

引用格式: 孙文峰,王世强.数据质量合规的概念、问题及对策[J].标准科学,2025(12):51–55.
SUN Wenfeng,WANG Shiqiang. Concepts, Problems, and Solutions to Data Quality Compliance [J].Standard Science,2025(12):51–55.

数据质量合规的概念、问题及对策

孙文峰¹ 王世强²

(1.中国标准化研究院; 2.国能朔黄铁路发展有限责任公司)

摘要: 【目的】近年来,在政策推动下数据产业快速发展,“数据合规”成为重要领域,确保数据合规是数据治理工作中的一项重要目标。然而数据质量合规问题隐患普遍存在,应用领域对于数据质量合规的问题并未给予足够的重视。对相关概念问题进行明确揭示将有助于应用领域中数据质量的提高。【方法】通过引用标准进行概念辨析、为概念中涉及的“合规性”列举法律依据、通过举例列举出按照这一概念内涵下存在的问题及其可能产生的影响。【结果】通过这一过程,证明了数据作为一种商品,同样存在质量方面的合规问题。【结论】建议在数据管理的工作中应引入数据质量合规的概念,使用外部规范的数据字典并加强自身数据的规范性。

关键词: 数据质量; 数据合规; 数据标准

DOI编码: 10.3969/j.issn.1674-5698.2025.12.007

Concepts, Problems, and Solutions to Data Quality Compliance

SUN Wenfeng¹ WANG Shiqiang²

(1.China National Institute of Standardization; 2.National Energy Group Shuohuang Railway Development Co.,Ltd.)

Abstract: [Objective] In recent years, the data industry has developed rapidly with the policy incentives. "Data compliance" becomes an important field, and ensuring data compliance is a crucial goal in data governance work. However, data quality compliance issues are prevalent, and application domains do not give sufficient attention to data quality compliance problems. Clearly revealing related conceptual issues will help improve data quality in application domains. [Methods] This paper conducts conceptual analysis by referencing standards, lists legal bases for "compliance" involved in the concepts, and identifies problems existing under this conceptual framework and their potential impacts through examples. [Results] Through this process, it demonstrates that data, as a commodity, also has compliance issues in terms of quality. [Conclusion] This paper ultimately recommends introducing the concept of data quality compliance in data management work, using externally standardized data dictionaries and strengthening the standardization of one's own data.

Keywords: data quality; data compliance; data standards

作者简介: 孙文峰,本科,副研究员,研究方向为数据标准化。

王世强,硕士,高级工程师,研究方向为智慧重载运输。

1 研究背景

近年来,以数据为核心对象的业务已经兴起。政府部门已经发布《“十四五”数字经济发展规划》(2021)、《中共中央 国务院关于构建数据基础制度更好发挥数据要素作用的意见》(数据二十条,2022)等相关政策,不断刺激数据产业的发展。因此,相关技术领域也在不断通过标准制修订、发布指南等工作,支撑数据产业走向结构清晰、产权明确的规范化道路。近三年来,我国已提出一批与数据产业直接相关的国家标准、行业标准或地方标准,如国家标准计划《资产管理 数据资产登记指南》(20242781-T-469)、GB/T 44109—2024《信息技术 大数据 数据治理实施指南》DB33/T 1329—2023《数据资产确认工作指南》。

在这一工作趋势中,“数据合规”的提法和用语逐渐开始被相关的研究领域使用。截至2025年4月底,在以“数据合规”为关键词且相关度最高的100篇CNKI文献中,标题直接使用“数据合规”一词的有17篇,标题中既包含“数据”又包含“合规”的有38篇。从上述17篇标题中直接使用“数据合规”的文献来看,目前数据合规方面的研究主要集中在3个方面:(1)从价值和利益等角度进行数据立法、现有法律的进一步解释等研究;(2)从满足现有法律法规对数据安全、个人隐私保护、版权等突出问题进行的具体技术研究;(3)从满足法律法规要求的角度进行合规管理方案的研究,包括管理制度的建立、责任的落实和持续改进等。“数据合规指引”方面的研究则更具有系统性和落地性,可能涉及以上3个方面。

