2025年“省统筹传染病监测预警平台数据治理与管理服务项目”项目需求

一、**项目名称、数量和金额**

项目名称：省统筹传染病监测预警平台数据治理与管理服务项目

金额：144万元

二、**用途概述**

针对省统筹传染病监测预警平台开展数据治理与管理服务。

**三、型号规格、技术要求**

**（一）服务内容**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 采购内容 | 数量 | 政府采购编号 | 预算（元） | 单一来源论证 | 进口论证 |
| 省统筹传染病监测预警平台数据治理与管理服务 | 1项 |  | 1440000 | 否 | 否 |

**1.项目背景**

项目旨在针对省统筹传染病监测预警平台开展数据治理与管理服务，主要包括四个方面：

（1）支撑省统筹平台与国家平台对接，做好EMR、EDR数据治理，涵盖平台对接的持续优化和评估（接口、数据和安全），EMR数据整合汇聚、清洗、脱敏、标化等全周期管理，EDR数据实时采集与传输优化、数据关联融合、监测指标体系完善、数据质量监控与评估。

（2）人工智能语料治理，建立传染病语料全生命周期持续更新优化技术支撑体系，涵盖语料采集与筛选、标注分类、清洗预处理、扩充增强以及质量评估更新等核心内容。

（3）疾控综管平台数据接入和治理服务，从不同业务系统中采集数据，确保数据能够顺利进入综管平台，同时确保数据的质量、一致性和安全性。

（4）综合业务监管和监测预警指标模型信息服务，通过对疾控中心各业务系统的数据进行整合与分析，实现对业务运行状态的实时监控和评估，通过数据建模和算法分析，构建预警指标体系。

