

计量科学大数据分级分类

智峰, 田锋, 赵若凡

中国计量科学研究院国家计量科学数据中心, 北京 100029

摘要

基于我国数据共享开放的发展趋势以及科研数据安全的相关政策, 对我国计量行业数据分级分类的现存方案进行研究。对国内20个国家级数据共享平台以及美国的相关数据分级分类方法展开调研, 讨论了我国计量科研数据分级分类方法的不足, 并从数据安全以及数据管理的角度提出了计量数据安全等级划分模型以及计量数据分级分类编码方法。

关键词

计量学; 分级分类; 数据安全; 开放共享; 数据管理

中图分类号: TP311.1

文献标志码: A

doi: 10.11959/j.issn.2096-0271.2022006

Classification of big data in metrology

ZHI Feng, TIAN Feng, ZHAO Ruofan

National Metrology Data Center, National Institute of Metrology, China, Beijing 100029, China

Abstract

Based on the tendency of development around data sharing in China and the relevant policies of safety management of scientific research data, the existing schemes of data classification of the metrology industry in China were studied. After investigating data classification methods of 20 national data-sharing platforms in China and relevant rules in the United States, the defects of the classification methods of metrological scientific research data in China were discussed, and the data security classification model and the data classification encoding method of metrological data were put forward.

Key words

metrology, classification, data security, opening and sharing, data management

0 引言

随着信息技术的发展,数字化的发展模式几乎遍布各行各业,由此也带来了数据管理以及数据安全等相关问题。2020年4月9日,中共中央、国务院发布了《关于构建更加完善的要素市场化配置体制机制的意见》,该意见将数据要素与土地要素、劳动力要素、资本要素等传统生产要素一并为完善要素市场化配置的关键因素^[1]。该意见所倡导的“推进政府数据开放共享”与“加强数据资源整合和安全保护”揭示了数据开放共享与数据安全对推动数字化发展的重要性^[2]。《中华人民共和国数据安全法》于2021年9月1日起施行,其进一步强调了数据安全对国家安全与经济发展的重要性。该法案第二十一条提出的“建立数据分类分级保护制度”,指出了开展数据分级分类对维护数据安全的必要性,为未来各行各业的信息安全工作提出了宝贵的指导意见。

对于科学研究和创新发展而言,科学数据是一种基础性的战略资源。而科学数据的分级分类则是维护数据安全、实现数据开放共享的必要方法。在计量领域,随着计量单位制的量子化和量值传递扁平化的变革,以及随之而来的计量基准、标准和标准物质的数字化等变革,以设备数字图谱、电子原始记录、数字证书为基础的扁平化量传体系,以及以区块链为主要技术的法制计量大数据联盟链机制将成为未来计量行业的发展趋势^[3-4]。计量行业数字化的发展趋势将会导致数据产出量和积累量迅速上升。同时,由于计量学涉及的领域十分广泛且数据量庞大,对计量数据进行合理的分级分类对于实现计量数据的收集、

存储、分析、共享以及安全维护是十分必要的^[5]。

1 国内关于数据分类的相关政策

科学数据主要指在自然科学、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^[6]。随着信息技术的发展,科学数据对科研工作的推进作用愈发重要,科学数据已然成为一种重要的科学基础资源。针对这一发展趋势,自21世纪以来,我国陆续发布多条相关政策,以加强对科学数据的管理,并逐步形成以行业机构、领域数据中心和国家层面的科学数据中心为主体的科学数据政策体系^[7-8]。

2015年,国务院印发《促进大数据发展行动纲要》,明确部署了与发展科学大数据、知识服务大数据应用有关的战略决策,提出了“构建科学大数据国家重大基础设施,实现对国家重要科技数据的权威汇集、长期保存、集成管理和全面共享”以及“对各领域知识进行大规模整合,搭建层次清晰、覆盖全面、内容准确的知识资源库群,建立国家知识服务平台与知识资源服务中心,形成以国家平台为枢纽、行业平台为支撑,覆盖国民经济主要领域,分布合理、互联互通的国家知识服务体系,为生产生活提供精准、高水平的知识服务”等主要任务。

2018年,国务院办公厅发布的《科学数据管理办法》成为我国首个国家层面的科学数据管理办法。该办法进一步针对目前我国科学数据管理中的工作人员职责,数据的采集、汇交与保存,数据的共享与利用以及数据的保密与安全等

工作进行系统部署。其中第二十条明确要求“法人单位要对科学数据进行分级分类,明确科学数据的密级和保密期限、开放条件、开放对象和审核程序等,按要求公布科学数据开放目录,通过在线下载、离线共享或定制服务等方式向社会开放共享”。该办法首次将数据分级分类作为保障数据安全的主要手段,旨在在保障数据安全的前提下,实现以“开放为常态、不开放为例外”为原则的数据管理环境。

