

标准数字化应用 第3部分：数据质量规范

Standard digital applications
Part 3: Data quality specification

2024-08-23 发布

2024-08-26 实施

湖北省软件行业协会 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 职责要求	2
5 标准数字化应用数据质量总体要求	2
6 技术要求	4
7 数据质量评价	4
附录 A（资料性） 数据规范字段示例	6

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。

本文件是T/HBSIA 001-2024的第3部分。T/HBSIA 001-2024已经发布了以下部分：

- 第1部分：元数据
- 第2部分：数据交换规范
- 第3部分：数据质量规范

本文件由湖北省标准化与质量研究院提出。

本文件由湖北省软件行业协会归口。

本文件起草单位：湖北省标准化与质量研究院、武汉市道玄科技有限公司、武汉盛锦汇科技有限公司、武汉爱迪科技股份有限公司、湖北华中电力科技开发有限责任公司、武汉金档科技有限公司、武汉达梦数据库股份有限公司、武汉百智诚远科技有限公司。

本文件主要起草人：徐术坤、韩阳昱、华振楠、余梅、舒成、赵亮清、邵璇、莫颜君、康维、彭涛、王豪、周志强、马哲贵、刘红玲、赵艳丽、吴颖波、吴锴、李庄庄、孙莉莉。

引 言

为了更加有效地利用标准数字资源，促进标准数字资源的共享开放，编制可信、易于理解的标准数字化应用标准已成为使用标准数字资源的首要任务。《标准数字化应用》由以下部分构成。

- 第1部分：元数据。目的在于确定元数据描述方法、模型、不同种类元数据的描述及拓展要求。
- 第2部分：数据交换规范。目的在于确定标准数据交换体系、交换流程及交换方式。
- 第3部分：数据质量规范。目的在于确定标准数据质量的总体要求、技术要求及质量评价。

标准数字化应用

第3部分：数据质量规范

1 范围

本文件规定了标准数字化应用中数据质量管理职责要求、总体要求、技术要求及数据质量评价。本文件适用于本文件适用于标准数字化应用的数据质量管理。

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 5271.1-2000 信息技术 词汇 第1部分：基本术语

GB/T 5271.17-2010 信息技术 词汇 第17部分：数据库

GB/T 19000-2016 质量管理体系 基础和术语

GB/T 20000.1-2014 标准化工作指南 第1部分：标准化和相关活动的通用术语

T/HBSIA 001.1-2024 标准数字化应用 第1部分：元数据

3 术语和定义

下列术语和定义适用于本文件。

3.1

数据 data

信息的可再解释的形式化表示，以适用于通信、解释或处理。

[GB/T 5271.1-2000, 定义01.01.02]

3.2

元数据 metadata

关于数据或数据元素的数据（可能包括其数据描述），以及关于数据所有权、存取路径、访问权和数据易变性的数据。

[GB/T 5271.17-2010, 定义17.06.05]

3.3

质量 quality

客体的一组固有特性满足要求的程度。

[GB/T 19000-2016, 定义3.6.2]

3.4

数据质量 data quality

在指定条件下使用时，数据的特性满足明确的和隐含的要求的程度。

3.5

数据集 dataset

可以识别的数据集合。

3.6

数据模型 data model

对分析的图像和文本表述,该分析识别了组织为完成其使命、功能、目标、目的和战略,以及管理和评价组织所需要的数据。

3.7

标准 standard

通过标准化活动,按照规定的程序经协商一致制定,为各种活动或其结果提供规则、指南或特性,供共同使用和重复使用的文件。

[GB/T 20000.1-2014, 定义5.3]

