

ICS 点击此处添加 ICS 号

CCS 点击此处添加 CCS 号

团 体 标 准

T/CIIA XXXX—XXXX

公共数据资源社会价值与经济价值 评价标准

Social and Economic Value Evaluation Standards for Public Data Resources

草案版次选择

在提交反馈意见时，请将您知道的相关专利连同支持性文件一并附上。

XXXX - XX - XX 发布

XXXX - XX - XX 实施

中国信息协会 发布

目 次

前 言	II
引 言	III
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
4 概述	1
5 评价指标体系	1
5.1 总体框架	1
5.2 公共数据质量评价	2
5.3 公共数据应用价值评价	4
6 评价流程	4
6.1 评价准备	4
6.2 评价实施	5
6.3 评价得分计算	5
6.4 评价报告编制	5
附 录 A （资料性附录） 数据质量得分计算方法	7
附 录 B （资料性附录） 应用价值得分计算方法	9
附 录 C （资料性附录） 指标权重	10
附 录 D （资料性附录） 数据质量及数据价值等级	11

前 言

本文件按照GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由中国信息协会提出并归口。

本文件起草单位：国家信息中心、国家发展改革委价格监测中心、交通运输部科学研究院、中国软件评测中心（工业和信息化部软件与集成电路促进中心）、中国科学院科技战略咨询研究院、青岛数据资产登记评价中心有限公司、工信通（北京）信息技术有限公司、杭州数据交易所、福建省数据流通控股有限公司、贵阳大数据交易所有限责任公司、厦门大数据有限公司、湖州市数字集团、江苏钟吾大数据发展集团有限公司、包头市数字产业有限责任公司、甘肃联合大数据要素市场有限公司、广州广电运通信息科技有限公司、潍坊市数据发展集团有限公司、数字日照有限公司、武威市天马行大数据运营有限公司、北京海新域城市更新集团有限公司、中电云计算技术有限公司、武汉东湖大数据科技股份有限公司、软通智慧科技有限公司、云津智慧科技有限公司、北京联合永道软件股份有限公司、上海零数众合信息科技有限公司、中移互联网有限公司、广东省电信规划设计院有限公司、珠海市卓轩科技有限公司、深圳壹键数智科技有限公司、上海尖晶数睿数据科技有限公司、云基华海信息技术股份有限公司、明曦金道（北京）数字科技有限公司、广东翼启数据产业有限公司。

本文件主要起草人：王璟璇、王建冬、赵传启、潘菲、崔佳佳、刘礼勇、王闯、冯海红、赵公正、崔雪峰、张树成、刘琦、徐舒扬、黄佩、滕梓、窦悦、郭夏琳、高杰、潘凯伟、卞羽、韩坤洁、姜山、王震、刘思佳、郭润仙、左家云、喻泽文、刘成、吴卫增、王涛、胡博、火兴盛、朱江、代强强、赵烁、杜乐、林镇阳、文博、张春、兰春嘉、杨珍、邱浚漾、余瑶、李柯辰、施美英、周卉章、鲁胜强、郝成金、林树生、李克

引 言

公共数据作为重要的社会资源，既是政府决策的重要依据，也是社会创新和经济发展的支撑。通过科学地评价公共数据资源的社会价值与经济价值，识别高价值公共数据资源，可以更好地发挥其在社会治理、经济发展和社会服务中的作用。

公共数据资源规模庞大，形式多样，其质量和价值也存在较大差异。如何科学评价公共数据资源的社会价值与经济价值，成为亟待解决的问题之一。制定公共数据资源评价标准，有助于明确公共数据资源的质量和价值要求，促进公共数据资源的有效利用和科学管理，为公共数据资源开放、共享和运营的效益评估提供依据，为数据资源管理和数据资产评估提供依据。

公共数据资源社会价值与经济价值评价标准

1 范围

本文件规定了公共数据资源社会价值与经济价值评价的指标体系和评价流程。

本文件适用于公共数据资源供给、管理和运营单位及其相关服务机构，在开展公共数据资源开放、共享、运营等公共数据资源的开发利用和流通活动中，评价公共数据资源的社会价值与经济价值。评价结果可用于评定高质量数据集，评测公共数据资源开放、共享和运营的效益，以及作为数据资源管理和数据资产评估的依据。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 35274-2017 信息安全技术 大数据服务安全能力要求

