贵州省政府数据质量评价规范

编制说明

云上贵州大数据产业发展有限公司

贵州中软云上数据技术服务有限公司

二〇一八年九月

目 录

[一、标准编制背景及必要性 1](#_Toc529547958)

[二、标准编制目的和意义 2](#_Toc529547959)

[三、标准编制的原则、依据和方法 3](#_Toc529547960)

[（一）编制原则 3](#_Toc529547961)

[（二）编制依据 3](#_Toc529547962)

[（三）编制方法 4](#_Toc529547963)

[四、主要编制过程 4](#_Toc529547964)

[五、与有关的现行法律、法规和强制性国家标准关系 5](#_Toc529547965)

[六、规范的结构及适用范围 5](#_Toc529547966)

[七、采用国际标准、国外先进标准程度 5](#_Toc529547967)

[八、标准主要内容及技术依据 5](#_Toc529547968)

[（一）标准主要内容 5](#_Toc529547969)

[1.标准适用范围 5](#_Toc529547970)

[2.数据质量评价原则 6](#_Toc529547971)

[3.数据质量评价一般流程 6](#_Toc529547972)

[4.数据质量评价维度 6](#_Toc529547973)

[5.数据质量评价指标 7](#_Toc529547974)

[6.质量评价方法 7](#_Toc529547975)

[（二）技术依据 8](#_Toc529547976)

[九、专利及涉及知识产权情况 9](#_Toc529547977)

[十、重大分歧意见的处理经过 9](#_Toc529547978)

[十一、标准作为强制性或推荐性地方标准的建议 10](#_Toc529547979)

[十二、 废止现行有关标准的建议 10](#_Toc529547980)

[十三、预期效益分析 10](#_Toc529547981)

《贵州省政府数据质量评价规范》编制说明

云上贵州大数据产业发展有限公司（以下简称“云上贵州公司”）牵头，贵州省公共大数据重点实验室、中软云上数据技术服务有限公司、贵州商学院联合编制的《贵州省政府数据质量评价规范》。现对其编制情况作相关说明。

一、标准编制背景及必要性

2015年9月，国务院印发《促进大数据发展行动纲要》，明确提出“建立标准规范体系。推进大数据产业标准体系建设，加快建立政府部门、事业单位等公共机构的数据标准和统计标准体系，推进数据采集、政府数据开放、指标口径、分类目录、交换接口、访问接口、数据质量、数据交易、技术产品、安全保密等关键共性标准的制定和实施。”2016年3月，中华人民共和国国民经济和社会发展第十三个五年规划纲要，指出要“完善大数据产业公共服务支撑体系和生态体系，加强标准体系和质量技术基础建设”。2016年6月中共贵州省委、贵州省人民政府联合发布《关于实施大数据战略行动建设国家大数据综合试验区的意见》，提出了“建立关键共性标准。制订政府数据采集、开放、共、分类、质量、安全管理等关键共性标准”的要求。

为贯彻落实国家《行动刚要》、《十三五规划》以及贵州《关于实施大数据战略行动建设国家大数据综合试验区的意见》有关精神，结合《贵州省大数据发展应用促进条例》配套实施要求，编制数据质量评价标准是贵州省发展大数据战略的重要任务之一。

二、标准编制目的和意义

数据正在成为新的经济增长点,面对数据时代的来临,集中掌握资金、人力、政策等资源优势的政府部门扮演着重要角色。政府部门既是推进政府大数据发展的主体,也是自身需要经现代化信息技术关联分析后发现新知识、创造新价值的研究客体,其数据质量更是重中之重,关系到大数据最终的运算结果。质量评价体系的建立有助于确定数据质量标准化,规范政府大数据科学性管理,引导新的契合大数据技术手段的信息基础建设。

数据作为一种重要资源,在诸多学科领域中都有重要的影响,数据的质量对过程管理、决策支持、合作需求分析等活动均有重要的引导作用,高质量的数据能够提供可靠、准确的服务。然而,当数据获取量飞速增加时,由于容错标准不完善、存储数据格式不一致、信息来源可靠性低、数据更新周期过长等原因,造成数据的错误率和混乱程度增加。在数据使用过程中，数据质量不够优质很可能会给诸多领域带来严重的负面影响。规范化数据质量评价标准能更有效的保证优质数据质量。

