

# DB2301

## 黑龙江省哈尔滨市地方标准

DB2301/T 113—2022

### 公共数据采集规范

Specifications for public data acquisition

2022- 11- 14 发布

2022- 12- 13 实施

哈尔滨市市场监督管理局 发布

# 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 总体要求.....	1
5 数据采集准备.....	2
6 数据采集方案.....	7
7 数据映射关系.....	7
8 数据采集方式.....	7
9 数据质量控制.....	9
10 绩效评价.....	9
附录 A（规范性） 数据协议方申请/变更表.....	10
附录 B（资料性） 数据采集绩效考核表.....	12
参考文献.....	13

## 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。

本文件由哈尔滨市大数据中心提出并归口。

本文件起草单位：哈尔滨市大数据中心。

本文件主要起草人：张文程、刘承启、张茁、顾鸣声、姜荣昌、刘桓、国晓宇、蒋彤影、吴琼、裴伟明、蒋月英、张长龙、杨拓。

# 公共数据采集规范

## 1 范围

本文件规定了公共数据采集的术语和定义、总体要求、数据采集准备、数据采集方法、数据映射关系、数据采集方式、数据质量控制、绩效评价。

本文件适用于规范哈尔滨市公共数据开放平台的数据采集工作。

## 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 22239-2008 信息安全技术 信息系统安全等级保护基本要求

GB/T 35274 信息安全技术 大数据服务安全能力要求

GB/T 38888 数据采集软件的性能及校准方法

DB2301/T 91 公共数据资源目录工作规范

## 3 术语和定义

下列术语和定义适用于本文件。

### 3.1

#### 公共数据 common data

国家机关和法律、法规授权的具有管理公共事务职能的组织以及供水、供电、供气、供热、通讯、公共交通等公共服务运营单位(以下统称公共管理和服务机构)在依法履职或者提供公共管理和服务过程中收集、产生的，以一定形式记录、保存的各类数据及其衍生数据。

注：包含政务、公益事业单位数据和公用企业数据。

### 3.2

#### 公共数据平台 common data platform

统一规划、统筹建设，用于承载市级大数据业务的平台。

### 3.3

#### 数据采集 data acquisition

将特定原始数据收集进入公共数据平台的过程。

### 3.4

#### 数据协议方 data protocol party

通过公共数据平台提供资源的公共管理和服务机构。

## 4 总体要求

- 4.1 公共数据平台方应建立统一的采集管理制度，满足公共数据管理总体框架的要求，并配备系统管理员，实行逐级授权管理，实现管理角色的划分。
- 4.2 公共数据平台方应制定具体操作流程，按流程进行数据采集，保障公共数据平台能正确执行数据采集任务。
- 4.3 公共数据平台方应保证采集数据运行环境的安全，公共数据平台安全等级保护基本要求满足 GB/T 22239 三级的规定。
- 4.4 公共数据平台方通过对数据协议方的数据及数据接口进行分析，优化数据采集方案，确定数据采集方式。
- 4.5 数据协议方应遵守公共数据平台方制定的管理制度和操作流程。
- 4.6 数据协议方应梳理其公共数据资源目录，明确公共数据更新频率。
- 4.7 数据协议方应配备部门操作员，负责数据采集对接工作，按公共数据平台数据采集流程采集、上传数据。

## 5 数据采集准备

### 5.1 注册和变更

公共管理和服务机构应在公共数据管理平台用户注册信息，并填写附录 A 中表 A.1 《平台用户申请表》，将申请材料提交至公共数据管理部门，由公共数据管理部门按图 1 的流程，对申请表内容进行审核。填写内容完整、清晰的，审核予以通过。若审核未通过，应告知公共管理和服务机构未通过的原因，由公共管理和服务机构修改后重新提交，完成申请后，成为数据协议方，并按照公共数据平台方的要求进行管理与授权。