GB/T 35295-2017 信息安全 大数据 术语

GB/T 36344-2018 信息技术 数据质量评价指标

3 术语和定义

GB/T 35274-2017、GB/T 35295-2017和GB/T 36344-2018界定的以及下列术语和定义适用于本文件。

3.1 公共数据 public data

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。

3.2 公共数据资源 public data resource

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具有利用价值的数据集。

3.3 公共数据资源授权运营 authorized public data resource operation

将县级以上地方人民政府、国家行业主管部门持有的公共数据资源，按照法律法规和相关要求，授权符合条件的运营机构进行治理、开发，并面向市场公平提供数据产品和技术服务的活动。

3.4 公共数据产品和服务 public data product and service

基于公共数据加工形成的，可满足特定需求的公共数据加工品和公共数据服务。

4 概述

公共数据资源社会价值与经济价值的评价对象包括各类公共数据资源。一个评价对象，可以是一项数据资源，或是由多项数据资源组成的一个数据资源集。

公共数据资源社会价值与经济价值评价的指标体系见第5章，评价流程见第6章。

5 评价指标体系

5.1 总体框架

公共数据资源社会价值与经济价值评价指标体系框架见图1。

评价指标体系中包含社会价值和经济价值两个评价维度，各评价维度下均包含数据质量和应用价值两个一级指标。

数据质量是保证数据应用的基础，是实现数据价值的重要保障。数据质量评价包括数据内容质量和信息丰度两个二级指标。内容质量，主要对数据的完整性、唯一性、规范性、一致性、准确性和及时性做出定量测评，测评指标可依据数据源的特性进行调整；信息丰度主要考察数据所涵盖信息维度的丰富性，数据在地理、时间、行业等方面的广度，以及数据来源的稀缺性。公共数据质量评价指标详见表1。

应用价值是从该数据资源的应用场景和效益两个二级指标，分析公共数据的应用价值。应用场景主要考察数据可应用的场景广度和应用频率；效益包括社会价值中社会效益和经济价值中的经济效益，社会效益主要考察数据在公共服务、政府决策、科学研究等方面体现的价值，经济效益主要考察该数据资源以及利用该数据资源加工形成的数据产品和服务的经济收入。公共数据应用价值评价指标详见表2。

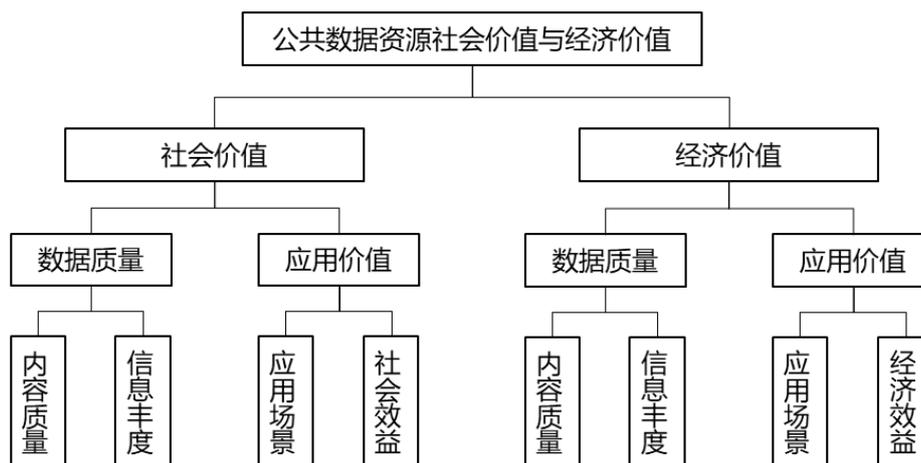


图1 公共数据资源社会价值与经济价值评价指标框架

5.2 公共数据质量评价

5.2.1 内容质量

(1) 完整性

完整性主要考察数据信息的缺失状况，数据缺失包括整条数据记录缺失和一条数据中某个应被赋值的字段信息缺失。评价数据的完整性时，主要关注的指标项包括：

- 记录完整性；
- 内容完整性；
- 字典完整性；
- 字段完整性。

(2) 唯一性

唯一性主要考察某个数据项或某组数据，没有重复的数据值。主要关注的指标项包括：

- 主键唯一性；
- 编码表唯一性；
- 数据记录唯一性。

(3) 规范性

规范性主要考察某个数据项的数据格式规范。主要关注的指标项包括：

- 代码值域约束；
- 长度约束；
- 内容规范约束；
- 取值范围约束；
- 标志位取值约束。

(4) 一致性

一致性是指数据元素的类型、含义以及逻辑必须一致和清晰。主要关注的指标项包括：

- 逻辑一致性；
- 语义一致性；
- 编码一致性。

(5) 准确性

准确性是指数据正确描述其对象实际情况的程度。与一致性不同的是，准确性主要考察数据记录的信息是否存在异常或错误，而不仅是规则上的不一致。准确性主要关注的指标项包括：

