

湘雅医学大数据 数据集规范

编制说明

（征求意见稿）

**内部资料，注意保密**

**2016年3月**

* **文件信息**

|  |  |
| --- | --- |
| **名称** | 湘雅医学大数据数据集规范 编制说明 （征求意见稿） |
| **作者** |  |
| **状态** |  |
| **版本** | 1.0 |

* **签署清单**

|  |  |  |  |
| --- | --- | --- | --- |
| **职务** | **姓名** | **签名** | **日期** |
|  |  |  |  |
|  |  |  |  |

* **相关会议及研讨会**

|  |  |  |  |
| --- | --- | --- | --- |
| **日期** | **参与人员** | **主题** | **备注** |
|  |  |  |  |
|  |  |  |  |

* **其他相关文档**

|  |  |
| --- | --- |
| **相关文档** | **注释** |
|  |  |
|  |  |

* **版本变更记录**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **编号** | **更新日** | **变更描述** | **变更者** | **签署者** | **变更日期** | **注释** |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

目录

一、 任务来源 1

二、 目的和意义 1

三、 业务分析 2

1 业务需求 3

1.1 辅助决策 3

1.2 科研需求 4

1.3 辅助诊疗 4

1.4 新药研发 5

2 管理需求 5

2.1 医院质量控制，减少医疗纠纷 5

2.2 主管部门的管理需求 5

3 服务对象需求 6

四、 建模 7

1 健康档案的三维系统模型 7

2 湘雅医学大数据的信息来源 10

3 湘雅医学大数据的业务域划分 10

五、 规范性参考文件 11

1 卫生信息标准 12

2 国际标准 13

六、 规范管理工作流程 14

# 任务来源

为加快落实湘雅大数据平台的顶层（概念）设计，加快推进中南大学2015年医学大数据建设项目（一期）的建设，全面实现湘雅系医院数据资源的统一汇聚，保证数据资源的充分使用，特编制湘雅大数据平台系列标准及规范。

“湘雅大数据平台”标准规范体系将包括数据类、技术类、服务类以及管理类等四大类系列标准。此次征求意见为《湘雅医学大数据数据集规范 基本信息数据集 （征求意见稿）》和《湘雅医学大数据数据集规范 医疗服务数据集（征求意见稿）》。

# 目的和意义

“十三五”规划纲要中指出，“实施国家大数据战略。把大数据作为基础性战略资源，全面实施促进大数据发展行动，加快推动数据资源共享开放和开发应用，助力产业转型升级和社会治理创新”，“提升健康信息服务和大数据应用能力”。

作为百年知名品牌，湘雅大数据平台的建设将为中南大学湘雅系医院的新百年谋划坚实的基础，为医疗体制改革提供关键支持并保证自身的健康发展，将大幅度提高患者就医水平与就医质量，突破现在的医疗服务瓶颈。

中南大学2015医学大数据建设项目将统一汇聚校附属医院已保存了多年的历史医疗相关数据，数据包括：就诊（住院）记录、影像记录、检验记录等，按照统一的数据格式和要求进行采集、加工、存储、交换和共享，完成标准的湘雅医学大数据库的建设将为充分发挥数据的潜在价值打下坚实基础。

由于标准体系涵盖的标准内容繁多，而目前医疗卫生信息化建设的规范标准只涵盖了少数几个方面，还没有形成一套完整的、系统化的规范标准体系，没有能直接供湘雅大数据平台使用的标准和规范。所以，湘雅大数据平台系列规范标准需要结合实际不断摸索、不断总结和完善。

由此可见，湘雅大数据平台系列规范标准的制定是湘雅大数据平台建设的基础，在遵循国家卫生部等已有的各类规范标准、充分参考国际相关标准基础上，我们将突出重点、狠抓关键，配合平台建设分析相关标准的轻重缓急，将《湘雅医学大数据数据集规范 基本信息数据集 （征求意见稿）》和《湘雅医学大数据数据集规范 医疗服务数据集（征求意见稿）》先行制定，并逐步运用在系统的设计和建设中；其它标准将根据系统建设的不同阶段，有计划地、循序渐进地制定并运用。逐步完善和建立一套在国家现有规范标准基础之上的，有湘雅特色的湘雅大数据系列规范标准。