2021年6月10日,第十三届全国人民代表大会常务委员会第二十九次会议通过《中华人民共和国数据安全法》,该法案将数据安全提升至国家安全层面。除了在数据安全与发展、数据安全保护义务、政务数据安全与开放以及数据安全的相关法律责任方面做出详细规定,该法案第三章“数据安全制度”的首条规定特别强调了数据分级分类保护制度对维护国家核心数据安全的重要性。其中,第二十一条指出“各地区、各部门应当按照数据分类分级保护制度,确定本地区、本部门以及相关行业、领域的重要数据具体目录,对列入目录的数据进行重点保护”。该条款明确指出,以数据分级分类为主要方法的数据安全管理是各行各业有关单位必须重视的核心工作,同时也是数据安全建设得以实施的基础。

从近几年我国有关科学数据分级分类的管理办法以及《中华人民共和国数据安全法》中不难看出,在以大数据为技术基础的数字化发展背景下,数据安全方面的治理刻不容缓。在“开放为常态、不开放为例外”的原则下,数据安全是实现数据安全共享的前提。由于不同行业领域中数据的类型及特点具有一定差异,针对计量学制定的数据分级分类方法则需要依据计量学的特点进行研究。

2 计量科学大数据的分级分类

2.1 计量数据分级分类的必要性

计量学涉及的领域十分广泛且数据分类方法多种多样,制定统一而合理的数据分类规则是实现计量数据资源有效利用的基础。根据《通用计量术语及定义》(JJF1001—2011),计量学是“测量及其应用的科学”^[9],这样的定义意味着计量学涵盖了各个学科领域中与测量的理论和应用有关的各个方面。按社会服务功能划分,计量学可分为法制计量、科学计量和工业计量。按专业领域划分,可分为几何量计量、热学计量、电磁学计量等十大类。而在国家计量科学数据中心的数据共享平台,计量学则被分为标准参数数据、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五大类。由此可见,计量数据根据分类视角的不同会产生不同的分类方法,而目前国内计量科学领域对元数据的分类标准尚未统一。同时,移动互联时代的大数据与PC时代数据的本质区别在于其量大、多维且格式多样。在没有充分挖掘之前,大量孤立的原始计量数据价值不高,而经过采集、加工后形成的衍生数据以及数据挖掘产生的分析结果则会价值倍增^[10]。随着数据利用率的提高,对计量科学数据依照统一且合理的规则进行分类,更有利于科研工作者从种类繁多、分散在各个研究机构且数量庞大的原始计量数据中选取有效数据进行分析,为之后的数据挖掘工作提供便利。

安全的数据环境是保证数据交换和数据挖掘的基础,而数据的分级分类则是维护数据安全的有效方法。计量学与国家法

律法规以及行政管理有着紧密的联系^[11]，这在其他学科是少有的。计量学的特性使得计量科学数据多数被国家和地方的计量研究院所掌控且部分数据关乎某些个人或组织的合法权益，重要数据的泄露甚至会危害国家安全以及公共利益。与此同时，在数字化的发展背景下，计量数据的开放共享是必然的趋势，它对推动计量学的进步也是至关重要的。在这样的发展趋势下，要求计量数据的管理者能兼顾数据的安全性和开放性。由此可见，数据管理和保护能力是开展计量科学研究工作的基础。而只有做好数据分级分类，才能将需要保护的重要数据分离出来，将可以共享的数据分享到大数据平台，以供各行各业充分利用^[12-13]。

由此可见，对于存储着重要计量数据的各研究机构而言，以数据分级分类为核心内容的数据安全管理是各单位必须重视的核心工作。

2.2 我国计量数据分级分类的现状

正如前文提到的，对于计量科研数据而言，无论是按照社会服务功能将数据分为法制计量、科学计量和工业计量这三大类，按照专业将其分为几何量计量等十大类，还是如国家计量科学数据中心的数据共享平台将数据分为标准参数数据、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，目前的分类方法都是从不同的角度根据数据种类进行划分的，且分类方法缺乏统一性。这样缺乏统一性的分类方法会给科研工作者以及其他行业中对计量数据有使用需求的工作人员带来不便。而在全国范围内施行统一的数据分类标准则会大大提高计量数据的共享及利用效率，减少相关工作人员在数据检索时的负担。

《中华人民共和国数据安全法》第二十一条明确指出，为了加强对重要数据的保护，有关机构应依据数据的重要程度以及出现意外事件后其后果的严重程度，通过分级分类的方法对数据进行保护。对于涉及多个领域且关乎众多行业发展的计量学，其科研相关数据中很大一部分属于“关系国家安全、国民经济命脉、重要民生、重大公共利益”的国家核心数据，按规定应受到更高层次的重视。在数据驱动发展的背景下，数据的共享开放是必然的趋势，而数据安全则是保证发展可持续性的前提。由此可见，为了维护数据安全，推进计量数据共享开放，计量行业需要尽快针对数据的安全性制定统一的分级分类标准。