- 取值准确性；
- 计算准确性；
- 逻辑准确性。

(6) 及时性

及时性是指数据的更新情况满足应用需求的程度。对及时性有要求的数据资源，主要关注的指标项包括：

- 更新频率；
- 延时时长。

5.2.2 信息丰度

信息丰度主要考察数据的信息维度、内容的覆盖范围和资源的稀缺性。主要关注的指标项包括：

- 信息维度；
- 覆盖范围；
- 稀缺性。

表1 公共数据质量评价指标

二级指标	三级指标	指标描述	评价方法
内容质量	完整性	(1) 记录完整性：数据表中记录的数量与实际发生的记录数量一致，不存在缺失情况； (2) 内容完整性，应被赋值的字段内容中不存在空值； (3) 字典完整性，对数据字典的描述完整； (4) 字段完整性，数据资源中包含应用所需的必要字段。	参考附录A中A.1的要求
	唯一性	(1) 主键唯一性，主键数据不存在重复； (2) 编码表唯一性，编码含义，编码规则不存在重复； (3) 数据记录唯一性，数据记录中不存在重复内容。	
	规范性	(1) 代码值域约束，数据符合数据值域约束，无超出情况； (2) 长度约束，数据符合数据长度要求； (3) 内容规范约束，数据符合数据类型约束要求； (4) 取值范围约束，数据符合数据范围的约束要求； (5) 标志位取值约束，标识数据满足编码规则，如：身份证号、统一社会信用代码等。	
	一致性	(1) 逻辑一致性，存在关联的数据字段间满足逻辑约束关系； (2) 语义一致性，数据定义与内容在不同数据资源中是一致的； (3) 编码一致性，数据内容和编码方式是一致的。	
	准确性	(1) 取值准确性，数据取值准确，与实际状态及真实特征或对应事物无偏差； (2) 计算准确性，经计算得出的数据，计算方式合理，计算结果准确； (3) 逻辑准确性，数据无逻辑错误，如时间顺序正确、数值大小关系正确。	
	及时性	(1) 更新频率，数据按一定频率定期更新，并确保数据在两次数据更新间隔期间的时效性； (2) 延时时长，数据从产生到可以访问、查看、使用的时间间隔在可接受范围内。	参考附录A中A.2的要求
信息丰度	信息维度	数据中是否包含丰富的信息要素，可用于多类应用场景。	参考附录A.3的要求
	覆盖范围	数据是否覆盖较多的行业领域、较广的时间范围和广阔的地域范围。	

	稀缺性	数据来源是否具有不可替代性，是否难以获取。	
--	-----	-----------------------	--

5.3 公共数据应用价值评价

5.3.1 应用场景

数据的应用场景主要考察基于该数据资源开发的数据产品和服务的应用场景数量与广度，以及数据的使用频率，具体可从数据资源被应用的产品、服务与应用场景的数量、用途类别、行业领域等角度进行分析。主要关注的评价指标包括：

- 场景广度；
- 应用频率。

5.3.2 社会效益

数据的社会效益，主要考察数据资源本身或基于该数据资源开发的产品和服务，在应用中所能产生的社会价值。主要关注的评价指标包括：

- 公共服务；
- 政府决策；
- 科学研究。

5.3.3 经济效益

数据的经济效益，主要考察数据资源本身或基于该数据资源开发的模型、产品、应用所能产生的经济价值。主要关注的评价指标包括：

- 数据资源收入；
- 数据产品收入。

表2 数据应用价值评价指标

二级指标	三级指标	指标描述	评价方法
应用场景	场景广度	数据是否在多个领域和场景中得到应用。	参考附录B的要求
	应用频率	数据是否得到持续、大规模使用。	
社会效益	公共服务	数据资源是否有效地支撑公共服务，例如提升公共服务的质量与效率，增强社会安全与应急响应能力，促进教育和知识普及、可持续发展等。	
	政府决策	数据资源是否有效地支撑政府决策，例如以数据支撑提升政府决策的效率与科学性，提高决策的透明与公平性，增强政府的社会问题解决能力等。	
	科学研究	数据资源是否有效地支撑科学研究，推动技术进步与模式创新，通过数据挖掘与分析，衍生出新的商业模式、服务和应用，促进社会发展。	
经济效益	数据资源收入	数据资源供给、管理和运营单位，通过公共数据授权运营等方式，提供数据资源服务取得的收入。	
	数据产品收入	数据运营单位及其相关服务机构，通过利用该数据资源开发的公共数据产品和服务取得的收入。	

