SATOL

imit 浙江数字医疗卫生技术研究院
Institute of Medical-care Information Technology

树兰互联网医院 Shulan Health Online

2022世界生命科技大会 | WORLD SCIENCE & TECHNOLOGY OF LIFE CONFERENCE

数字医疗与互联网高峰论坛

探索 EXPLORATION　融合 FUSION　创新 INNOVATION

2022-12-17　中国·杭州·良渚

指导单位： 中华医学会、中国医师协会、浙江省科学技术协会

主办单位： 中国工程院、杭州市人民政府、浙江大学

承办单位： 浙江数字医疗卫生技术研究院、树兰互联网医院（杭州）有限公司

2022世界生命科技大会 WORLD SCIENCE & TECHNOLOGY OF LIFE CONFERENCE

数字医疗与互联网高峰论坛

SATOL

关于健康医疗数据质量管理与控制的一些思考

2022-12-17 中国·杭州·良渚

指导单位： 中华医学会、中国医师协会、浙江省科学技术协会

主办单位： 中国工程院、杭州市人民政府、浙江大学

承办单位： imit 浙江数字医疗卫生技术研究院 Institute of Medical-care Information Technology 树兰互联网医院 Shulan Health Online

前言

- 2022年9月2日，国家卫生健康委员会新闻发布会介绍了党的十八大以来我国全民健康信息化建设取得的显著成效，但也强调了由于医疗服务的复杂性、服务流程的多样性、服务对象的广泛性、医疗数据的敏感性，医疗健康信息的互通共享应该说仍然还是一个有待攻克的难题，我国目前仍处在探索发展、摸索经验的关键时期。而其中，务实管理和控制数据质量是提升医疗健康信息互联互通水平的重要内容。
- 关于数据治理

  数据确权

  数据安全

目录

- 第一部分：全民健康信息化建设与数据质量问题
- 第二部分：健康医疗数据质量管理的关键环节
- 第三部分：关于健康医疗数据质量控制的思考

第一部分：
全民健康信息化建设与数据质量问题

一、业务系统建设阶段

- 三要素：业务规则，业务流程，业务场景
- 数字化转型：从业务仿真到业务模式创新与业务模型重构
- 问题：业务系统的异构，导致数据异构，一旦参与交互，数据质量就会失控
- 难点：异构数据从数据结构到语义的对齐，不能低估语义对齐的难度与意义

第一部分：
全民健康信息化建设与数据质量问题

二、集成平台建设阶段

- 建设目的：解决异构业务系统的互联互通
- 点对点接口：N平方问题
- 集成代理：HUB与总线，EDI到ESB
- 互联互通：通易联难，
- 难点：语义对齐（术语与值域代码）
- 语言与语义：沟通交流，思维意识，知识发现，知识构建，知识表达……
- 语言的固有问题：语义的模糊性、主观性、领域性
- 本体论：概念、属性、关系、符号
- 基础支持：术语集、叙词表、知识图谱、语义网

第一部分：
全民健康信息化建设与数据质量问题

三、数据价值发现与数据应用阶段

- 业务系统里的数据与集成平台数据管理中心里的数据
- 平台提供一整套数据标准来约束异构数据在交互过程中数据格式及语义的映射转换，凡通过平台实现与其他系统进行数据交互和业务协同的业务系统，只需要与平台数据标准建立一对映射转换的机制即可，避免了 N 平方陷阱。
- 平台建设与管理运行质量，决定了平台数据交互质量，也决定了平台数据中心汇聚的数据质量及数据应用质量与应用价值。

第二部分：
健康医疗数据质量管理的关键环节

- 多源、异构、分散、复杂、海量
- 采集定位
- 数据源环境变化的感知
- 接口规范与映射转换
- 数据湖、数据仓库、数据集市
- 数据资产目录

第二部分：
健康医疗数据质量管理的关键环节

1，数据采集过程的数据质量控制

- 采集工具
- 采集定位
- 清洗转换：

标准转换 数据是符号 数据的含义是语义 语义是数据在某个领域的解释和逻辑表示 语义具有领域性特征 语义对齐是消除同一事物在不同领域理解不同的语义异构的必要过程

- 数据校验：

是为保证数据的完整性、合规性而进行的一种验证操作

- 数据稽核：

实现数据的完整性和一致性检查。从数据采集、预处理、比对、分析、预警、通知到问题修复的数据质量控制机制

第二部分：
健康医疗数据质量管理的关键环节

2，数据管理与应用过程中的质量控制

- 数据存储：

  数据湖、数据仓库、数据集市

- 数据标准管理：

  数据集标准、数据元标准、数据元值域代码标准、主索引、主数据、元数据

- 数据资产目录与元数据管理
- 统计分析指标的标准化——指标库

设计要点包含但不限于统一标识符、名称、定义、使用范围、统计口径、算法、计算引用数据的约束、例外声明等。

第二部分：
健康医疗数据质量管理的关键环节

3，集成平台控制数据质量的主要基础功能

- 主索引管理
- 主数据管理
- 元数据管理（以及基于元数据管理的数
- 据资产目录）
- 数据质量指标体系与测量方法
- 指标库管理
- 数据标准管理
- 注册服务
- 特别指出：平台缺乏有效的运行管理

第三部分：
关于健康医疗数据质量控制的思考

一、评估数据质量——数据质量指标体系

至少应该包括以下指标：

- 数据完整性：数据是否存在缺失记录或者缺失字段
- 数据及时性：数据的产生以及供给是否及时
- 数据有效性：数据是否满足用户定义的约束条件或者是否在一定的取值范围之内
- 数据准确性：数据是否与其对应的客观实体特征保持一致
- 数据一致性：同一实体同一属性的数据在不同的系统中是否一致