**2.具体服务内容**

（1）个体疾病档案数据治理服务

##### 多数据源汇聚整合

###### 数据源识别与接⼊

数据整合：将传染病监测数据、疾控业务系统、电子病历（EMR）、疫苗接种记录等结构化与非结构化数据进行汇聚、清洗和标准化处理。

数据源扩展性：系统支持灵活接入新数据源（如新增医疗信息化系统），适应业务发展需求。

###### 数据采集方式

实时采集：实现实时数据采集获取个人诊疗过程动态数据，确保信息及时更新。

批量采集：支持批量数据采集，实现历史病历档案、检验报告等非实时数据，优化采集效率。

###### 数据格式处理

异构数据转换：支持不同格式数据统一标准化，制定转换规则确保准确性。

非结构化数据结构化：支持非结构化数据转换，实现提取病历文书中的关键信息，便于后续分析。

##### 数据治理

目标是将接收到的多源数据按照EDR数据形式，进行实时清洗、治理和数据关联融合，包括：

###### 数据清洗

完整性检查：识别缺失值并补充或标记，确保数据完整可用。

准确性校验：基于医学规则验证数据合理性。

一致性保障：消除患者信息冲突，通过主数据管理统一身份标识。

敏感数据处理：识别隐私信息并脱敏，保障共享安全。

###### 数据标准化

数据模型标准化：支持统一档案模型，实现定义患者主索引、诊断记录等核心实体属性。

数据元标准化：支持行业标准规范字段名称和格式。

主数据管理：支持统一患者、机构等基础数据标识，实现跨系统互联。

###### 数据质量监控与评估

监控机制：支持对数据采集、传输、存储、处理等环节进行监控，实现及时发现数据质量问题。

定期评估与报告：支持定期对数据质量进行全面评估，生成评估报告，分析问题原因并提出改进建议，推动数据质量持续提升。

##### 平台对接

###### 接口开发与优化

国家平台对接接口：支持按国家数据标准上传接口和下载接口，确保数据格式与平台兼容。

其他平台对接接口：支持根据不同平台需求接口，实现通过患者主索引调取所需健康数据，实现跨平台共享。

接口性能优化：支持保障大容量数据交互时的可靠性。

###### 数据交互安全

数据加密传输：支持数据加密，实现接口传输内容加密处理，防止信息泄露或被篡改。

身份认证与授权：支持限制接口调用权限，实现根据角色分配数据访问范围。

数据审计与追溯：实现记录接口调用日志，便于安全事件追溯与责任判定。

###### 接口评估与持续改进

接口稳定性评估：统计接口故障频率与恢复效率，分析稳定性趋势。

数据准确性验证：定期比对上传数据与原始数据，确保准确性。

持续改进机制：根据评估结果和业务需求，修复故障接口并优化性能。

##### 数据整合汇聚

###### 数据整合与存储

数据整合规则：支持按患者主索引关联整合不同 EMR 模块和科室的异构数据，形成以患者为中心的电子病历视图，实现解决数据分散、重复问题。

数据存储方案：支持海量历史病历存储，频繁访问的现症患者数据情况。

数据冗余与⼀致性管理：适当保留数据冗余提高查询分析效率，建立机制确保冗余数据更新时一致。

###### 数据融合与应用

数据融合算法应用：运用数据融合算法处理多源数据，解决冲突互补问题，提升数据质量价值。

临床知识图谱构建：基于关联融合后的数据，构建临床知识图谱，以个人主索引为核心节点，将疾病、症状、诊断、治疗、药物、疫苗接种史等实体进行关联，形成结构化知识网络，为数据分析和检索提供知识基础。

可视化展示界面：提供可视化展示界面，以图表、仪表盘等形式直观呈现指标数据。

（2）人工智能语料治理建设服务

对标疾控高质量数据集建设任务，为了更好的训练传染病模型，建立传染病相关语料全生命周期持续更新优化技术支撑体系，涵盖语料采集与筛选、标注分类、清洗预处理、扩充增强以及质量评估更新等核心内容。

##### 数据采集

支持从医院的HIS、电子病历系统（EMR）、实验室信息系统（LIS）、区域卫生信息平台、疫苗接种记录、随访信息等系统中采集数据。确保数据的全面性和准确性。同时，收集检验报告的原始数据文件和医嘱信息。 从疾病预防控制中心获取传染病疫情监测数据。

##### 数据治理

包括数据完整性检查：检查数据是否存在缺失值、空值等问题，对缺失数据进行补充或标记，以便重新进行采集，保障数据完整性。数据准确性校验：基于医学知识和业务规则，校验数据准确性，对异常数据进行纠正或标注。数据⼀致性保障：消除数据冲突和不⼀致情况。敏感数据识别：精准识别患者姓名、身份证号、联系方式等敏感信息，以及病历中的隐私数据如性病、精神疾病等诊断信息。根据数据使用场景和隐私要求，选择合适的脱敏算法，对敏感数据进行脱敏处理，确保数据在共享和分析过程中的隐私安全性。

##### 标注分类

实体标注：对文本数据中的传染病相关实体进行标注。标注时需遵循统⼀的标注规范，确保标注的准确性和⼀致性。

语义关系标注：标注症状与疾病之间的关联关系。通过标注语义关系，为后续的知识图谱构建和数据分析提供基础。

数据分类：按照传染病的类型，分为病毒性传染病、细菌性传染病、真菌性传染病等。依据传播途径，分为飞沫传播传染病、接触传播传染病、空气传播传染病、血液传播传染病等。还可根据疾病的严重程度，分为普通传染病、重大传染病、烈性传染病等。通过多维度的分类，方便用户根据不同需求快速检索和分析数据。

##### 数据挖掘

聚焦于传染病相关数据的深度挖掘与结构化处理，以满足高质量数据集建设需求。通过重点病例数据的采集与筛选，结合诊疗数据的后结构化处理，将对主诉、现病史等关键字段进行精细化解析，提取与传染病诊断、治理和防控高度相关的特征信息。数据治理将涵盖从原始数据的清洗与预处理到标注分类的全流程优化，确保数据的准确性、一致性和可用性。同时，通过数据扩充与增强技术，构建动态更新的语料库，为传染病模型的训练提供持续的技术支撑。

##### 数据存储与安全

数据存储：搭建存储架构，将结构化、半结构化和非结构化数据进行统⼀存储。

访问控制：建立完善的用户角色和权限管理体系。根据用户的职责和需求，定义不同的用户角色。为每个角色分配相应的权限，包括数据查询权限、数据修改权限、数据删除权限、数据导⼊导出权限等。通过基于角色的访问控制机制，确保只有授权用户能够访问和操作相应的数据，保护语料库的数据安全。

