

CY

中华人民共和国新闻出版行业标准

CY/T 239—2021

ISLI 数据质量检测

ISLI data quality inspection

行业标准信息服务平台

2021 - 09 - 22 发布

2021 - 11 - 01 实施

目 次

前言	III
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
4 基本原则	2
5 检测对象	2
6 数据质量要求	2
6.1 数据完整性	2
6.2 数据规范性	2
6.3 数据正确性	2
6.4 数据一致性	2
6.5 数据提交及时性	2
7 检测项目	3
8 检测方法流程	3
8.1 检测方法	3
8.2 检测流程	3
9 检测规则	4
9.1 检测项目合格判定	4
9.2 检测结果判定	4
10 质量检测报告	4
附录 A (资料性) ISLI 数据质量检测报告格式	5

行业标准信息服务平台

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。

本文件由全国新闻出版标准化技术委员会（SAC/TC527）提出并归口。

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。

本文件起草单位：中国音像与数字出版协会、化学工业出版社有限公司、河北省新华书店有限责任公司、北京中启智源数字信息技术有限责任公司、北方工业大学、CNONIX国家标准应用与推广实验室。

本文件主要起草人：张传静、余叶烨、唐贾军、冯思然、丁瑜、丁汀、李鹏、韩丽璞、靳力英、李钢、王强、何丽、付晓玲、刘高军、黄玉萌、郑南、张晓明、陈银莉、王淮、黄兆晨。

行业标准信息服务平台

ISLI 数据质量检测

1 范围

本文件规定了ISLI数据质量检测的基本原则、检测对象、数据质量要求、检测项目、检测方法及流程、检测规则，以及质量检测报告。

本文件适用于新闻出版领域ISLI标准的应用推广及数据质量检测。

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 32867 中国标准关联标识符（ISLI）

CY/T 238—2021 ISLI服务编码和ISLI编码申请管理

3 术语和定义

GB/T 32867界定的以及下列术语和定义适用于本文件。

3.1

ISLI编码 ISLI code

国际标准关联标识编码。

[来源：GB/T 32867—2016，2.3，有修改]

3.2

数据质量 data quality

对数据的完整性、规范性、正确性、一致性和及时性进行量化约束的指标。

3.3

空值率 null data rate

在数据库记录中规定的可选数据字段显示为空的记录数在所有记录中占的比率。

3.4

申请者 applicant

申请为一项ISLI服务申请服务编码和要求为一个关联分配ISLI编码的请求方。

3.5

源 source

作为关联起点的实体。

[来源：GB/T 32867—2016，2.6]

3.6

目标 target

作为关联终点的实体。

[来源：GB/T 32867—2016，2.7]

3.7

服务 service

由注册机构定义的类型化的源和类型化的目标之间的通用应用关联类别。

[来源：GB/T 32867—2016，2.8]

4 基本原则

数据质量检测遵循以下原则：

- a) 通过技术手段降低检测过程中的人工干预；
- b) 检测内容应符合业务需求；
- c) 检测结果可量化，明确质量问题及位置，设计合理的阈值及量化标准。

5 检测对象

检测对象为CY/T 238—2021定义的数据，包括：

- a) 服务编码管理数据（服务申请者信息、ISLI 服务信息）；
- b) ISLI 编码管理数据（源信息、目标信息）；
- c) 源和目标实体之间的关联关系。

6 数据质量要求

6.1 数据完整性

检测数据的完整性，包含以下内容：

- a) 遵循CY/T 238—2021表1和表2必选数据元素的要求，无遗漏；
- b) 遵循CY/T 238—2021表1和表2对可选数据元素的要求，可选元素的空值率应控制在50%以内。

6.2 数据规范性

检测数据的规范性，包含以下内容：

- a) 遵循CY/T 238—2021表1和表2规定的的数据标签、定义、属性及使用规范；
- b) 遵循CY/T 238—2021附录A中的代码表，代码取值使用正确。

6.3 数据正确性

检测数据的正确性，包含以下内容：

- a) 数据正确：源、目标信息的数据应正确；
- b) 必选和有则必选数据元素表达不应有误。

6.4 数据一致性

检测数据的一致性，包含以下内容：

- a) 概念一致性：数据概念规则应一致无歧义；
- b) 值域一致性：取值应在规定的值域范围内；
- c) 关联关系一致性：源、目标信息之间的关联关系与实体关联关系应一致。

