**中国消化系统肿瘤基础数据资源智能管理平台构建**

# 技术规格要求

## 1.项目背景

多年以来，我国十分重视消化系统肿瘤的防控工作，开展了大量人群水平的病因学预防及早诊早治项目，但是各研究间缺乏共享平台与机制，限制了宝贵的生物资源和数据库的利用效率。为了解决上述问题，实现数据资源整合与管理，急需建设消化系统肿瘤基础数据整合、管理及分析系统。

## 系统目标

本项目依托已开发建设的中国癌症防控信息管理平台，将涵盖我国高发、低发不同区域的食管癌、胃癌、肝癌和结直肠癌的发病、死亡及人群危险因素监测信息，及病理、影像、检验等临床诊疗信息、生物样本资源信息等电子数据集，基于统一的标准和规范，一并纳入到中国癌症防控信息管理平台资源库，并通过技术手段形成一套唯一的、标准的、权威的和高价值的专业化主题数据资源库，进一步实现消化系统肿瘤数据的挖掘、可视化分析、动态监测与共享利用。

## 3.项目建设内容

## 3.1项目建设内容一览表

|  |  |  |  |
| --- | --- | --- | --- |
| 序号 | 内容名称 | 数量 | 单位 |
| 1 | 中国消化系统肿瘤基础数据资源智能管理平台 | 1 | 个 |
| 2 | 中国消化系统肿瘤基础数据资源信息库 | 1 | 个 |
| 3 | 中国消化系统肿瘤数据采集标准规范 | 4 | 套 |
| 4 | 数据质控体系 | 1 | 套 |
| 5 | 相关数据统计指标 | 1 | 套 |
| 6 | 数据共享标准接口 | 1 | 套 |

## 3.2建设内容功能参数要求

#### 3.2.1中国消化系统肿瘤基础数据资源智能管理平台

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，包括诊断、病理、检查、检验、费用等数据，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服务等数据格式的传输。

支持各种数据交换组件、数据处理组件，如表交换、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、加工与装载等内容

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等；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、诊疗信息等，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等。

支持对各类数据进行标准化处理，包括数据脱敏处理、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、数据异常值处理等内容。

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、性别等指标；各指标权重配置；关联度配置等内容。

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

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

1. 根据项目实际情况并结合数据挖掘平台实现各主题库、数据可视化、数据共享API的搭建

支持对海量的肿瘤基础数据进行一般统计分析，如按照癌种、时间、区域的维度统计癌症病例数量、生物样本数量等。

支持对数据的数学分析模型，对数据进行二次深度挖掘分析，发现隐藏的数据信息，为完善肿瘤防控提供科学支撑。如基于肿瘤大数据进行中国高发、低发地区食管癌危险因素分布分析、发病和死亡分析、就诊路径分析、癌症带来的经济负担分析等。

支持饼图、柱状图、折线图等常见图形；支持气泡图、面积图、省份地图、词云、瀑布图、漏斗图等高级图表类型。

支持多主题（如食管癌、胃癌、肝癌和结直肠癌）、多场景、多层次、多维度，多样式的展示，支持多级下钻，友好的展示界面。

支持数据权限配置，数据权限精确到研究对象、研究对象的具体指标等内容。支持数据权限按组划分、按人划分等。

#### 3.2.2中国消化系统肿瘤基础数据资源信息库

1. 资源信息库建设

建设抑郁症，乳腺癌，肾脏病，结直肠癌，食管癌等临床数据库，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；

1. 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。用户可以使用患者信息、诊断、用药、检验结果等各种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。

#### 3.2.3中国消化系统肿瘤数据采集标准规范

依据国际、国家、行业等规范标准及卫生信息标准（如国家卫生统计标准、ICD-10、AJCC分期等），并依照中国消化系统肿瘤基础数据资源智能管理业务实际需求，完成肿瘤数据采集标准规范（数据表单）的编制，为肿瘤信息互联互通和数据传输提供统一权威的标准，达到以网络信息化的服务为手段，对我国高/低发地区的食管癌、胃癌、肝癌和结直肠癌的电子数据集进行采集与整合，为各地区卫生管理机构、科研机构等提供数据共享协同服务。