审计与监控：对用户的数据访问行为进行记录和审计。记录每个用户对数据的查询、修改、删除等操作。通过审计日志，可以追溯数据的使用情况，发现潜在的数据安全风险。同时，设置数据监控机制，实时监测数据的访问流量、存储容量、数据⼀致性等指标，当出现异常情况（如数据访问量突然激增、数据丢失或损坏等）时，及时发出警报通知系统管理员进行处理，保障语料库的稳定运行和数据安全。

##### 质量评估与优化

建立评估指标体系：从数据准确性、完整性、标注⼀致性、查询效率等方面建立评估指标体系。数据准确性通过对比标注数据与权威标准数据，计算准确率和召回率来评估；完整性检查数据集中是否存在关键信息缺失的情况；标注⼀致性通过计算不同标注人员对同⼀数据标注结果的⼀致性程度来衡量；查询效率通过模拟实际应用中的查询场景，测试查询响应时间和结果准确性来评估。

评估与优化：定期对语料库进行评估，根据评估结果发现问题并及时进行优化。对于数据准确性问题，加强数据审核和标注质量控制；对于完整性问题，完善数据收集流程和缺失值处理方法；对于标注⼀致性问题，加强对标注人员的培训和管理，优化标注规范；对于查询效率问题，优化数据库索引、调整存储架构或采用缓存技术等。通过持续的评估和优化，不断提高语料库的质量和实用性。

（3）疾控中心综管平台数据中心接入和治理服务

##### 系统对接服务内容

###### 对接系统梳理

系统清单管理：识别并分类需对接的业务系统，建立系统对接台账。

接口规范更新：根据接口标准，更新接口文档模板，统一字段命名规则、数据格式、时间戳精度。

###### 调用方式提供

实时交互：支持数据实时交互实现高频数据的秒级同步。

批量处理：支持批量数据处理实现传输慢性病随访记录等大容量数据。

多源数据接入实施异构协议兼容：支持数据库直连、文件解析、消息队列等多种接入模式。

##### 核心治理服务内容

###### 数据清洗服务

缺失值处理：规则核查并补全关键字段，采用业务规则补缺。

重复数据剔除：按主键合并重复记录，保留最新版本并标注历史变更轨迹。

异常值修正：设定阈值规则，规则复核后修正或剔除。

###### 数据标准化服务

格式统一：日期、行政区划、性别标准化转换。

单位规范：统一血压单位、病毒载量单位等关键计量单位。

###### 数据质量管理服务

质控规则制定：定义完整性、一致性、及时性等6类质量指标。

质量监控执行：按月生成质量报告。

###### 指标标准化服务

统一指标库:提供核心指标元数据服务,确保不同的系统指标定义一致.

