

DB 4201

武汉市地方标准

DB 4201/T XXX—XXXX

武汉市全民健康信息平台医疗机构数据采集规范

Data collection standard of Wuhan National Health Information Platform
for medical institutions

XXXX—XX—XX 发布

XXXX—XX—XX 实施

武汉市市场监督管理局 发布

目 次

前 言	II
1 范围	3
2 规范性引用文件	3
3 术语和定义	3
4 数据采集原则	4
5 数据采集内容	4
5.1 采集数据集	4
5.2 数据基本属性	5
6 数据采集方法	5
6.1 医院端推送数据采集解析	5
6.2 医院端镜像库采集解析	5
6.3 云端全量数据采集解析	6
7 数据采集流程	6
7.1 主动推送数据	6
7.2 自动采集	7
8 数据质量控制	9
9 数据采集安全	10

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。

本文件由武汉市卫生健康信息中心提出。

本文件由武汉市卫生健康委员会归口。

本文件起草单位：武汉市卫生健康信息中心、万达信息股份有限公司。

本文件主要起草人：XX。

武汉市全民健康信息平台医疗机构数据采集规范

1 范围

本文件规定了武汉市全民健康信息平台医疗机构数据采集的内容、方法、流程和安全等要求。
本文件适用于武汉市全民健康信息平台医疗机构数据采集。

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。

WS 363 卫生信息数据元目录

WS 364 卫生信息数据元值域代码

WS 365 城乡居民健康档案基本数据集

WS 370 卫生信息基本数据集编制规范

WS 445 电子病历基本数据集

WS/T 448 基于居民健康档案的区域卫生信息平台技术规范

WS/T 482 卫生信息共享文档编制规范

WS/T 483 健康档案共享文档规范

3 术语和定义

下列术语和定义适用于本文件。

3.1

数据域

指具有统一的人员主索引、基础字典，以及业务逻辑和数据一致性的数据集合。

3.2

医疗机构

指具有唯一的统一社会信用代码以及组织机构代码的实体机构。

3.3

分院

指医疗机构开展的另一物理位置且具有独立的卫生资源但还未具有统一社会信用代码和组织机构代码的实体机构。

3.4

数据表

指数据标准中，要求医疗机构数据上传所参照标准的具体表现形式，每张数据表应有明确的业务解释，其中应包含有具体的数据表字段，字段属性及字段描述等。

3.5

上传参照时间

指数据表在业务上要求数据上传的时间节点，与实际上传时间的差异用于评价上传时效性。

3.6

必填

指每张数据表中的数据表字段的上传填报要求之一，该字段要求为必填时，表示该字段有明确的业务需求，是业务所需要的核心字段，应充分理解字段说明，上传时应不能为空，有定义值域时，应在值域范围内。

3.7

鼓励填

指每张数据表中的数据表字段的上传填报要求之一，该字段要求为鼓励填时，表示该字段亦有业务需求，是业务所需要的重要字段，但考虑医疗机构信息化水平和业务覆盖范围等因素，采用评价体系中的鼓励分子以奖励。

3.8

数据元

即数据元素，是通过定义、标识、表示和允许值等一系列属性描述的数据单元，在一定语境下，构建一个语义正确、独立且无歧义的特定概念语义的信息单元。

3.9

数据元值域

是数据元允许值的集合。

4 数据采集原则

主要内容包括：

- a) 可靠性:指数据的获取源可靠，在一定的条件下，系统对数据的获取过程包括数据源的选择、数据审核、数据筛选、传输、数据存储，确保采集数据的可靠性。
- b) 可追溯性:指数据在采集过程中对数据采集的系统、时间等因素进行监控，确保数据采集过程的可追溯性。时效性:根据数据的特点，设置其上传参照时间，使数据从产生到结束的流程能及时更新，发挥数据在传输过程中的及时性和有效性。

5 数据采集内容

5.1 采集数据集

医疗机构数据采集内容可分为主数据字典数据、门急诊数据、住院数据、临床诊疗综合性数据、统计汇总类数据、业务运营数据六种类型，主要包括：

a) 主数据字典数据:各医疗机构基础性主数据字典信息，包括但不限于就诊人基本信息数据、医疗机构基本信息数据、科室数据、医护人员数据、住院，观察，手术床位信息数据、挂号科别与专科数据、院内收费明细项目对照字典数据、材料明细字典数据、药品明细数据、库房字典数据、院内诊断字典数据、临床检验指标院内字典等数据表。

b) 门急诊数据:反映门急诊、医疗体检等医疗应用数据，包括但不限于个人基本信息数据、签到挂号信息数据、接诊信息数据、诊断明细数据、处置单信息数据、医嘱信息数据、门急诊收费明细数据、门诊病历信息数据、门急诊留观入/出观记录信息数据、体检报告信息等数据表。

