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

前言

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。

一、任务由来

本标准由贵州省大数据管理局提出，由省质监局《关于下达贵州省2017年第一批标准化项目的通知》批准立项，云上贵州大数据产业发展有限公司牵头，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、贵州商学院联合起草制定。

二、本标准的主要起草单位及人员

《贵州省政府数据质量评价规范》主要起草单位及人员如下表所示。

表2-1 主要起草单位及人员一览表

| **主要起草单位** | **主要起草人员** | **任务分工** |
| --- | --- | --- |
| 云上贵州大数据产业发展有限公司 | 秦晓东 | 标准起草 |
| 中软云上数据技术服务有限公司 | 杨建国 | 标准起草 |
| 中软云上数据技术服务有限公司 | 王迪江 | 标准起草 |
| 中软云上数据技术服务有限公司 | 韦超 | 标准起草 |
| 中软云上数据技术服务有限公司 | 徐凯琳 | 标准起草 |
| 贵州商学院 | 李静 | 标准起草 |
| 贵州商学院 | 袁华 | 标准起草 |
| 贵州商学院 | 刘梦珠 | 标准起草 |

三、本标准起草过程

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。

四、标准编制原则

（一）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。

（二）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。

（三）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。

（四）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。

五、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。

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。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。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。

六、主要验证分析和预期经济效果

1.主要论证分析

（1）关于“3 数据质量评价原则”的说明

**科学性原则：**质量评价的结果应能正确反映数据资源的质量状况。主要体现在正确的质量指标选择，以及采用科学合理的评价方法等方面。评价必须有一定的理论作为基础，但又不能够脱离实际。另外，科学性还反映适度的简单，评价不可能穷尽所有因素，也不能过于简单。

**客观性原则：**评价应是符合实际、客观可信的。评价指标的选择须考虑当前数据资源环境的总体水平，反映出不同领域的差异。

**系统性原则：**由于评价对象的广泛性、复杂性、必须使用若干指标来衡量，同时指标间可能相互联系、相互制约。但是，在评价中，每个指标又必须是独立的，不互相包容的，需考虑指标的层次性、系统性，避免指标间冲突。

**可操作性原则：**科学合理的评价体系应该是可行的、操作方便的，指标的设计避免过于繁琐，还要考虑指标体系所涉及指标的量化及数据获取的难易程度和可靠性，注意选择能够反映数据质量状况的综合指标和具有代表性的指标。

**针对性原则：**政府数据资源种类繁多，数据积累具有连续性，各种资源除了具有与其他资源相同的共性之外，也具有其自身的特殊性。数据质量评价应能充分考虑各类政府数据资源所特有的类型特征并能将其揭示出来，要在指标的权重和分值上予以区分，以体现其针对性的导向作用。

**引导性原则：**进行政府数据质量评价，目的在于了解政府数据资源的质量情况，为有关的取舍提供判断依据，以帮助用户快速选择有针对性的信息。因此，必须以方便专业人员快捷而有效的选择和获取有价值的信息资源为导向。

（2）关于“4 数据质量评价流程”的说明

数据质量评价过程是评价者将数据质量评价程序应用于目标数据或数据集并最终获取评价对象质量状态的一系列步骤。本规范提出的数据质量评价的一般流程,具体于一个具体的执行过程中根据数据对象或业务背景有所不同仍可根据具体情况适当增删。

数据质量评价包括以下一般流程：



数据质量评价一般流程

数据质量评价过程是一个迭代过程,各个过程的先后顺序仅表达阶段活跃的大致顺序,根据实际执行情况一些过程可能需要重复执行或省略。

数据质量评价的一般流程具体流程步骤说明如下：

需求分析

数据质量评价要以业务需求为中心进行。数据需求是人们在各项实践活动过程中,为解决所遇到的问题而产生的对数据的不足感和求足感。数据不同于实体产品, 用途具有个性化、多样化、不稳定等特点,因此,必须首先了解具体业务对特定数据资源的需求特征才能选择针对性的评价指标体系。

确定评价对象及范围

数据质量要素受行业领域、数据类型和应用目的等因素影响，不存在面向所有领域和资源类型的普适性数据质量框架，所以要根据需求分析结论选定当前用户为解决业务问题的数据作为评价对象，明确数据在属性、数量、时间等维度的具体界限。需要说明的是，评价对象一定是一个确定的静态集合。

选取评价维度及指标

数据质量维度是质量的特征，为数据质量评价提供明确的途径。因此,在确定数据质量评价的对象和范围后, 首先要依据具体业务需求选择适当的数据质量维度和评价指标。其次，要选取可测、可用的质量维度作为评价指标准则项，在不同的数据类型和不同的数据生产阶段，同一质量维度有不同的具体含义和内容，应该根据实际需要和生命阶段确定质量维度。

