

DB4401

广 州 市 地 方 标 准

DB4401/T 304.1—2024

卫生健康数据采集规范 第1部分： 采集与交换

Data collection specification for health data
—Part 1: Collection and exchange

2024 - 12 - 20 发布

2025 - 01 - 20 实施

广州市市场监督管理局 发 布

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 总体要求 2

 4.1 通用要求 2

 4.2 质量要求 2

 4.3 安全要求 2

 4.4 数据范围 2

5 数据元属性描述规则 3

 5.1 数据集分类代码 3

 5.2 数据元名称、字段名和填报要求 3

 5.3 数据元标识符 3

 5.4 数据元标识类属性 4

6 数据采集 5

 6.1 采集方式 5

 6.2 采集要求 5

7 数据交换 5

 7.1 数据组成 6

 7.2 数据脱敏 6

 7.3 数据格式 6

8 采集交换验证 7

 8.1 API 方式 7

 8.2 数据库方式 7

 8.3 文件方式 7

附录 A（资料性） 数据采集交换范例 8

参考文献 11

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。

本文件是DB4401/T 304《卫生健康数据采集规范》第1部分。DB4401/T 304已发布以下部分：

- 第1部分：采集与交换；
- 第2部分：居民健康档案；
- 第3部分：卫生资源；
- 第4部分：妇幼保健；
- 第5部分：慢性病管理；
- 第6部分：卫生监督管理；
- 第7部分：公共卫生管理；
- 第8部分：医疗服务与运营管理。

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。

本文件由广州市卫生健康委员会提出并归口。

本文件起草单位：广州市卫生健康委员会、中国电信股份有限公司广州分公司、广东亿迅科技有限公司、广州市卫生健康技术鉴定和人才评价中心、广州市标准化研究院、广东省电信规划设计院有限公司、万达信息股份有限公司、中通服中睿科技有限公司、卫宁健康科技集团股份有限公司、东华医为科技有限公司。

本文件主要起草人：吴颖琳、刘玉转、余文家、刘林泉、戴卓骅、高红潮、陈燕燕、王一丰、陈维熙、陈勇岐、李佳杰、兰泽勇、丁疆、王博、穆斌、王贺珍。

卫生健康数据采集规范 第1部分：采集与交换

1 范围

本文件规定了数据采集与交换的术语和定义、总体要求、数据元属性描述规则、数据采集、数据交换和采集交换验证等内容。

本文件适用于卫生健康数据的采集与交换。

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 7408.1 日期和时间 信息交换表示法 第1部分：基本原则

GB/T 22239 信息安全技术 网络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 36344 信息技术 数据质量评价指标

WS/T 303—2023 卫生健康信息数据元标准化规则

WS/T 363.2 卫生健康信息数据元目录 第2部分：标识

3 术语和定义

下列术语和定义适用于本文件。

3.1

数据采集 data acquisition

从数据源中得到原始数据，通过标准化处理并转化为满足数据共享与利用需求的过程。

[来源：GB/T 36625.3—2021，3.2]

3.2

数据元 data element; DE

用一组属性描述其定义、标识、表示和允许值的数据单元。

[来源：GB/T 18391.1—2009，3.3.8]

3.3

基本数据集 base dataset

在特定主题下，由必需、基本的数据元组成的数据集。

3.4

数据元标识符 data element identifier

数据元在卫生业务信息基本数据集中的唯一标识代码。

3.5

数据交换 data exchange

各类实体之间相互提供数据的过程。

[来源：GB/T 36107—2018，3.1]

4 总体要求

4.1 通用要求

4.1.1 数据采集应记录并保留归集过程中历史数据的变化和移动情况，确保数据的可追溯性。数据采集过程中不应造成数据的缺失和遗漏，确保数据的完整性。应如实准确处理并记录数据，不应存在异常或错误数据，不应虚构或篡改数据。

4.1.2 应根据数据采集的结构、内容、关系和来源以及数据的质量和安全要求进行数据探查。综合考虑数据源网络环境、数据采集工具技术路线选型和现有数据采集通道建设情况，从而确定数据采用以下采集方式之一：

