

罕见病数据智能化管理系统功能需求

（一）建设目标

在长期的临床、科研及健康管理过程中，本院积累了大量患者就诊信息、用药记录、检验结果及检查报告数据。这些数据目前分散存储于多个系统、文件夹或本地设备中，格式不一、结构混乱，缺乏统一标准和集中管理机制，严重制约了数据的可访问性、互操作性与后续价值挖掘。为夯实数字化基础，支撑未来临床决策、科研分析与精准医疗应用，需对历史数据进行全面梳理、标准化处理，并构建一个安全、可靠、可扩展的“核心医疗数据仓库”。

罕见病数据智能化管理系统主要实现罕见病临床数据结构化入库、编辑、检索等功能；基于 NLP 算法实现罕见病临床病历标准 HPO 词条转化、自定义编辑等功能；实现罕见病阳性患者基因检测数据结构化入库、编辑及检索等功能，完成罕见病数据中台建设（临床数据&基因数据），实现数据导入、编辑、检索、导出等核心功能。

（二）主要技术参数

1. 数据库建设需求

1). 数据字典标准：针对诊疗业务规范使用到的数据字典，能够规范化管理（如联系人关系字典参照《家庭关系代码》国家标准（GB/T4761）、性别字典参照国家标准《个人基本信息分类与代码》（GB/T 2261.1-2003）等）；同时对于业务标准尚未明确标准的，可以自定义的增加和调整。

2). 诊疗数据规范化：将所有历史病历、医嘱、检验、检查等诊疗数据转换为符合行业规范（如 HL7 FHIR、OMOP CDM 或国家/地区医疗数据标准）的统一标准格式。

3). 诊疗数据可追溯：识别并登记已采集数据源的位置、格式、体量及元信息。

4). 数据治理工具可配置：配置自动化工具，可将异构数据批量转换为目标标准

格式；同时对于数据字典的转换可以提供对应的数据对照，可自由进行对照，对其对照修改进行记录追溯。

5). 数据结构化存储：设计并部署符合医疗数据安全规范的数据库架构，支持结构化存储、索引、查询与审计。

▲6). 数据的集中管理：建立一个统一管理、高可用、高效率、权限可控的“核心医疗数据库”，将患者诊疗信息全面入库管理，从相关各个业务系统及时抽取罕见病所需数据集，并对所抽取数据进行格式化的校验，明确查看抽取过程，实现过程可质控、数据可质控、结果可质控。

▲7). 建立标准化的数据集：针对罕见病的研究，按照其诊疗规范以及诊疗数据，分析并制定标准罕见病数据集，针对不同的标准数据集记录其特有的名称、版本信息、使用范围、内容、结构，对于数据集的修改可以查查看发布时间以及更新记录。

8). 数据集质控：建立完善的数据集校验规则，能够对其数据集的完整性、一致性、准确性、时效性进行多方面监控，保证数据的高可用性，为后续的数据治理、AI 建模、临床研究及患者服务提供高质量、结构化、可追溯的数据底座。

▲9). 数据访问控制：基于角色实现细粒度数据访问控制；权限模型应支持动态角色、数据集的灵活授权，为科研协作场景下的虚拟研究组管理预留扩展能力。满足合规要求（如 HIPAA、GDPR 或《个人信息保护法》）。

10). 数据传输：数据传输与存储全程加密（数据存储在本地），符合等保三级及以上要求；

11). 数据集数据追溯：记录数据集中每条数据的来源、处理过程与变更历史；

12). 数据库架构支持未来新增数据类型与容量增长；支持百万级患者记录的高效查询与批量导入；

▲13). 数据库建设能力要求：投标方具备不少于 20 万中国人群、上亿变异位点的数据库建设能力并已成功完成产品上线，支持提供云平台查询服务，须提供系

统截图及彩页证明。

2. 项目功能建设需求

▲1).用于实时展示以下信息：系统数据总览，包括但不限于临床数据总数量、基因数据总数量、基因诊断阳性数量；年度数据总览，包括但不限于临床数据总数量、基因数据总数量、基因诊断阳性数量；用于以下信息快捷入口：ID 信息查询、临床信息录入；基因信息录入；

2).用于以患者标识（如门诊号、住院号等）的临床信息与基因信息合并后的展示界面，同时支持多维度检索查询、导出等功能。展示界面字段包括但不限于：序号、门诊号、姓名、科室、主诉/临床诊断、异常系统、检测项目、检测结论、突变类型、基因、诊断疾病、更新日期；可支持自定义勾选展示字段。