6 评价流程

6.1 评价准备

6.1.1 与评价委托方明确开展数据资源评价的目的、评价对象、数据探查时间。

6.1.2 协助评价委托方完成数据资源梳理，明确评价对象的内容和范围，形成数据资源清单。针对评价对象，即待评价的数据资源，应明确包括但不限于：

- 数据资源名称；
- 来源类型；
- 数据来源；
- 信息项；
- 更新频率；

- 数据量；
- 数据范围；
- 存储形式。

6.1.3 核验委托方主体及其待评价数据资源的合法合规性。

6.1.4 针对评价对象编制评价方案，方案内容包括但不限于数据资源清单、评价范围、拟选用的评价指标与方法、人员与分工安排等。

6.2 评价实施

6.2.1 评价实施主要包括数据质量测评和应用价值分析两个环节。

6.2.1 评价实施过程应保证评价委托方的数据安全性，以风险最小化的方式进行数据探查、抽样和测评。

6.2.2 应确保评价对象的真实性和完整性，若以抽样样本作为评价对象，应采用合理的抽样方式确保样本的代表性。

6.2.3 根据第5章的数据质量评价维度，对各数据资源的各信息项选定适宜的评价指标，经相关方确认后，开展相关指标的测评，并作为数据质量评价的输入值。

6.2.4 根据数据资源的应用特点，从第5章的数据应用评价维度中选择适宜的评价指标，经相关方确认后，收集数据资源的应用情况，开展相关指标的测评，并作为数据应用价值评价的输入值。

6.3 评价得分计算

6.3.1 根据评价实施过程中采集的信息，计算指标体系中各项三级指标得分。

6.3.2 设置各指标权重，依次计算数据质量得分、社会价值得分、经济价值得分和综合得分。评价得分计算流程如图2所示，各指标的权重设定参照附录C。

6.3.3 根据得分评定数据质量等级、社会价值等级、经济价值等级和综合等级。各项评价得分的对应评级参照附录D。

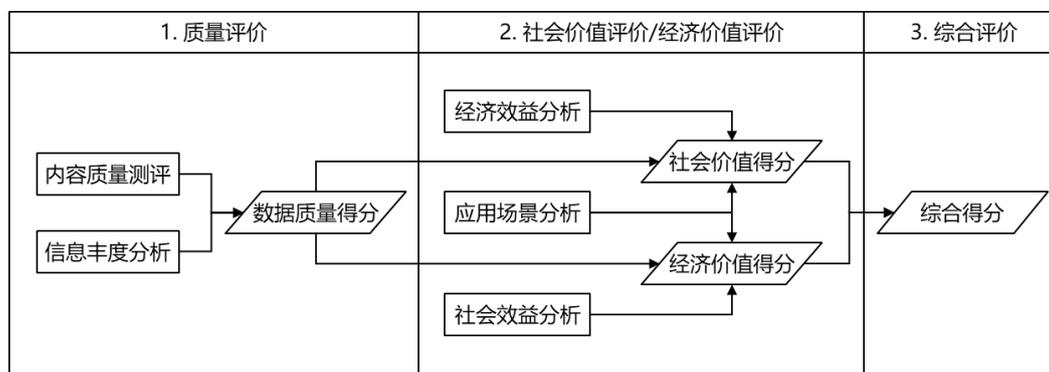


图2 评价得分计算流程

6.4 评价报告编制

6.3.1 完成评价实施后，应根据实际测评和分析结果编制评价报告，出具评价报告前可在不影响评价结果的客观性和分析判断的独立性的前提下，与评价委托方就报告相关内容进行必要沟通。

6.3.2 评价报告包括但不限于以下内容：

- 数据资源基本信息，应包括数据资源名称、资源权属单位、数据类型、主要存储方式、主要来源、资源简介、主要数据范围、数据总量、数据探查时间；
- 数据资源构成，包括但不限于各项数据资源的名称、数据量、测评权重、数据来源、更新频率、信息项；
- 测评方法，包括但不限于测评权重的测定方法、各资源中各指标的测定内容等；
- 数据质量评价结果，包括但不限于总体得分、各资源得分、质量问题分析等；
- 应用价值评价结果，包括但不限于应用场景与效益分析、应用价值得分等；
- 评价总结，包括对数据质量、数据应用、价值情况的概述；