- a) 结构单一、数据量相对较小的结构化数据且数据实时性高要求的应用场景，宜采用 API 数据采集；
- b) 数据量相对较大且数据实时性中等要求的应用场景，宜采用前置机批量数据采集；
- c) 数据量相对较大且数据实时性低要求的应用场景，宜采用文件批量数据采集。

4.1.3 实施数据采集的原则包括：

- a) 在待采集数据的源系统业务不繁忙时间段进行，避免读取动作影响源系统正常运行；
- b) 对于数据量较大、单批量采集可能会造成系统故障的，应支持分批或增量读取，并采用分布式方式对数据源进行读取；
- c) 使用源系统的备份库作为采集对象时，应保证数据一致性和可用性。

4.2 质量要求

4.2.1 数据采集质量要求应符合 GB/T 36344 的要求，在数据采集周期内，应保证数据的规范性、完整性、准确性、一致性、时效性、可访问性和可追溯性。

4.2.2 应根据国家、省级及本级政府数据质量控制要求，选择适合的数据质量评价方法对数据质量进行定量评价，一般可分为全数检查和抽样检查。

4.2.3 针对国家强制要求、特殊要求和其他可能导致严重影响的数据质量项目进行全数检查。

4.3 安全要求

4.3.1 数据采集安全保护要求应符合 GB/T 35274 和 GB/T 22239 的规定，对采集数据进行分级分类管理，并根据数据安全措施和技术手段，对数据采集过程进行有针对性的保护，个人信息、敏感数据和重要数据应加强安全管控措施。

4.3.2 数据采集安全贯穿于数据采集整个过程，在数据采集周期内，应符合数据所属或主管部门的安全要求，按数据安全及安全控制要求，实现授权访问、可定位溯源、数据加密、安全审计及监测等。

4.3.3 对来源于不同网络环境的原始数据，应根据以下要求进行：

- a) 原始数据位于电子政务外网环境的，对数据直接采集；
- b) 原始数据位于业务专网环境的，建立与电子政务外网之间的安全传输通道，采取必要的安全措施保障数据传输安全性；
- c) 原始数据位于互联网环境的，建立与电子政务外网之间的安全传输通道，采取必要的安全措施保障数据传输安全性，对互联网环境数据的采集。

4.3.4 应对数据采集传输过程中的敏感数据进行监测。

4.3.5 应对公共数据进行加密传输，加密算法应符合国家密码相关法律、法规要求。

4.4 数据范围

数据范围包括居民健康档案、卫生资源、妇幼保健、慢性病管理、卫生监督管理和公共卫生管理和医疗服务与运营管理七个大类，其中：

- a) 居民健康档案包括健康档案和健康体检等数据；
- b) 卫生资源包括医疗机构、卫生人力、医疗基础设备、信息化建设和地理信息建设等数据；
- c) 妇幼保健包括婚前孕产产前检查、产前产时产后检查、出生缺陷登记、计划生育登记和儿童保健等数据；
- d) 慢性病管理包括高血压病病例管理、糖尿病病例管理、心脑血管病病例管理、肿瘤病病例管理、口腔病病例管理、麻风病病例管理、结核病病例管理、性病病例管理等数据；
- e) 卫生监督管理包括卫生监督组织、卫生监督管理相对人（机构类）、卫生监督管理相对人（个人类）、卫生监督管理相对人（设备类）、卫生监督管理相对人（产品类）、卫生行政许可、卫生监督检查、卫生行政处罚、卫生行政强制、卫生投诉处理、卫生监督协管、卫生监督非现场监管和卫生监督资源等数据；
- f) 公共卫生管理包括传染病管理、慢阻肺防治、精神病防治、免疫规划管理、学校卫生管理、疾病预防与控制、老年人管理、残疾人管理、家庭医生管理和健康教育等数据；
- g) 医疗服务与运营管理包括医疗服务基础信息、临床诊疗记录、电子病历、病案首页、检验检查、用血管理、家庭病床、手术麻醉、体检信息、转诊（院）记录、法定医学证明及报告、院感管理、血透质控、腹透管理、院前急救、重症医学和中医药等数据。