3.8

标准数字化 standard digitization

将标准化的信息、流程、产品或服务转化为数字形式,以便更好地管理、分析、优化和交付。

3.9

标准数据管理者 standard data manager

统筹管理标准数字化应用数据资源的部门或者人员。

4 职责要求

4.1 标准收集和处理主体

4.1.1 遵守国家法律法规要求,确保标准数据采集全过程的合法合规性。

4.1.2 确保收集的标准数据准确、完整和可靠,采取措施来验证数据的来源和有效性。

4.1.3 负责开展标准数据资源目录编制、数据汇集、数据分类分级、清单编制、合规开放和数据安全约束等工作。

4.1.4 负责标准数据的数字化处理工作。

4.2 标准数字化应用数据管理者

4.2.1 采取措施来保护收集的标准数据免受未经授权的访问、滥用或损坏。

4.2.2 设定合理的数据保留期限,并及时删除不必要的数据库。

4.2.3 依据适用的法律和法规共享或转移数据,并与相关方达成协议。保证数据在处理和传输过程中的安全和保密性。

4.2.4 制定和执行数据管理策略,制定合理的数据治理流程和规范,数据质量控制、元数据管理和数据文档化等方面的工作。

4.3 标准数字化应用数据利用主体

4.3.1 确保数据的合法获取,遵守数据隐私、知识产权和其他相关法律规定。

4.3.2 遵守与标准数字化应用数据管理者的协议,处理数据时获取必要的同意或授权。

5 标准数字化应用数据质量总体要求

5.1 数据质量规范

标准数字化应用中数据质量的规范性应包括数据的准确性、完整性、一致性、可用性、可维护性、合理性、安全性、时效性，见图1。

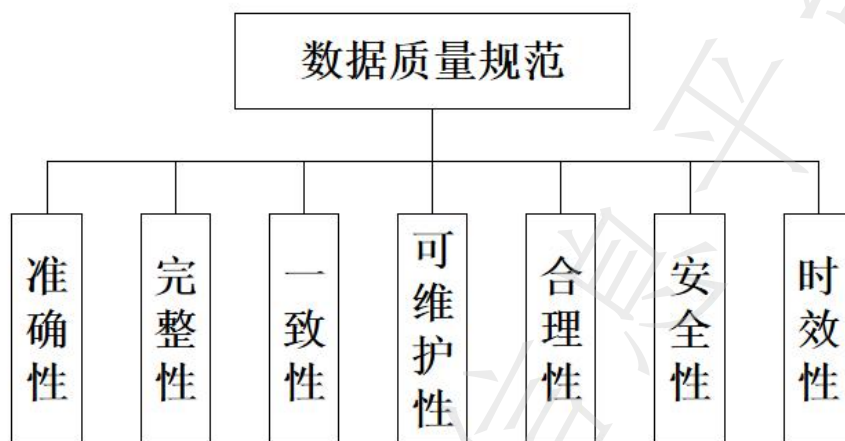


图1 标准数字化应用数据质量规范

5.2 准确性

5.2.1 标准数字化应用数据应准确地反映标准文献的信息。

5.2.2 标准数字化应用数据应经过验证和审查以确保其准确性。

5.3 完整性

5.3.1 标准数字化应用数据应包含所有必要的字段和信息。

5.3.2 标准数字化应用数据应符合预期的结构、格式、属性要求等内容。

5.4 一致性

5.4.1 标准数字化应用数据应与数据库结构保持一致。

5.4.2 标准数字化应用数据存储格式应与存储要求保持一致。

5.4.3 标准数字化应用数据应在不同的系统和应用程序之间保持一致。

5.5 可用性

5.5.1 标准数字化应用数据应以结构化和可管理的方式存储，例如数据库或数据仓库。

5.5.2 标准数字化应用数据应在需要时可以进行传输和访问。

5.5.3 标准数字化应用数据应在数据丢失、硬件故障或灾难性事件等情况下，可以快速恢复数据可用性。

5.6 可维护性

5.6.1 标准数字化应用数据应易于维护和更新。

5.6.2 标准数字化应用数据应有适当的版本控制和历史记录，以便跟踪数据的变化。

5.6.3 标准数字化应用数据应有适当的清理和备份策略，以确保数据的准确性和完整性。

5.7 合理性

- 5.7.1 标准数字化应用数据应符合逻辑和实际情况。
- 5.7.2 标准数字化应用数据应有适当的校对策略，以确保数据的合理性。

5.8 安全性

- 5.8.1 标准数字化应用数据应确保只有授权用户才能访问。
- 5.8.2 标准数字化应用数据应具备加密、访问控制、数据备份和恢复策略等特性。

5.9 时效性

- 5.9.1 标准数字化应用数据应与标准发布公告信息保持一致，确定标准数据是否现行有效。
- 5.9.2 标准数字化应用数据应与标准文献状态保持一致，确保使用标准数据准确可用。
- 5.9.3 标准数字化应用数据应有适当的核校策略，以确保数据的时效性。

6 技术要求

6.1 数据质量控制规则

在标准数据处理过程中，应该遵守的数据处理规则：

- a) 标准号中存在错误年份。
- b) 标准号含有修改单的标准数据。
- c) 标准号含有%的标准数据。
- d) 标准号中有调号、原标准号、由...转化、后面带括号，括号里面有标准号的标准数据。
- e) 标准号中带有~的标准数据。
- f) 标准号中带括号中文字符的和处理状态为X的标准数据。
- g) 标准号中含有被*替代的标准数据。
- h) 标准号中带有增补的标准数据。
- i) 标准号后面带括号的标准数据。
- j) 标准号中含有XG的标准数据。
- k) 标准号中含有COR的标准数据。
- l) A461, A462中出现“%”“被...代替”的标准数据。
- m) 标准号中含有.pdf的标准数据。
- n) 存在标准重复性数据。