以上研究中,均未提到数据合规中的一个重要方面:数据的规范性,即数据及其元数据的语义和语法均符合法律法规或标准的规定。按照数据质量的定义,也可称其为数据的质量合规。数据合规方面的重要性在于,一方面,虽然数据在语义和语法方面的规定均为推荐性标准,但推荐性标准可以被法律法规所引用,从而具有强制性效力^[1],违反了数据规范性同样可能面临严重的后果。另一

方面,数据的语义即数据的内容,内容错误可能导致很严重的法律责任或舆情;数据的语法错误可能导致数据不可用,在造成经济或社会损失的情况下同样具有合规风险。

2 数据质量合规的概念

国家标准对数据质量有两种不同的定义。在GB/T 36344—2018《信息技术 数据质量评价指标》中,数据质量指数据的规范性、完整性、准确性、一致性、时效性等方面指标。在当前提及数据质量的应用领域中,也往往指这个方面的含义。例如数据的质量直接决定数据的可用性^[2],数据质量缺陷主要包括客观性、完整性、时效性、数据粒度不佳^[3]等。所针对的工作也侧重于数据的采集、加工、存储、交换和应用的各个环节,主要通过加强岗位责任管理,应用数据清洗等技术手段提高数据的质量。

而国家标准对数据质量的另一种定义则与“质量”的定义一脉相承。GB/T 19000—2016《质量管理体系 基础和术语》将“质量”定义为“客体的一组固有特性满足要求的程度”。而在GB/T 24040—2008《环境管理 生命周期评价 原则与框架》中,数据质量被定义为数据在满足所声明的要求方面的能力特性。“所声明的要求”即标准规定的技术要求。“满足要求”则指可以根据标准开展合格评定活动,形成质量管理和监管的闭环。

在生产与流通中,后者所定义的数据质量即标准的一致性,属于质量基础设施体系^[4]中的一个重要环节,更具有合规性意义。在这个语境下,数据被看作一种产品。如同产品具有质量一样,数据也具有数据质量。对于产品好不好用来说,产品质量更重要的方面是有害物质的残留值应符合标准。同理,在数据好用不好用方面,数据质量中更重要的是数据中的语义和语法是否符合标准的规定。

通过以上概念的比对,数据质量合规指数据的标准被法律法规所引用,或者法律法规中直接

包含相关技术性条款时,因数据对标准或技术条款的符合程度带来的数据合规方面的状态。

3 数据质量合规的法律依据

与数据质量合规相关的法律依据(法规、规章和其他规范性文件)可以分为如下3种情况。

一是《中华人民共和国产品质量法》中对产品质量具有明确的规定。当数据作为一种产品时,该法同样适用。该法律规定了产品质量的监督,生产者、销售者的产品质量责任和义务,以及赔偿责任等。尤其是,数据类企业同样应该开展全面质量管理^[5],并通过质量体系认证。数据的生产者和销售者应该确保数据具有使用性能,且正确标注数据(类似于传统的产品标识等)。

二是法律法规中包含了技术性条款。经过统计,进一步可至少包括如下几种较为突出的情况。

(1) 法律法规中通过其定义条款^[6],规定了规范的用语及其解释,相关领域的数据应采用这些规范的用语。例如,《中华人民共和国高等教育法》规定了高等学校的定义。如果学校相关的数据库中仅将大学赋值为“高等学校”,则可能具有合规风险。

(2) 法律法规中规定了业务对象的类型或分类,相关领域的数据应采用这些规范的分类信息。例如,《中华人民共和国野生动物保护法》规定国家重点保护的野生动物分为一级保护野生动物和二级保护野生动物。如果某个数据库中存在“特级”或“三级”保护野生动物,则该数据库具有合规风险。

(3) 法律法规中指定了规范名录、目录等,相关领域的数据中应采用这些名录或目录中的列项,同一字段中不应出现其他的列项。例如,《中华人民共和国体育法》指定了体育项目目录,《中华人民共和国职业病防治法》指定了职业的分类和目录。

