

团体标准

T/CABC XXX—XXXX

商品源数据 数据质量评价规范

Product source data - specification for data quality assessment

(征求意见稿)

XXXX-XX-XX 发布

XXXX-XX-XX 实施

中国条码技术与应用协会

发布

目 次

前 言 II

商品源数据 数据质量评价规范 1

1 范围 1

2 规范性引用文件 1

3 术语和定义 1

4 数据质量指标 2

4.1 完整性指标 2

4.2 规范性指标 2

4.3 准确性指标 3

5 数据质量评价内容 3

5.1 数据质量缺陷等级 3

5.2 数据质量评价内容 3

6 数据质量评价方法 6

6.1 数据质量缺陷分级 6

6.2 数据质量评分方法 6

7 数据质量等级 7

7.1 质量等级划分规则 7

7.2 质量等级判定方法 7

8 数据质量评价报告 7

附 录 A （资料性） 商品源数据质量评价报告 8

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。

本文件由中国物品编码中心提出。

本文件由中国条码技术与应用协会归口。

本文件起草单位：XXXX。

本文件主要起草人：XXXX。

商品源数据 数据质量评价规范

1 范围

本文件规定了商品源数据质量评价的术语和定义、数据质量指标、数据质量评价内容、数据质量评价方法、数据质量等级、数据质量评价报告等要求。

本文件适用于商品源数据质量评价工作。

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 2828.1—2012 计数抽样检验程序 第1部分：按接收质量限(AQL)检索的逐批检验抽样计划

T/CABC XXX—XXXX 商品源数据 数据质量核查实施规范

3 术语和定义

下列术语和定义适用于本文件。

3.1

商品源数据 product source data

是商品实物的数字化信息。以全球贸易项目代码为关键字，包括商品图片、标签文字、包装尺寸、商品重量等，并符合GS1全球数据模型（GS1 Global Data Model）定义的数据结构和规则。

3.2

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。

注：可以通过人工或自动手段处理数据。

[GB/T 5271.1—2000, 定义01.01.02]

3.3

单位数据项 data item

是组成商品数据的最小单元。例如每个属性值视为一个单位数据项，图片的每一个质量要求视为一个单位数据项。

3.4

数据质量 data quality

在指定条件下使用时，数据的完整性、规范性、准确性满足明确和隐含要求的程度。

3.5

致命缺陷 the most serious defect

商品源数据中的极重要单位数据项极严重不符合规定，以致不经返工处理不能正常使用。

3.6

严重缺陷 high defect

商品源数据中的极重要单位数据项不符合要求，对数据应用有严重影响。

3.7

重缺陷 medium defect

商品源数据中的重要单位数据项不符合要求，对数据应用有重大影响。

3.8

轻缺陷 low defect

商品源数据中的一般单位数据项不符合要求，对数据应用有轻微影响。

4 数据质量指标

4.1 完整性指标

数据质量完整性指标描述见表1。

表1 完整性指标描述

一级质量元素	描述	二级质量元素	描述
完整性	单位数据项的多余、重复或者遗漏	冗余	数据中出现多余或重复的单位数据项。
		缺失	数据中缺失要求的单位数据项。

4.2 规范性指标

数据质量规范性指标描述见表2。

表2 规范性指标描述

一级质量元素	描述	二级质量元素	描述
规范性	单位数据项的规范程度	合规性	数据符合GB 12904《商品条码、零售商品编码与条码表示》、《商品源数

			据采集规范》、《商品源数据 商品属性信息规范 第1部分：通用属性》、《商品源数据 数据质量控制实施规范》等规范性文件的程度。
--	--	--	--

4.3 准确性指标

数据质量准确性指标描述见表3。

表3 准确性指标描述

一级质量元素	描述	二级质量元素	描述
准确性	单位数据项的准确程度	非量化单位数据项的正确性	非量化单位数据项的正确与否。
		量化单位数据项的正确性	量化单位数据项的值与客观真值之间的接近程度。