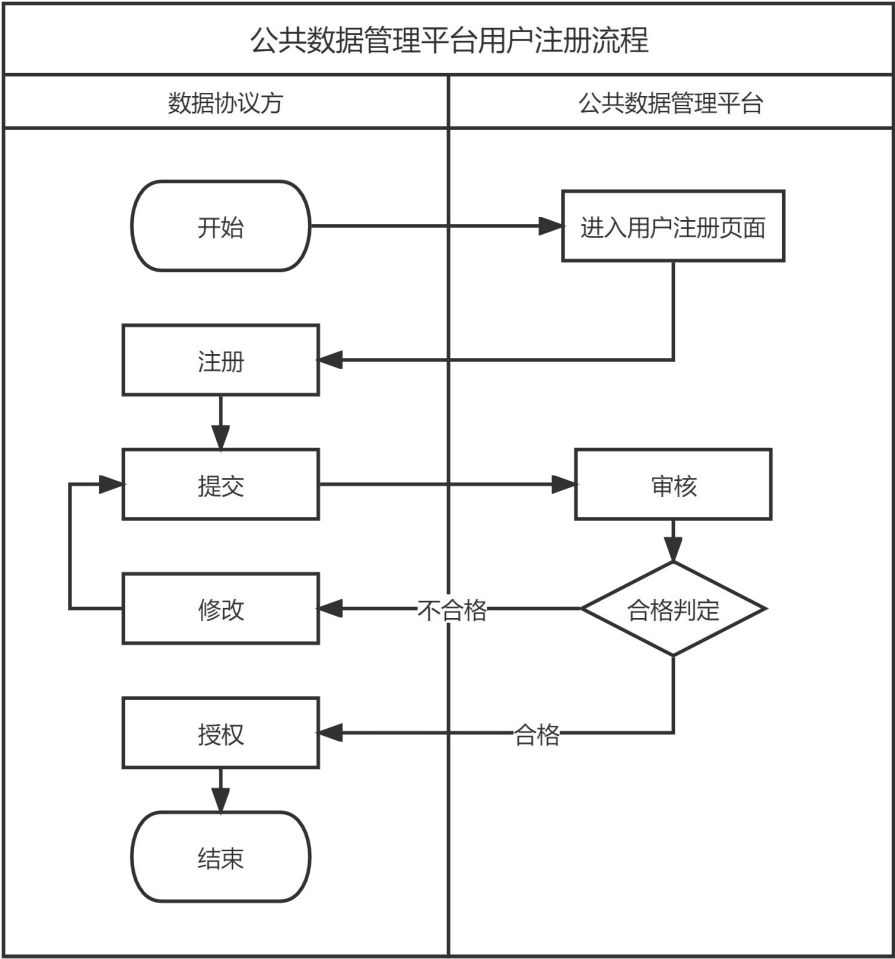


图 1 平台用户注册流程图

5.1.1 数据协议方如需要变更用户信息，应在公共数据管理平台填写附录 A 中表 A.2 《平台用户变更申请表》，将申请材料提交至公共数据管理部门，由公共数据管理部门按图 2 的流程变更用户信息及权限。