(4) 法律法规中指定了相关的信息平台,相关领域的数据应与这些信息平台中的元数据规范

保持一致。例如,《中华人民共和国海洋环境保护法》规定了国家组织建设统一的入海排污口信息平台。

三是虽然标准本身没有法律效力,但是法律中明确规定应符合相关标准,这些标准因法律本身的效果而发挥了作用^[7]。如果此时相关标准对数据的质量具有明确规定,则数据质量在法律效力的要求下具有强制约束力。例如,《中华人民共和国城市房地产管理法》规定,房地产开发项目的设计、施工必须符合国家的有关标准和规范。而GB 50352—2019《民用建筑设计统一标准》除了规定各类构筑物和设备的功能和性能参数外,还规定了类型、评价指标等数据类内容,建筑设计软件中的数据应符合这些数据方面的规定。

4 数据质量合规存在的典型问题及其归纳

4.1 典型问题

如下3个示例显示了数据质量合规方面存在的典型问题。

示例1: 错误使用地区的代码

一些贸易地点在国家标准GB/T 15514—2021《中华人民共和国口岸及相关地点代码》和GB/T 2260—2007《中华人民共和国行政区划代码》中规定的规范名称和代码,及其适用范围如表1所示。

以上代码和名称都具有特定的含义。如果在相关的数据集中,对于表达以上相关地区数据项使用了错误的名称或代码,则可能违反地方行政管理方面的规定,产生数据合规风险。

对于具有敏感性的地点的标识代码,如果数据集未使用规范的代码和名称,或者使用了不适用的标准中所用的代码和名称,则会面临更为严重的合规后果。

示例2: 错误使用数据元组合

对于数据库中的“金额”类数据元,一般对于国内业务需要与“计量单位”配合使用即可;而相同的数据共享于涉外的应用场景中时,还需要使

表1 同一地区可使用两种不同的国家标准地区代码举例

标准号	三峡库区翻坝运输码头		后石港区		洋山港		由来及其适用范围
	规范名称	代码	规范名称	代码	规范名称	代码	
GB/T 15514—2021	重庆港(部分码头)	CNCHQ	厦门港(部分港区)	CNXMG	上海洋山港	CNYSA	联合国下属UNECE制定, 采 用为中国国标。特指市内的机 场、车站、港口或其综合体
GB/T 2260—2007	宜昌市秭归县	420527	漳州市龙海区	350604	舟山市嵊泗县	330922	我国相关行政主管部门制定, 特指行政区划

用“币种”数据元的组合。否则会造成金额理解错误,具有数据合规风险。

示例3: 不规范使用日期时间表示

对于Unix时间戳1743358200(2025-04-29 14:30:00 UTC的秒数),如果缺少标准声明或者未进行格式转换,则无法正确解析,导致数据同步错误。在法律要求数据传输具有一定时限的情况下,则产生数据合规风险。

4.2 问题的归纳

从上述3个示例来看,如果聚焦于数据对技术要求的符合性方面,数据质量合规问题可具体包括对数据质量问题的认识、数据技术要求的体系化、数据自身的规范这3个方面。

在对数据质量问题的认识方面,应用领域往往缺乏合规性视角。对传统产品来说,在某些情形下,同一类别的产品质量有高有低,只是影响了其价格、销量和品牌等市场价值,并不涉及合规问题;而在另一些情形下,产品的质量与环境、健康等敏感领域相关,这样就会产生合规性问题^[8]。与传统产品的质量类似,数据的质量在很多时候属于数据的价值等市场范畴,但在上文所述的法律法规对数据提出要求的情形下,数据的质量也将涉及合规性问题。然而应用领域往往未能将数据质量纳入合规管理的范畴。

在数据技术要求的体系化方面,问题的难点在于数据集可能包含了海量的数据内容。这些数据可能被用于很多不同的行业、领域,导致不同的行业领域的监管活动中可能都对数据提出了相关要求或规范化的建议。与其相关的法律法规和标准的类型和数量可能较多,且随着时间的推移,相关的技术要

求会不断更新。应用部门对某类数据甚至某个数据需要符合哪些技术要求,往往缺少全面而系统的收集,最终导致形成数据不规范的隐患。

在数据自身的规范方面,一是语法的规范,主要指数据字段的结构、类型、长度等属性是否符合该要求。如果不符要求,一个系统的数据则往往难以被另一个系统所识别,将影响数据的共享与交换,最终造成决策有效性受到影响、数据混乱和冲突等后果^[9]。二是语义的规范,主要指数据字段的含义是否规范。数据的语义主要作用于对人的界面中,对外主要呈现为自然语言,由单独的词语或词语的组合构成,其承载的是业务知识和逻辑,判定其是否合规具有复杂性。其难点首先在于语义的判定具有较强的主观性;词语的含义受语境、语用影响较大,存在一词多义和语义偏移问题^[10],导致实际使用的语言与所要求的语言存在一定差别,但是不能对这种差别进行“正确”或者“错误”的两极化判定。其次,在很多情况下,自然语言的使用理应具有丰富的情感并采用多样化的表达,在不造成严重后果的情况下,数据相关活动的监管方不应对对此进行限制。

