

一体化智能化公共卫生大数据平台构建策略与应用实践^{*}

江 涛¹ 单超群¹ 孙梦莎² 谢诚明¹ 叶旭辉² 蔡俨薇¹ 蒋 震¹

(¹ 浙江省疾病预防控制中心 杭州 310050 ² 阿里云计算有限公司 杭州 310030)

〔摘要〕 目的/意义 构建一体化智能化公共卫生大数据平台, 支撑传染病监测预警、公共卫生决策和疫情防控等工作。方法/过程 利用 OneID、标签体系、智能模型等关键技术, 构建一体化智能化公共卫生大数据平台, 并进行应用实践。结果/结论 一体化智能化公共卫生大数据平台有效支撑疾控数据跨部门归集治理, 提升了数据质量和共享效率, 为推动公共卫生数字化转型和加快疾控现代化建设提供支撑。

〔关键词〕 公共卫生大数据; 一体化; 智能化; 数据共享

〔中图分类号〕 R-058 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2025.05.005

Construction Strategies and Application Practices of an Integrated and Intelligent Public Health Big Data Platform

JIANG Tao¹, SHAN Chaoqun¹, SUN Mengsha², XIE Chengming¹, YE Xuhui², CAI Yanwei¹, JIANG Zhen¹

¹Zhejiang Province Center for Disease Control and Prevention, Hangzhou 310050, China; ²Alibaba Cloud Computing Co. Ltd., Hangzhou 310030, China

〔Abstract〕 **Purpose/Significance** To build an integrated and intelligent public health big data platform, so as to support infectious disease monitoring and early warning, public health decision-making, and epidemic prevention and control. **Method/Process** Key technologies such as OneID, tag systems, and intelligent models are utilized to build an integrated and intelligent public health big data platform, which is then applied in practice. **Result/Conclusion** The integrated and intelligent public health big data platform effectively supports the cross-departmental collection and governance of disease control data. Data quality and sharing efficiency are enhanced, providing support for promoting the digital transformation of public health and accelerating the modernization of disease control.

〔Keywords〕 public health big data; integration; intelligent; data sharing

1 引言

近年来我国政府高度重视公共卫生体系建设, 出

台了一系列政策和措施, 为公共卫生事业的发展提供了有力保障。数字化和智能化技术的应用, 也在重大传染病防控工作中发挥了重要作用^[1]。近年来, 国内公共卫生信息化建设取得了显著进展, 传染病疫情报告管理系统实现了传染病病例的实时报告与分析, 突发公共卫生事件管理系统提高了突发事件应对能力, 各地建立的省级和市级公共卫生信息平台也提高了数据的集中管理和利用效率^[2]。2023 年 1 月国家疾病预防控制中心综合司印发《加快建设完善省统筹区域

〔修回日期〕 2025-05-06

〔作者简介〕 江涛, 高级工程师, 发表论文 10 余篇; 通信作者: 蒋震。

〔基金项目〕 浙江省疾病预防控制中心科技计划项目 (项目编号: 2025JK007)。

传染病监测预警与应急指挥信息平台实施方案》，提出开展省级疾控云、大数据协同等基础设施建设，整合现有传染病监测和应急相关信息系统，建立“全国一网统管、平台两级建设，数据统一采集、业务分级应用”的一体化省统筹区域平台。当前，公共卫生信息化面临数据分散、多源异构、标准不统一等挑战，亟待通过技术创新实现跨部门、跨区域数据共享^[3]，建成集数据采集、治理、共享为一体的智能化数据平台。本研究基于浙江省实践，探讨一体化智能化公共卫生大数据平台的构建策略，重点解决数据归集治理、跨域共享等关键问题，支撑传染病监测预警、公共卫生决策和疫情防控等工作。