3).系统检索功能字段包括但不限于：门诊号、姓名、主诉/临床诊断、异常系统、科室、检测项目、检测结论、突变类型、基因、诊断疾病等。

▲4).检索的患者列表，点击其特殊标记（如门诊号、住院号等）能够显示患者详细信息，如患者的门诊号、姓名、性别、年龄、孕周、民族、联系方式等基本信息；科室、主诉/临床诊断、现病史、既往史、家族史、异常系统、HPO 词条、检验报告、影像报告等临床信息；检测项目、检测结论、突变类型、基因、转录本号、变异信息 HGVS、染色体位置、遗传模式、变异致病性、诊断疾病等基因信息。

5).对于患者临床信息信息的展示，仅展示最新临床信息，历史临床信息支持跳

转至临床信息界面查看，实现系统的互联互通，方便查看相关历次诊疗信息。

6).对于基因信息的展示，可以根据 SNV/InDel、CNV、ROH、SV、线粒体 DNA、动态突变、附表、次要发现等不同模块类型进行展示相关内容。

▲7).临床信息管理模块，提供通过患者相关标识（如门诊号、住院号等）的检索功能，支持对已有临床信息导入、编辑、检索、导出等功能，支持新增临床信息、批量导入功能；同时展示界面展示内容包括但不限于：序号、门诊号、姓名、科室、主诉/临床诊断、异常系统、HPO 词条、检验报告、影像报告、分组、更新时间等；检索功能提供高级检索，其检索字段包括但不限于：门诊号、姓名、科室、主诉/临床诊断、异常系统、HPO 词条、分组、更新时间等。

8).新增临床信息：包含基本信息、病历信息、检验报告、影像报告四个模块信息，其中病历信息支持在线编辑、Excel 结构化数据导入；检验报告、影像报告支持电子附件上传形式；其中病历信息（包括主诉/临床诊断、现病史、既往史）导入后启动 NLP 自然语义算法自动转化为标准 HPO 词条进行入库；且加入人工标记/矫正提示，补充添加【异常系统】【分组】字段信息；

9).对于已经纳入的患者诊疗数据，能够针对其自定义的分组规则进行智能分组，支持同一样本纳入不同分组。

10).分组列表管理：支持新增分组功能；界面展示字段包括但不限于：序号、分組组名、标签、备注、入组数量、操作【查看、删除】等；点击新增分组，添加分組组名、标签、备注等信息，完成分组创建后透视临床信息界面支持入组案例检索及选择入组；点击操作【查看】，展示该分組下所有收录的患者列表，列表内容可以自定义选择；添加新增入组案例标记，点击后透视临床信息界面支持案例检索及选择入组。

▲11).基因数据管理模块：提供通过患者相关标识（如门诊号、住院号等）的检

索功能，支持已有基因信息导入、编辑、检索、导出等功能，支持新增基因信息、批量导入功能；同时展示界面展示内容包括但不限于：包括但不限于：序号、门诊号、姓名、主诉/临床诊断、检测项目、检测结论、突变类型、诊断疾病、分组信息、更新时间等；检索功能提供高级检索，其检索字段包括但不限于：门诊号、姓名、主诉/临床诊断、检测项目、检测结论、突变类型、诊断疾病、分组信息、更新时间等。

12).新增基因信息：包含基本信息、检测信息【主诉/临床信息、异常系统、HPO 词条、检测项目、检测结论、诊断疾病】、突变信息【包括 SNV/InDel、CNV、ROH、SV、线粒体 DNA、动态突变、附表、次要发现不同模块展示】，以上信息支持在线编辑、Excel 结构化数据导入。

13).将病历信息（包括主诉/临床诊断、现病史、既往史）自动转化为标准 HPO 词条的 NLP 自然语义算法需基于神经网络模型开发。

▲14). 须具备基因检测数据分析能力，具有实现 ACMG 自动化评级及变异位点智能化排序的核心自研算法，须提供专利或文献等证明文件。

3. 部署服务要求

1). 系统软硬件安装部署、调试。

2). 系统对接服务。

3). 系统软件包升级服务，包括软件、数据库及知识库升级服务等。

4). 系统运营与集成等相关服务。

▲5). 以上第（1）-（4）项部署服务要求技术团队不少于 2 人。（投标时须提供承诺函）。

6). 包含软件功能个性化需求调整（工作量 30 人天内）。

4. 服务器硬件要求

*1). 机架式服务器：配置 ≥ 2 颗处理器，主频 $\geq 2.0\text{GHz}$ ，单颗处理器性能不低于16核X86架构或性能不低于以上的其他架构处理器；配置 $\geq 256\text{GB}$ DDR5 内存；配置 ≥ 2 块600GB硬盘， ≥ 12 块4TB硬盘；配置独立RAID卡，支持RAID0/1/5/10/6等；配置双口万兆电口网卡；配置 ≥ 2 块1300W冗余热插拔电源，并提供配套的电源连接线(投标时须提供承诺函)。