统一计算逻辑:提供统一的指标计算口径,确保不同系统的指标计算方式一致。

（4）综合业务监管和监测预警指标模型信息服务

##### 业务数据监管

###### 数据规范性与标准化监管

元数据统一治理

字段级规范定义：建立业务数据字典，强制约束9大核心业务系统的字段命名、编码规则、单位制式，消除“一数多义”。

数据血缘可视化

支持动态血缘图谱追踪数据流转路径。

动态质量稽核

实时合规校验：建立多种类校验规则，对非法字符、超范围值、逻辑矛盾实施三级拦截。

###### 数据采集全面性监管

多维度覆盖度评估

时空完整性审计：

空间维度：实现行政单元网格化核验，按省-市-县-社区四级划分监管网格，每日比对相邻区域数据量差异。

时间维度：校验时间序列连续性。

数据缺口预防

历史数据趋势建模

周期性波动分析：统计过去各业务数据的月度/季度采集规律，识别固定时段的采集低谷。

关联性缺口预测

跨业务波动监测：建立业务数据关联矩阵，当监测到关联业务异常波动时，自动检查目标业务数据完整性。

采集失败根因分析：统计各类采集失败原因占比，针对高频问题制定预防清单。

##### 指标定义监管

###### 指标分类

构建三类指标：实时指标，批量指标，复合指标。

###### 动态更新与验证

支持灰度发布机制以及异常熔断回退机制。

##### 预警指标体系构建

###### 阈值设定

历史数据驱动：基于3年历史数据滚动计算波动区间，例如传染病周报数超过均值2倍标准差时触发预警。

业务规则库：内置专家经验规则，如“慢性病药品库存低于7天用量”自动列为红色风险。

###### 权重动态优化

统计辅助：通过分析历史预警效果，调整指标权重。例如，误报率高的指标权重逐步下调，精准预警的指标权重提升。

分级预警机制：按严重程度分为红、橙、黄三级，红色预警需12小时内响应，黄色预警纳入周期性分析报告。

##### 监测与预警

###### 监控体系

全链路追踪：从数据采集、指标计算到预警触发，全程监控。

###### 多渠道触达

预警通知：红色预警通过邮件、短信通知等方式。

系统集成：实现预警信息的系统集成，可在其他相关平台查看到。

**（二）服务要求和交付**

1.总体要求：（1）深刻理解本项目实施的背景，同时密切结合该项目的实际情况，开展行业及业务背景情况调研，提交项目实施方案和计划，在以便充分保障本项目按期完成；

（2）供应商不得将该项目进行分包或转让。

2.安全保密要求：（1）中标单位必须遵循有关国家信息安全保障体系要求，按照国家法规实施安全等级保护，加强系统信息安全管理。同时按照个人信息保护法及保密法对数据进行保密，未经过采购方授权不能提供给第三方数据，不得在网络上随意发布个人信息，建立健全的信息管理及保护制度。

（2）中标单位须与采购方签订保密协议，需承诺未经采购方书面许可，中标单位不得以任何形式向第三方透露本项目的任何内容。所有参与项目开发、测试以及能接触到相关数据的人员均须签订保密协议，保密内容包含但不限于以下内容：技术资料及文档（含源代码及执行码）、系统数据等。

（3）泄密责任：赔偿招标方损失。

3.项目验收要求：交付的数据质量符合预定指标，分析结果具有科学价值，供应商提供完整文档及代码。

4.企业综合资质要求：投标人须具备数据安全及隐私保护合规能力，提供IS09001质量管理体系认证证书、IS020000信息技术服务管理体系认证证书、ISO27001信息安全管理体系认证证书、IS027017云服务信息安全管理体系认证证书、IS027701隐私信息管理体系认证证书、ITSS信息技术服务标准符合性证书优先考虑。

5.其他要求：供应商需提供技术支持及培训，确保数据处理流程可复用；数据交付需符合国家医疗数据安全管理规定，确保数据保密性；允许双方协商调整技术方案，以适应数据特征和研究需求。

**（三）人员要求**

为保证服务质量，要求供应商提供一支为项目建设提供全面解决方案、数据运维管理、数据库建设和数据分析项目经验丰富、配合默契且稳定的专业化团队。人员配置科学合理、分工明确。参与此项目的技术人员必须具有相关开发经验，能够与用户进行良好的沟通，掌握各专项技术领域的相关基础知识，具备项目管理、系统分析、软件评测等相关专业技术资格证书。

1.项目经理：具有丰富的项目开发经验、熟悉信息安全相关法规，具备信息系统项目管理师（高级）和注册信息安全专业人员（CISP）证书优先考虑，负责项目整体项目实施及进度把控。负责项目全面管理，包括制定计划并监督计划执行情况、控制项目进度、评审阶段性成果及协调解决疑难问题等。

2.供应商应明确各岗位的职责和任职资格。供应商所提供的项目团队人员不少于6人，配备系统分析师证书、软件设计师（中级）证书、软件评测师（中级）证书、注册信息安全专业人员(CISP)等证书等相关专业资格证书优先考虑。

3.供应商应安排经验丰富的培训人员，为系统使用人员提供系统配置、开发、安装、使用和维护等方面的培训。

1. **服务期限**

2025年11月30日前完成服务并通过验收。