态确定安全状态, 是一种旨在提高数据安全性的模型, 特别是在科学研究和大数据环境中。此模型通过动态评估系统组件或参与者的信任度来调整安全控制措施的强度, 从而优化安全性能和资源利用率。

2.3 评估流程

基于评估指标体系和DSM模型^[8], 给出科学数据的安全评估方法如图4所示。评估流程包括: 评估准备、制定评估计划、开展安全评估、结果分析、报告编写。评估将从科学数据生成、传输、处理、管理和存储全生命周期展开。在DSM

模型基础上, 从主客观角度对科学数据的安全性进行分析(图4)。从主客观分析法出发, 分别使用层次分析法和熵权法对多维指标进行主客观权重计算, 之后采用基于组合权重的灰色关联度决策方法, 通过将定量化表征的主客观权重整合到安全评估中, 实现对科学数据安全的综合评估。对数据的安全评估计算如下:

$$S = \sum_{i=1}^{10} S_i \lambda_i$$

$$= \sum_{i=1}^{10} \sum_{j=1}^n \sum_{k=1}^m \lambda_i E_{i,j,k} \alpha_{i,j,k} \lambda_{i,j,k}, \quad (1)$$

式中, S 表示安全状态, 经过归一化处理, 处于 $\{0-0.5-1\}$ 数据值空间; λ 表示安全要素权重系数; E 代表安全要素危害程度; α 代表安全要素风险概率。 λ_i , $\lambda_{i,j,k}$ 被评估数据的用户或评估发起者在填写评估任务时分配, 结合层次分析法去设定; $E_{i,j,k}$ 通

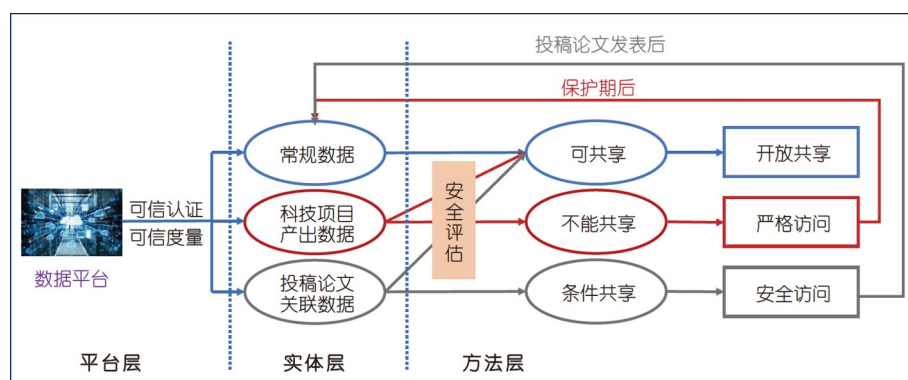


图3 (网络版彩色)动态科学数据安全模型(DSM)模型
Figure 3 (Color online) Dynamic Security Model

过科学数据评估库中的危害程度表获取。根据 S 的数值,建立与安全状态的对应关系,其中“0”属于可共享状态,“0~0.5”属于条件共享状态,“0.5~1”为不可共享状态。在数据类型发生变化时,将触发数据的重新评估机制,重新确定数据安全状态,采取不同的安全控制形式。

3 国家青藏高原科学数据安全评估的实践

国家青藏高原科学数据中心是我国唯一针对青藏高原及周边地区的数据中心,负责青藏高原及周边地区各类科学数据的收集/汇交、存储、管理、集成、挖掘、分析、共享和应用推广,是该区域科学数据门类最全、最权威的数据中心^[9-11]。国家青藏高原科学数据中心对科学数据进行动态评估,是科学数据安全管理的成功实践。在评估环节方面,包括科学数据的生成、处理、传输和管理等环节。在评估对象方面,主要包括常规数据、科技项目产出数据和投稿论文关联数据。其中,投稿论文关联数据,尤其是涉及论文未被正式接收之前的关联数据,是科学数据安全管理的对象。如图3所示,一般而言,常规数据对应科学数据安全评估中的开放共享;科技项目产出数据对应开放共享或者严格访问(设置了保护期, $S \in (0.5, 1]$);投稿论文关联数据对应开放共享或者安全访问(审稿中论文关联数据, $S \in (0, 0.5]$)。随着时间的推移,严格访问的科技项目产出数据在保护期过后(设置 $S = 0$)、安全访问的投稿论文数据在文章接收之后(设置 $S = 0$)会转换为开放共享,会根据改变动态触发数据的重新评估机制,重新确定数据安全状态,采取不同的安全控制形式。

