### 癌症筛查数据管理

### 1.项目背景

目前，中国的癌症发病和死亡形势越发严重，是国民健康的主要威胁。癌症的防控工作受到了党和国家的高度重视。中国医学科学院肿瘤医院一直在癌症防治工作中发挥领军作用。

自2012年“城市癌症早诊早治项目”启动以来，中国医学科学院肿瘤医院领导各项目参与单位，在全国城市地区针对城市高发的五大类癌症——肺癌、结直肠癌、乳腺癌、上消化道癌和肝癌，开展危险因素调查、高危人群风险评估、临床筛查、卫生经济学评估和随访工作。至当前，项目已覆盖全国20个省份45个城市，累计完成问卷调查300余万例，评估出的癌症高风险人群130余万例，临床筛查90万余人次，阳性或疑似癌症检出8余万例。在我国癌症防控领域积累了海量的科学数据，数据已逐渐成为制定癌症防治策略、开展癌症研究的重要基础信息。但目前各数据库之间均为独立孤岛，不同数据库直接不能进行有效的数据整合与互通，收集到的资源数据存在数据不规范不统一等问题，为肿瘤数据收集、共享和利用带来极大的不便；此外卫生行政部门，临床科研以及社会各界对监测数据的需求不断增加。

为适应我国社会经济发展的形势，满足肺癌和结直肠癌防治工作的需求，建立癌症筛查数据管理项目，实现肿瘤大数据资源集中存储和统一管理、数据挖掘与共享的基础平台搭建为当务之急。

### 2.项目目标

本项目旨在建立和完善覆盖肺癌和结直肠癌筛查全周期的肿瘤研究队列，通过基于数据资源有限共享、多维国家级肿瘤防控数据库链接，构建涵盖多癌种、可扩充的病种研究临床数据库和肿瘤大数据资源系统，实现肿瘤大数据资源集中存储和统一管理、实现数据挖掘与共享的基础平台搭建。最终所建立的肺癌和结直肠癌高危人群筛查动态队列资源和技术体系作为医科院特色性科研资源，实施有效的开放共享，推动多学科融合集成，形成高效和系统的生物医学研究和创新发展机制。

### 3.项目建设内容

#### （1）项目建设内容一览表

|  |  |  |
| --- | --- | --- |
| 序号 | 建设内容名称 | 数量 |
| 1 | 对数据进行系统链接 | 1 |
| 2 | 云端大数据存储 | 1 |
| 3 | 肿瘤大数据资源系统 | 1 |
| 4 | 构建基础数据资源管理系统 | 1 |

#### （2）建设内容规格参数要求

##### 1.对数据进行系统链接

建立用户权限及数据权限管理模式，通过角色与数据访问授权方式将数据共享给用户使用。

本课题用户功能：实现下载本任务相关数据、并能够查看本任务相关统计图表；不需要项目PI及数据管理员审核直接可以浏览查看和使用。

本课题用户功能：查看其他任务相关数据或下载/查看其他任务的相关统计图表，需要相关任务课题负责人审核，同时需要项目PI及数据管理员审核方可浏览查看和使用。

本课题以外的其他科研用户功能：按照国家科学数据汇交和共享相关规定执行。

##### 2.云端大数据存储

要求在保证安全的基础上实现大数据高效存储和查询，具体要求如下：

* 支持大数据分布式存储以及关系型数据库混合使用的方案。
* 支持跨数据中心存储，热门数据多个数据中心缓存，确保业务稳定、持久可信。
* 支持基于自建节点的大规模分布式加速网络，上传速度快、稳定，支持大流量并发上传。
* 支持业务应用要求的高效查询能力.为了应对不同的业务应用场景，平台应选择合适的存储形式。