5 数据元属性描述规则

5.1 数据集分类代码

数据集分类代码由数据集一级分类代码、二级分类代码、三级分类代码组成，结构见图1。

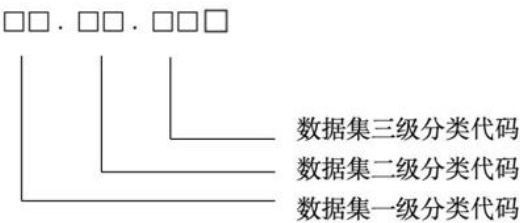


图 1 数据集分类代码机构

注：一级、二级代码都采用两位数字表示，代码为01-99，按顺序编码，三级代码采用三位数字表示，代码为001-999，按顺序编码用小数点（.）隔开。

5.2 数据元名称、字段名和填报要求

- 5.2.1 卫生健康数据元名称、定义按照WS/T 303—2023中的第6和7章的要求进行命名和定义。
- 5.2.2 卫生健康数据元字段名，按数据元名称汉字的拼音首字母进行编码。
- 5.2.3 卫生健康数据元填报要求分为必填和非必填，非必填可空。根据每一类数据集的数据项业务要求，判断数据项内容是否可空。

5.3 数据元标识符

卫生健康数据元标识符编码应符合WS/T 363.2的要求。

5.4 数据元标识类属性

5.4.1 数据元值的数据类型

数据元值的数据类型的描述规则见表1。

表 1 数据元值的数据类型描述规则

数据类型	表示符	描述
字符型(string)	S1	S1 表示不可枚举的，且以字符描述的形式通过字符形式表达的值的类型。可包含字母字符（a~z，A~Z）、数字字符等
字符型(string)	S2	S2 表示枚举型，且列举值不超过 3 个。可包含字母字符（a~z，A~Z）、数字字符等
字符型(string)	S3	S3 表示代码表的形式
布尔型(boolean)	L	又称逻辑型，采用 F（False）或 T（True）形式表示的逻辑值的类型
数值型(number)	N	通过“0”到“9”数字形式表示的值的类型
日期型(date)	D	采用 GB/T 7408.1 中规定的 YYYYMMDD 格式表示的值的类型
日期时间型(datetime)	DT	采用 GB/T 7408.1 中规定的 YYYYMMDDThhmmss 格式表示的值的类型（字符 T 作为时间的标志符，说明日的时间表示的开始）
时间型(time)	T	采用 GB/T 7408.1 中规定的 hhmmss 格式表示的值的类型
二进制(binary)	B	上述无法表示的其他数据类型，如图像、音频、视频等二进制流文件格式

5.4.2 数据元值的表示格式

数据元值的表示格式中，字符含义描述规则见表2所示，字符长度描述规则见表3所示。

表 2 数据元值字符含义描述规则

字符	含义
A	字母字符
N	数字字符
AN	字母或(和)数字字符
D8	采用 YYYYMMDD 的格式表示，其中，“YYYY”表示年份，“MM”表示月份，“DD”表示日期
T6	采用 hhmmss 的格式表示，其中“hh”表示小时，“mm”表示分钟，“ss”表示秒
DT15	采用 YYYYMMDDThhmmss 的格式表示，字符 T 作为时间的标志符，说明日的时间表示的开始；其余字符表示与上同

表 3 数据元值字符长度描述规则

类别	表示方法
固定长度	在数据类型表示符后直接给出字符长度的数目

表 3 数据元值字符长度描述规则（续）

可变长度	1)可变长度不超过定义的最大字符数 在数据类型表示符后加“..”后给出数据元最大字符数目。 2)可变长度在定义的最小和最大字符数之间 在数据类型表示符后给出最小字符长度数后加“..”后，再给出最大字符数。
有若干字符行表示的长度	按固定长度或可变长度的规定给出每行的字符长度数后加“×”后，再给出最大行数。
有小数位	按固定长度或可变长度的规定给出字符长度数后，在“,”后给出小数位数。字符长度数包含整数位数、小数点位数和小数位数

6 数据采集

6.1 采集方式

6.1.1 API 方式

以Restful数据接口服务作为数据资源进行汇集，采用JSON等格式进行数据传输。数据提供方调取业务应用系统或数据库中的数据，并封装提供数据接口服务。数据需求方通过数据接口调用获取数据。