2 平台构建框架

2.1 总体架构设计

公共卫生大数据平台应兼具数据采集、存储、

计算、共享、安全管控等一体化集成能力，同时可为公共卫生业务活动及其应用系统提供数据治理、质量评价、模型搭建等智能化应用支撑能力。因此在设计平台总体架构时，采用分层架构多级应用的模式，自下而上，分为基础设施层、数据层、交换层和应用层，见图 1。基础设施层采用云计算平台，提供应用部署环境、数据存储环境和基础网络服务。数据层通过多部门数据采集，建立数据仓库，采用大数据离线计算、大数据实时计算和大数据流式计算工具，支撑数据归集和治理，并为应用层提供多样化的数据来源。交换层统一管理数据交换接口和交互规范，打通与公共卫生相关信息平台的数据通道。应用层基于数据层向各类业务应用提供数据交互服务，实现一数一源，一处采集，多方利用，减少数据冗余，提供统一数据服务，保证数据的一致性、准确性和完整性。

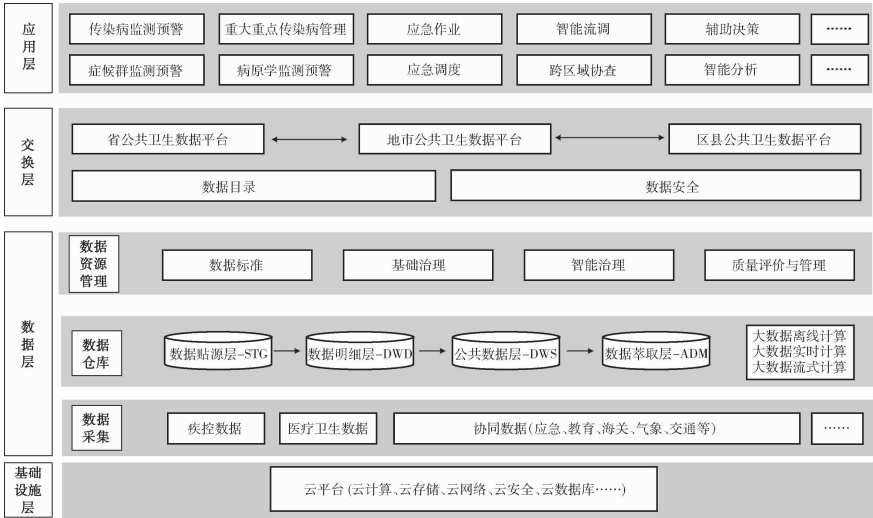


图 1 一体化智能化公共卫生大数据平台总体架构

2.2 功能模块

疾控部门亟需以数据为驱动加快转型发展^[4]，一体化智能化公共卫生大数据平台应具备广泛采集数据、高效治理数据、安全管理数据、便捷共享数据的能力，因此以下功能模块成为平台建设中的关键要素。

2.2.1 数据采集 数据采集来源涉及疾控、卫生

健康、教育、环保、气象、网信等多个部门，数据格式和采集方法不尽相同，因此数据采集模块应具备数据库对接、数据交互接口对接和离线数据导入等多种功能，同时适配多种数据格式。

2.2.2 数据仓库搭建 数据仓库是平台的核心组成^[5]，用于完成数据对接、存储、治理与服务，可进一步细分为贴源层、明细层、公共数据层、萃取层，通过逐层治理分析，最终形成高可用的标准化数

据和业务专题库。数据贴源层：完成各类公共卫生数据的汇聚与存储（与原始数据保持一致）。数据明细层：合并贴源层数据，并完成数据主索引和数据更新。公共数据层：按业务维度分解形成标签集，对明细层数据进行标签化处理。数据萃取层：按业务标签提取公共数据层数据，形成多维度的业务专题数据库，如应急资源专题库、流感监测专题库等。

2.2.3 数据质量管理 为提升数据价值，提高数据挖掘效率，平台应结合具体应用场景，针对核心数据开展质量评价与管理^[6-7]。根据实际业务需求及数据监管要求，设置不同规则、权重的数据质量评分体系，开展数据质量评分分析。评分规则可按日、周、月等配置周期自动执行，并出具评分报告。

2.2.4 数据资源管理 数据资源管理主要用于规范数据资源的登记和流通，促进数据共享利用，包括数据目录、元数据、数据字典、业务数据集等。其核心功能是数据编目管理，支持按统一的规范和流程完成归集数据编目。