#### 3.2.4数据质控体系

1. 实现整个数据质控平台框架搭建

支持重复数据检查，包括使用组合字段检查，MD5检查等方式。

支持数据完整性检查，包括每个字段是否填充、字段根据一定的条件是否填充。

支持数据规范性检查，包括日期格式、数值格式、值域内容等。

支持数据合理性检查，包括业务逻辑的检查、数据波动检查、数据平衡检查、数据离散值异常值检查。

支持结果集比对，可以根据不同的系统进行数据比对，从而判断数据的合理性。

1. 根据采集数据的数据模型，设计具体的质控规则

支持规则可以配置到具体的字段、表、数据库，实现整个数据全方位检查。

支持规则可以配置到内容级别，统计级别，实现多维度的数据检查。

支持每个规则的权重配置。

支持每个规则的级别配置，对于必须通过的规则，可以反馈到数据生产源，整改。

#### 3.2.5相关数据统计指标

运用肿瘤大数据可视化图表展示手段（除了传统的饼图、柱状图、折线图等常见图形，还有气泡图、面积图、省份地图、词云、瀑布图、漏斗图等图表类型），结合动画及图形特效技术，展示特色业务和工作成果。根据不同的对象，采用多主题（如食管癌、胃癌、肝癌和结直肠癌）、多场景、多层次、多维度，多样式的展示，支持多级下钻，友好的展示界面，所见即所得。

数据可视化不仅仅是将数字变成直观的图表，数据可视化的目标是快速发现问题，识别问题，分析原因。系统可以将中国消化系统肿瘤基础数据资源库中每一个数据项作为单个图元元素表示，大量的数据集构成数据图像，同时将数据的各个属性值以多维数据的形式表示，可以从不同的维度观察数据，从而对数据进行更深入的观察和分析。

#### 3.2.6数据共享标准接口

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科研机构等用户使用，即保证了数据的访问权限，也保证了开放共享。

本课题用户：查看本任务相关数据或下载/查看本任务的相关统计图表，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：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：按照国家科学数据汇交和共享相关规定执行。

## 总体技术要求

（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等软件平台，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）技术应用架构要求

需采用业界近年发展的各项新技术，支持分布式存储、SOA（面向服务的体系结构），支持分层设计，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等。保证整体信息基础设施的灵活性、可扩展性、厂商兼容性、可维护性等要求。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。

## 总体安全要求

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。

## 项目实施要求

项目实施交付时间：合同签订之日起4个月。

项目实施地点：采购人指定的地点。

项目团队要求：要求投入的人员配置科学合理、分工明确，至少应包括项目经理、技术负责人、系统设计开发及测试、部署实施、运维等角色。

必须定期向采购人作进度汇报，并对任务执行过程中存在的问题进行沟通，共同提出解决方案。

## 7.项目验收要求

在规定的时间内完成系统建设，经投标人自测试通过并提交相关文档后，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，组织对系统功能、相关技术与安全性进行验收，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。

项目验收后投标人须保证提供详细的项目交付验收物,包括但不限于系统设计文档、系统安装手册及测试报告，交付物的形式不限于电子版及纸质版。

## 8.人员培训要求

投标人须提供详细、合理、完整的培训方案，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资料。

## 9.售后服务要求

投标人应建立专业化、快速响应的售后技术服务团队，及时解决售后服务问题，并提供完整可行的售后技术服务方案。

免费质保期限：验收合格之日起3年。

技术服务响应时间：7\*24小时实时响应并提供技术支持。

技术服务内容包括但不限于技术咨询服务、系统故障诊断与故障排除、故障分析服务等。

服务方式应至少包含现场、远程、热线电话等方式，灵活解决反馈的问题和故障。