

河南省地方标准

DB41/T 2847—2025

商品源数据 数据质量控制规范

2025 - 04 - 21 发布

2025 - 07 - 20 实施

目 次

前言 II

1 范围 1

2 规范性引用文件 1

3 术语和定义 1

4 基本要求 1

5 数据质量检查 1

6 数据质量评价 2

7 纠正与控制措施 3

附录 A（规范性） 检查指标表..... 4

附录 B（规范性） 评价指标表..... 5

参考文献 7

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。

本文件由河南市场监督管理局提出并归口。

本文件起草单位：河南省标准化和质量研究院、中国物品编码中心河南分中心、方圆标志认证集团河南有限公司、北京化工大学文法学院。

本文件主要起草人：韩颖、杨士民、武松霞、周树甫、刘娟、杨艳丽、梁慧芳、吴娜、李丽、刘强、王婉、潘立龔、张荣花、李智、丁蕾、徐巧莹、赵茜、王诗源、韦甜甜、周晓丽、马晓霞、张艺馨。

商品源数据 数据质量控制规范

1 范围

本文件规定了商品源数据的数据质量控制的基本要求、质量检查、质量评价、纠正与控制措施。
本文件适用于商品源数据的数据质量控制。

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 12904 商品条码 零售商品编码与条码表示
DB41/T 2215 商品源数据采集规范

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

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

数据质量

商品源数据的数据完整性、规范性、准确性符合GS1全球数据模型（GS1 Global Data Model）定义的数据结构和规则及相关技术特性的程度。

4 基本要求

4.1 按照 GB 12904 和 DB41/T 2215 的要求对商品源数据的数据规范性、准确性、完整性和可用性进行检查，对不符合要求的数据采取措施。具体检查要求如下：

- 源数据描述应符合标准规定的字段格式、命名规则及编码体系，确保数据结构和元数据与行业规范一致；
- 商品属性（如尺寸、重量）应通过标准化工具（测量仪或电子秤）进行验证，误差范围应符合相关行业标准；
- 数据记录内容应覆盖所有数据，不存在多余、重复、缺失现象；
- 数据字段值应在不同场景下有效可用。

4.2 商品源数据评价报告应客观、准确地反映商品源数据的质量状况，为数据的改进和优化提供数据。

5 数据质量检查

5.1 数据检查方式

- 对采集的数据进行分析，选择检查方式：
- 全量检查：对所涉及的所有数据进行逐一检查数据质量；
 - 抽样检查：按照抽样方案（见表1），对抽取的数据进行逐一检查数据质量。

表1 抽样方案

批量	样本量	接收数	拒收数
2~8	2	0	1
9~15	3	0	1
16~25	5	0	1
26~50	8	0	1
51~90	13	1	2
91~150	20	1	2
151~280	32	1	2
281~500	50	2	3
501~1 200	80	3	4
1 201~2 200	125	5	6
2 201~10 000	200	7	8
10 001~25 000	315	14	15
25 001~150 000	500	21	22
150 001~500 000	800	21	22
500 001及以上	1250	21	22

注：当批量小于2时，样本量等于批量，为全量检查，接收数为0。

5.2 数据检查指标

按照表A.1给出的检查指标进行数据检查，检查指标表按附录A的规定。

6 数据质量评价

6.1 数据收集

由商品源数据采集的相关技术机构按照既定的抽样方案，收集用于评价的数据样本。抽样方案可参照表1。

6.2 评价实施

6.2.1 数据审查

评价人员对收集的数据样本按照附录B的规定进行审查，检查数据的各项指标是否符合要求，并记录审查结果。

6.2.2 评分计算

根据表B.1的评分细则，对每个样本的各项指标进行打分，然后按照式1计算出每个样本的综合得分。

$$T = \sum_1^x(t \times m)$$

..... (1)

式中：
T ——综合得分；
x ——覆盖的一级指标价数；
t ——一级得分指标；
m ——对应一级指标权重。

6.2.3 评价结果判定

6.2.3.1 合格判定

设定综合得分合格分数线为80分（可根据实际情况调整）。综合得分达到或超过80分的数据样本判定为合格；得分低于80分的数据样本判定为不合格。

6.2.3.2 结果判定

对所有样本的评价结果进行汇总，统计合格样本数量 and 不合格样本数量，计算合格率。

$$M = \frac{n}{N} \times 100\%$$

..... (2)