5 数据质量评价内容

5.1 数据质量缺陷等级

本文件中用数据质量缺陷来量化数据质量评价中出现的各类不符合要求的问题，数据质量缺陷分为四级：致命缺陷、严重缺陷、重缺陷、轻缺陷。

5.2 数据质量评价内容

5.2.1 概述

数据质量评价内容包括商品图片信息、商品测量信息、商品其他属性信息三部分内容。

5.2.2 商品图片缺陷等级

表4 商品图片数据质量缺陷分级

项目编号	项目	要求	缺陷编号	缺陷	数据指标分类	缺陷等级
TP-A	陈列图视图	必选正面、左面和顶面视图各 1 张，产品其他有内容面应拍摄图片；扁形商品必选的是正面和背面。	TP-A1	默认正面判断错误	完整性	致命缺陷
			TP-A2	缺少必选图 1 张	完整性	重缺陷
			TP-A3	缺少必选图 2 张	完整性	严重缺陷
			TP-A4	缺少其他有内容面图超过 2 张	完整性	重缺陷
			TP-A5	缺少其他有内容面图 1 张	完整性	轻缺陷

TP-B	商品主图俯角视图	除了扁平的特殊商品之外，应该拍摄3个不同的15°俯角3D视图，俯角拍摄时镜头中心对准产品正面中心点。所拍摄的角度误差不超过±5°。	TP-B1	缺少1张俯角视图	完整性	轻缺陷
			TP-B2	缺少2张及以上俯角视图	完整性	重缺陷
			TP-B3	俯角角度大于15°+5°	规范性	轻缺陷
			TP-B4	左右俯角视图角度误差超过±5°	规范性	轻缺陷
TP-C	背景	纯白（RGB 255/255/255）	TP-C1	背景非纯白	规范性	重缺陷
TP-D	背景宽高比	1:1	TP-D1	背景宽高比非1:1	规范性	轻缺陷
TP-E	商品图位置	在背景上居中	TP-E1	商品图在背景上未居中	规范性	轻缺陷
TP-F	商品图边距	商品图最长边的顶点与背景边界的空白距离应占背景边长的2%~5%	TP-F1	商品图最长边的顶点与背景边界的空白距离占背景边长总长<2%	规范性	轻缺陷
			TP-F2	商品图最长边的顶点与背景边界的空白距离占背景边长总长>5%	规范性	重缺陷
TP-G	商品轮廓	商品轮廓应完整、清晰	TP-G1	正面图商品轮廓不完整或不清晰	规范性	重缺陷
			TP-G2	非正面图商品轮廓不完整或不清晰	规范性	轻缺陷
TP-H	色差	图片无偏色，与商品颜色一致	TP-H1	图片偏色严重，与商品颜色严重不一致	规范性	轻缺陷
TP-I	对比度	对比度应整体平衡，高光均匀，不造成标签信息模糊	TP-I1	对比度严重不平衡	规范性	轻缺陷
			TP-I2	高光不均匀，标签信息模糊	规范性	重缺陷
TP-J	反光和阴影	处理反光与阴影时，应注重细节和形状，不造成标签信息丢失	TP-J1	出现商品变形失真	规范性	重缺陷
			TP-J2	反光与阴影造成标签信息丢失	规范性	重缺陷
TP-K	多余内容	无指纹、污渍或水印等	TP-K1	出现与商品拍摄无关的物品（营销图片除外）	规范性	重缺陷
			TP-K2	有指纹、污渍或水印	规范性	重缺陷
TP-L	修图痕迹	图片缩放至100%时无明显修图痕迹	TP-L1	100%比例下有明显修图痕迹	规范性	轻缺陷
TP-M	失真	压缩或放大时无失真	TP-M1	压缩或放大时失真	规范性	轻缺陷
TP-N	像素	2400*2400 ~ 4800*4800	TP-N1	最低像素未达到2400*2400	规范性	重缺陷