2.2.5 数据交换共享 数据交换共享为跨部门、跨区域的监测预警、辅助决策、应急响应、资源配置等工作提供全面的数据支持。平台向外提供数据赋能的服务接口，以及标准统一的数据共享实施路径^[8]。面向不同服务类型、数据类型，提供应用程序编程接口（application programming interface, API）、数据库表对接等多种服务形式。

2.2.6 数据分析 数据分析模块通过整合人工智能技术驱动的辅助工具，集成统计分析、自动化建模及可视化组件等核心模块，提供数据预处理、模型构建和分析结果呈现的全流程智能化支持，提升数据分析效率，实现高效精准的数据价值挖掘。

3 平台构建关键技术

3.1 OneID 体系

基于人群、疾病、环境因素等核心维度，通过算法模型构建主索引体系，为同一个体或实体分配唯一标识符（如患者 ID、疾病名称、值域字典等），主要构建流程如下。一是定义标识规则，明确生成唯一标识符的数据字段（如身份证号、手机号、姓名等）；

二是开发标识符生成算法，确保唯一性和稳定性；三是实施数据清洗与标准化流程，消除重复及错误数据，统一数据格式；四是采用模糊匹配算法识别并合并相同个体的多源记录；五是应用机器学习模型提高数据匹配精度；六是创建主索引表作为体系枢纽，存储唯一标识符和原始数据映射关系；七是设置定期更新机制，确保新增数据能够及时纳入主索引；八是实施数据质量监控机制，持续修正数据错误。

3.2 精细化治理算法

在传统数据治理技术基础上，引入人工智能技术，提升数据处理效率与准确性。一是医疗文本标准化，基于自然语言处理技术，实现诊断、检验项目、检查项目、药品、手术操作等短文本数据的标准化处理，提取主诉、检查结论等长文本数据的结构化特征；二是地址标准化，基于全国标准地址库，通过字符串处理、语义识别与相似度计算技术，实现行政区划、家庭住址、单位地址、学校名称、医疗机构名称等地址实体的标准化映射。

3.3 标签管理

数据标签是提升数据管理和分析效率的重要手段之一，构建智能标签核心管理功能包括如下步骤。一是建立符合业务数据需要的逻辑模型；二是提供标签数据管理、加工与查询的可视化交互界面；三是开放业务对象的标签查询，授权公共数据标签的查询权限；四是支持类 SQL 实体关系关联查询；五是实现查询结果一键封装数据服务 API。

3.4 智能模型

通过模型和算法的构建和应用，建立多维传染病监测预警技术。一是开发特征提取算法，快速提取监测数据中的特征信息，并转化为预警信号；二是建立时空多维分析模型，从时间、空间等多维度对传染病暴发的潜在风险进行预警^[9]；三是设置分级预警规则，按照病种差异配置预警时效^[10]；四是搭建模型工具引擎，支持算子、数据、流程等关键要素配置；五是构建疫情处置知识图谱，辅助突发公共卫生事件规范处置；六是探索传染病暴发预测

模型, 基于疾病特征、地域分布、人群特征等多维度参数, 实现疫情态势感知。

4 平台应用成效

省级一体化智能化公共卫生大数据平台的建设在数据归集、数据治理、数据共享、数据利用等方面均取得了良好效果, 同时与省级传染病监测预警与应急指挥平台联动, 推动公共卫生业务高效运行。

4.1 数据归集

通过交换接口或数据集成导入方式实现实时或非实时归集传染病、慢性病、免疫规划等疾控业务相关数据集 712 个, 涉及 41 644 个字段, 约 11.7 亿条数据; 获取教育、环保、气象、网信等部门数据集 16 个, 共计 1.4 亿余条数据; 建立卫生健康数据共享数据库, 数据共计 12.9 亿条。

4.2 数据治理

通过持续推进数据治理, 已构建 8 365 万个 OneID 实体标识。数据标签体系建设取得了阶段性成果, 累计完成 114 个业务标签开发。其中: 症候群类标签 44 个, 覆盖数据量 108 326 021 条; 疾病种类标签 8 个, 涉及数据量 29 149 272 条。