##### 3.肿瘤大数据资源系统

（1）资源信息库建设

建设包括但不限于乳腺癌、肾脏病、结直肠癌、食管癌的临床数据库，建立可扩充的病种研究系统。

支持数据综合概览，包括专病数据库集成病历信息数据；专病疾病特征信息统计；支持专病数据库的元数据管理。

支持新增患者，支持设置专病数据库的入排条件，实现符合条件的患者自动加入专病库。

（2）资源信息库检索

支持使用关键字对平台上所有医疗数据进行模糊搜索。搜索范围可以是平台所有的数据（对于有权限的用户），结果返回速度不超过3秒。

支持高级搜索功能。用户可以使用包括但不限于患者信息、诊断、用药、检验结果特征值组合查询条件。

支持对搜索返回结果进行可视化展现，针对不同的数据类型，可以选择各种不同的展示图形。

支持按照配置的特征参数，对查询结果进行过滤筛选，并且同步进行图形化展示。

支持灵活的数据查询和导出权限设置。可以基于不同层级的管理单元（科室、小组）进行权限设置，可以支持设定权限的有效期限。

对于单个特定病人，可以支持该病人最新临床数据的调阅；对于按照查询条件搜索的病人，查询到的信息是前一天晚上的更新的数据。

##### 4.构建基础数据资源管理系统

1. 结合数据采集标准及范围，实现数据交换与共享平台与元数据平台搭建。

支持数据采集的范围包括但不限于高/低发区域发病、死亡数据；危险因素数据；临床诊疗资料，包括但不限于诊断、病理、检查、检验、费用数据，影像资料和生物样本。

支持对数据交换进行可视化配置，监控节点的数据传输状态；支持各种主流数据库、文本文件、Excel文件、API接口、WebService服务数据格式的传输。

支持各种数据交换组件、数据处理组件，包括但不限于表交换、文件传输、SFTP上传下载、Http组件、清洗与转换组件，满足各种不同类型的数据传输交换需求。

支持拖拽式方式设计流程，通过可视化操作界面自动生成满足业务需要的交换任务。支持定时和事件触发，通过检测执行时间、监听数据库、指定文件变化来触发调度，灵活进行数据抽取和交换。

支持数据加密传输、断点续传、各种脱敏算法、数据权限与功能权限双重控制，让数据传输更安全可控，数据分区、并行装载技术让数据传输更高效快速。

1. 结合数据标准实现采集数据的转换、加工与装载

支持对采集的各种类型的数据进行质控，包括数据的唯一性检查，从而避免数据重复；数据的完整性检查，如身份证号码为空检查；数据的规范性检查，如性别的值（男、女、其它、未知）；数据的合理性检查，如年龄不大于200岁、体温不高于50；数据关联性检查，如以受试者为中心能关联上相关的发病信息、诊疗信息，最终实现受试者全周期的数据链条。

支持对采集的各类数据的标准进行统一管理，包括各采集指标的标准化业务命名管理、数据集版本管理、数据指标的类型管理、数据指标的值域管理。

支持对各类数据进行标准化处理，包括但不限于数据脱敏处理、数据格式标准化处理、数据值标准化处理、数据拆分合并处理、数据异常值处理内容。

支持受试者主索引配置管理，包括主索引指标设置，如身份证号码、出生日期、姓名、性别指标；各指标权重配置；关联度配置内容。

支持受试者主索引的管理，包括主索引的增加、删除、编辑功能；主索引的注册服务；主索引的人工合并与拆分管理。

1. 实现数据挖掘与共享的基础平台搭建

支持敏捷开发模式，集指标定义、数据建模和数据落地于一体，完成对现有指标进行统一管理工作；支持大量批量创建的方式完成指标元数据的定义、指标数据的处理以及指标数据落地的数据处理过程；支持不同数据库数据的落地工作，无需进行多余的ETL抽取数据处理。

支持跨数据库类型和版本的算法兼容性，同时提供了多种创建数据字典（数据整合）、指标定义（基础指标、公式指标、采集指标）以及创建指标主题（自动、手动）的方式来保证工具的扩展性。

支持数据并行计算和并行装载技术来保证数据落地的高效性。与此同时，指标管理支持大量的基于数据库或操作系统脚本语言和技术，来提供数据的抽取和装载性能。

支持合理的JAVA内存回收机制，用户操作结束后，系统不允许仍然占有该连接的内存；有完善的错误诊断和恢复机制，不会出现因用户误操作而导致系统崩溃；指标数据及元数据都会持久化在数据库中，不会因为应用服务器崩溃或病毒入侵而丢失数据。

### 4.系统技术要求

（1）总体要求

总体上要求项目采用先进、成熟的技术，保证技术先进性，保证投资的有效性和延续性，支持常用的操作系统、数据库、应用服务器和开发工具软件平台，应能够保证系统的安全、可靠稳定的运行，可伸缩、可扩展、方便移植，具有高可用性和高响应速度，并易于维护，开发部署灵活。

（2）技术应用的架构先进性

需采用业界近年发展的各项新技术，包括不限于分布式存储、SOA（面向服务的体系结构）。不仅要进行应用环境的综合分析评价，还要对应用软件、平台系统、底层架构开发所用的数据库系统及开发工具及所选择的开发方法、开发策略、开发规范、软件开发所贯穿的管理思想诸方面做出全面的考虑。另外，需要以全生命周期的观点对系统投运后的持续修改完善、功能扩充、维护升级、技术更新、集成协同方面进行统筹设计。

（3）注重整体架构设计

将采购人信息化所涉的硬件、软件、辅助设施在整体服务体系建设的顶层设计下，作为有机整体进行建设设计与优化，以信息化基础设施、企业应用架构的观点进行统筹考虑，而不是传统单一应用系统的建设方式。

（4）平台化的体系架构

将信息化基础设施进行分层设计，包括具体应用与功能应用的分层、应用与数据分层、不同类型数据的分层，杜绝传统建设中只以具体应用系统建设为重点的“单层”建设方式，保证整体信息基础设施的灵活性、可扩展性、厂商兼容性、可维护性要求。

（5）系统性能要求

查询类任务≤3秒；计算类任务≤10秒。

系统需要支持同时并发数≥500个（同时在线）。

### 5.项目安全要求

要求投标人参考国家相关法律法规、政策标准、行业规范开发本项目相关系统，并建立完整有效的项目系统以及安全管理方案，保障项目系统的安全稳定运行。能够对系统数据的机密性、完整性、可用性以及个人隐私保护等相关安全技术进行描述。