虽然在全国范围内并未实现计量数据分类标准的统一，但作为国家级的计量数据中心——国家计量科学数据中心在《中华人民共和国数据安全法》正式实施以前就制定了比较详细的数据分级分类方案。在2020年11月发布的《国家计量科学数据中心数据分级分类管理办法》（以下简称《计量分类》）中，国家计量科学数据中心根据数据的内容和形式，按照层次分类法将计量科学数据分为两级并对其编码，具体见表1。

同时，国家计量科学数据中心基于数据安全和利用价值将数据分为公开数据、内部数据、受保护数据3类，并对其采取不同的保存策略和共享方式，具体见表2。

同时，国家计量科学数据中心也公布了各类数据的界定标准，具体如下。

- 完全开放共享数据：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：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：不宜共享利用的科学数据资源。

表1 计量科学数据分级及编码

一级分类	一级分类代码	二级分类	二级分类代码
支撑项目形成 相关科研成果 的科学数据	J10000	实验装置相关数据	J11000
		技术方法相关数据	J12000
		研究报告相关数据	J13000
		检测报告相关数据	J14000
		规程规范相关数据	J15000
		标准相关数据	J16000
支撑项目形成 科技文献等的 相关科学数据	J20000	其他	J17000
		论文相关数据	J21000
		论著相关数据	J22000
		其他	J23000
影视资料、 照片图表	J30000	视频、音频文件	J31000
		照片	J32000
		图表	J33000

表2 计量科学数据安全分级

数据级别	共享方式
公开数据	完全开放共享
内部数据	协议共享
受保护数据	不予共享

从国家计量科学数据中心制定的分级分类标准可以看出,当前国家计量科学数据中心施行的分级分类方法既依据数据类型对数据进行了分类编码,也从数据的安全性角度考虑将数据分为三大类以对特定数据进行分级保护,其分级分类标准已经初见雏形并形成系统。为了进一步改善现有的分级分类标准,本文将参考国内外其他领域的数据分级分类方法,择其善者而从之,其不善者而改之。

3 国内其他领域数据分级分类的现状

从2018年发布的《科学数据管理办法》第十九条中提到的“开放为常态、不开放为例外”的原则,以及持续推进国家数据交换平台建设的发展趋势中不难看出,近年来我国越来越重视科学数据共

享。自2004年科学技术部启动国家科技基础条件平台建设工作到2011年首批23个国家科技基础条件平台获批并公布,我国在科研领域的数字化发展一直没有停止前进的脚步。为了响应《科学数据管理办法》和《国家科技资源共享服务平台管理办法》,除了在计量领域于2019年成立了国家计量科学数据中心,在农业科学、地球科学和气象科学等19个领域也成立了相应的国家科学数据中心。通过对19个其他领域数据中心的数据分级分类方法进行调研,可以为计量领域的数据分级分类提供参考^[14-16]。笔者在调研后发现了如下现象。

(1) 计量行业缺乏详细的元数据分类标准

在对其他数据中心进行调研的过程中,笔者发现部分中心制定的数据分级分类方法更详细具体,值得计量行业效仿。例如在国家林业和草原科学数据中心发布的《林业科学数据分类与编码(V1.0)》中,林业科学数据依据数据性质被分为三大门类。在每个门类中,数据根据其学科领域以及其子学科的数据内容被分为一级分类和二级分类。其中,一级分类共36项,二级分类共116项。相比之下,国家计量科学数据中心制定的分级分类方法并没有针对具体学科领域以及相应子学科对元数据进行分类。对于涉及较多学科领域的计量学,制定更详细的元数据分类标准不仅有利于数据的汇交和管理,更有利于提高数据搜索效率。

(2) 计量数据安全分类缺乏量化的分类标准

随着对数据安全需求的提高,国内已经有学者以ISO27001体系为基础提出了数据资产分级模型。ISO27001体系将数据安全性拆分为保密性(confidentiality)、完整性(integrity)、可用性(availability)

(简称CIA)^[17]。陈驰等人^[18]以CIA为基础,建立了数学模型:

$$V = \text{Round1} \{ \text{lb}[(A \times 2^{\text{Conf}} + B \times 2^{\text{Int}} + C \times 2^{\text{Ava}}) / 3] \} \quad (1)$$

其中, V 代表数据资产价值; Conf、Int、Ava分别代表数据在保密性、完整性、可用性3个方面具有的资产价值; A 代表保密性的权值, B 代表完整性的权值, C 代表可用性的权值,具体数值可依据其应用的行业特征进行调整。此表达式通过对CIA三方面的资产价值进行幂运算,并乘以相应权值之后再行对数运算(lb代表以2为底的对数),最终得出数据资产价值的估值, Round1表示保留1位小数。该表达式反映了数据的业务价值,进而可以结合数据的涉密性完成对数据保密等级的划分,数字资产价值与数据资产保密等级的关系见表3^[18-19]。

