

团体标准《信息技术 公共数据质量评价规范》（征求意见稿）

编制说明

一、工作简况

1.任务来源

根据中国电子工业标准化技术协会下达的2024年第十批团体标准制修订项目，团体标准《信息技术 公共数据质量评价规范》（计划编号CESA-2024-180）由国家信息中心组织起草，由中国电子工业标准化技术协会（CESA）负责归口。

2.主要工作过程

国家信息中心牵头成立了标准编制工作组，制定详细的标准编制计划，工作组对目前公共数据资源使用和开放利用过程中，公共数据开放的数量、质量及使用支撑情况进行调研，确定了适用于公共数据质量评价主要技术内容，于2024年8月形成标准草案。

（1）标准需求调研

根据2022年12月19日出台的《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》，公共数据是各级党政机关、企事业单位依法履职或提供公共服务过程中产生的数据。公共数据以其鲜明的公共性、规模性、多样性和权威性蕴藏着巨大价值。随着我国数据要素化和价值化进程加快，公共数据的重要性日益凸显，公共数据高质量的有效供给是数据共享和开发利用的基础，也是释放公共数据价值的关键。根据工作组前期的研究基础以及在各地合作调研的情况，目前各地政府及相关单位均积累了大量的公共数据，并随着公共管理和服务范围扩大而不断增加，但是当前公共数据的规模优势不能掩盖数据质量不高、有效供给不足的现实，一方面公共数据资源整体呈现分散化、碎片化、无序化状态，内涵边界模糊，存在家底不清、权责不明、标准不一、统筹不力等问题。另一方面因多系统、多平台、多来源导致的数据结构各异而产生的数据缺失值、异常值、重复值、错误值等问题，大量公共数据难以挖掘复用，使公共数据开放共享和有效流通动力不足，数据资源使用和开发利用方面效率不高，亟需一套行

之有效的标准化方法，对公共数据质量进行评价，以评促治、以评提质，提升公共数据质量，促进公共数据开发利用。

（2）标准预研

在2024年4月至2024年8月期间，工作组组织行业专家代表参与起草研讨，分别进行了多次线上技术研讨，对于标准草案的技术框架和范围进行了初期的明确。

（3）标准立项

2024年9月，在工作组的协调沟通下，标准预研起草核心团队参加了中国电子工业标准化技术协会组织的团体标准立项评审会议。本标准预研单位代表对该标准技术框架和主要内容等方面进行了汇报阐述，最终顺利通过立项评审。会后，综合各单位专家意见，对标准草案进行了补充修改和完善，形成了征求意见稿及相关资料。

（4）成立标准编制组，明确标准对象和范围

2024年10月，工作组基于前期预研阶段参与单位代表，以及通过多种渠道方式征集标准参编单位，成立了标准编制组，建立联络方式和工作机制，确定将通过线上线下、专题等形式开展标准讨论和意见处理。同期进行了标准研讨，进一步明确了标准对象、范围及框架等。

（5）编制形成标准草案

2024年10-11月，工作组组织召开标准研讨会议，主要针对迭代修改后的草案内容以及征集到的意见建议展开讨论，对标准草案完成了优化。

（6）形成标准征求意见稿

2024年12月，工作组组织召开线上研讨会议，会上专家代表对标准草案进行评审，对标准内容基本达成一致，形成了该标准征求意见稿。

3.参加单位

本标准的起草单位有：国家信息中心、深圳市润鹏华通创新科技有限公司、中国科学院深圳先进技术研究院、中电数据产业集团有限公司、广东南方通信建设有限公司、广东省电子商务认证有限公司、天山材料股份有限公司、深圳市大数据资源管理中心、深圳市公安局视频警察支队、深圳大学、深圳技术大学、中国联合网络通信有限公司智能城市研究院、中电数创（北京）科技有限公司、沈

阳市数据局、北京国信恒达智慧城市科技发展有限公司、深圳市智慧城市规划设计研究院有限公司、中电科新型智慧城市研究院有限公司、深圳数据交易所有限公司、深圳市华傲数据技术有限公司、深圳市低空经济产业协会、深圳市软件行业协会、北京理工大学（珠海）、腾讯云计算（北京）有限责任公司、浪潮云信息技术股份公司、奇安信科技集团股份有限公司、中认北斗安全检测中心有限公司、福建省数易科技有限公司、福建博思数据技术有限公司、国信天数（上海）数据科技有限公司、招商局集团有限公司、萍乡市安源数字投资有限公司