c) 住院数据：反映医疗保健、住院、出院和其他医疗服务情况的数据，包括但不限于个人基本信息数据、入院信息数据、住院诊断数据、住院医嘱数据、出/退院登记信息数据、住院收费明细数据、新生儿信息数据、住院病案首页信息数据、病程记录数据、护理记录数据、转诊（院）记录等数据表。

d) 临床诊疗综合性数据：包括但不限于手术记录数据、麻醉记录数据、输血记录数据、用药信息数据、出院小结数据、临床辅助检查报告数据、实验室检验检查数据、不良事件等数据表。

e) 统计汇总类数据：指医疗机构填报的各类周期性统计指标报表类数据，包括但不限于统计日、月周期指标数据、就诊人失信数据、医院分担超指标数据、双评议信息等数据表。

f) 业务运营数据：包括但不限于药品库存信息数据、材料型号规格及价格信息数据、血液产品信息数据、医疗设备信息数据等数据表。

5.2 数据基本属性

本文使用表1中所示的8个属性对医疗机构数据采集信息实体和信息元素进行描述：

表1 信息实体和信息元素描述属性

序号	描述属性 ¹	要求 ²	定义及说明 ³
1	数据项	M	数据元其中一个信息元素，由各项字段集合形成医疗机构业务域
2	字段名称	M	信息元素的英文名
3	外部标识符	M	遵照《WS 363.2 卫生信息数据元目录》第2部分:标识 对分项信息元素进行识别不同实体的符号
4	数据类型	M	对信息实体和信息元素的有效值域和允许对该值域内的值的规定
5	数据长度	M	对信息实体和信息元素的有效值域和允许对该值域内的值长度的规定
6	填报要求	M	对信息实体和信息元素的上报力度规定，主要包括必填、选填、鼓励填
7	外键	0	信息元素与其他数据项数据元素之间的关联关系
8	说明	0	对信息实体和信息元素含义的解释

注：“M”是“Mandatory”的缩写，表示必备属性；“0”是“Optional”的缩写，表示可选属性

6 数据采集方法

6.1 医院端推送数据采集解析

采集方法如下：

- 医疗机构根据数据采集接口标准将本机构数据解析为符合标准要求的数据集，主动推送至平台端质控区；
- 配合完成未通过质控数据集数据质量提升，直至数据质量达到平台端要求；
- 数据采集方对医疗机构上传的数据集进行质控，通过质控的数据集入库至平台端。

6.2 医院端镜像库采集解析

¹ 描述属性：描述信息实体和信息元素的属性

² 要求：描述该属性是必备属性还是可选属性

³ 定义及说明：对属性的定义和补充说明

采集方法如下：

- a) 医疗机构配备前置机并搭建前置机所需软件环境；
- b) 通过医疗机构推送或数据采集方抽取等方式将机构全量数据镜像库复制到前置机端；
- c) 向数据采集方提供本机构业务系统的库表结构、数据字典和镜像库访问权限；
- d) 数据采集方在镜像数据中解析获得符合标准要求的数据集，上传至平台端。

6.3 云端全量数据采集解析

采集方法如下：

- a) 医疗机构提供本机构全量数据镜像库和访问权限，提供本机构业务系统的库表结构和数据字典；
- b) 数据采集方将机构全量数据镜像库复制到云端；
- c) 数据采集方在云端镜像数据中解析获得符合标准要求的数据集，入库至平台端。