6.5 数据提交及时性

检测数据提交的及时性，包含以下内容：

- a) 数据应按 ISLI 标准注册机构规定的时间提交；
- b) 数据变更时，应按 ISLI 标准注册机构的要求提交变更。

7 检测项目

ISLI数据质量检测项目如表1所示。

表1 数据质量检测项目

质量要求	检测项目
数据完整性	a) 必选数据元素非空
	b) 可选元素空值率小于 50%
数据规范性	c) 数据标签/定义/属性/格式规范
	d) 代码表使用正确
数据正确性	e) 无非法数据和无效数据
	f) 源和目标信息的数据正确
	g) 元素差错率低于 5%, 差错率指被检数据记录中数据字段的错误记录数在所提供记录中占的比率
数据一致性	h) 数据概念一致, 无歧义
	i) 关联关系一致, 实体信息与实体一致
	j) 数据元素取值在值域范围内
数据提交及时性	k) 已分配 ISLI 编码的数据完成创建、修改至数据上传到 ISLI 标准注册机构的时间应在规定时间范围内

8 检测方法及流程

8.1 检测方法

数据质量检测方法包括：

- 自动检测：数据的规范性、完整性、值域一致性、及时性由程序自动检测；
- 人工检测：数据的正确性、概念一致性、关联关系一致性由人工检测。人工检测分为定期检测和随机抽测。ISLI标准注册机构每季度抽测被检数据不少于入库数据的50%，当数据检测结果不合格时，应向申请者反馈检测结果（格式见附录A），并在5个工作日内由申请者对数据进行修订，对修订后的数据应再次进行检测。

8.2 检测流程

数据质量检测的基本流程如图1所示。

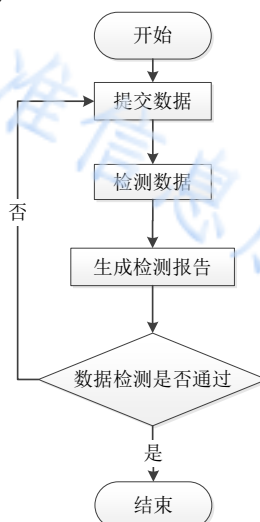


图1 检测流程

9 检测规则

9.1 检测项目合格判定

数据质量要求、检测项目合格判定和检测方法如表2所示。

表2 质量检测项目合格判定

质量要求	检测项目合格判定	检测方法
数据完整性	满足表1 a)、b) 则合格	程序自动检测
	表1 a)、b)有一项不满足则不合格	
数据规范性	满足表1 c)、d)合格	程序自动检测
	表1 c)、d)有一项不满足则不合格	
数据正确性	满足表1 e) 则合格	程序自动检测
	表1 e) 不满足则不合格	
	满足表1 f)、g) 则合格	人工检测
	表1 f)、g)有一项不满足则不合格	
数据一致性	满足表1 h)、i) 则合格	人工检测
	表1 h)、i)有一项不满足则不合格	
	满足表1 j)则合格	程序自动检测
	表1 j)不满足则不合格	
数据提交及时性	满足表1 k)则合格	程序自动检测
	表1 k)不满足则不合格	

9.2 检测结果判定

当数据完整性、数据规范性、数据正确性、数据一致性、数据提交及时性的检测结果全部合格时，数据质量综合检测结论为合格；存在一项不合格，则数据质量综合检测结论为不合格。

10 质量检测报告

数据质量检测完成后应给出数据质量检测报告。

数据质量检测报告应包括送检日期、检测日期、被检数据描述、数据来源单位、数据检测单位、检测人员、检测问题清单、检测结论等信息。数据质量检测报告样表见附录A。

行业标准信息服务平台

附 录 A
(资料性)
ISLI 数据质量检测报告格式

A.1 ISLI 数据质量检测报告样例

ISLI 数据质量检测报告样例见表 A.1。

表A.1 ISLI 数据质量检测报告样例

编号		数据来源单位	
送检日期		检测日期	
数据检测单位			
被检数据描述	被检数据数量、类型等信息		
问题清单	数据完整性		
	数据规范性		
	数据正确性		
	数据一致性		
	数据及时性		
检测结论	<input type="checkbox"/> 合格 <input type="checkbox"/> 不合格		
检测人员 (签字)			
备注			

中华人民共和国新闻出版行业标准

ISLI 数据质量检测

CY/T 239—2021

*

中国书籍出版社出版发行

北京市丰台区三路居路 97 号

邮政编码：100073

电话：（010）52257143 52257140

北京睿和名扬印刷有限公司

各地新华书店经销

*

开本 880 毫米 × 1230 毫米 1/16 印张 0.75 字数 11 千字

2021 年 10 月第 1 版 2021 年 10 月第 1 次印刷

*

书号：35068 · 227

定价：12.00 元

如有印装差错 由本社发行部调换

版权专有 翻印必究

举报电话：（010）52257140