二、标准编制原则和确定主要内容的论据及解决的主要问题

1. 编制原则

本标准制定过程中，主要遵循了如下几个原则：

- （1）符合国家的有关政策法规要求；
- （2）与已颁布实施的相关标准相协调；
- （3）结合实际应用需求；
- （4）围绕确定评价目标范围、收集数据与资料、制定评价标准、执行评价、分析评价结果、报告与反馈全流程，遵循指标设置和评价方法选取的科学性、客观性、系统性、可操作性以及可比性；
- （5）充分考虑加快公共数据资源开发利用，充分释放公共数据要素潜能的发展需求，满足应用基本要求，确保本标准可以作为各级党政机关、企事业单位在公共数据采集、汇聚、治理、开放、共享和授权运营等全生命周期活动的质量管理中切实可行的评价规范。

2. 确定主要内容的依据

（1）本标准参考了《信息技术 数据质量评价指标》（GB/T 36344—2018）中给出的评价指标体系，参考了GB/T 25000.24—2017《系统与软件工程 系统与软件质量要求和评价（SQuaRE） 第24部分：数据质量测量》中的相关指标体系和要求。

（2）本标准公共数据的术语充分参考借鉴了国家数据局发布的《数据领域常用名词解释（第一批）》及相关国家标准，公共数据质量评价充分学习借鉴了深圳市等先进地市在公共数据分级分类、共享开放等领域的实践工作和地方标准。

(3) 本标准广泛收集数据质量评价相关的已发布的国际和国家、地方标准成果,吸收各参编单位提出的成熟而先进的技术成果、实践成果,结合当前国内对公共数据研究的最新进展,对公共数据质量评价的范围、评价流程、指标体系、评价方法等进行了拓展和细化,保证评价指标的可得性、评价流程的合理性、评价方法的可操作性。

3.编制解决的主要问题

随着国家数据要素市场化配置改革的深化,要求不断扩大公共数据资源供给,破除公共数据流通使用壁垒,推进跨层级、跨地域、跨系统、跨部门、跨业务数据共享和业务协同,有序推动公共数据开放,提高开放数据的完整性、准确性、及时性和机器可读性,丰富数据应用场景,加快推动公共数据资源开发利用。

一是目前国内关于数据质量相关标准不多,已发布的数据质量国家标准多为转化国际标准,暂无涉及公共数据质量评价的内容。国内部分地市对公共数据,特别是政务类公共数据的治理、分级分类、数据开放等进行了地方标准探索,但多集中在建立公共数据质量管理制度和规范方面,对公共数据质量评价的研究不足,对公共数据数据项、元数据、数据集等数据元素质量的评价尚缺乏一套科学系统的统一评价体系和评价标准,迫切需要构建一套科学可靠的评价指标体系及评价方法,来指导公共数据质量评价工作,确保公共数据资源规模化高质量供给和有效利用。

二是本标准为了使公共数据质量评价更加有效落地,对公共数据质量的特性和当前的一些评价方法进行了更加深入的分析 and 评估,在国家标准(GB/T 36344—2018)的基础上,在设计公共数据质量评价指标框架时,基于公共数据产生过程、管理方式、内容特点等区别于其他数据类型的特点以及公共数据特有的公共属性、共享属性、开放属性、价值属性、安全属性等属性,更加注重公共数据在存储、开放、访问、流转、使用等过程中安全性,对公共数据质量评价从规范性、完整性、准确性、一致性、时效性、可访问性、安全性7个维度提出了具体要求。同时,充分考虑到公共数据分级分类管理的要求,对重要性不同的公共数据集,在计算质量得分时赋予不同的调节系数,为公共数据质量评价提供科学的评价方法。

三、主要试验[或验证]情况分析

本标准编制工作紧密结合了公共数据质量评价业务开展情况及服务内容，并由中电数创（北京）科技有限公司利用数据质量评价工具对标准中评价指标的可得性、可测性进行验证：

1. 团队建立

由中电数创数据质量评估团队组成验证团队，中电数创质量评估团队成员皆为取得了工信部“数据评价师（中级）”证书人员，团队成员具备丰富的数据评估实施经验，能够保证验证工作的专业性和全面性。团队按照《信息技术 公共数据质量评价规范》中数据资产质量评价的要求，进行标准的验证工作。

2. 数据准备

确定数据评价范围，采用抽样测试的方式，按照抽样方案从数据存储系统中提取城市生命线相关数据，为后续质量评价工作奠定基础。

3. 评估指标选取与权重确定

依据本标准指标体系，结合城市生命线相关数据特点和业务需求，依据规范性、完整性、准确性、一致性、时效性、可访问性、安全性7个一级指标以及对应的二级指标建立数据质量评价体系，并通过城市生命线相关数据的使用场景确定指标权重。