4.3 数据共享

平台提供 24 个数据交互接口和 30 个业务专题库, 实现不同业务之间的数据共享服务, 同时根据数据安全管理和分类分级要求, 对敏感数据字段实施脱敏处理, 确保数据应用安全。截至 2025 年 2 月, 数据被调用 3 656 563 次。数据共享服务主要应用场景如下。一是传染病监测预警 (占比 53%): 通过调取传染病报告数据、中西医电子病历、实验室检测数据及影像学检查结果, 帮助公共卫生部门实时了解传染病的发生情况, 迅速识别潜在疫情。通过整合传染病报告和学生因病缺课监测数据, 快速、精准评估学校聚集性疫情风险。借助多源数据融合分析, 动态追踪传染病发病趋势, 自动识别聚集性疫情, 及时触发预警机制, 实现疫情早期识别

和快速响应, 有效降低疾病传播风险。二是重点传染病管理 (占比 32%): 基于门急诊电子病历和电子处方数据, 建立传染病症状监测系统, 实时监测分析病原体的传播特征, 提升传染病早期监测与预警能力, 通过动态监测治疗效果和病程进展, 实现重点传染病患者从发病至结局的全流程追踪管理, 为制定精准防控策略提供依据, 提升传染病管理的科学决策能力。三是应急处置 (占比 15%): 依托临床诊疗和实验室检测数据, 构建疫情风险评估模型, 精准评估疫情的风险与发展趋势, 及时调整防控策略。通过调取门急诊电子病历和传染病报告数据, 建立便捷的智能流调系统。通过快速生成传播链, 精准定位传染源与传播路径, 支持精准的疫情溯源和高风险人群筛查。通过多维数据的自动化流转和分析, 显著提升应急响应效率, 以信息共享与协同流程支持跨部门协同作战。

4.4 数据利用

一体化智能化公共卫生大数据平台有助于赋能科研和业务数据挖掘。通过智能搜索引擎提供高效的数据查找功能, 帮助用户基于模糊、碎片化的条件, 在海量数据中快速精准定位个体医疗健康信息; 可视化引擎配备丰富的数据大屏模板与多元化组件库, 可快速构建数据看板和驾驶舱, 缩短建设周期并降低维护成本; 智能科研引擎集成多种科研工具组件, 支持数据清洗、加工分析、数据建模、结果应用等科学研究全流程管理; 空间分析引擎内置多种时空分析模型, 支持传染病数据时空轨迹解析与风险热区定位。智能化分析引擎体系已实现与平台数据仓库的无缝对接, 实时解析多渠道监测业务数据, 支撑传染病监测预警与应急指挥决策的快速响应。

5 结语

本研究提出了公共卫生大数据平台的总体架构和关键功能, 通过 OneID、核心指标体系及数据标签系统构建数据治理体系, 提炼公共卫生领域数据归集治理的核心技术。基于一体化智能化公共卫生大数据平台开发实践, 实现了海量多源数据的标准化归集与共

享应用，形成了支撑传染病监测预警、重大传染病管控与慢性病动态监测的数字化底座，成为推动公共卫生数字化转型和加快疾控现代化建设的重要基础设施。后续研究将重点围绕以下方向：一是深化数据治理与标准化体系建设，提升数据质量与管理效能；二是完善跨部门、跨区域数据共享与协作机制，构建省域高效联动的公共卫生数据服务生态^[11]；三是引入大模型驱动的智能分析算法，增强传染病预测预警模型的精度；四是制订公共卫生数据分类分级标准，健全信息安全隐私保护机制。

作者贡献：江涛负责论文撰写与修订；单超群、孙梦莎负责文献检索与分析；谢诚明、叶旭辉负责平台应用成效分析；蔡俨薇负责论文修订；蒋震负责研究设计、论文审核。

利益声明：所有作者均声明不存在利益冲突。

参考文献

1 MORGAN O W, ABDELMALIK P, PEREZ - GUTIERREZ E, et al. How better pandemic and epidemic intelligence will prepare the world for future threats [J]. *Nature medicine*, 2022, 28 (8): 1523 - 1526.

2 邹晓辉, 朱闻斐, 杨磊, 等. 谷歌流感预测——大数据

在公共卫生领域的尝试 [J]. *中华预防医学杂志*, 2015, 49 (6): 581 - 584.