6.1.2 数据库方式

以数据库表作为数据资源进行汇聚，通过在数据交换两端部署数据交换组件及交换库。数据提供方数据发生更新后，实时通过交换组件推送至前置机交换库，由两端交换组件协调双方交换库的同步。数据需求方通过交换组件从前置机交换库提取数据。

6.1.3 文件方式

以电子文件作为数据资源进行汇聚，常用的电子文件类型有wps、xml、txt、doc、docx、html、csv、xls、xlsx等。通过文件服务器的共享目录或FTP服务，实现共享文件数据组装、数据传输、数据解析和数据使用，达到数据采集的目的。

6.2 采集要求

6.2.1 API 方式

采集需传以下参数：X-APP-ID为账号参数，X-APP-KEY为密码参数，X-Request-Id为请求流水号。业务参数按各业务场景进行组装。

6.2.2 数据库方式

需获取数据库服务器实例、账号密码信息。账号按最小化权限原则进行授权。

6.2.3 文件方式

需获取文件服务器的地址端口、账号密码、路径等信息。账号按最小化权限原则进行授权。

7 数据交换

7.1 数据组成

数据组成包括元数据和数据说明。数据说明包括接口协议文档和数据库表、字段等注释信息。

7.2 数据脱敏

数据提供方在数据共享过程中应对数据脱敏，包括：

- a) 对敏感数据建立数据脱敏安全策略，并按照安全策略进行脱敏；
- b) 对数据脱敏操作过程进行记录，记录内容至少包括操作时间、操作人、操作对象等；
- c) 特殊情况下，可保留敏感数据的原数据格式、属性或关联。

7.3 数据格式

7.3.1 API 方式

交换方式包括通用交换和HL7 FHIR两种方式，范例参见附录A中A.1和A.2。

7.3.2 数据库方式

数据库表命名格式由区域(2位，区域的拼音缩写)、WJ+数据集分类(2位)和数据集分类编码(9位)三部分组成，其要求如下：

- a) 数据集分类编码参见表 A.1；
- b) 名称的各部分之间以下划线()连接；
- c) 数据库表中应含有公共字段，具体要求和含义见表 4。

示例：GZ_WJ01_01.01.004 表示广州卫健居民健康档案个人基本信息。

表 4 公共字段要求

属性名称	属性编码	数据类型	是否可空	属性备注
唯一标识	S_GUID	字符类型，长度38位	非空	交换数据记录的物理唯一标识
创建时间	S_CREATION_TIME	时间类型，精确到秒	非空	交换数据记录在当前节点交换库的物理创建时间
修改时间	S_LAST_UPDATED	时间类型，精确到秒	非空	交换数据记录在提供方交换库的最后修改时间
记录的状态	S_STATUS	字符类型，长度1位	非空	交换数据记录的状态 0：添加；1：修改；2：废置

7.3.3 文件方式

文件命名格式由区域(2位，区域的拼音缩写)、WJ+数据集分类(2位)、数据集分类编码(9位)、更新频率(1位)和日期五部分组成，其要求如下：

- a) 文件流在文件传送过程中以“~文件名.tmp”命名，完成传送后改回原文件名；
- b) 数据集分类编码参见表 A.1；
- c) 名称的各部分之间以下划线()连接；
- d) 更新频率：每小时生成的文件为 H 编码，每天生成文件为 D 编码，每月生成文件为 M 编码，每年生成的文件为 Y 编码；
- e) 日期：每小时生成的文件 yyyyMMddhh，每天生成文件 yyyyMMdd 格式，每月生成文件为 yyyyMM 格式，每年生成的文件为 yyyy 格式。