7 数据采集流程

7.1 主动推送数据

7.1.1 采集流程

采集流程见图1。

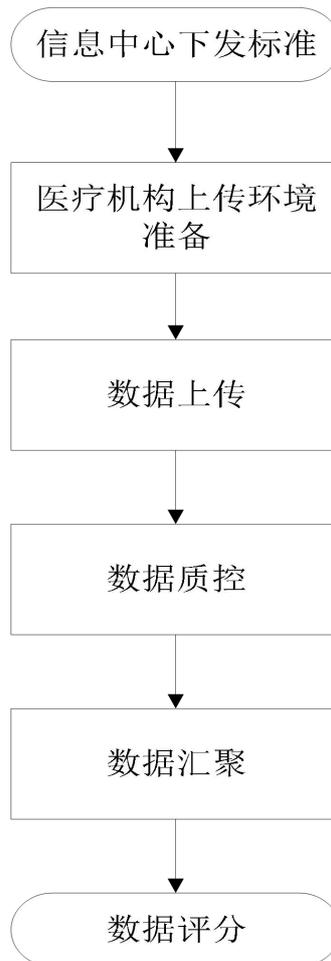


图1 主动推送数据采集流程图

7.1.2 信息中心标准下发

信息中心统一下发数据采集标准给各医疗机构。

7.1.3 医疗机构上传环境准备

医疗机构根据数据采集标准准备上传的相关环境。

7.1.4 数据上传

医疗机构根据数据采集标准上传对应数据集。

7.1.5 数据质控

市平台根据数据采集标准对医疗机构上传的数据进行校验。

7.1.6 数据汇聚

市平台将通过校验的数据和校验日志上传汇聚到市平台中心端。

7.1.7 数据评分

根据汇聚的数据和校验日志对各医院上传的数据进行评分。

7.2 自动采集

7.2.1 采集流程

采集流程见图2。

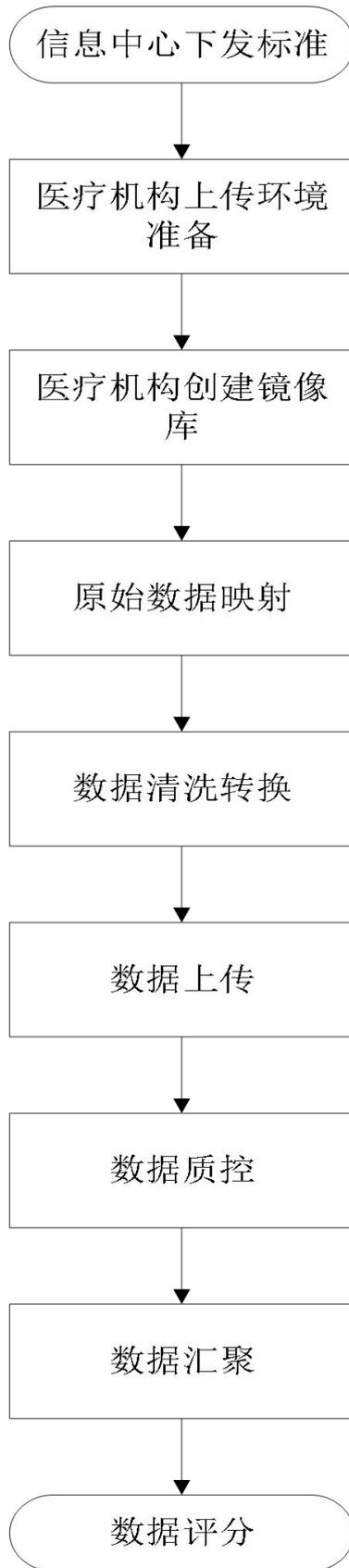


图2 自动采集数据采集流程图

7.2.2 信息中心标准下发

信息中心统一下发数据采集标准给各医疗机构。

7.2.3 医疗机构创建镜像库

医疗机构根据数据采集标准创建相关业务系统的镜像库。

7.2.4 原始数据映射

采集方从医疗机构镜像库中将相关业务系统的数据原样映射到清洗库。

7.2.5 数据清洗转换

采集方根据数据采集标准将医疗机构原始业务数据清洗转换为平台标准数据集。

7.2.6 数据上传

采集方将清洗转换好的数据集上传到平台质控库。

7.2.7 数据质控

市平台根据数据采集标准对医疗机构上传的数据进行校验。

7.2.8 数据汇聚

市平台将通过校验的数据和校验日志上传汇聚到市平台中心端。

7.2.9 数据评分

根据汇聚的数据和校验日志对各医院上传的数据进行评分。

8 数据质量控制

数据质量控制是将数据资源采集的基础上通过对数据清洗、对码、转换等过程整合、优化全民健康信息数据，保障机构数据的完整性、规范性、及时性、一致性和关联性，主要包括：

——数据清洗，对采集的数据进行处理，并与平台已有的数据进行比对，将格式不对、重复数据进行标识。

——数据关联，将区域内不同机构提供的同一居民医疗信息进行关联合并。

——错误数据处理，按照“谁提供，谁负责”的原则，将错误数据返回给各医疗卫生机构。

——质量控制规则，为保障医疗机构上传数据的质量，建立鼓励填，选填等概念，通过动态可扩展性，可适应一些特殊的数据质量管理需求。

——数据质量评估，数据质量评估主要包含对医疗机构上传到市平台数据的完整性、规范性、及时性、一致性和关联性五个维度，能够科学、高效提升中心端的数据质量，并支持扩展、可适应一些特殊的数据质量管理需求。对医疗机构上传数据进行综合评分，有利于医疗机构进行数据整改。

——考评模型，由于各业务线条关注的数据集不同，可建立各自的专项评分模型，同一批数据从多种维度进行评分，医疗机构可以根据专项评分进行专项整改提升数据质量。

9 数据采集安全

数据采集安全贯穿于数据采集整个过程中,应符合数据所属或主管部门的安全要求,包括但不限于:

- a) 数据在整个采集过程中应依据授权使用,不被非法冒充、窃取、篡改、抵赖;
- b) 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;
- c) 应能够对采集的数据进行定位溯源,对数据采集过程进行安全审计及监测;
- d) 遵循数据最小化原则,只采集满足业务所需的最少数据;
- e) 遵循确保安全原则,对采集的数据进行分类分级标识,并对不同类和级别的数据实施相应的安全管理策略和保障措施。对数据采集环境、设施和技术采取必要的安全管控措施;
- f) 应采用通过商用密码认证的产品,来保障采集过程中的安全性。