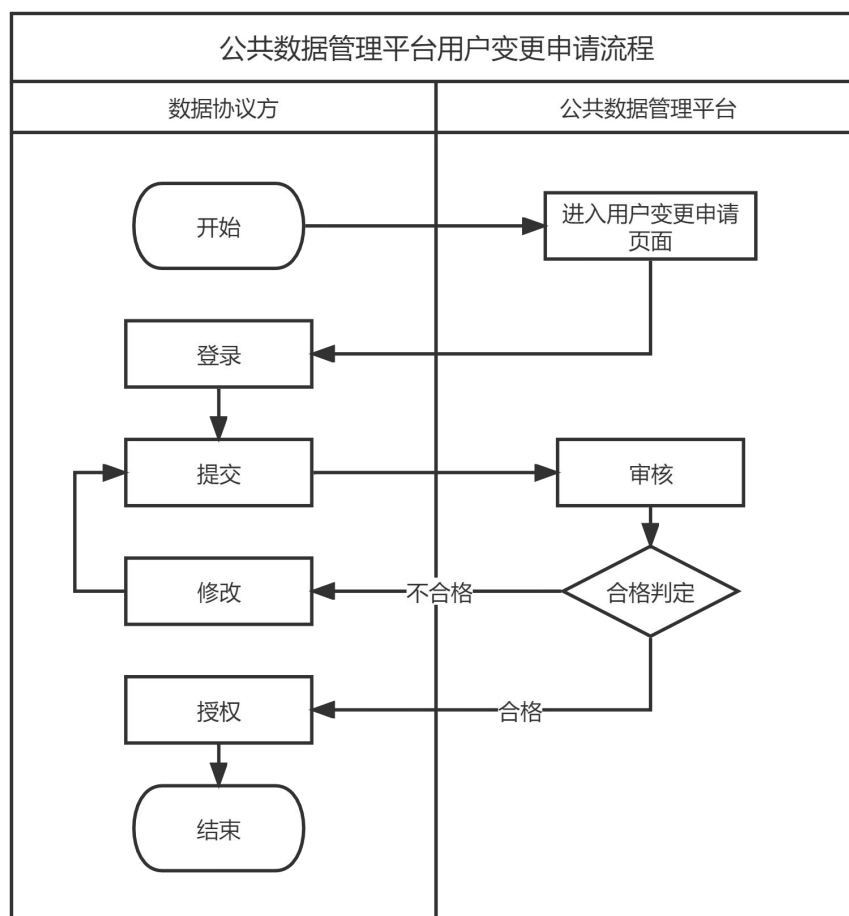
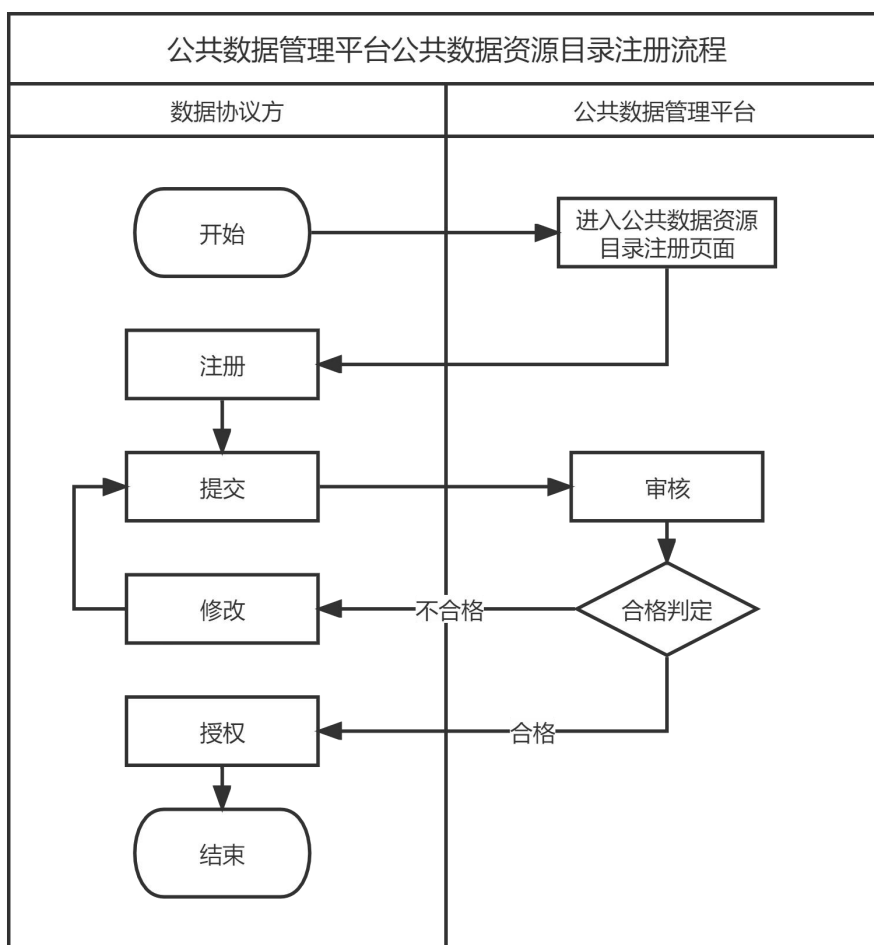


图2 平台变更用户流程

## 5.2 数据资源目录管理

5.2.1 数据协议方应分析本单位涉及的公共数据，按照 DB2301/T 91 的规则，编制本单位公共数据资源目录，按图 3 的流程在公共数据平台注册。公共数据平台方对注册的公共数据资源目录进行审核，符合 DB2301/T 91 的要求，审核予以通过。若审核不通过，应告知原因。数据协议方修改后，需重新提交审核。