6.2 数据规范字段

标准数字化处理的标准数据应符合T/HBSIA XXXX.1规定的格式，字段示例见附录A。

7 数据质量评价

7.1 指标评价

标准数字化应用数据质量指标描述符合第5章的要求，并满足表1规定的相应评价指标。

表 1 评价指标

评价指标	指标名称	指标描述	评价值域
准确性	数据内容正确性	反映标准文献信息的准确程度	低/中/高
	数据格式合规性	标准数据各项内容的格式准确程度	低/中/高
完整性	数据完整性	标准数据的各项内容完整程度（是否存在数据缺失）	低/中/高
一致性	数据一致性	标准数据在不同副本中能够动态保持一致的程度	低/中/高
可用性	数据可访问性	标准数据的访问便利程度（等待时长）	低/中/高
	数据分类分级程度	标准数据分类分级维度、层级和成熟度	低/中/高
可维护性	数据可扩展性	标准数据可扩展情况	低/中/高
合理性	数据是否合理	标准数据是否符合逻辑和实际情况	低/中/高
安全性	数据安全性	数据是否具备加密、访问控制、数据备份和恢复策略	低/中/高
时效性	数据时效性	数据是否与标准状态保持一致	低/中/高

附录 A
(资料性)
数据规范字段示例

表A.1-表A.4给出了标准要素数据规范字段的示例。

A.1 题录数据规范字段示例

字段名称	字段描述	示例
标准号	标准编号	GB/T 43715-2024
中文名称	标准的中文名称	薏仁米
英文名称	标准的英文名称	Coix seed
适用范围	标准适用的领域或对象的范围	本文件适用于以薏苡为原料，经干燥、清理、脱壳等工艺加工制成的薏仁米。
标准状态	标准实施状态，包含“现行”、“废止”、“未生效”	未生效
发布日期	标准正式发布或发行的日期	2024-03-15
实施日期	标准正式生效的日期	2024-10-01
废止日期	标准正式作废的日期	YYYY-MM-DD
发布单位	标准制定机构或标准化组织	国家市场监督管理总局、国家标准化管理委员会
中国标准分类	CCS代码	B22
国际标准分类	ICS代码	67.060
代替标准	被当前标准所取代的旧版标准	GB/T XXXX-XXXX
被替代标准	取代当前标准的新版标准	GB/T XXXX-XXXX
提出单位	标准的提出单位	国家粮食和物资储备局
归口单位	标准的归口单位	全国粮油标准化技术委员会
起草人	标准的起草人员	张XX
主要起草单位	标准的主要起草单位	XX研究院
参与起草单位	标准的参与起草单位	XX公司
标准类型	标准文件的类别	产品标准
引用标准	标准的规范性引用文件	GB/T XXXX-XXXX

A.2 全文信息数据规范字段示例

字段名称	字段描述	示例
标准号	标准编号	GB/T 43715-2024
章节信息	标准文献中当前段落的章节编号及名称	8 标签标识
段落信息	标准文献中各段落内容信息	8.1 产品名称应按照本文件规定的名称标注。
文件内排序	标准文献中当前段落全文顺序序号	32
正文页码	标准文献中当前段落所处页码	3

A.3 术语信息数据规范字段示例

字段名称	字段描述	示例
标准号	标准编号	GB/T 43715-2024
术语中文名称	标准文献中术语的中文名称	薏苡
术语英文名称	标准文献中术语的英文名称	coix
术语定义	标准文献中术语的说明	禾本科草本植物栽培薏苡的颖果
术语注释	标准文献中术语的补充说明	颖壳呈黄、淡褐、黑、白等色，有光泽，籽粒呈椭圆形
来源	标准文献中术语出现的标准编号	GB/T 22515-2008, 2.2.5.10

A.4 指标信息数据规范字段示例

字段名称	字段描述	示例
标准号	标准编号	GB/T 43715-2024
标准化对象	标准所规范的对象	薏仁米
指标类型	内容指标类型	薏仁米质量
体例	内容指标所在的章、条	糙薏仁米
指标项	指标约束项	不完整善粒含量
指标值	指标取值	≤5.0