			TP-N2	最高像素超过4800*4800	规范性	重缺陷
TP-O	分辨率	300dpi	TP-O1	分辨率非 300dpi	规范性	轻缺陷
TP-P	色彩模式	RGB	TP-P1	色彩模式非 RGB	规范性	轻缺陷
TP-Q	大小	≤7M	TP-Q1	大小>7M	规范性	轻缺陷
TP-R	文件格式	JPEG/PNG/TIF	TP-R1	文件格式非 JPEG/PNG/TIF	规范性	轻缺陷
TP-S	命名	命名规范	TP-S1	命名不规范	规范性	重缺陷

5.2.3 商品测量缺陷等级

表5 商品测量信息数据质量缺陷分级

项目编号	项目	要求	缺陷编号	缺陷	数据指标分类	缺陷等级
CL-A	高宽深关系	高宽深关系正确	CL-A1	零售单元高、宽、深关系错误	规范性	严重缺陷
			CL-A2	非零售单元高、宽、深关系错误	规范性	严重缺陷
CL-B	数量级	尺寸、重量数量级与商品（包装）相符	CL-B1	尺寸数量级与商品（包装）不符	准确性	重缺陷
			CL-B2	重量数量级与商品（包装）不符	准确性	重缺陷
CL-C	数据完整	高宽深、重量数据完整	CL-C1	高宽深数据缺失	完整性	致命缺陷
			CL-C2	重量数据缺失	完整性	致命缺陷
CL-D	数值修约	按规则修约	CL-D1	未按规则修约。长度未精确到毫米，重量未精确到克。	准确性	重缺陷

5.2.4 商品其他属性信息缺陷等级

表6 商品其他属性信息数据质量缺陷分级

项目编号	项目	要求	缺陷编号	缺陷	数据指标分类	缺陷等级
SX-A	必填项	必填项不应漏填	SX-A-(1)	必填单位数据项漏填 1 次	完整性	致命缺陷
SX-B	必填项	必填项不应多余或重复	SX-B-(1)	必填单位数据项多余或重复填写 1 次	完整性	重缺陷
SX-C	必填项	必填项属性值应符	SX-C-	必填单位数据项不符	准确性	重缺陷

		合包装标签描述	(1)	合包装标签描述 1 次		
SX-D	必填项	必填项单位应正确	SX-D-(1)	必填单位数据项单位错误 1 次	准确性	严重缺陷
SX-E	条件必填项	条件必填项不应漏填	SX-E-(1)	条件必填单位数据项漏填 1 次	完整性	重缺陷
SX-F	条件必填项	条件必填项不应多余或重复	SX-F-(1)	条件必填单位数据项多余或重复填写 1 次	完整性	重缺陷
SX-G	条件必填项	条件必填项属性值应符合包装标签描述	SX-G-(1)	条件必填单位数据项不符合包装标签描述 1 次	准确性	重缺陷
SX-H	条件必填项	条件必填项单位应正确	SX-H-(1)	条件必填单位数据项错误 1 次	准确性	重缺陷
SX-I	选填项	选填项不应多余或重复	SX-I-(1)	选填单位数据项多余或重复填写 1 次	完整性	轻缺陷
SX-J	选填项	选填项属性值应符合包装标签描述	SX-J-(1)	选填单位数据项不符合包装标签描述 1 次	准确性	轻缺陷
SX-K	选填项	选填项单位应正确	SX-K-(1)	选填单位数据项错误 1 次	准确性	重缺陷

6 数据质量评价方法

6.1 数据质量缺陷分级

收到数据质量核查人员提供的《核查结果明细表》后，按照 5.2 中的规定对核查结果进行缺陷分级，并将结果汇总到表 7《数据质量评价汇总表》中。

6.2 数据质量评分方法

按以下步骤对样本数据集进行评分：

第一步：若样本数据集中出现致命缺陷，则直接判定对应的 GTIN 数据得分为 0，且该样本数据集得分亦为 0。

第二步：计算单个 GTIN 数据质量缺陷率 U 。缺陷率 U =轻缺陷量/涉及单位数据项数 m 。严重缺陷量和重缺陷量应换算为轻缺陷量后纳入缺陷率计算。严重缺陷、重缺陷、轻缺陷的换算比例为 6: 3: 1。