数据资产分级模型与表2的计量科学数据安全分级方法均依据数据安全等级对数据进行分级划分。相比之下,数据资产分级模型通过数学模型量化了数据资产的价值,并依据量化结果对数据的安全等级进行划分。相比《计量分类》中描述性的分级标准,数据资产分级模型对安全等级的评估更为明确且精准,而且便于数据工作人员操作。计量数据分级也可以参照数据资产模型,根据不同计量领域的学科特点制定公式化的分级标准,以提高分级的精确性。

(3) 部分数据平台的数据组织规范性不足

部分平台存在同一平台出现多种分类方式的现象,容易降低数据搜索效率。例如国家气象科学数据中心官网首页的“数据分类体系”中将数据分为地面气象资料、高空气象资料、卫星探测资料、天气雷达探测资料以及数值预报模式产品这五大类,而在“数据服务”这一栏目中则将数据分为共享目录、地面资料、高空资料、数值

表3 数据资产价值与数据资产保密等级的关系

数据资产价值	价值分类	数据资产保密等级
≤ 2.5	较低	公共数据
(2.5, 3]	中等	内部数据
(3, 3.5]	高	敏感数据
> 3.5	很高	高敏感数据

预报、雷达资料、卫星资料和专题服务七大类。两种分类方式存在一定重叠但又有所区别,这样的分类方法并不是必需的,且会给浏览网站的人员造成数据检索困难。

(4) 部分数据平台元数据分类不严谨

在国家地球系统科学数据中心的共享平台中,多数降水数据被归入一级类目“大气圈”下的二级类目“降水”中。这使得大量本应该在“陆地水圈”中存储的陆地降水数据被归入“大气圈”。由于数据分类存在概念重叠,对陆地水循环有研究需求的用户无法在“陆地水圈”找到足够的降水数据,给相关工作人员的数据检索工作造成了较大困难^[20]。在对其他共享平台的调研中发现的上述问题可以为国家计量科学数据中心之后制定更详细的计量数据分级分类方案起到警示作用。

4 国外数据分级分类的现状

和国内相比,美国在数据共享、数据管理以及数据安全方面的工作开展得更早。早在1991年,白宫科技政策办公室就发布了关于研究数据管理的政策声明,要求对全球变化研究项目所产生的科研数据实行以“完全与开放”(full and open)为原则的开放共享政策^[21],如今已建立了比较完善的科学数据管理体系,十分值得我国从事数据管理的相关工作人员借鉴^[22-23]。根据《联邦信息安全管理法案》(Federal Information Security

Management Act, FISMA)^[24], 美国联邦政府将信息系统分为联邦信息系统和国家安全系统两类, 并分别针对这两类信息系统制定了《联邦信息和信息系统安全分类》(FIPS 199)与《国家安全系统的安全分类和控制选择》(CNSSI 1253)^[25-26]。这两套数据分级分类体系有相似之处, 但因适用领域不同而又有所区别, 对这两套美国经典的信息安全体系进行调研有助于帮助我国完善计量领域的的数据分级分类方法。

4.1 FIPS 199与CNSSI 1253的诞生

在2002年美国国会会议上由总统签署颁布的电子政务法案(公共法令107-347)明确了信息安全对美国国家安全的重要性。该法案第三章“联邦信息安全管理法案(FISMA)”成为美国信息安全发展的基石^[27]。为了实现信息安全发展, FISMA将实施步骤分为开发标准和指南、形成安全能力、运用自动化工具3个阶段。由美国国家标准与技术研究院(National Institute of Standards and Technology, NIST)制定的FIPS 199完成了FISMA的第一个目标: 为所有联邦机构(除国家安全系统外)提供对信息和信息系统进行分级分类的标准^[28], 并与之后推出的《将各类信息和信息系统映射到安全类别的指南》(SP 800-60)配合使用, 为联邦信息系统的的核心安全管理提供了基本依据^[29]。

与《计量分类》不同的一点是, FIPS 199的分级分类标准是完全针对数据的安全性进行评级, 并没有根据数据的形式、类别或学科进行分类。FIPS 199从保密性、完整性以及可用性3个维度对信息的安全性进行评级, 并针对每一个维度均赋予了发生意外事件的负面影响程度的评级, 分别为低、中、高或不可用。每种信息类型以及整个信息系统均可以通过安全分级SC的

通用表达式进行描述^[30-31]:

$$SC_{\text{信息类型/信息系统}} = [(\text{保密性, 影响}), (\text{完整性, 影响}), (\text{可用性, 影响})] \quad (2)$$

在FIPS 199的综合评估过程中, 需要依照“就高不就低”的原则依次对每个信息类型以及整个信息系统的安全性进行评级。为了配合FIPS 199的实施, 美国国家标准与技术研究院后续推出的SP 800-60详细介绍了联邦信息系统中可能运行的所有信息类型, 针对每一种信息类型介绍了如何选择影响级别的标准, 并给出了推荐采用的级别^[32]。