# 业务分析

湘雅大数据平台以信息化手段整合湘雅系医院的数据资源，以及院外与医院管理、服务以及病人相关的各种数据，对数据资源进行有效的数据分析和数据挖掘，从中取得大量的有价值的信息，将来还可以汇聚移动终端、智能穿戴设备数据，未来还可与交通、气象、人口等多个领域数据源交互，开展更多维度的分析处理，从而帮助医生、医院的管理层、患者等，促进医院更好地提供医疗服务和提升治疗质量，促进各病种的科学研究。将来能对疾病早期发现、早期干预、早期治疗，促进患者对医疗资源便利化利用和控制医疗费用，提高资源效率，缓解“看病难、看病贵”问题，推进人人健康目标的实现。

湘雅大数据平台的用户主要包含医院PI、院方管理者、卫生行政管理者、患者/对健康有需求的人、教学科研人员、业务相关方，范围涉及区域内各医院、医疗卫生服务机构和其它业务相关的企业或部门。

湘雅医学大数据将满足研究人员用于科研、病友/普通大众的个人健康管理外，其另一个重要价值因素在于提供权威数据的二次使用，为诊疗、教学、科研，为管理决策、医疗健康和社会经济发展提供服务。下面对这些需求进行描述。

## 业务需求

### 辅助决策

大数据平台提供统一可视化分析展示平台，为相关决策提供数据依据。透过数据下探可深入分析挖掘医院现状，挖掘分析潜在信息及其各个数据之间的相关性。

透过过去与现在的数据规律，更可结合其它数据源综合分析，给出未来一定时间段内的预测。

临床辅助决策系统可以提高工作效率和诊疗质量，提醒医生防止潜在的错误，如药物不良反应。

在将来临床辅助决策系统深入的融合到医院的业务中，优化医疗流程还可以使医疗流程中大部分的工作流流向护理人员和助理医生，使医生从耗时过长的简单咨询工作中解脱出来，从而提高治疗效率。

### 科研需求

为102个临床病种项目（一期为15个临床示范项目，将来可扩展更多的临床病种项目），提供基于病种的科研数据采集平台；

* 病症关联性研究：通过多维度分析大量病症，用图形化方式展示不同病症之间的关联性，更可深入洞察病症之间的关联性；
* 长期跟踪：关联展示同一病人不同时期的检查检验结果，结合大数据平台汇聚的可穿戴设备等跟踪数据，长期深入的跟踪病人，分析疗效；也可以对慢性病人的远程监控系统收集数据，并将分析结果反馈给监控设备(查看病人是否正在遵从医嘱)，从而确定今后的用药和治疗方案；
* 比较效果研究：通过湘雅大数据平台中的数据，可以通过全面分析病人特征数据和疗效数据，然后比较多种干预措施的有效性，可以找到针对特定病人的最佳治疗途径。对同一病人来说，医疗服务提供方不同，医疗护理方法和效果不同，成本上也存在着很大的差异。精准分析包括病人体征数据、费用数据和疗效数据在内的大型数据集，可以帮助医生确定临床上最有效和最具有成本效益的治疗方法。减少过度治疗和治疗不足带给病人的负面影响；
* 更可深入洞察病症诊疗手段与成果，为相关病症研究提供数据支撑。

### 辅助诊疗

对于辅助诊疗，大数据平台借助过往各类病例和各类数据源，深入分析相关病症并需找、推荐最优治疗方案，为个性化诊疗提供基础。

### 新药研发

在临床用药过程中，药物使用可能会引起病人的不良反应。这种不良反应会导致治疗作用减弱甚至失败，严重的可能会导致患者死亡，同时不合理用药也会使患者医疗费用大大增加，给患者带来更多的经济负担。因此研究药物副作用对于提高患者疾病的治疗质量，指导临床用药以减少药物对患者的伤害，降低药物费用以及指导新药研发都具有重要的意义。