3.1 对常规数据的安全治理

除科技项目产出数据和投稿论文关联数据外的常规数据,国家青藏高原科学数据中心平台为其提供了安全可靠的大数据汇交系统,可由科研人员自行提交。数据的质量由相关领域的专家进行审核,审核通过后再由系统管理人员进行发布,实现数据全生命周期规范化管理,并对数据安全起到

数据作者、数据专家 and 平台管理人员的多重把控。

3.2 科技项目产出数据的安全治理

2018年国务院办公厅印发《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)项目科学数据向科学数据中心汇交。自2019年6月正式成立以来,国家青藏高原科学数据中心已承接了多个科研项目的科学数据汇交与管理,包括第二次青藏高原综合科学考察国家专项、中国科学院“泛第三极环境变化与绿色丝绸之路建设”A类先导科技专项。国家青藏高原科学数据中心还为青藏高原研究相关的项目提供数据汇交共享服务,涉及10多个国家重点研发计划,共计汇交3200多个科学数据集,涵盖青藏高原及其周边地区一系列多尺度、多学科、多类型的科学数据。国家青藏高原科学数据中心对数据用户始终秉承数据开放获取原则。为兼顾项目内科研人员在数据使用权益上的优先级,国家青藏高原科学数据中心对项目数据作者进行了分级分类,不同分级分类对应不同的数据访问权限。项目数据作者具体分为以下五类:项目负责人、课题负责人、子课题负责人、科研骨干和一般用户。另外,根据作者对特殊数据保护的诉求,国家青藏高原科学数据中心对相关数据的共享附加额外条件,包括设置数据申请审批流程(申请获取, $S \in (0, 0.5]$)和可设置不超过两年的数据保护期($S \in (0.5, 1]$),保护期数据访问权限类型在一定期限后自动转换为由数据作者设置的开放获取类或申请获取类。基于上述两种方式,即项目数据作者分级分类和不同数据共享方式(即开放共享、申请获取和保护期),国家青藏高原科学数据中心实现对科技项目产出数据的安全治理。

3.3 对投稿论文关联数据的安全保护措施

投稿论文关联数据是宝贵的前沿科研数据,拥有得越多越有话语权。同时,越来越多的国际出版社和学术期刊要求发表论文时要将数据公开发布到它们推荐的公共数据仓储

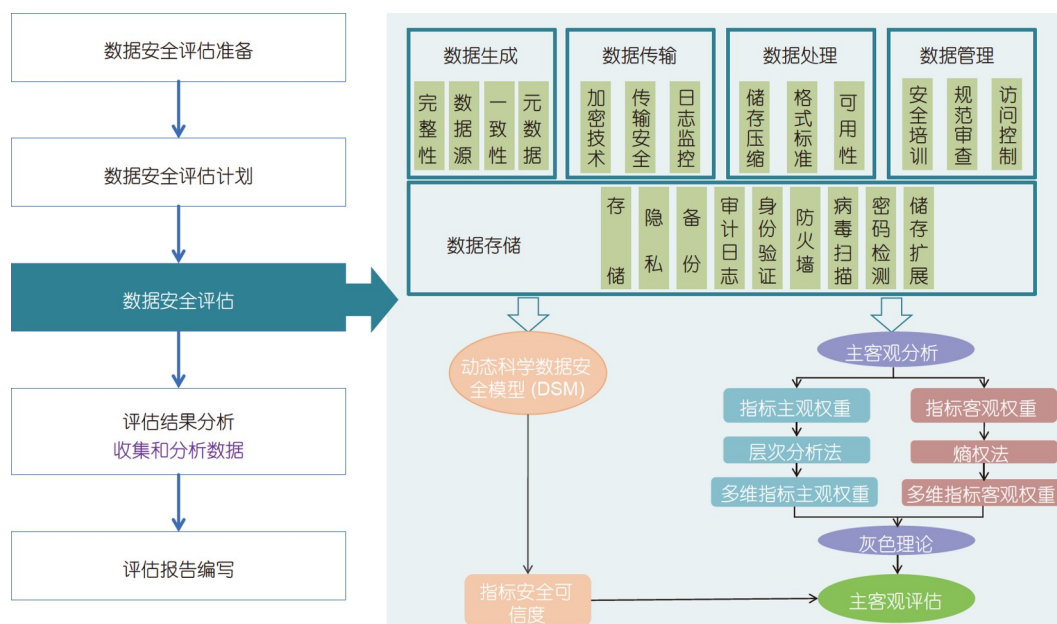


图4 (网络版彩色)数据安全评估流程图

Figure 4 (Color online) Data security assessment flowchart

中,但是我国鲜有科学数据中心在它们的推荐名单上。国家青藏高原科学数据中心于2020年先后成为国内首个Springer Nature旗下*Scientific Data*期刊、AGU(American Geophysical Union)、EGU(European Geosciences Union)和ESSD(Earth System Science Data)等国际重要地球科学出版商和地球科学数据期刊信任的数据仓储中心,为我国地球科学前沿科研数据的安全提供仓储保障和保驾护航。