在FIPS 199的基础上, 美国国家安全系统委员会于2012年出台了《国家安全系统的安全分类和控制选择》, 对国家安全系统的信息分类和相应的技术防护要求做出了规定。其分级方法以联邦信息系统的分级原则为基础, 定级程序也与FIPS 199相似。CNSSI 1253与联邦信息系统不同的两点分别如下。

(1) 系统安全级别矢量化

CNSSI 1253没有把系统的安全性简单划分为高、中、低3个级别, 而是分别考虑了系统的保密性、完整性、可用性。其安全级别体现为一个具有3个元素的矢量, 分为{(保密性, 高), (完整性, 高), (可用性, 高)}、{(保密性, 中), (完整性, 高), (可用性, 高)}、{(保密性, 低), (完整性, 高), (可用性, 高)}等27个类别。

(2) 保密性分级精细化

对于保密性分级, CNSSI 1253提出了更细致的要求, 不仅要依据信息本身的保密性, 还要根据系统的使用环境、相关人员等因素对保密性级别进行调整。

4.2 FIPS 199与CNSSI 1253对计量数据分级分类的启示

FIPS 199对数据安全性的衡量有更

丰富的维度和更具体的标准。与《计量分类》相比,这种更加具体的分级分类方法更有利于相关工作人员开展针对数据安全层面的分级分类工作且分类的标准更加统一,值得我国计量行业借鉴。

通过对比FIPS 199和《计量分类》可以发现,二者均根据数据安全等级将数据分为3个安全等级,然而FIPS 199在得出最终等级判定之前,从保密性、完整性以及可用性3个维度依次对每个信息类型以及整个信息系统进行了判定,多维度、多步骤的分类过程有利于增加数据分级的准确性。相比之下,《计量分类》中的分级标准则比较模糊,过于粗略,在数据分类的过程中主观性较强。同时,SP 800-60对FIPS 199起到了补充说明的作用,针对各种信息类型分别提出了安全等级划分的建议。对于拥有庞大数据种类的计量学,数据工作人员往往很难准确地把握每种信息类型的安全性分级,因此针对计量学制定类似SP 800-60的数据分级分类指南是十分必要的。

相比之下,采取了安全级别矢量划分并对保密性分级更加精细的CNSSI 1253在安全等级分类方面精确度更高,适用于对数据安全性要求较高的领域。对于可能涉及重大财产损失以及国家安全机密的计量学数据,可以考虑采取类似的分级方式。同时可以考虑针对不同学科领域的特点以及数据存储环境等因素的差异对保密性的评估进行调整。

5 对我国计量数据分级分类方法改进方案的建议

5.1 计量数据安全等级划分

通过前文对国内外数据分级分类方法

的调研与分析可以看出,对数据安全级别进行矢量化分析可以提高分级分类的准确性,数学模型的建立可以降低人为赋值的主观性。同时,可以创建类似于陈驰等人^[18]提出的数据资产量化模型,通过数学表达式计算出精确的数据资产价值,并结合数据可能承受的威胁等级对安全分级进行综合考量^[33-34]。数据安全等级的分类分为以下3步。

(1) 数据资产价值的计算

正如前文提到的,ISO27001中规定数据资产具有保密性、完整性和可用性3个安全属性。同时也可以将其理解为,资产价值是由数据在这3个安全属性未达成时所造成的影响程度决定的。因此,笔者参考式(1)提出了用于计算计量数据资产价值的数学模型,并实现了对式(1)的简化:

$$V(A)=Wc \times Vc + Wi \times Vi + Wa \times Va \quad (3)$$

其中, $V(A)$ 代表特定数据种类 A 的资产价值; Wc 、 Wi 、 Wa 分别代表保密性、完整性、可用性对资产价值重要性所占权重, Wc 、 Wi 、 Wa 的和为1; Vc 、 Vi 、 Va 分别代表数据在保密性、完整性、可用性这三方面具备的资产价值。其中资产的价值并不以资产的经济价值为衡量标准,而是取决于其对关键业务的开展以及对实现目标的重要程度。对于计量学的不同学科而言,资产价值权重及 Vc 、 Vi 、 Va 的赋值均有所不同,应由相应领域的数字工作人员依据学科特点及数据类型进行调整,最终形成涵盖整个计量学的资产价值评估体系,并以SP 800-60为模板编撰相应的资产价值评估指南。

(2) 数据威胁等级的计算

同时,在对数据进行安全等级的分类时应该考虑数据的威胁等级,即安全事件发生的可能性及后果的严重性。威胁等级的划分同样依照FIPS 199与CNSSI 1253的分类方法,分别从保密性、完整性、可用

性这3个维度进行分析。同时,除了考虑数据在以上三方面出现威胁的严重程度,还应考虑威胁可能出现的频率。综合以上因素,得出威胁等级的数学模型:

$$T(A) = (F_c \times T_c + F_i \times T_i + F_a \times T_a) / 3 \quad (4)$$

其中, $T(A)$ 代表数据类别 A 的数据威胁等级; F_c 、 F_i 、 F_a 分别代表保密性、完整性、可用性三方面威胁发生的频率,取值为1~5的整数;而 T_c 、 T_i 、 T_a 分别代表该数据类型在保密性、完整性、可用性三方面可能出现的威胁强度,即安全事件的严重程度,取值同样为1~5的整数。与数据资产价值的计算过程类似,式(4)中的变量也应由相关领域数据的工作人员进行调整并形成威胁等级评估指南,以提高评估标准的一致性。