示例1: GZ_WJ01_01.01.004_D_20240828 表示2024年8月28日广州卫健居民健康档案个人基本信息交换文件。

示例2: GZ_WJ01_01.01.004_M_202408 表示2024年8月广州卫健居民健康档案个人基本信息交换文件。

8 采集交换验证

8.1 API 方式

通过返回编码code来判断是否采集或交换成功，失败原因存放在字段desc中，字段data用来存放交换过程中的业务信息，范例参见附录A.3。

8.2 数据库方式

通过日志来判断是否采集或交换成功，日志包括采集或交换开始时间、结束时间、成功和失败数据量、成功编码信息。

8.3 文件方式

通过回执文件来判断是否成功，回执文件包括采集或交换日期、成功和失败数据量信息。

附 录 A
(资料性)
数据采集交换范例

A.1 通用交换范例

以居民建档场景(包括健康档案的个人基本信息、体检信息)为范例:

POST /XXX/XXX HTTP/1.1

Host: localhost:9309

X-APP-ID: id

X-APP-KEY: key

X-Request-Id: 123456

Content-Type: application/json

Content-Length: 821

```
{
  "GRJBXXBSH": "个人基本信息标识号",
  "ZZJGDM": "组织机构代码",
  "JKDABH": "健康档案编号",
  "JGMC": "机构名称",
  "JDJGDH": "建档机构电话",
  "JDYSGH": "建档医生工号",
  "JDYSXM": "建档医生姓名",
  "JDRQ": "建档日期",
  "TJJL": [ //体检记录信息表
    {
      "TJJGBM": "体检机构编码",
      "YYBH": "预约编号",
      "SG_CM": "身高 (cm) ",
      "TZ_KG": "体重 (kg) ",
      "JDRQ": "为服务对象建立相关档案当日的公元纪年日期",
      "ZRYSM": "责任医生姓名"
    },
    {
      "TJJGBM": "体检机构编码",
      "YYBH": "预约编号",
      "SG_CM": "身高 (cm) ",
      "TZ_KG": "体重 (kg) ",
      "JDRQ": "为服务对象建立相关档案当日的公元纪年日期",
      "ZRYSM": "责任医生姓名"
    }
  ]
}
```

A.2 HL7 FHIR 交换范例

以患者到医疗机构进行就诊场景(包括门诊信息、检验报告、处方等信息)为范例:

```
POST /XXX/XXX HTTP/1.1
Host: localhost:9309
X-APP-ID: id
X-APP-KEY: key
X-Request-Id: 123456
Content-Type: application/json
Content-Length: 821
{
  "Organization": {
    "OrganizationID": "*****",
    "OrganizationName": "XXXX医院",
    "Patient": {
      "IdType": "证件类型",
      "IdNo": "证件号",
      "PatientName": "测试患者姓名",
      "Encounter": {
        "VisitID": "就诊号",
        "VisitType": "0",
        "VisitDate": "2023-07-01",
        "DepName": "骨科",
        "DoctorName": "宋医生",
        "Documentreference": [
          {
            "Title": "门诊记录单",
            "Doctype": "fhir",
            "mimeType": "json",
            "FilePath": ".document/fhirdoc/QHCG/111/123/100101_门诊记录
单.xml"
          },
          {
            "Title": "检验报告",
            "Doctype": "fhir",
            "mimeType": "json",
            "FilePath": ".document/fhirdoc/QHCG/111/123/1000102_检验报
告.xml"
          }
        ]
      }
    }
  }
}
```

```
        "FilePath": ".document/fhirdoc/QHCG/111/123/1000103_处方.xml"
    }
}
]
```

A.3 API 返回信息范例

```
{
    "code": X,          --采集成功失败编码 0 表示成功、-1 表示失败
    "data": XXX,        -- 返回的业务信息，例如主键、编码等
    "desc": XXX,        -- 采集失败原因
    "requestId": XXX    -- 请求流水号
}
```

A.4 数据集分类编码

数据集分类编码参见表 A.1。

表 A.1 数据集分类编码

编码	名称
01	居民健康档案
02	卫生资源
03	妇幼保健
04	慢性病管理
05	卫生监督管理
06	公共卫生管理
07	医疗服务与运营管理

参 考 文 献

- [1] GB/T 18391.1—2009 信息技术 元数据注册系统(MDR) 第1部分：框架
 - [2] GB/T 36107—2018 法人和其他组织统一社会信用代码数据交换接口
 - [3] GB/T 36625.3—2021 智慧城市 数据融合 第3部分：数据采集规范
-