## 管理需求

### 医院质量控制，减少医疗纠纷

医疗改革领域中的临床转型实际上就是在品质提升、风险管理、绩效考核和成本控制几者之间保持微妙的平衡。通过湘雅医学大数据项目，将大数据技术应用于湘雅系医院将来扩展为其它的医院，可以从一个新的高度分析医院的运营情况。通过大数据的分析和挖掘，对医院服务质量进行控制，如对某些病人的出院决定是否太草率了，又或许是病人出院后的用药指导不够明确等，可以协助医院发现问题，并尝试解决，更好地进行质量控制，减少医疗纠纷。

### 主管部门的管理需求

卫生计生相关主管部门可以基于湘雅大数据平台中的数据，对公共卫生舆情进行监控预警；对传染病进行研究（如传染病的扩散范围等）；对公共卫生应用提供助力，如结合环境数据与人群健康大数据，分析环境对健康的影响，分析环境污染与某些疾病的关联性；可以对医疗资源进行合理的调配和适当的分配。

## 服务对象需求

湘雅大数据平台的用户主要包含医院科研人员、院方管理者、卫生行政管理部门、患者/对健康有需求的人、教学人员、业务相关方，范围涉及区域内各医院、医疗卫生服务机构和其它业务相关的企业或部门。不同用户对基于健康档案的区域卫生信息平台需求有不同的关注点，具体如下：

**湘雅系医院的科研人员：**主要是利用湘雅医学大数据项目进行科学研究，为其在诊断过程中，进行辅助诊断。通过湘雅医学大数据项目中的从医院业务系统采集的数据和基于病种的临床采集系统采集的数据，通过能够缩小诊断基础条件、症状、用药、副作用，病史和其它影响因素的大数据技术，让医生能够分离、发现和研究疾病罕见的细微差别，并相应的治疗病人，可以通过预测建模和机器学习，帮助医生进行辅助诊断。

**院方管理者：**可以通过湘雅医学大数据，通过分析医生、设备等资源信息，合理安排人员、设备等资源。能优化诊疗流程，将大数据分析结果融合到医院的业务中，优化医疗流程还可以使医疗流程中大部分的工作流流向护理人员和助理医生，使医生从耗时过长的简单咨询工作中解脱出来，从而提高治疗效率。

可以优化医院服务，对出院病患进行实时访问及跟踪回访，对病人的院后情况进行收集，通过数据平台进行对比分析，为患者提供院后的专业服务等。

**卫生行政管理部门：**可以将区域卫生信息平台与湘雅医学大数据相互结合，进行数据交互，为公共卫生政策、医疗纠纷处理、进一步提高区域资源共享水平、提高监督管理能力，化解疾病风险等方面。

**患者/对健康有需求的人：**主要关注的是如何能获得可及的、优质的卫生服务；获取连续的健康信息、全程的健康管理等方面需求。

# 建模

## 健康档案的三维系统模型

健康档案的系统架构是以人的健康为中心，以生命阶段、健康和疾病问题、卫生服务活动（或干预措施）作为三个维度构建的一个逻辑架构，用于全面、有效、多视角地描述健康档案的组成结构以及复杂信息间的内在联系。通过一定的时序性、层次性和逻辑性，将人一生中面临的健康和疾病问题、针对性的卫生服务活动（或干预措施）以及所记录的相关信息有机地关联起来，并对所记录的海量信息进行科学分类和抽象描述，使之系统化、条理化和结构化。