——评价人和审核人签字，评价机构盖章。

6.3.3 评价报告中不应披露数据资源中涉及国家安全、企业商业秘密和个人隐私信息的数据。

6.3.4 未经评价委托方许可，不得向第三方提供或公开评价报告。

附录 A
(资料性附录)
数据质量得分计算方法

A.1 除及时性以外的，数据内容质量各评价指标得分的计算方法为：

$$X = \frac{A}{B}$$

式中：

X ——该维度的质量测评得分；

A ——满足要求的数据量

B ——被评估的数据总量。

A.2 及时性的赋分标准见表 A.1。

表A.1 及时性赋分标准

更新频率与延时时长	得分 (%)
两次数据更新之间的时间间隔不超过1小时，且延时时长不超过1小时。	100
两次数据更新之间的时间间隔不超过1小时，且延时时长大于1小时，但不超过24小时。	95
两次数据更新之间的时间间隔不超过1日，且延时时长不超过1日。	90
两次数据更新之间的时间间隔不超过1日，且延时时长大于1日，但不超过7日。	85
两次数据更新之间的时间间隔不超过1周，且延时时长不超过1周。	80
两次数据更新之间的时间间隔不超过1周，且延时时长大于1周，但不超过5周。	75
两次数据更新之间的时间间隔不超过1个月，且延时时长不超过1个月。	70
两次数据更新之间的时间间隔不超过1个月，且延时时长大于1个月，但不超过3个月。	65
两次数据更新之间的时间间隔不超过1个季度，且延时时长不超过1个季度。	60
两次数据更新之间的时间间隔不超过1个季度，且延时时长大于1个季度，但不超过4个季度。	55
两次数据更新之间的时间间隔不超过1年，且延时时长不超过1年。	50
数据更新频率或延时时长超过1年，或数据不更新但能够满足当前的应用要求。	25 ~ 50
数据不更新，且既有数据难以满足应用要求。	0 ~ 25

A.3 信息丰度的赋分标准见表 A.2。

表A.2 及时性赋分标准

指标	得分 (%)	标准
信息维度	0	缺乏有效信息，难以用于数据产品的开发，以及在场景中的使用。
	25	有效信息较少，数据仅能满足在来源业务系统的原有规划场景中使用。
	50	有效信息充足，数据可基于来源业务系统，应用于同类业务应用场景。
	75	有效信息比较丰富，可用于来源业务系统外的同行业应用场景。
	100	数据信息非常丰富，可用于多类业务场景。
覆盖范围	0	数据内容零散，难以有效覆盖单一领域或地区（覆盖率低于50%）；对于需要持续更新、记录的数据，时间跨度较短（少于1年）。
	25	数据内容较零散，难以全面覆盖单一领域或地区（覆盖率低于90%），且难以有效覆盖多个领域或地区的组合（覆盖率低于50%）；对于持续更新、记录的数据，时间跨度较短（少于1年）。

指标	得分 (%)	标准
	50	数据能够全面覆盖单一领域或地区（覆盖率高于90%），或能有效覆盖多个领域或地区的组合（覆盖率高于50%）；且对于持续更新、记录的数据，时间跨度满足应用要求（一般不少于2年）。
	75	数据能够全面覆盖单一领域或地区（覆盖率高于90%），且能有效覆盖多个领域或地区的组合（覆盖率高于50%）；且对于持续更新、记录的数据，时间跨度满足应用要求（一般不少于2年）。
	100	数据覆盖广泛，能全面覆盖多个领域或地区的组合（覆盖率高于90%）；且对于持续更新、记录的数据，时间跨度满足应用要求（一般不少于2年）。
稀缺性	0	公开信息，数据可从多渠道自由获取。
	25	非公开信息，但数据可从多渠道获取。
	50	公开信息，数据仅可从单一渠道获取。
	75	非公开信息的唯一数据来源，但有近似可替代渠道。
	100	非公开信息的唯一数据来源，无可替代渠道。