第三步：计算单个 GTIN 数据得分。单个 GTIN 数据得分 $S=100-U \times (40/AQL)$ ，其中 AQL 为该批数据核查抽样方案中确定的接收质量限。单个 GTIN 数据得分最低为零分，低于零分按零分计算。

第四步：计算样本数据集得分。样本数据集得分为样本中所有 GTIN 数据得分的算数平均数。

表7 《数据质量评价汇总表》

批次名称/编号: XXXX-XXXX

送审单位				核查单位				N/AQL/n(批量/接收质量限/样本量)					
序号	GTIN	样本单位数据项数 M	涉及单位数据项数 m	核查方法	核查类别	缺陷编号	缺陷描述	缺陷数量				缺陷率 U%	得分 S
								致命缺陷	严重缺陷	重缺陷	轻缺陷		
样本数据集得分:													
核查范围和内容包括:													

7 数据质量等级

7.1 质量等级划分规则

质量等级的划分规则采用优秀、良好、合格、不合格 4 个等级，按百分制划分质量等级。S 为得分，数据质量等级的划分为：

- 1) 优秀：S≥90 分；
- 2) 良好：90>S≥75 分；
- 3) 合格：75>S≥60 分；
- 4) 不合格：S<60 分。

7.2 质量等级判定方法

按以下步骤判定批数据的质量等级：

第一步：若样本数据集中出现致命缺陷，则直接判定该批数据不合格；

第二步：若样本数据集中未出现致命缺陷，则按 6.2 中的评分方法和 7.1 中的质量等级划分规则计算样本数据集中不合格的 GTIN 数量。若不合格数大于等于核查抽样方案中限定的拒收数（参见《商品源数据 数据质量核查实施规范》附录 A），则判定该批数据为不合格。若不合格数小于等于核查抽样方案中限定的接收数（参见《商品源数据 数据质量核查实施规范》附录 A），则判定该批数据为合格。

第三步：若该批数据判定为合格，则按《数据质量评价汇总表》中的样本数据集得分和 7.1 中的质量等级划分规则判定该批数据的质量等级。

8 数据质量评价报告

数据质量评价报告是数据质量检查与评价过程、方法及结果的综合描述和评价，是对商品源数据质量特征的综合评价。数据质量评价报告由正文和数据质量评价附表组成，正文应是质量评价方法、过程和结果的全面记录和描述，包括质量检查与评价的组织、数据集概况、检查方法、评价依据、评价规则、质量评述、存在问题及结论等。

商品源数据质量评价报告示例参见附录A。

附 录 A
(资料性)
商品源数据质量评价报告

A. 1数据质量评价报告封面格式

编号: _____

商品源数据 数据质量评价报告

批次名称/编号: _____

采编单位: _____

送评单位: _____

评价单位: _____

评价日期: _____

A.2 数据质量评价报告文本格式

一、数据的整体描述

1. 数据提交和评价单位。

相关组织或机构的人员、时间、地点、形式等。

2. 数据概况

数据来源、数据(集)范围、提交时间、格式等。

3. 数据质量评价的依据

列出相关依据文件。

二、数据质量评价过程描述

1. 数据质量的评价步骤

2. 数据质量的评价方法

采用定量评价法的应包含:抽样方案、评价项目、评价规则、缺陷等级、权重等。

3. 数据质量的评价结果

对数据质量评价的过程和结果进行描述(包括对致命缺陷、严重缺陷、重缺陷、轻缺陷的说明)。

三、数据质量评价结论

包括评价结果、等级和(或)合格与否等结论。

四、附表

《数据质量核查表》、《数据质量评价汇总表》、评审组名单等。