同时应满足但不限于如下要求：

提供满足国家标准的安全管理体系框架和安全管理方案，包括不限于安全管理制度、应用安全、数据存储、数据脱敏、数据传输、数据访问控制等安全。安全管理体系框架和安全管理方案需遵循国家法律法规、政策标准、行业规范。

系统对账号进行权限管理，要求必须具备分级权限管理，同时对超级用户和管理员用户特权进行安全管理，保证信息使用安全，保证防止非授权用户的侵入。

采用国际主流的不可逆加密算法MD5或国家标准SM3加密方式加密密码等信息，确保信息安全。

对敏感数据进行脱敏转换，保证个人隐私安全。

采用Https安全技术，保证数据传输安全。

数据访问控制需采用分级权限管理，系统对账号进行权限管理，同时对超级用户和管理员用户特权进行安全管理，保证信息使用安全，保证防止非授权用户的侵入。

### 6.项目实施要求

投标人应提供详细的项目实施与管理方案，内容包括但不限于项目管理、项目实施及进度计划等内容，并给出进度计划控制办法以及保障措施，确保能够按照项目时间要求完成项目实施与管理。

（1）项目实施交付时间：合同签订之日起3个月内交付。

（2）项目实施地点：采购人指定的地点。

（2）项目实施地点：采购人指定的地点。

（3）项目团队要求：至少应为本项目配备项目管理、技术负责人、设计开发及测试、部署实施、培训、运行维护等人员，明确各岗位的职责和任职资格，并确保人员配置科学、合理。

投标人配备的项目经理须专职并全程负责本项目实施与管理，应具有较强的组织协调能力，有足够的人力、物力及财力调配权，做好日常资源管理工作，并直接控制项目管理计划的各个要素。项目经理应具有4年以上类似项目管理经验。

投标人配备的技术负责人，须专职并全程负责本项目技术管理与指导工作。应制定本项目相关技术方案，识别本项目实施中关键技术及风险，并指导技术方案的具体实现。技术负责人须具有4年以上相关技术经验。

投标人配备的软件开发人员应具备系统、接口设计开发能力，须有2年以上软件开发与实施经验。

### 7.人员培训要求

针对本项目的人员培训，投标人需提供详细、合理、完整的培训方案，培训方案至少包括培训目标、培训计划、培训对象、用户答疑渠道，并按照上述要求进行详细说明。通过培训使系统使用人员能够熟练的掌握系统使用和维护方法，使其能够独立管理、使用和维护系统。同时投标人应提供培训所需计划和教材，培训教材格式包括视频教材、用户操作手册、培训PPT，培训所使用的语言和教材必须是中文。

### 8.项目验收要求

（1）项目验收需达到的标准与验收流程

在合同规定的时间内完成项目所有内容建设，经投标人自测试通过后，由投标人提交项目验收申请，采购人在接到书面验收申请后，组织对项目系统功能、技术文档等进行整体验收，并根据与采购人确定的验收核验清单提交相关交付物，全部验收通过后，双方签署《验收单》。验收不合格的，投标人需根据采购人意见整改至达到验收要求。

（2）验收交付物要求

投标人须保证提供详细的项目交付物，交付物包括但不限于项目技术文档及软件需求说明书、设计文档、测试报告、部署文档，以及项目相关过程管理文档及其电子版材料。

### 9.售后服务要求

投标人应建立响应及时、专业化的技术服务团队，并提供完整可行的售后技术服务方案，技术支持服务应具有一整套规范的技术支持服务运作体系和流程，指定稳定的技术服务队伍，提供故障诊断、技术咨询等全方位的技术支持服务。

质保期限：项目验收合格之日起3年。

当系统发生故障时，及时响应并进行故障处理。投标人应提供实时在线的客户服务，具备灵活多样的通讯手段，提供7\*24小时实时响应服务。

服务要求如下：

* 响应时间

投标人在接到故障报修要求时，0.5小时内做出明确响应，在4小时内给出处理方案。如需到现场处理故障的，投标人应于8小时内到现场提供故障处理服务并于24小时内解决问题。

* 技术咨询服务

投标人应提供7\*24小时技术咨询服务，在接到技术支持及业务咨询时，应为系统使用人员提供业务交流及技术支持服务。

* 故障诊断与排除服务

当接到用户申告故障服务请求后，投标人应提供远程或上门进行故障诊断与排除服务，技术服务执行过程接受客户的监督和检查。

* 软件升级优化服务

质保期内对于因政策变动、关联系统调整、管理规范调整、应用需求变更原因进行的升级优化服务，必须对采购人做驻场培训和指导。

* 系统BUG修复

在维护合同有效期内，免费提供BUG问题的修复。

* 备份与还原服务

投标人应提供系统备份服务，必要时须提供系统还原服务。以防止由于操作失误、系统故障等因素或意外原因导致数据丢失。

* 季度巡检服务

每季度不少于1次现场对系统进行全面巡检服务，及时发现和排除潜在问题或故障隐患，保证系统的稳定运行。