4. 评价工具配置和应用

运用数据质量评价工具“好数评”，根据质量要求和质量评价指标体系，在数据质量评价工具中配置相应的质量评价规则和算法，对城市生命线相关数据进行自动化质量检查和统计分析，同时结合人工审核的方式进行数据质量检核，确保数据质量评价结果的准确性和可靠性。

5. 数据质量问题了解和分析

通过数据质量评价工具记录的数据质量问题和分析结果，了解数据质量情况，包括问题类型和出现频率等。

6. 结果计算与报告编制

根据设定的评估指标计算方法和权重，通过数据质量评价结果自动计算数据质量评价总体评分和各维度得分，并生成数据质量评价报告。

以上应用验证证实了标准的有效性、科学性和可操作性，为后续标准实施落地做好前置准备工作。

四、知识产权情况说明

未发现本标准涉及知识产权问题。

五、产业化情况、推广应用论证和预期达到的经济效果

开展公共数据质量评价是推动数据资源标准化、规范化建设的必要步骤，也是促进公共数据合规高效流通使用、提高资源开发利用水平的关键环节。开展公共数据质量评价工作，将有助于各级地方政府、企事业单位强化数据源头治理和质量监督检查。

在国家深化数据资源市场化改革、加快公共数据资源开发利用改革的政策要求下，本标准进一步明确了公共数据开发、开放、共享和质量管理工作方向和重点，切实指导公共数据质量评价工作的开展，促进公共数据的开放共享和开发利用，助力公共数据要素价值释放。

六、采用国际标准和国外先进标准情况

未采用国际标准和国外先进标准。

七、与现行相关法律、法规、规章及相关标准的协调性

本标准与我国现行相关法律、法规、规章及相关标准的协调一致。

本标准文件第六章“评价指标”从规范性、完整性、准确性、一致性、时效性、可访问性、安全性七个方面对公共数据质量进行评价。其中：

1、规范性指标共计 5 个，均根据 GB/T 36344—2018《信息技术 数据质量评价指标》的 5.2 规范性指标进行修改而来。

(1) 数据标准符合度、数据模型符合度、元数据符合度、业务规则符合度、权威参考数据符合度等 5 个指标根据 GB/T 36344—2018《信息技术 数据质量评价指标》5.2 规范性指标（0101-0105）的内容修改，主要修改为指标名称和指标描述。指标名称的修改旨在使指标更具指向性，更能体现指标考察的主要方向；指标描述的修改旨在使指标更加符合公共数据的评估要求，进一步强调指标的评估对象范围。

(2) 规范性指标未继承 GB/T 36344—2018《信息技术 数据质量评价指标》5.2 规范性指标中“0106-安全规范”的内容，相关内容集中体现在本标准“6.8 安全性”一节中。

2、完整性指标共计 4 个，其中 2 个指标继承了 GB/T 36344—2018《信息技

术《数据质量评价指标》中完整性指标的内容，2个指标是本标准文件中创新提出的。

(1)数据元素完整性、数据记录完整性等2个指标继承了GB/T 36344—2018《信息技术 数据质量评价指标》中完整性指标的内容，GB/T 36344—2018中对于数据元素和数据记录的评价仍适用于对公共数据的完整性评价。

(2)元数据标识子集完整性、元数据内容子集完整性等2个指标是本标准中创新提出的指标，元数据能够帮助使用者了解数据集的结构、背景和意义等内容，将元数据纳入指标体系中能够更加全面地评价公共数据的完整性。

3、准确性指标共计5个，其中4个指标继承了GB/T 36344—2018《信息技术 数据质量评价指标》中准确性指标的内容，1个指标是本标准文件中创新提出的。

(1)数据内容正确性、数据格式合规性、数据唯一性、脏数据出现率等4个指标继承了GB/T 36344—2018《信息技术 数据质量评价指标》中准确性指标的内容，GB/T 36344—2018中对于数据内容、数据格式，以及数据是否唯一和无效数据出现频率的评价仍适用于对公共数据的准确性评价。

(2)元数据正确性等1个指标是本标准中创新提出的指标，旨在从元数据是否按照需求对数据进行描述方面更好地评价公共数据的准确性。

4、一致性指标共计4个，其中1个指标继承了GB/T 36344—2018《信息技术 数据质量评价指标》中一致性指标的内容，1个指标继承了GB/T 25000.24—2017《系统与软件工程 系统与软件质量要求和评价（SQuaRE）第24部分：数据质量测量》中一致性指标的内容，1个指标是在GB/T 36344-2018《信息技术 数据质量评价指标》中一致性指标的基础上进行了修改，1个指标是本标准文件中创新提出的。