式中：
M ——合格率；
n ——合格样本数量；
N ——样本总数。

6.2.4 评价报告编制

评价完成后，编制详细的评价报告，报告内容包括评价目的、评价范围、评价依据、评价过程、评价结果（包括样本得分情况、合格率）、存在问题等。

7 纠正与控制措施

7.1 不合格数据控制

对不合格数据进行标识，能改善的数据进行纠正，并对改善过的数据对其进行再一次的检查，使其符合要求；无法改善的不合格数据，予以剔除。

7.2 不合格数据控制措施

对采集过程中涉及的人、机、料、法、环等全过程进行质量控制，当出现不符合要求的情况时，宜采取以下纠正措施：

- 对采集人员进行能力再确认，审核人员采取双人复审机制；
- 对采集设备进行再一次溯源，确保溯源结果正确；
- 对采集的样品采取替代采样机制，确保样品完整无破损；
- 对采集的环境采取实时监控，并制定应急预案，确保贮藏室环境符合样品存贮要求；
- 对采集完成的数据进行初审，确保数据符合要求；若数据初审不合格，找到导致不合格数据的出现原因，消除不合格数据的原因，防止不合格数据再发生。

附 录 A
(规范性)
检查指标表

检查指标见表A. 1。

表A. 1 检查指标

一级指标	二级指标	三级指标	四级指标	五级指标
1 条码符合性	系统成员名称与产品责任主体一致	—	—	—
	使用集团公司条码，子公司应备案	—	—	—
2 图片	图片清晰完整、顺序方向正确无日期	图片数量及顺序	规则商品6张图，不规则商品2张图；其中正背面图必有	—
		默认正面	应正确	—
		图片质量	真实、全面、清晰、正立、无异物、角度误差不超过10度	—
		背景全白、像素、分辨率、大小	RGB为255:255:255；像素为2400×2400~4800×4800；分辨率为300 dpi；大小为7M	—
		边缘轮廓、抠图痕迹	完整清晰无明显修图痕迹	—
		边界距离	商品居中，最长边顶点与图片边界的距离应占图片边长总长的2%~5%	—
3 测量	高宽深及重量判断	默认正面	默认正面正确	—
		高宽深逻辑	高宽深关系正确	—
		数值逻辑	尺寸、重量数量级与商品相符，按规则修约	—
4 文字信息采编	录入数据逐字检查无误，内容准确	营养成分表分类填写正确	—	—
		生产企业信息正确	—	—
		—	—	有机标识、地理标识、税务编码、内装中包、GPC分类、包装类型酒品分类、品牌、净含量、配料
注：角度误差不超过10度测量方法可选用量角器比对商品主轴线与图片边框夹角。				

附 录 B
(规范性)
评价指标表

评价指标见表B. 1。

表B. 1 评价指标

一级指标	权重	二级指标	评分细则
条码符合性	20%	系统成员名称与产品责任主体一致性	系统成员名称与产品责任主体完全一致得50分；差异但提供工商备案证明得25分；不一致且无证明得0分
		集团公司条码使用备案情况	子公司使用集团公司条码且备案齐全得50分；备案部分缺失得25分；未备案得0分
图片质量	20%	图片数量及顺序	规则商品有6张图、不规则商品有2张图且正背面齐全得25分；每缺少一张关键图片或图片顺序错误扣5分；扣完为止
		默认正面准确性	默认正面正确得25分；错误得0分
		图片质量特性	图片真实、全面、清晰、正立、无异物、角度误差不超过10度得15分；每出现一项不符合要求的情况扣3分；扣完为止
		背景、像素、分辨率及大小	图片背景、像素、分辨率、大小均符合标准（RGB为255:255:255；像素为2 400*2 400~4 800*4 800；分辨率为300dpi；大小为7M）得25分；每一项不符合标准扣5分；扣完为止
测量准确性	15%	边缘轮廓与抠图痕迹	边缘轮廓完整清晰无明显修图痕迹得10分；有明显修图痕迹扣10分
		默认正面判断准确性	默认正面的高宽深及重量判断正确得50分，错误得0分
		数值逻辑合理性	尺寸、重量数量级与商品相符且按规则修约得50分；数值不合理或修约错误，每一项不符合标准扣15分；扣完为止

表B.1 评价指标（续）

一级指标	权重	二级指标	评分细则
文字信息准确性	20%	录入数据准确性	录入数据逐字准确、内容无误得50分，出现错别字或信息错误，每个错误扣10分；扣完为止
		各类信息填写正确性	营养成分表、生产企业信息、有机标识、地理标识、税务编码、内装中包、GPC分类、包装类型、酒品分类、品牌、净含量、配料等信息填写正确得50分；每出现一项错误扣5分；扣完为止
数据完整性	15%	数据项完整性	数据记录内容覆盖所有规定数据项得50分；每缺失一个关键数据项扣20分；扣完为止
			数据记录无多余、重复现象，且不存在数据缺失得50分；出现多余或重复记录，每条扣10分；扣完为止
数据一致性	10%	数据格式一致性	数据集中每个数据项格式符合要求且保持一致得50分；格式不一致，每出现一处扣10分；扣完为止
		数据内容一致性	同一数据在不同场景下表述一致得50分；出现不一致情况，每一次扣10分；扣完为止

参 考 文 献

- [1] 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
-