3 王玉琢, 马红霞, 靳光付, 等. 大数据时代的流行病学研究: 机遇、挑战与展望 [J]. *中华流行病学杂志*, 2021, 42 (1): 10 - 14.

4 赵嘉, 苏雪梅, 赵自雄, 等. 新形势下疾病预防控制信息化需求及应用场景分析 [J]. *医学信息学杂志*, 2024, 45 (2): 65 - 70.

5 查文中, 兰希, 比确子拉. 医院信息系统中的跨平台数据集成技术研究 [J]. *信息与电脑 (理论版)*, 2024, 36 (18): 66 - 68.

6 姚成, 胡伊玢. 数据挖掘技术在医学信息中的应用 [J]. *国际医药卫生导报*, 2008, 14 (7): 122 - 124.

7 张晨, 计虹, 金昌晓. 医疗数据质量分析与治理对策探讨 [J]. *中华医院管理杂志*, 2020, 36 (9): 747 - 750.

8 刘阳, 李亚子, 郭珉江, 等. 基于“惠民保”的医保数据共享模式研究 [J]. *中华医院管理杂志*, 2022, 38 (12): 901 - 905.

9 马宇航, 殷一, 王凯, 等. 传染病多阶段、多场景监测预警多样化方法体系的研究与思考 [J]. *中华预防医学杂志*, 2023, 57 (10): 1529 - 1535.

10 赖圣杰, 冯录召, 冷志伟, 等. 传染病暴发早期预警模型和预警系统概述与展望 [J]. *中华流行病学杂志*, 2021, 42 (8): 1330 - 1335.

11 徐艳晴, 姚洪. 供需视角下重大突发公共卫生事件数字治理研究 [J]. *中国行政管理*, 2023, 39 (10): 131 - 139.

(上接第 23 页)

7 KOCH R T, ERAZO D, FOLLY A J, et al. Genomic epidemiology of West Nile virus in Europe [EB/OL]. [2024 - 11 - 09]. <https://doi.org/10.1016/j.onehlt.2023.100664>.

8 周李承, 文哲, 罗伟权, 等. 建立全球传染病疫情监测、预警及应对的 3M - SPR 体系 防范和应对国门生物安全风险研究 [J]. *口岸卫生控制*, 2021, 26 (5): 27 - 30.

9 YUE Z J, SHI X Y, ZHANG H L, et al. The viral trends and genotype diversity of norovirus in the wastewater of Shenzhen, China [EB/OL]. [2024 - 11 - 09]. <https://doi.org/10.1016/j.scitotenv.2024.174884>.

10 焦敏, 高倩倩, 刘金妹, 等. 新发 (重大) 传染病“全哨点”概念界定 [J]. *中国公共卫生*, 2021, 37 (10): 1459 - 1462.

11 JIANG Y Y, YUAN Z Y, WANG Y X, et al. Wastewater - based intestinal protozoa monitoring in Shanghai, China [J]. *Microbiology spectrum*, 2024, 12 (11): e0403223.

12 詹启敏, 张路霞, 韩鸿宾, 等. 健康数据科学导论 [M]. 北京: 北京大学医学出版社, 2022.

13 柳东如. 湖北省统筹区域传染病监测预警体系建设实践与思考 [J]. *公共卫生与预防医学*, 2023, 34 (6): 1 - 4.

14 王宇. 新发传染病风险的智能化监测与评估 [J]. *中国数字医学*, 2023, 18 (10): 6 - 10.

15 袁广林. 人工智能时代高等教育变革 [J]. *国家教育行政学院学报*, 2019 (8): 11 - 17.

16 张诚, 夏天, 毛丹, 等. 基于医疗健康大数据的重大传染病监测预警标准体系构建设想 [J]. *预防医学情报杂志*, 2024, 40 (4): 430 - 434.

17 杜明梅, 刘运喜. 我国传染病监测预警系统的发展与应用 [J]. *中华医院感染学杂志*, 2022, 32 (6): 801 - 804.

18 程潇, 黄利群, 陈丹, 等. 我国口岸城市呼吸道传染病多源数据预警综合指标体系建立 [J]. *中国公共卫生*, 2022, 38 (9): 1160 - 1163.