**图 3 公共资源目录注册流程**

注：数据协议方在完成公共数据资源目录申请后，由公共数据平台方对公共数据资源目录对应的数据库表类信息资源、文件类信息资源、接口类信息资源进行审核。

5.2.2 如发生机构改革或者法定职责调整，数据协议方应在有关法律法规作出修订或行政职能发生变更之日起，按图 4 的流程，对相应的公共数据资源目录名称、信息项、描述等内容进行变更，确保公共数据资源目录与实际情况的一致。公共数据平台方对变更的公共数据资源目录进行审核，符合 DB2301/T 91 的要求，审核予以通过。若审核不通过，应告知原因。数据协议方修改后，需重新提交审核。



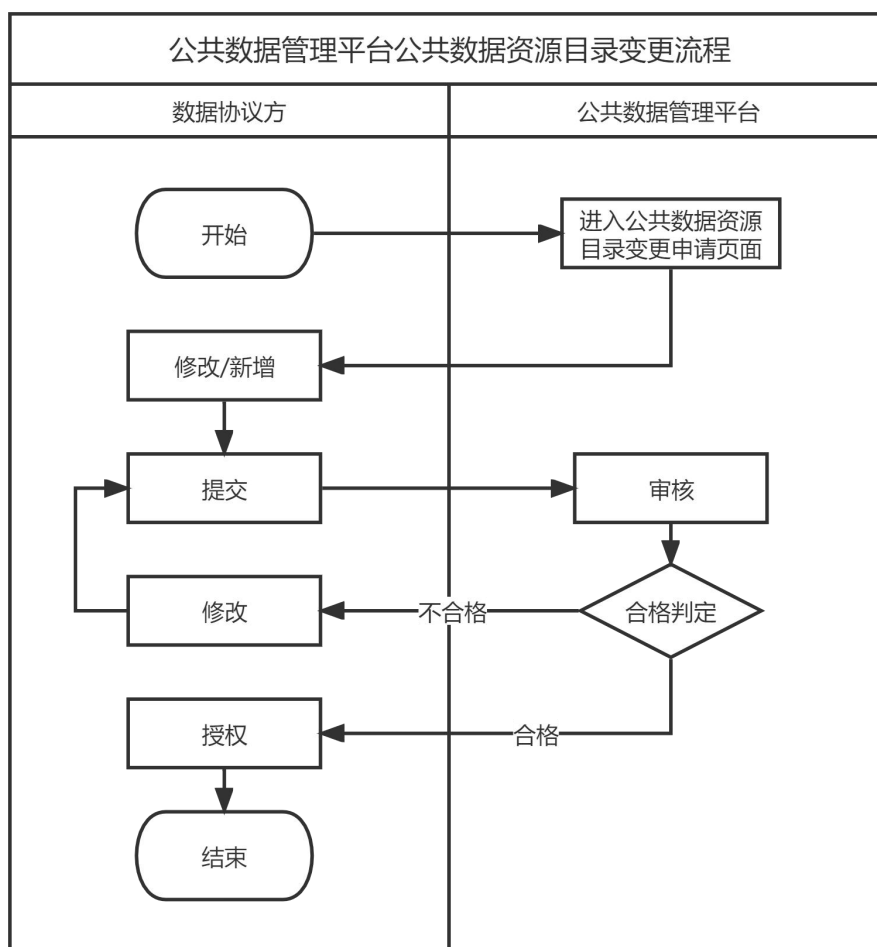


图4 公共资源目录变更流程

注：数据协议方在完成公共数据资源目录变更申请后，应保证数据库表资源随目录变更的生效同步完成变更。

### 5.3 数据安全保障

5.3.1 公共数据平台方应建立公共数据资源安全管理制度，分级管理、按流程申请、按流程修改。确保数据来源的合法性、正当性、可定位溯源性，满足公共数据资源采集审计及监测的要求。

5.3.2 公共数据平台方应建立数据访问权限管理制度，明确公共数据平台方和数据协议方分级权限安全管理职责分工和工作要求，对数据采集环境、设施和技术采取安全管控措施，数据在整个采集、转化、传输过程中应依据授权使用，不被非法冒充、窃取、篡改、抵赖。

5.3.3 公共数据平台方应建立公共数据安全监测、安全审计、数字签名、数据加密等技术或手段，对不同数据进行分类并标识，采用安全技术进行安全维护。

5.3.4 公共数据平台方应建立应急事件响应机制，编制应急预案，开展应急演练，编制应急演练报告，持续优化应急预案等。

5.3.5 公共数据平台方数据服务安全能力应符合 GB/T 35274 的规定，定期对公共数据采集的安全性进行风险评估，并据此制定相应的风险处理计划，及时排查安全漏洞、加固安全技术。