健康档案的三维系统架构如图所示：

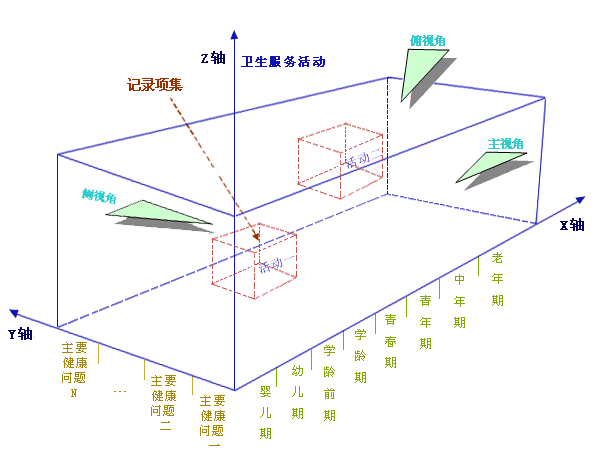


图 健康档案的三维系统模型

第一维（X 轴）：生命阶段

按照不同生理年龄可将人的整个生命进程划分为连续的若干生命阶段，如：婴儿期（0～1 岁）、幼儿期（1～3 岁）、学龄前期（3～6 岁）、学龄期（6～12岁）、青春期（12～20 岁）、青年期（21～45 岁）、中年期（46～60 岁）、老年期（60 岁以上）等八个生命阶段。也可以根据基层实际工作的需要，将人群化分为：儿童、青少年、育龄妇女、中年和老年人。

第二维（Y 轴）：健康和疾病问题

每一个人在不同生命阶段所面临的健康和疾病问题不尽相同。确定不同生命阶段的主要健康和疾病问题及其优先领域，是客观反映居民卫生服务需求、进行健康管理的重要环节。

第三维（Z 轴）：卫生服务活动（或干预措施）

针对特定的健康和疾病问题，医疗卫生机构开展一系列预防、医疗、保健、康复、健康教育等卫生服务活动（或干预措施），这些活动反映了居民健康需求的满足程度和卫生服务利用情况。

健康档案的三维概念模型，可以清晰地反映出不同生命阶段、主要疾病和健康问题、主要卫生服务活动三者之间的相互联系。同时，坐标轴上的三维坐标连线交叉所圈定的空间位置（域），表示了人在特定生命时期、因特定健康问题而发生的特定卫生服务活动所需记录的特定记录项集。由于三维空间中的任意一个空间位置都对应着某个特定的健康记录，从而构成了一个完整、立体的健康记录，这些健康记录全面地反映了个人健康档案内容的全貌。

健康档案的三维概念模型为健康档案内容的规划与设计提供了一个科学、合理、灵活的指导框架。由于人的健康状况及健康危险因素很大程度上受到社会经济和环境因素条件的影响，因此在不同的社会经济发展阶段、不同的地区和环境条件下，所需重点关注的主要健康问题以及所需记录的主要健康信息必然存在差异。在进行健康档案的规划设计时，应因地制宜，在三维概念模型的指导下，根据不同环境条件和关注的重点选取适合本地需求的主要健康问题和记录项集；并可根据实际情况进行灵活的调整（更新、缩减或扩展），使有限的卫生资源得到合理的分配和充分利用。

另一方面，与特定健康问题和卫生服务活动相对应的记录项集的内容，即内部记录项也不是一成不变的。在所关注的健康问题及卫生服务活动的深度和广度不断调整、完善的过程中，健康记录的内容可以随着居民健康管理需求或干预措施的变化与改善而进行适时调整。