5 数据质量合规的对策建议

针对以上问题,应从如下3个方面加强数据合规工作。

5.1 在数据合规管理中加入数据质量合规的相关内容

目前在数据合规管理方面存在大量的数据合规方案、数据合规指引等技术成果。针对应用

领域中往往未能将数据质量纳入合规管理的问题,未来应在相关技术成果中加入数据质量合规方面的内容。主要聚焦于数据的生产、采集环节,对数据的结构设计和赋值方面的规则等提出指引。应要求相关数据按照其最初的采集和使用目的,建立清晰而规范的元数据和相关预置的值,且应最大限度地与现有的法律法规和相关标准保持一致。

5.2 建立并使用面向数据质量合规的元数据字典

针对技术要求的体系化方面的问题,主要应建立并使用面向数据质量合规的元数据字典。元数据字典主要由特定的数据服务商建立,专门针对数据方面的技术要求的体系化问题,合法地采集当前法律法规和标准中已规定的数据构件(术语、元数据、信息分类代码等)要求,以数据库服务的形式全面提供合规的基础数据构件。既可建立全国范围内具有公共属性的通用型字典,也可建立领域或机构内部的字典。

5.3 加强数据本身的规范

针对数据本身规范性的问题,一方面,从数据的语法入手,开展转换映射等工作,这类工作往往

较为容易实现,数据合规风险引起的代价也较小。另一方面,从数据的语义入手,最大程度地通过上述元数据字典使用已经发布的各类数据标准,选取相关的数据构件用来设计应用中所需的元数据,并在采集、存储、加工、交换等环节进行质量控制。针对数据在语义方面难以进行规范的问题,数据领域已经从各种角度制定了大量的信息分类代码、数据元等标准,其中包括机构、产品、人和各种抽象对象的描述。相关标准将代码赋予某类概念,这些概念预先使用尽可能准确、具有广泛适用性的自然语言表达,甚至辅助以图形符号,从而避免或大量减少了对自然语言的自由使用,但同时却保持了其含义,从而大大规避语义方面的数据合规风险。

6 结语

综上所述,数据质量的合规应通过加强相关的管理措施,并采用元数据字典和已有的数据标准等措施,进一步完善数据合规的技术体系,从各环节降低数据合规风险,促进数据产业的规范发展。

参考文献

- [1] 文松山. 推荐性标准在一定条件下具有强制性: 同《推荐性标准是否具有强制性》一文商榷[J]. 中国标准化, 1996(8):22-23.
- [2] 黄永军. S-100 系列国际标准数据质量研究[J]. 海洋测绘, 2024, 44(6):77-82.
- [3] 苏宇. 公共数据质量的制度保障[J]. 行政法学研究, 2025(4):85-100.
- [4] 刘洪生. 标准: 规则的一致性和国家质量基础设施[J]. 上海质量, 2018(5):48-50.
- [5] 连旭, 裴童心, 魏琪. 数据中心项目全面质量管理要素评价研究[J]. 科技与创新, 2025(4):184-187.
- [6] 汪全胜, 张鹏. 法律文本中“定义条款”的设置论析[J]. 东方法学, 2013(2):13-21.
- [7] 陈伟. 作为规范的技术标准及其与法律的关系[J]. 法学研究, 2022, 44(5):84-100.
- [8] 蒋苏华. 产品质量风险的合规性控制[J]. 中国质量, 2013(4):42-44.
- [9] 吴花平, 翟如月, 汤麒朋. 数智化转型背景下企业数据治理框架构建研究[J]. 财会通讯, 2025(12):27-31.
- [10] 陈智. 面向经典语义现象的循环神经网络研究[D]. 天津: 天津财经大学, 2024.