- 5.3.6 公共数据平台方应开展公共数据管理安全培训，编制公共数据安全培训规划，培训公共数据专业知识，开展公共数据安全工作经验交流，宣传公共数据安全知识等。
- 5.3.7 公共数据平台方应明确数据采集过程中个人信息和重要数据的知悉范围和安全管控措施。并采取必要的技术手段和管理措施保证数据不被泄露。
- 5.3.8 数据协议方应梳理接入应用系统和终端的公共数据，确定接入终端数量、网络带宽和接入地点等信息，制定接入实施方案，明确访问控制、接入认证和审计等安全措施。

## 6 数据采集方案

- 6.1 采集方案应包括采集内容、采集频率、采集方式等。
- 6.2 采集内容由数据协议方填报，分类、分项符合 DB2301/T 91 的规定。
- 6.3 公共数据平台应支持增量更新、全量更新、定时更新、事件触发更新和手动更新等方式。
- 6.4 公共数据平台提供多种标准协议的服务接入方式，包括但不限于数据库抽取、服务网关、消息队列、文件传输、直报系统、标准协议接口。
- 6.5 公共数据采集实时性要求低的数据可采用定时批量采集的方式，实时性要求高的数据应采用实时采集的方式。
- 6.6 公共数据量较大、单批量采集可能会造成系统故障的，应使用分批采集。
- 6.7 宜使用数据协议方系统的备份库作为采集对象，使用备份库时应保证数据一致性和可用性。对于海量数据，宜支持分批或增量读取，宜采用分布式方式对数据源进行读取。

## 7 数据映射关系

- 7.1 数据协议方通过数据库表方式向公共数据平台提供数据时，应建立数据协议方数据库表与平台数据库表之间的存储结构映射关系，通过数据库表对接。
- 7.2 数据协议方以接口方式向公共数据平台提供数据时，如返回数据为结构化或半结构化的接口类型，应建立接口返回数据结构与目标数据库之间存储结构映射关系，通过接口进行数据采集。
- 7.3 数据协议方以接口方式向公共数据平台提供数据时，如返回数据为非结构化的接口类型，可将文件通过接口上传后建立文件映射关系表。