目前贵州发展大数据仍存在一些不足,对影响数据质量的性质分析得不够全面和深入,未能对各种性质的满足或违反情况进行明确的定义说明。此外,对于几种重要性质也只是对部分问题进行了分析处理，未能全面说明这几个数据质量性质之间的关联。故本标准的目的在于找出多种数据质量性质间的明确的关联关系,对数据质量进行更为全面的描述,并避免数据质量管理过程中的重复性工作。

三、标准编制的原则、依据和方法

（一）编制原则

1.适用性原则：在确定标准项目时首先要注意标准的适用范围，既不要让标准所涵盖的领域过宽，使编制的标准没有实际技术内容；也不要让标准所涵盖的领域过窄，造成对标准的肢解。

2.先进性原则：编写标准草案时要在充分调查研究的基础上，认真分析国内外同类技术标准的技术水平，在预期可达到的条件下，积极地把先进技术纳入标准。

3.统一性和协调性原则：编制过程中要注意符合法律法规的规定以及与相关标准协调，避免与法律法规、相关标准之间出现矛盾，给标准的实施造成困难。

4.经济性原则：要以满足实际需要出发，避免指标、要求造成不必要的浪费。

（二）编制依据

GB 18030 信息技术 中文编码字符集

GB/T 21063.4—2007 政务信息资源目录体系 第4部分：政务信息资源分类

GB/T 26816-2011 信息资源核心元数据

GB/T 21063.3-2007 政务信息资源目录体系 第3部分：核心元数据

GB 11714 全国组织机构代码编制规则

GB/T 19486 电子政务主题词表编制规则

GB/T 19488.1 电子政务数据元 第1部分：设计和管理规范

本规范按照GB/T1.1-2009《标准化工作导则 第1部分：标准的结构和编写》的要求进行起草。

（三）编制方法

采用调研并结合意见反馈制定本标准。根据对贵州省20朵云单位数据梳理工作的实际结果，并根据国标GB/T1.1-2009要求，结合贵州省各部门单位反馈意见及我省的实际政府数据应用需求制定本标准。

四、主要编制过程

1.2017年4月：成立编制小组，制定了编制工作计划和实施方案，明确工作分工；

2.2017年4月确定标准主要技术内容；

3.2018年2月：对省发改委、省国资委、省工商局、省质监局等单位调研。

4.2018年3月-9月：完成《贵州省政府数据质量评价规范》初审稿和编制说明。

5.2018年9月下旬：整理相关资料，准备标准评审。

五、与有关的现行法律、法规和强制性国家标准关系

完全符合现行法律、法规和强制性国家标准规范内容，无冲突或矛盾内容。

六、规范的结构及适用范围

本标准适用于贵州全省范围内，用于指导政府部门对采集的结构化数据进行质量评估，规范数据质量评价指标和评价方法。目的在于促进建立政府数据质量评价体系，在实现贵州省政府数据价值最大化的同时，确保不会影响到国家安全、社会稳定、公众和个人安全。同时促进政府部门改进提供数据的质量。

七、采用国际标准、国外先进标准程度

《贵州省政府数据质量评价规范》是根据贵州省大数据发展战略需要，建立质量评价体系，有助于确定数据质量标准化,规范政府大数据科学性管理,目前暂无适用的国际标准可以借鉴。

八、标准主要内容及技术依据

（一）标准主要内容

1.标准适用范围

本标准适用于贵州全省范围内，用于指导政府部门对采集的结构化数据进行质量评估，规范数据质量评价指标和评价方法。

2.数据质量评价原则

评价原则包括：科学性原则、客观性原则、系统性原则、可操作性原则、针对性原则和引导性原则。

3.数据质量评价一般流程

本规范提出数据质量评价的一般流程，包括：需求分析、确定评价对象及范围、选取评价维度及指标、确定评价方法、进行质量评价、分析结果及报告

4.数据质量评价维度

数据质量是一个多维度的概念，涉及数据及其生产服务过程的多个方面。一般而言，对数据质量的认识通过将其分解为多个质量维度，并逐个识别实现。

按数据以及质量需求，本规范将数据质量评价要素划分为基本层、准则层和指标层三层。

基本层指标（一级指标）：按照数据的形式、内容和效用三个维度来认识。即形式质量、内容质量和效用质量。

准则层指标（二级指标）：为数据质量评价活动中常见的质量指标，不保证其完备性，具体于一次评价活动中使用的质量指标仍需按照质量评价工作流程由评价方做出决定。

指标层指标（三级指标）：是总体评价目标的具体体现，是评价目标的具体分解。但每个指标都应从一定的角度或侧面反映评价目标，与评价目标紧密相关，同时指标总体对目标应具有足够的覆盖面，与评价目标保持高度的一致性。