第三部分：
关于健康医疗数据质量控制的思考

二、数据质量的测量方法

三、数据质量报告

四、数据质量管理与控制体系架构设计要点

以数据资产目录为轴心，厘清并说明数据在平台生态环境中的各种关系与印迹，指引各种关于数据的操作。

表达数据资产目录下一个具体的数据元，至少要有以下内容：

数据资产目录ID，

数据元统一标识符（数据元标准，数据元值域代码标准），

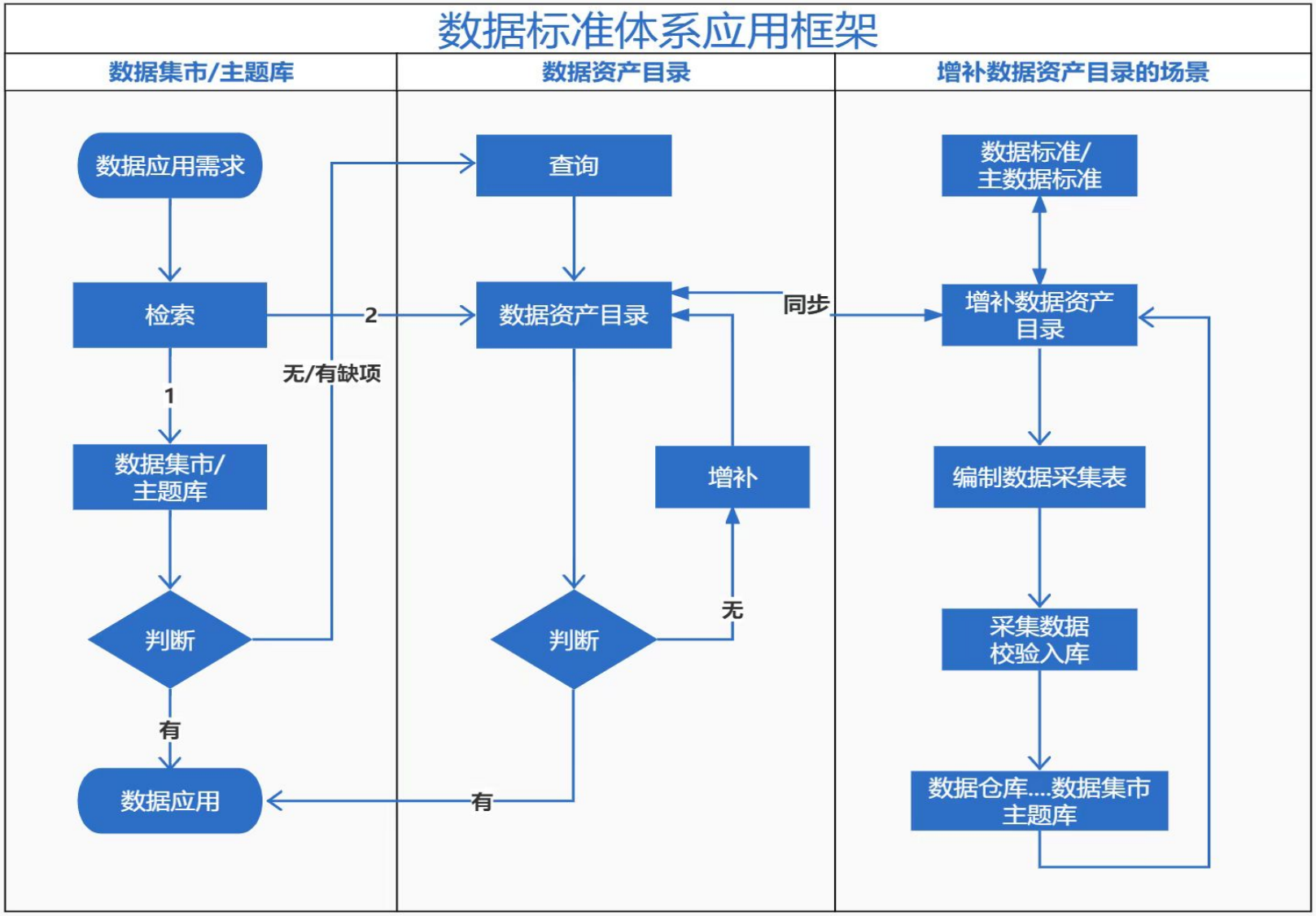
元数据标识符（元数据标准），

数据仓库物理存储位置，

查询及调用参数，

指标库、数据集市主题库数据集等引用记录，变更记录，生成记录等，并在实践中逐步完善。

四位一体应用场景


数据标准体系应用框架
数据集市/主题库
数据资产目录
增补数据资产目录的场景
数据应用需求
检索
1
数据集市/
主题库
判断
有
数据应用
2
无/有缺项
查询
数据资产目录
判断
有
增补
无
同步
数据标准/
主数据标准
增补数据资产
目录
编制数据采集表
采集数据
校验入库
数据仓库....数据集市
主题库

2022世界生命科技大会 | WORLD SCIENCE & TECHNOLOGY OF LIFE CONFERENCE

数字医疗与互联网高峰论坛

SATOL

谢谢！
欢迎交流与指导

2022-12-17　中国·杭州·良渚

指导单位： 中华医学会、中国医师协会、浙江省科学技术协会

主办单位： 中国工程院、杭州市人民政府、浙江大学

承办单位： imit 浙江数字医疗卫生技术研究院 Institute of Medical-care Information Technology　树兰互联网医院 Shulan Health Online