## 8 数据采集方式

### 8.1 数据采集方式分类

数据采集方式包括在线数据采集和离线数据提交。

### 8.2 获取公共数据采集权限

确定公共数据采集方案，获取公共数据的采集权限，包括但不限于：

- 数据量；
- 数据范围；
- 采集时间；
- 采集区间；
- 数据使用权限。

### 8.3 在线数据采集

8.3.1 在线数据采集符合图 5 的流程。

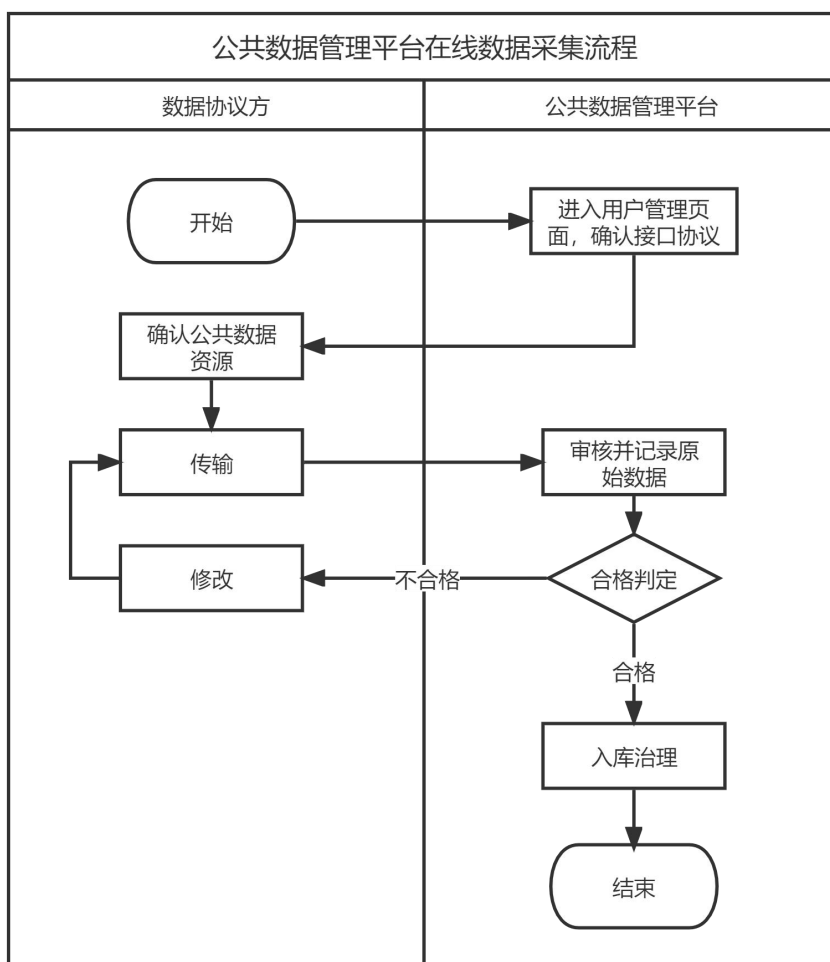


图 5 线数据采集流程

8.3.2 确定数据项间的逻辑关系，通过数据服务接口协议进行数据传输，实现对接采集。

8.3.3 数据协议方如发生数据变更，可根据业务需要申请增加数据信息项目，由公共数据平台方依申请负责创建数据项。

8.3.4 公共数据平台方应维护数据采集的业务管理规范，可以查询、查看采集数据信息，但不能新增、删除数据信息。

8.3.5 公共数据平台方应确认写入数据与读取数据数量一致。判定一致后，按流程将数据入库待治理；发现数据存在差异，联系数据协议方重新获取加载信息，并将修改和原始记录都存储下来，以供校验，获取差异数据应符合 GB/T 38888 的规定。

### 8.4 提交离线数据

8.4.1 离线数据可由表格数据、文件材料、信息扫描录入，通过移动介质提交，保证数据的完整性和准确性。

8.4.2 离线数据文件采集应与公共数据平台的数据库表结构一一对应，确定数据分类和格式，实现数据分类入库存储。

## 8.5 数据清洗管理

8.5.1 对原始数据开展数据清洗管理，定义清洗规则，分析、检查和处理数据源存在的质量问题。对定义的清洗方法的正确性和效率进行验证与评估，对不满足清洗要求的清洗方法进行调整和改进。

8.5.2 数据转换应对数据的标准代码、格式、类型等进行转换。必要时，可建立数据转换规则表。

## 9 数据质量控制

对采集的原始数据进行清洗、转换、分析等处理，确保数据的完整性、准确性、一致性、可访问性、时效性。

- 完整性：应包含数据规则要求的数据的必要元素；
- 准确性：应真实反映数据所描述的实体；
- 一致性：应保证数据与其他特定上下文中使用的数据无矛盾；
- 时效性：应保证数据发生变化后及时被更新；
- 可访问性：应保证数据在需要时能被安全访问；
- 可追性：应保证数据能够被跟踪和管理。

## 10 绩效评价

### 10.1 总体要求

10.1.1 应通过数据的逐年积累，各绩效指标各年度数据的对比分析，建立标准统一、数据准确、普遍认可的绩效指标体系和评分标准。

10.1.2 绩效评价设计应能反映数据提供方提供数据的基本情况。

10.1.3 绩效评价设计支持自动生成数据报表模式。

10.1.4 可实行指标库动态化管理，应用于数据采集评价。

10.1.5 绩效评价数据统计方式，宜具备易用性和易见性，实现数据易于展示和调取，操作步骤简洁合理。

### 10.2 绩效指标

10.2.1 按数据协议方注册的数据资源目录内容，分为共性绩效指标和个性绩效指标，见附录 B。

10.2.2 绩效指标分为三级，包括一级指标、二级指标、三级指标，见附录 B。

### 10.3 绩效指标管理

指标库的维护是指数据采集绩效指标的入库、调整、退库等操作，应在既定管理规则约束下实施：

- 对共性绩效指标，由平台进行维护；
- 对个性绩效指标，由数据协议方提出申请，填报需要新增、调整或删除的数据采集绩效指标信息，同时提供相关的理由或文件依据，经平台审核，不合格返回数据协议方，合格后进行调整更新。