科研数据的安全问题在很大程度上表现为对审稿中论文关联数据的保护。因涉及的期刊编辑、论文评审专家具有很大不确定性,正在审稿中的论文关联数据面临着被其他数据用户盗用并被抢先用于发表论文的风险。为此,国家青藏高原科学数据中心专门设计了其保障系统——专家免登录设置($S \in (0, 0.5]$),将相应数据归类到“专家免登录预览数据组(审稿中论文所关联数据)”。该设置兼顾保护数据作者权益和专家评审需要,一方面,相应数据在数据中心暂未对公众开放,同时也屏蔽了百度、Google等搜索引擎,保护了数据作者的未发布数据;另一方面,掌握数据链接的数据作者、期刊编辑和评审专家可以免登录浏览和下载数据,投稿期刊的数据评审要求也得以满足。待文章正式接收后,数据作者可以将其关联数据根据期刊要求设置共享模式。

综上,国家青藏高原科学数据中心的科学数据安全评估机制,有助于科研团队在处理各类数据上兼顾共享和安全,使各类科研人员能高效获取和使用数据的同时,让其在数据使用的过程中得到基本的数据知识产权保障。

4 结论

科学数据的安全评估体系对于保护数据安全具有重要意义,本文在梳理科学数据安全评估研究进展及挑战的基础上,构建了安全评估指标、给出了动态评估模型和评估方法,进一步阐述了在当今数据开放共享大背景下,对国家青藏高原科学数据进行评估得出的对常规数据、科技项目产出数据和投稿论文关联数据的评估结果,以及所采取的数据安全措施,为科学数据实现开放和共享作出了探索与贡献。

随着开放科学的边界不断延伸,科学数据面临更严格更精细化的管理要求。就我国而言,科学数据管理基础设施的建设仍在起步,尽管不少数据中心已投身积极实践,但是针对科学数据管理仍缺乏系统性的理论基础、全局性的规划指导及利益主体权责界定。本文提出的科学数据安全评估模型及方法仍需在实践中完善改进,同时,科学数据全生命周期安全管理中仍存在激励不足、投入不足、专家不足等问题亟待解决。科学数据管理是一个长期工程,实现开放科学既是我国新时期科技发展的新要求,也是世界各国携手应对共同挑战的新趋势。在科学数据安全评估的基础上,进一步推动我国科学数据管理升级及开放科学发展,需要从认识层面、政策层面、基础设施层面、经验和技能层面着手多维度综合考虑。坚持问题导向,识别出关键且亟须解决的问题;坚持规划引领,合理设定实施路径和阶段性重点任务;坚持效率优先,着力提升相关工作的针对性和有效性,以科学数据管理为抓手切实推进国家科学现代化进程。

致谢 感谢国家重点研发计划(2023YFF0804901)、拉萨市中电科技计划(LSKJ202407)、西藏自治区科学技术厅“基于仿生鹰眼的高原广域生态环境智慧监测系统”项目(XZ202201ZY0015G)和中国科学院网络安全和信息化专项咨询研究项目(CAS-WX2023ZX02-02)资助。

推荐阅读文献

- 1 Li X, Cheng G, Wang L, et al. Boosting geoscience data sharing in China. *Nat Geosci*, 2021, 14: 541–542
- 2 Li J. Problems and insights of scientific data management from the perspective of archival management (in Chinese). *Office Oper*, 2022, 380: 98–99 [李洁. 档案管理视角下科学数据管理存在的问题及启示. *办公室业务*, 2022, 380: 98–99]
- 3 Liao F Y, Hu L L, Wang J, et al. Research and suggestions on scientific data security standards (in Chinese). *Chin Sci Bull*, 2024, 69: 1142–1148 [廖方宇, 胡良霖, 王健, 等. 科学数据安全标准研究与工作建议. *科学通报*, 2024, 69: 1142–1148]
- 4 Song L Y. Quantitative analysis of domestic and foreign scientific data management policy texts from the perspective of policy tools (in Chinese). Doctor Dissertation. Nanchang: Nanchang University, 2022 [宋李叶. 政策工具视角下的国内外科学数据管理政策文本量化分析. 博士学位论文. 南昌: 南昌大学, 2022]
- 5 Wu W Q, Shi K, Wu C H, et al. Research on the impact of information security certification and concealment on financial performance: Impact of ISO 27001 and concealment on performance. *J Glob Inf Manag*, 2022, 30: 128–143
- 6 Alonge C Y, Arogundade O T, Adesemowo K, et al. Information asset classification and labelling model using fuzzy approach for effective security risk assessment. In: 2020 International Conference in Mathematics, Computer Engineering and Computer Science (ICMCECS), 2020
- 7 Ahmadi H, Arji G, Shahmoradi L, et al. The application of internet of things in healthcare: A systematic literature review and classification. *Univ Access Inf Soc*, 2019, 18: 837–869
- 8 Nie X W, Feng D G. Modified security model based on dynamic trusted degree (in Chinese). *J Commun*, 2008, 29: 37–44 [聂晓伟, 冯登国. 基于动态可信度的可调节安全模型. *通信学报*, 2008, 29: 37–44]
- 9 Li X, Che T, Li X, et al. CASEarth Poles: Big Data for the Three Poles. *Bull Am Meteorol Soc*, 2020, 101: E1475–E1491
- 10 Pan X, Guo X, Li X, et al. National Tibetan Plateau Data Center: Promoting Earth system science on the Third Pole. *Bull Am Meteorol Soc*, 2021, 102: E2062–E2078
- 11 Li X, Feng M, Ran Y, et al. Big Data in Earth system science and progress towards a digital twin. *Nat Rev Earth Environ*, 2023, 4: 319–332