附录 B
(资料性附录)
应用价值得分计算方法

应用价值的赋分标准见表 B。

表 B 应用价值赋分标准

二级指标	三级指标	分值 (%)	标准
应用场景	场景广度	0	数据难以在场景中使用。
		25	数据已在来源业务系统的原有规划场景中使用。
		50	数据已基于来源业务系统，应用于同类业务应用场景。
		75	数据已应用于来源业务系统外的同行业应用场景。
		100	数据已应用于多个行业领域的多类应用场景。
	应用频率	0	数据几乎未得到应用（月均调用频次<10）
		25	--（月均调用频次 10-100）
		50	--（月均调用频次 100-500）
		75	--（月均调用频次 500-1000）
		100	数据应用频繁，被持续大规模使用（月均调用频次>1000）
社会效益	公共服务	0-100	利用该数据，①每应用于 1 项公共服务得 10 分，上限 80 分；②应用的部门数每增加一个得 5 分，上限 20 分。
	政府决策	0-100	利用该数据，①每应用于 1 项决策得 5 分，上限 60 分；②每应用于 1 个部门得 10 分，上限 40 分。
	科学研究	0-100	利用该数据，①每完成 1 项创新性应用得 10 分，②每形成 1 项专利得 15 分，③每发表 1 篇核心期刊论文得 25 分；单项得分无上限，满分 100 分。
经济效益	数据资源收入	0-100	$X = M \sqrt{\frac{I}{G}}$ 式中： X ——数据资源收入测评得分，满分 100 分； M ——得分系数，初始取值为 25； I ——数据资源产生的经济收入，单位元； G ——年人均 GDP，单位元。
	数据产品收入	0-100	$X = M \sqrt{\frac{I}{G}}$ 式中： X ——数据产品收入测评得分； M ——得分系数，初始取值为 25； I ——数据产品产生的经济收入，单位元； G ——年人均 GDP，单位元。

附 录 C
(资料性附录)
指标权重

各指标的权重见表 C.1。

表 C 指标权重

评价维度		一级指标		二级指标		三级指标	
指标项	权重 (%)	指标项	权重 (%)	指标项	权重 (%)	指标项	权重 (%)
社会价值	60	数据质量	30	内容质量	70	完整性	15
						唯一性	15
						规范性	20
						一致性	20
						准确性	15
						及时性	15
		应用价值	70	信息丰度	30	信息维度	30
						覆盖范围	30
						稀缺性	40
				应用场景	40	场景广度	50
						应用频率	50
						社会效益	60
政府决策	30						
科学研究	30						
经济价值	40	数据质量	30	内容质量	60	完整性	15
						唯一性	15
						规范性	20
						一致性	20
						准确性	15
						及时性	15
		应用价值	70	信息丰度	40	信息维度	30
						稀缺性	30
						覆盖范围	40
				应用场景	40	场景广度	50
						应用频率	50
						经济效益	60
产品服务收入	50						

注：各级指标权重，均为该指标在上一级指标中的权重；如三级指标“完整性”的权重为15%，表示“完整性”在二级指标“内容质量”中的占比为15%。

附 录 D
(资料性附录)

数据质量及数据价值等级

D.1 数据质量得分的对应数据质量等级见表 D.1。

表D.1 数据质量得分与等级

强度等级	等级名称	分值区间	等级描述
S	卓越级	95 分 \leq X \leq 100 分	在评价时点, 无明显的数据质量问题。
A	优秀级	90 分 \leq X<95 分	在评价时点, 有较少的数据质量问题, 对数据应用一般无明显影响。
B	良好级	85 分 \leq X<90 分	在评价时点, 有一定的数据质量问题, 会影响数据的应用效果。
C	发展级	80 分 \leq X<85 分	在评价时点, 有较多数据质量问题, 严重影响数据的应用效果。
D	基础级	0 分 \leq X<80 分	在评价时点, 数据质量需得到改善后才能满足数据应用的基本要求。

D.2 社会价值、经济价值和综合得分的对应等级见表 D.1。

表D.2 社会价值、经济价值和综合价值得分与等级

强度等级	等级名称	分值区间	等级描述
S	卓越级	90 分 \leq X \leq 100 分	在评价时点, 处于细分领域同业领先水平。
A	优秀级	80 分 \leq X<90 分	在评价时点, 处于细分领域同业优异水平。
B	良好级	70 分 \leq X<80 分	在评价时点, 处于细分领域同业良好水平。
C	发展级	60 分 \leq X<70 分	在评价时点, 处于细分领域同业平均水平。
D	基础级	0 分 \leq X<60 分	在评价时点, 低于细分领域同业平均水平。