附 录 A  
(规范性)  
数据协议方申请/变更表

A.1 平台用户申请表

表 A.1 为平台用户申请表。

表 A.1 平台用户申请表

申请单位名称				
统一社会信用代码				
单位性质		单位经济类型		
单位所属行业		单位隶属关系		
单位地址				
单位负责人		联系方式		
单位电子邮箱		单位邮编		
联络人职务及姓名		联系方式		
类别 角色	所在处室	姓名	手机号码	邮箱
部门管理员				
部门操作员				
.				
单位负责人意见：  单位负责人签字： <div style="float: right; margin-top: 20px;">             (公章)              日期：           </div>				

## A.2 平台用户变更申请表

表 A.2 为平台用户变更申请表。

表 A.2 平台用户变更申请表

申请单位名称				
统一社会信用代码				
单位性质		单位经济类型		
单位所属行业		单位隶属关系		
单位地址				
单位负责人		联系方式		
单位电子邮箱		单位邮编		
变更事项		变更理由		
联络人职务及姓名		联系方式		
类别 角色	所在处室	姓名	手机号码	邮箱
部门管理员				
部门操作员				
.				
单位负责人意见：				
单位负责人签字：				
				(公章)
				日期：

**附 录 B**  
**(资料性)**  
**数据采集绩效考核表**

- B.1** 绩效考核指标分为共性绩效指标和个性绩效指标。
- 共性绩效指标是适用于所有绩效管理对象、常规性工作的绩效指标；
- 个性绩效指标是适用于不同业务主管部门或行业领域、专业性较强的指标，主要根据业务主管部门或行业特点设定。
- B.2** 绩效指标分为三级，一级指标为最上级，下级指标从属于上级指标：
- 一级指标为采集数据项；
- 二级指标对应一级指标分为支持自动采集数量指标、人工采集通过指标、时效指标；
- 三级指标由数据协议方申请设置，由平台收录，包括针对性强的绩效指标。
- B.3** 公共数据平台方对数据协议方开展数据采集的绩效评价，见表 B.1。

一级指标	二级指标	三级指标	分值	得分
采集数据项	支持自动采集数量指标 (40分)	超出数据协议方申请自动采集数量	40	
		完成自动采集数量达到 90%	25	
		完成自动采集数量达到 80%	15	
	人工采集通过指标 (30分)	占自动采集小于等于 2%	30	
		占自动采集小于等于 5%	20	
		占自动采集小于等于 10%	10	
	时效指标 (30分)	实时采集大于等于 80%	30	
		实时采集大于等于 60%	20	
		实时采集大于等于 40%	10	
总分				

**表 B.1 数据采集绩效考核表**

## 参 考 文 献

- [1] 《国务院办公厅印发政务信息系统整合共享实施方案》
  - [2] 《黑龙江省促进大数据发展应用条例》
  - [3] 《政务信息资源共享管理暂行办法》
  - [4] 《推动“数字龙江”建设加快数字经济高质量发展若干政策措施》
  - [5] 《黑龙江省“十四五”数字政府建设规划》
  - [6] 《哈尔滨市加快数字经济发展实施方案（2022—2024年）》
  - [7] 《哈尔滨市公共数据开放管理办法》
  - [8] 《哈尔滨市“十四五”数字经济“1+4”发展规划体系》
  - [9] GB/T 39440-2020 公共信用信息资源目录编制指南
  - [10] GB/T 21063.1-2007 政务信息资源目录体系 第1部分：总体框架
  - [11] GB/T 21063.2-2007 政务信息资源目录体系 第2部分：技术要求
  - [12] GB/T 21063.3-2007 政务信息资源目录体系 第3部分：核心元数据
  - [13] GB/T 21063.4-2007 政务信息资源目录体系 第4部分：政务信息资源分类
  - [14] GB/T 21063.6-2007 政务信息资源目录体系 第6部分：技术管理要求
  - [15] DB23/T 2839—2021 政务信息资源目录 第1部分：编码规则
  - [16] DB23/T 2840—2021 政务信息资源目录 第2部分：核心元数据
  - [17] DB23/T 2841—2021 政务信息资源目录 第3部分：编制指南
  - [18] DB23/T 2842—2021 政务信息资源数据交换规范
  - [19] DB23/T 3304—2022 大数据平台数据接入规范
-