Summary for “面向科学数据全生命周期的动态安全评估机制”

Dynamic security assessment mechanism for the entire lifecycle of scientific data

Xiaowei Nie^{1,2,3}, Xiaoduo Pan^{1,3*}, Xin Li¹, Shouyang Wang^{4*}, Jing Jin^{1,5} & Yang Yang^{2,3}

¹ National Tibetan Plateau Data Center, Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing 100101, China;

² School of Ecology and Environment, Tibet University, Lhasa 850000, China;

³ Science Center of Lingshan Forum of Guangdong Province, Guangzhou 511466, China;

⁴ Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China;

⁵ School of Environment and Natural Resources, Renmin University of China, Beijing 100872, China

* Corresponding authors, E-mail: panxd@itpcas.ac.cn; sywang@amss.ac.cn

Scientific data are strategic and fundamental resources for scientific research and innovation and socio-economic development. Data sharing under open science can accelerate scientific progress and innovation. Scientific data, generated from scientific research activities, require dynamic security access control throughout their entire lifecycle. Scientific data security is the adoption of measures to ensure that scientific data can be effectively protected and legally utilized in a state of continuous security. Based on the three characteristics of confidentiality, integrity and usability, scientific data security is more dynamic. Along with data processing activities and data security testing, the data processors, user credibility and data confidentiality level are undergoing dynamic changes, and data security management should be dynamically adjusted accordingly to achieve the purpose of protection. However, a corresponding security assessment mechanism has not been established yet, and there is a lack of pertinent assessment indicators, models, and methods. This paper provides a comprehensive analysis of the current state and shortcomings of scientific data security assessment research. For the first time, it proposes a dynamic security assessment model, a security metrics system, and an evaluation method applicable to the full lifecycle of scientific data, thereby establishing a scientific data security assessment mechanism. The scientific data security assessment index system proposed in this paper includes infrastructure security, data transmission security, data storage security, data backup and recovery, security management system and other aspects. The system evaluates scientific data security from multiple aspects such as data generation, processing, transmission, storage and management. Then this paper gives a dynamic security model (DSM). The DSM model is based on trustworthy authentication and measurement, defines the trustworthiness of indicators, and can dynamically analyze the dynamic changes of safety assessment information during the operation of scientific data systems. Based on the DSM model and the corresponding indexes, this paper gives the assessment method and the corresponding process of scientific data, including the life cycle of scientific data generation, transmission, processing, management and storage. Based on the DSM model, subjective and objective weights are calculated for multidimensional indicators using hierarchical analysis and entropy weighting methods, respectively, followed by a gray correlation decision-making method based on the combination of weights, which realizes a comprehensive assessment of the security of scientific data by integrating subjective and objective weights of quantitative characterization into the security assessment. These findings have been further applied to the practice of secure access to scientific data on the Tibetan Plateau to verify the scientific validity and feasibility of the proposed security assessment mechanism. The dynamic assessment of scientific data by the National Tibetan Plateau Data Center is a success of the scientific data safety assessment mechanism. The scientific data security assessment mechanism of the Tibetan Plateau Data Center helps scientific research teams to balance sharing and security in handling various types of data, so that various types of researchers can efficiently access and use the data while allowing them to obtain basic data intellectual property protection in the process of data use. The security assessment mechanism of scientific data proposed in this paper is of great significance for realizing the security management of scientific data.

scientific data, data security standards and systems, regional development, data service, National Tibetan Plateau Data Center

doi: [10.1360/TB-2023-1224](https://doi.org/10.1360/TB-2023-1224)