由此可见，用于描述健康记录的数据模型必须具备良好的可扩展性，在满足所记录的健康内容不断变化的同时，能够保持数据模型的稳定。

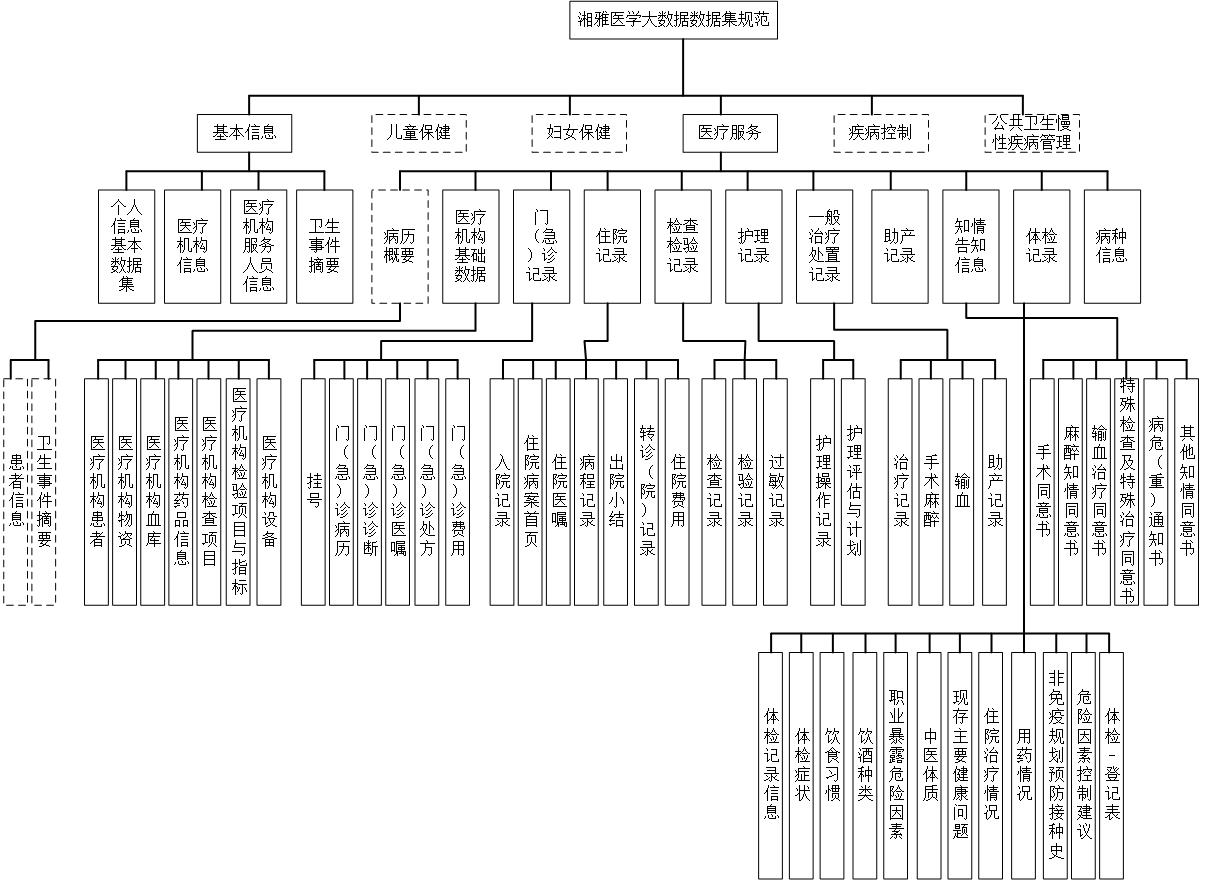
## 湘雅医学大数据的信息来源

湘雅医学大数据的信息主要来源为湘雅系医院的数据资源，以及院外与医院管理、服务以及病人相关的各种数据，对数据资源进行有效的数据分析和数据挖掘，从中取得大量的有价值的信息，将来还可以汇聚移动终端、智能穿戴设备数据，未来还可与交通、气象、人口等多个领域数据源交互。

中南大学2015年医学大数据建设项目（一期）将汇聚湘雅系医院的数据资源。

## 湘雅医学大数据的业务域划分

通过分析，湘雅医学大数据业务领域（主题）分为6个一级类目，即：基本信息、医疗服务、儿童保健、妇女保健、疾病控制、公共卫生慢性疾病管理。其中“医疗服务”包含10个二级类目：病历概要、医疗机构基础数据、门急诊记录、住院记录、检查检验记录、护理记录、一般治疗处置记录、助产记录、知情告知信息、体检记录和针对病种的病种信息。全部业务域包括业务子域共35个。



图：湘雅医学大数据的业务域目录

业务子域在整个健康管理过程中并不是独立的，之间存在着一定的关系。

如“出生医学登记”将产生“预防接种”服务；“儿童保健域组”、“妇女保健域组”、“公共卫生疾病卫生管理域组”的信息均可源于“医疗服务域组”；“疾病控制域组”中“传染病报告”的信息可来源于“医疗服务域组”、“疾病管理域组”、“儿童保健域组”或“妇女保健域组”的卫生服务活动所产生的信息。