(3) 数据风险值的计算及安全等级的划分

完成对 $V(A)$ 与 $T(A)$ 的计算后,可以得出数据的风险值:

$$R(A) = \text{Round}1[T(A) \times V(A)] \quad (5)$$

$R(A)$ 的计算综合考虑了数据的威胁等级和资产价值。对 $T(A)$ 及 $V(A)$ 的计算是基于数据在保密性、完整性、可用性三方面具有的资产价值及潜在安全威胁的估算。相比《计量分类》中由主观判定一次性得出分类结果的方法,风险值 $R(A)$ 的推导经历了3次计算,精确度更高且受主观因素影响较小。在得出风险值之后,数据工作人员可以依据表4中的分类方法以风险值为依据对数据进行安全等级划分。其中, X 、 Y 、 Z 的取值将由各学科的数据工作人员在协商后确定。

表4 数据资产安全分级

数据风险值	风险等级	数据资产安全等级
$\leq X$	较低	公共数据
$X \sim Y$	中等	内部数据
$> Z$	高	受保护数据

5.2 计量数据分级分类编码

为了更合理地组织计量数据资源,实现对数据的科学管理和高效检索利用,构建计量学的数据资源编码体系是十分必要的。《计量分类》中已经初步构建了以数据应用目标和数据格式为分类标准的两级分类方法,并对不同的数据类别进行了编码。然而由于计量学涉及的学科比较广泛,目前的分类方法依然比较简略,不利于针对学科进行管理和检索利用。同时,对计量数据的编码同样应考虑数据安全等级的划分。因此,在参考刘召栋等人^[35]提出的数据分类编码方法后,笔者建议在《计量分类》提出的二级分类法的基础上,将学科及安全等级划入数据编码体系,进而形成如下编码方法:

数据风险等级-学科代码+原一级分类编码+原二级分类编码

示例:H-C21000。其中,H代表数据风险等级高,数据风险等级中、低分别为M、L;C代表学科;2100则代表《计量分类》中的一、二级编码方式及分类标准。

6 结束语

在我国大力推进数据安全工作的背景下,通过数据分级分类的方法提升数据安全、促进数据共享将是计量行业数据人员的工作重点。通过对比计量行业目前的数据分级分类方案、国内其他领域数据中心的数据分级分类方法以及美国的FIPS 199可以看出,目前国内计量行业的数据分级分类方法尚未统一且仍需继续完善。

计量学作为一个应用广泛且数据量庞大的学科,其科研数据对于我国科研、工

业以及教育等多个领域而言都是重要的基础性战略资源。随着计量领域数字化的变革,数据分级分类对数据的管理及应用至关重要。制定统一、明确、高效的计量数据分级分类方法不仅有利于数据的检索且有助于提升数据的安全性,而数据安全则是实现数据共享开放的前提。对比国内其他19个国家级数据中心的数据分类方法可以发现,计量学缺乏更详细的元数据分类标准,不利于相关工作人员对数据开展搜索、管理、审核以及分析等工作。同时,在对其他数据中心的调研过程中,笔者发现了部分数据平台存在数据组织规范性不足以及数据分类不严谨的现象,在今后计量行业数据分级分类标准的完善工作中需要避免出现类似情况。在数据安全分类方面,美国制定的FIPS 199以及其补充方案SP 800-60相比《计量分类》而言更加详细且具体。在未来的工作中,计量行业相关人员可以尝试参照FIPS 199与CNSSI 1253,从多个角度对数据的安全性进行分析定位,并制定公式化的安全分类数学模型。

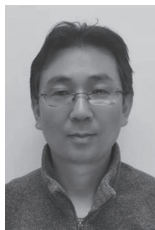
参考文献:

- [1] 袁博, 闫树. 数据要素市场化配置上升为国家战略[J]. 互联网天地, 2020(6): 34-37.
YUAN B, YAN S. The market-oriented allocation of data elements has become a national strategy[J]. China Internet, 2020(6): 34-37.
- [2] 李玉亮. 数据分类分级的现状与发展[J]. 中国信息安全, 2021(5): 55-56.
LI Y L. Current situation and development of data classification[J]. China Information Security, 2021(5): 55-56.
- [3] 黄松涛. 从《数据安全法》看计量检测数据的确权[J]. 中国计量, 2021(3): 61-63.
HUANG S T. Discuss the right of confirmation of metrological detection data from the perspective of the law of data security[J]. China Metrology, 2021(3): 61-63.
- [4] 李世岩, 刘金玲. 区块链技术在计量行业的应用方向[J]. 中国计量, 2020(11): 16-17.
LI S Y, LIU J L. Application direction of blockchain technology in metrology industry[J]. China Metrology, 2020(11): 16-17.
- [5] 宋超, 童树之, 石韡. 大数据技术在计量管理中的应用[J]. 工业计量, 2020, 30(S1): 46-47.
SONG C, TONG S Z, SHI W. Application of big data technology in metrology management[J]. Industrial Metrology, 2020, 30(S1): 46-47.
- [6] 杨晶, 康琪, 李哲. 推动科学数据开放共享的思考及启示[J]. 全球科技经济瞭望, 2019, 34(10): 37-43.
YANG J, KANG Q, LI Z. Thoughts and enlightenment on promoting the open sharing of scientific data[J]. Global Science, Technology and Economy Outlook, 2019, 34(10): 37-43.
- [7] 黄铭瑞, 李国庆, 李静, 等. 国家科学数据中心管理模式的国际对比研究[J]. 农业大数据学报, 2019, 1(4): 14-29.
HUANG M R, LI G Q, LI J, et al. International comparative study on management mode of national science data center[J]. Journal of Agricultural Big Data, 2019, 1(4): 14-29.
- [8] 闫倩, 马海群. 我国开放数据政策与数据安全政策的协同探究[J]. 图书馆理论与实践, 2018(5): 1-6.
YAN Q, MA H Q. Collaboration research on data opening policy and data security policy in China[J]. Library Theory and Practice, 2018(5): 1-6.
- [9] 中华人民共和国国家质量监督检验检疫总局. 通用计量术语及定义: JJF 1001—2011[S]. 2012.
General Administration of Quality Supervision, Inspection and Quarantine. General terms in metrology and their definitions: JJF 1001—2011[S]. 2012.

- [10] 程学旗, 靳小龙, 杨婧, 等. 大数据技术进展与发展趋势[J]. 科技导报, 2016, 34(14): 49-59.
CHENG X Q, JIN X L, YANG J, et al. Technological progress and trends of big data[J]. Science & Technology Review, 2016, 34(14): 49-59.
- [11] 局文. 2019年世界计量日主题活动在京举行国际单位制基本单位新定义在中国正式实施[J]. 中国计量, 2019(6): 7.
JU W. The theme event of 2019 world metrology day was held in Beijing, and the new definition of the basic unit of the international system of units was officially implemented in China[J]. China Metrology, 2019(6): 7.
- [12] 何珺. 工信部发布《工业数据分类分级指南(试行)》[J]. 今日制造与升级, 2020(3): 18-19.
HE J. The ministry of industry and information technology issued the guidelines for classification and classification of industrial data (for trial implementation)[J]. Manufacture & Upgrading Today, 2020(3): 18-19.
- [13] 周耀新, 贾启建. 计量数据的分类与采集[J]. 工业计量, 1992, 2(4): 17-18.
ZHOU Y X, JIA Q J. Classification and collection of metrological data[J]. Industrial Measurement, 1992, 2(4): 17-18.
- [14] 刘莉, 陈先来, 李忠民, 等. 精准医学大数据应用安全分类分级研究[J]. 医学信息学杂志, 2021, 42(1): 9-15, 35.
LIU L, CHEN X L, LI Z M, et al. Study on security classification of big data application in precision medicine[J]. Journal of Medical Informatics, 2021, 42(1): 9-15, 35.
- [15] 邓仲华, 黄雅婷. “互联网+”环境下我国科学数据共享平台发展研究[J]. 情报理论与实践, 2017, 40(2): 128-132.
DENG Z H, HUANG Y T. Research on the development of scientific data sharing platform at home in the “Internet +” environment[J]. Information Studies: Theory & Application, 2017, 40(2): 128-132.
- [16] 工信部信息技术发展司召开工业数据分类分级线上工作研讨会[J]. 工业控制计算机, 2021, 34(3): 82.
The information technology development department of the ministry of industry and information technology held an online workshop on industrial data classification[J]. Industrial Control Computer, 2021, 34(3): 82.
- [17] 黄水清, 陈双喜, 任妮. 基于ISO27001的数字图书馆信息安全风险评估模型研究[J]. 现代图书情报技术, 2009(6): 44-49.
HUANG S Q, CHEN S X, REN N. Research of risk assessment model of digital library information security based on ISO27001[J]. New Technology of Library and Information Service, 2009(6): 44-49.
- [18] 陈驰, 马红霞, 赵延帅. 基于分类分级的数据资产安全管控平台设计与实现[J]. 计算机应用, 2016, 36(S1): 265-268.
CHEN C, MA H X, ZHAO Y S. Data security control platform based on hierarchical classification: design and implementation[J]. Journal of Computer Applications, 2016, 36(S1): 265-268.
- [19] 王畅, 曾亚. 烟草行业数据的分类分级及安全防护方法探讨[J]. 内蒙古科技与经济, 2020(1): 31-32, 57.
WANG C, ZENG Y. Discussion on classification and safety protection methods of the data in tobacco industry[J]. Inner Mongolia Science Technology & Economy, 2020(1): 31-32, 57.
- [20] 王卷乐, 王明明, 石蕾, 等. 科学数据管理态势及其对我国地球科学领域的启示[J]. 地球科学进展, 2019, 34(3): 306-315.
WANG J L, WANG M M, SHI L, et al. The situation of scientific data management and its enlightenment to earth sciences of China[J]. Advances in Earth Science, 2019, 34(3): 306-315.
- [21] HIGMAN R, PINFIELD S. Research data management and openness[J]. Program: Electronic Library and Information Systems, 2015, 49(4): 364-381.
- [22] 张丽丽, 温亮明, 石蕾, 等. 国内外科学数据管理与开放共享的最新进展[J]. 中国科学院院