在此阶段要注意指标之间避免冲突,同时也要注意新增评价指标的层次、权重问题,以及与其它同层次指标的冲突问题。对三级评价指标的选择可根据评价对象的类别、评价要求进行量化处理。对当前技术条件无法量化的纬度可适当使用具有相关性的替代指标。

确定评价方法

数据质量评价在确定其维度和指标后,应该根据每个评价对象的特点,确定其评价方法,对于不同的评价对象需要选择不同的评价方法支持,所以应该根据质量对象的特点确定其评价方法。

进行质量评价

就是根据前面四步确定的质量对象、质量范围、评价方法实现质量评价的活动过程。评价对象的质量应当由多个质量维度的评价来反映,单个数据质量测量是不能充分、客观评价由某一数据质量范围所限定的信息的质量状况,也不能为数据的所有可能的应用提供全面的参考。多个质量维度的组合能提供更加丰富的信息。

分析结果及报告

评测后要对评测结果进行分析：对评价目标与结果进行对比分析,确定是否达到评价指标；对评价的方案的有效性进行分析,确认是否合适等。

根据评价结果确定对象的质量评价,如需要,可根据评价结果鉴定质量级别。确定评价对象的质量级别是建立在相应的质量分级方案基础上的,该分级方案是根据相应的质量规范或用户的需求确定的,也是判断数据质量成熟度的重要依据。

质量评价结果和评价报告是所有具体业务数据质量评价项目及其评测结果的合集， 最后要将质量评价结果和数据质量评价过程汇总并报告。在完整的数据质量评价结果和报告中，应该包括全部上述内容。

（3）关于“5 数据质量评价维度”的说明

参考了中国科学院计算机网络信息中心起草的，基础科学数据共享网项目标准的《数据质量评测方法与指标体系》

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。

数据质量的结构特性学术上目前仍无定论，为出于便于管理的需要，本规范根据数据以及质量需求的特征，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：该层提出适用于对数据进行评价的通用指标，通用指标归纳大部分数据共有的本质特征，以及普遍的技术特征，是数据质量评价的基本指标。基本层对数据质量，按照数据的形式、内容和效用，即语法、语义和语用的 不同层面，对其内在结构从形式、内容和效用三个层次来认识。

准则层：该层根据数据的内容特性提出适用于特定领域范围的质量指标，准则层是对基本层质量指标面向特定领域的细化指标，要求根据领域特点细化质量评价指标并分配权重。

指标层：是对准则层提出质量指标根据资源类型不同提出具体质量评测执行方法，要求根据资源特点指定评价指标评测方法。

（4）关于“6 数据质量评价指标”的说明

本规范6.1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。

（5）关于“7 数据质量评价方法”的说明

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》

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》

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》

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》

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》

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》

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》，主次分析法是根据评价指标中存在着一定相关性的特点，用较少的指标来代替原来较多的指标，并使这些较少的指标尽可能地反映原来指标的信息，从根本上解决指标间信息重叠问题，简化原指标体系的指标结构。

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》，层次分析法确定权重，是指采用1-9标度法，求出每层因素相对于评价目标的权重。首先将各层的指标的重要性比较判断，用1-9标度法构造标度判断矩阵；进行一致性检验，若检验通过，最大值特征值对应的特征向量即为权向量；若不通过，需重新构造标度判断矩阵。

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》

2.预期效果

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。

七、采用国际标准和国外先进标准的程度

《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、与有关现行法律、法规的关系

1.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。

2. 本标准的格式，编制和表达方法，按国家标准的要求制订。

九、是否涉及专利说明

本规范未涉及有关专利和知识产权。

十、作为强制性标准或者推荐性标准的建议

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。

十一、贯彻标准的要求、措施和建议

**1.贯彻本地方标准的要求**

本标准一旦发布实施，贵州省森林康养基地应严格按照本标准要求进行规划；全省林业主管部门应加强执行本标准的业务培训和技术指导，规划设计单位应自觉按此标准要求进行规划，确保森林康养基地规划的标准化、规范化、健康有序进行。

**2.贯彻本地方标准的措施建议**

（1）加大宣传力度，通过有关媒体发布、公告标准信息，扩大影响。

（2）建议在实施标准过程中对所发现的问题应及时反馈，以利于标准的修订和完善。

十二、其他应予说明的事项

1.本标准为首次制定，无废止现行有关标准的建议。

2.本标准的起草过程中未发生重大分歧意见。