# 规范性参考文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件，仅注日期的版本参考于本文件。凡是不注日期的引用文件，其最新版本（包括所有的修改单）参考于本文件。

## 卫生信息标准

卫生部《电子病历基本架构与数据标准（试行）》（卫办发〔2009〕130号）

卫生部《电子病历系统功能规范（试行）》（卫医政发〔2010〕114号）

国家中医药管理局《中医电子病历基本规范（试行）》（国中医药发〔2010〕18号）

卫生部《健康档案基本架构与数据标准（试行）》

卫生部《基于健康档案的区域卫生信息平台建设指南》

卫生部《基于健康档案的区域卫生信息平台建设技术解决方案（试行）》

卫生部《基于电子病历的医院信息平台建设技术解决方案（1.0版）》（卫办综发〔2011〕39号）

卫生部《三级综合医院医疗质量管理与控制指标2011版》

卫生部《医院管理评价指南（试行）》

卫生部《临床检验结果共享系统互操作规范》

卫生部《中国医院信息系统》标准代码

卫生部《医院信息管理系统（HIS）软件基本功能规范》

卫生部《全国卫生系统医疗器械仪器设备分类与代码》

卫生部《化学药品分类与代码YY0252-1997》

中国医院信息系统数据集。

《湖南省区域卫生信息化平台数据接入标准》

《国家卫生计生委医院监管系统数据采集范围说明-医疗服务分册》

## 国际标准

临床文档结构标准HL7 CDA V3

IHE（Integrating the Health Enterprise）医疗健康信息集成规范

ICD-9、ICD-10国际疾病分类

SNOMED、SNOMED RT，系统医学命名法

DRGs（Diagnosis Related Groups），诊断相关分组

DNC（National Drug Codes），国家药品编码

DICOM，医学数字化影像标准

ISO TC215，国际标准化组织

CEN TC 251，欧洲标准委员会

UMLS，统一的医学语言系统

# 规范管理工作流程

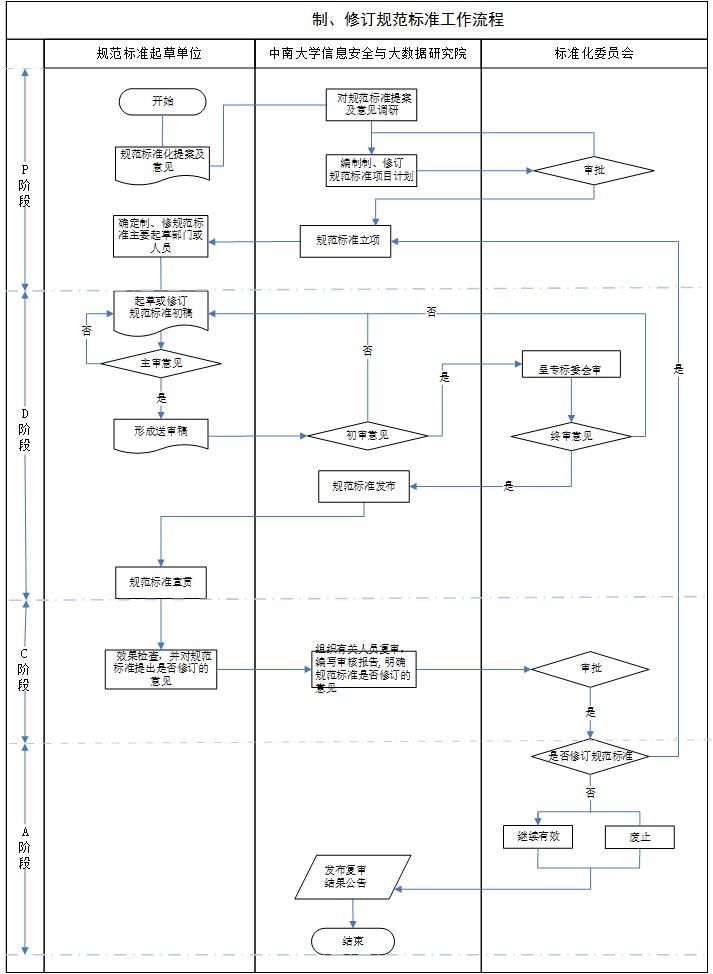


图 制修订规范工作流程