5.数据质量评价指标

基于上述数据质量评价维度，本规范提出一组数据质量评价常用指标，供评价方在数据质量评价过程中选择使用。

数据对象范围的复杂，本规范仅将基本层质量指标作为要求性内容；而准则层指标为参考性内容，仅供评价方参考；未对指标层的具体执行方法进行约束。具体的指标选取、权重确定和使用的评价方法由评价方在实施检查时依照质量评价流程确定。

6.质量评价方法

6.1定性方法

数据质量的定性评价方法是基于评价准则与要求，根据评价的目的和用户对象的需求，从定性的角度来对政府数据资源进行描述与评价。确定相关评价准则或指标体系，建立评价准则及各赋值标准，通过对评价对象大致评定，给出各评价结果，评价结果有等级制、百分制或其他表示。包括：第三方评测法、用户反馈法和专家评议法。

6.2定量方法

数据质量的定量评价方法是指按照数量分析方法，从客观量化角度对政府数据资源进行评价，主要用于评价结构化数据库。数据质量的定量指标描述数据集满足预先设定的质量标准要求及指标的程度，并提供定量的数据质量信息，保证评估结果的科学性和客观性。包括： 简单比率法、最小/最大值法和加权平均法。

6.3综合方法

综合方法是将定性和定量两种方法结合起来，从两个角度对政府数据资源质量进行评价。主要做法是在定性评价方法的基础上引入数学手段，定性问题通过人工设定的标准进行评分并做出量化处理，评分的过程都是针对事先建立的指标体系，利用如最小/最如大值法、加权平均法等进行定量计算，再结合如主次分析法、层次分析法、模糊综合评价法等做综合评价。包括：主次分析法、 层次分析法、模糊综合评价法和模糊层次综合评价法。

（二）技术依据

1.本规范中“表1 主要数据质量指标”参考了张胜的《数据质量评价指标和评价方法浅析》和马一鸣的《政府大数据质量评价体系构建研究》以及《云上贵州数据共享交换平台开放平台数据质量与应用评估报告》并结合贵州省数据共享交换平台实际数据和相关专家经验总结而成。

2.本规范中7.1.1第三方评测法参考了基础科学数据共享网项目标准的《数据质量评价方法与指标体系》

3.本规范中7.1.2用户反馈法参考了基础科学数据共享网项目标准的《数据质量评价方法与指标体系》

4.本规范中7.1.3专家评议法参考了基础科学数据共享网项目标准的《数据质量评价方法与指标体系》

5.本规范中7.2.1简单比率法参考了中国航空综合技术研究所王学良、商广娟的《多指标的数据质量评价方法综述》

6.本规范中7.2.2最小/最大值法参考了国防信息学院张胜的《数据质量评价指标和评价方法浅析》

7.本规范中7.2.3加权平均法参考了国防信息学院张胜的《数据质量评价指标和评价方法浅析》

8.本规范中7.3.1主次分析法参考了湖北大学资源环境学院何超，李萌，李婷婷，彭雪，李婕，赵锦慧的《多目标综合评价中四种确定权重方法的比较与分析》

9.本规范中7.3.2层次分析法参考了湖北大学资源环境学院何超，李萌，李婷婷，彭雪，李婕，赵锦慧的《多目标综合评价中四种确定权重方法的比较与分析》

10.本规范中7.3.3模糊综合评价法参考了湖北大学资源环境学院何超，李萌，李婷婷，彭雪，李婕，赵锦慧的《多目标综合评价中四种确定权重方法的比较与分析》

11.本规范中7.3.4模糊层次综合评价法参考了天津大学江高的论文《模糊层次综合评价法和其应用》

九、专利及涉及知识产权情况

本规范未涉及有关专利和知识产权。

十、重大分歧意见的处理经过

在本规范起草和编制过程中，采用调研并结合意见反馈和实际工作经验相结合的方法，未出现重大分歧意见。

十一、标准作为强制性或推荐性地方标准的建议

建议《贵州省政府数据质量评价规范》作为推荐性标准发布实施。

十二、 废止现行有关标准的建议

本规范与现行强制性国家标准规范内容无冲突或矛盾内容。

十三、预期效益分析

通过编制（制定）、发布、实施本标准，进一步规范政府数据质量管控，推动贵州省大数据价值的挖掘与利用，促进大数据产业发展，提升政府治理能力和商业创造能力，帮助打破数据可聚不可用的局面，从而最大程度的挖掘并实现政府数据的价值。