(1)关联数据一致性指标继承了GB/T 36344—2018《信息技术 数据质量评价指标》中一致性指标的内容，数据格式一致性指标继承了GB/T 25000.24—2017《系统与软件工程 系统与软件质量要求和评价（SQuaRE）第24部分：数据质量测量》中一致性指标的内容，国家标准中的相关指标也契合公共数据一致性评价的相关要求。

(2)数据记录一致性指标在继承GB/T 36344—2018《信息技术 数据质量

评价指标》中性指标在继承 GB/T 36344—2018 标准内容的基础上进行了修改，旨在更加准确地从数据记录层面对一致性进行定义，也便于使用者更加准确地理解指标内容。

(3) 计量单位一致性指标是本标准中创新提出的指标，旨在从数据元素的计量单位方面更好地评价数据的一致性要求。

5、时效性指标共计 4 个，其中 3 个指标继承了 GB/T 36344—2018《信息技术 数据质量评价指标》中时效性指标的内容，1 个指标继承了 GB/T 25000.24—2017《系统与软件工程 系统与软件质量要求和评价（SQuaRE） 第 24 部分：数据质量测量》中现时性指标的内容。

(1) 时间段的正确性、基于时间点的及时性、时序性 3 个指标继承了 GB/T 36344—2018《信息技术 数据质量评价指标》中时效性指标的内容，更新的及时性指标继承了 GB/T 25000.24—2017《系统与软件工程 系统与软件质量要求和评价（SQuaRE） 第 24 部分：数据质量测量》中现时性指标的内容，国家标准中的相关指标亦契合公共数据时效性评价的相关要求。

6、可访问性指标共计 3 个，其中：数据元素可使用率、数据元素可访问率 2 个指标根据 GB/T 36344—2018《信息技术 数据质量评价指标》中规范性指标的内容修改而来，数据格式可访问性根据 GB/T 25000.24—2017《系统与软件工程 系统与软件质量要求和评价（SQuaRE） 第 24 部分：数据质量测量》中可访问性的质量测度的内容修改而来。

(1) 数据元素可使用率、数据元素可访问率 2 个指标根据 GB/T 36344—2018《信息技术 数据质量评价指标》中规范性指标（0601-0602）的内容修改，主要修改为指标名称和指标描述，旨在突出指标考察主体，明确考察范围。

(2) 数据格式可访问性根据 GB/T 25000.24—2017《系统与软件工程 系统与软件质量要求和评价（SQuaRE） 第 24 部分：数据质量测量》中可访问性的质量测度（Acs-D-2）的内容修改，旨在更加符合公共数据的发展特性，更加全面地评估公共数据可访问性。

7、安全性指标共计 4 个，GB/T 36344—2018《信息技术 数据质量评价指标》未单独设置安全性作为评价维度，仅在规范性指标中设置了安全规范的具体指标项，故安全性指标是本标准文件中创新提出的。

(1) 相比 GB/T 36344—2018《信息技术 数据质量评价指标》所评价的数据，本标准文件中所评价的公共数据来源和作用决定其必然具备更高的敏感性，公共数据可以反映国家经济社会运行整体情况，涉及法人信息、自然人隐私，事关个人、企业权益以及国家安全。故而本标准文件单独设置安全性作为评价维度十分必要，也是本标准文件的重要创新之处。

(2) 敏感数据项脱敏、公共数据访问权限管控情况等 2 个指标参考了 GB/T 36344—2018《信息技术 数据质量评价指标》的规范性指标中安全规范指标，并在此基础上进行了补充修改，旨在更好地评价敏感数据项脱敏的过程与结果，评价公共数据访问权限的制度建设情况与落实效果。

(3) 公共数据防泄露管控情况、非脆弱性等 2 个指标是本标准参考行业和地方实践出创新提出的指标旨在从公共数据防泄漏、非脆弱性等方面更好地评价公共数据的系统安全建设情况，从而保障公共数据的合规授权访问。

八、重大分歧意见的处理经过和依据

无重大分歧意见。

九、贯彻标准的要求和措施建议

建议加快推进该标准的发布和实施工作。标准发布后，向各级政府、企事业单位开展标准的宣贯工作。

十、替代或废止现行相关标准的建议

无代替或废止现行相关标准。

十一、其它应予说明的事项

无。

团体标准《信息技术 公共数据质量评价规范》编制工作组

2025 年 1 月 16 日