- 刊, 2018, 33(8): 774-782.
- ZHANG L L, WEN L M, SHI L, et al. Progress in scientific data management and sharing[J]. Bulletin of Chinese Academy of Sciences, 2018, 33(8): 774-782.
- [23] 杨静, 赵俊杰. 国外科学数据管理情况研究[J]. 全球科技经济瞭望, 2019, 34(1): 26-31.
- YANG J, ZHAO J J. Research on foreign scientific data management[J]. Global Science, Technology and Economy Outlook, 2019, 34(1): 26-31.
- [24] HULITT E, VAUGHN R B. Information system security compliance to FISMA standard: a quantitative measure[J]. Telecommunication Systems, 2010, 45(2): 139-152.
- [25] 盛小平, 王毅. 利益相关者在科学数据开放共享中的责任与作用——基于国际组织科学数据开放共享政策的分析[J]. 图书情报工作, 2019, 63(17): 31-39.
- SHENG X P, WANG Y. The responsibilities and roles of stakeholders in the open sharing of scientific data: an analysis based on policies of open sharing of scientific data in international organizations[J]. Library and Information Service, 2019, 63(17): 31-39.
- [26] 卫军朝, 张春芳. 国内外科学数据管理平台比较研究[J]. 图书情报知识, 2017(5): 97-107.
- WEI J C, ZHANG C F. A comparative study of research data management platform domestic and abroad[J]. Documentation, Information & Knowledge, 2017(5): 97-107.
- [27] JACOBSEN A, DE MIRANDA AZEVEDO R, JUTY N, et al. FAIR principles: interpretations and implementation considerations[J]. Data Intelligence, 2020, 2(1/2): 10-29.
- [28] RADACK S M. Federal information processing standard (FIPS) 199, standards for security[R]. 2004.
- [29] 张明天, 王惠莅, 杨晨, 等. FIPS 199《联邦信息和信息系统安全分类》标准解读[J]. 信息技术与标准化, 2011(7): 52-55.
- ZHANG M T, WANG H L, YANG C, et al. Understanding and analysis on FIPS 199[J]. Information Technology & Standardization, 2011(7): 52-55.
- [30] 周济礼, 先学人. 美国信息系统分类分级保护的主要内容及启示[J]. 中国信息安全, 2013(11): 102-105.
- ZHOU J L, XIAN X R. Main contents and enlightenment of classified protection of the US information systems[J]. China Information Security, 2013(11): 102-105.
- [31] CAMPBELL P. An analysis of department of defense instruction 8500.2 information assurance (IA) implementation[R]. 2012.
- [32] STINE K, KISSEL R, BARKER W, et al. Guide for mapping types of information and information systems to security categories: appendices[R]. 2008.
- [33] 春增军. 基于ISO27001的企业信息安全保障体系的构建设想[J]. 情报杂志, 2009, 28(5): 155-158, 162.
- CHUN Z J. The imagination of constructing a enterprise information security system based on ISO27001[J]. Journal of Intelligence, 2009, 28(5): 155-158, 162.
- [34] 高磊, 赵章界, 林野丽, 等. 基于《数据安全法》的数据分类分级方法研究[J]. 信息安全研究, 2021, 7(10): 933-940.
- GAO L, ZHAO Z J, LIN Y L, et al. Research on data classification and grading method based on data security law[J]. Journal of Information Security Research, 2021, 7(10): 933-940.
- [35] 刘召栋, 周亿城. 科技大数据资源及分类分级研究[J]. 科技与创新, 2021(18): 123-126.
- LIU Z D, ZHOU Y C. Research on resources and classification of scientific big data[J]. Science and Technology & Innovation, 2021(18): 123-126.

作者简介



智峰 (1978-), 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, 主要研究方向为大数据、云计算、数据安全等。



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, 主要研究方向为大数据管理、数据安全等。



赵若凡 (1993-), 男, 中国计量科学研究院国家计量科学数据中心助理工程师, 主要研究方向为大数据、数据挖掘分析等。

收稿日期: 2021-09-18

通信作者: 田锋, tianf@nim.ac.cn