

医疗机构 医疗大数据平台建设指南解读

北京大学肿瘤医院

衡反修

2019年7月5日



- 1 引言
- 2 总体设计
- 3 建设要点
- 4 应用场景
- 5 专家论点
- 6 医疗大数据管理制度示例
- 7 医疗大数据平台建设案例
- 8 术语定义
- 9 参考文献
- 后记

一. 指南编写背景

二. 指南内容解读

三. 关于指南征求意见稿

1、国家政策背景

国务院办公厅于2016年发布了《关于促进和规范健康医疗大数据应用发展的指导意见（国办发[2016]47号）》、国家卫生健康委员会2018年出台了《国家健康医疗大数据标准、安全和服务管理办法（试行）》。

2、医疗数据利用现状

大部分医院医疗数据仍然属于沉睡状态，医院未建立比较好的机制和软件系统来实现数据利用，同时信息利用孤岛现象也存在。

3、医疗大数据平台发展现状

目前全国各地医院都在发展医疗大数据平台，但诸如数据标准、大数据平台技术、大数据平台安全及大数据应用等涉及平台建设无统一的规范化指导。

CHIMA课题：医疗大数据平台建设研究
课题产出：医疗大数据平台建设指南

目的：

- 1、为医疗机构医疗大数据平台建设提供规范和指导
- 2、推动我国医疗大数据行业应用发展

适用范围：

- 1、各级医疗机构：数据处理、管理和利用；
- 2、各级医疗管理机构：区域医疗机构数据应用
- 3、医疗信息化建设服务商：规划、研发、实施和运营

指南特点

- 1、以应用为主，技术为辅（不深入技术细节）
- 2、抓行业人员关注要点（不面面俱到）
- 3、强调易读性，尽量语言平实（不免存在俗语俗称）

核心读者：

医疗机构信息中心主任和技术工程师

医疗信息和数据服务企业工作人员

医疗大数据平台定义

什么是医疗大数据：

健康医疗大数据涵盖人的全生命周期，既包括个人健康，又涉及医药服务、疾病防控、健康保障和食品安全、养生保健等多方面数据的汇聚和聚合。

指南针对医疗大数据定义范围：医疗机构产生的以患者为中心的诊疗和服务数据。

什么是医疗大数据平台

国家卫计委发布的《医院信息化建设应用技术指引（试行）》中对医疗大数据平台的标准解释为：（1）数据交换汇集（2）数据存储（3）分布式计算（4）数据可视化。具有上述特征的即为医疗大数据平台。医疗大数据平台本身也属于一种具有上述特征的软件、系统。

医疗大数据平台的建设也可以理解为基于大数据技术，进行医疗数据处理、利用、分析和服务的软件系统。

医疗大数据平台建设模式

基于当前的医疗行业背景，医疗大数据平台建设存在两种模式：

1、医疗机构独立建设

- 对医院信息人员技术能力要求比较高
- 产出科研为主
- 主动、可控性强、可持续发展

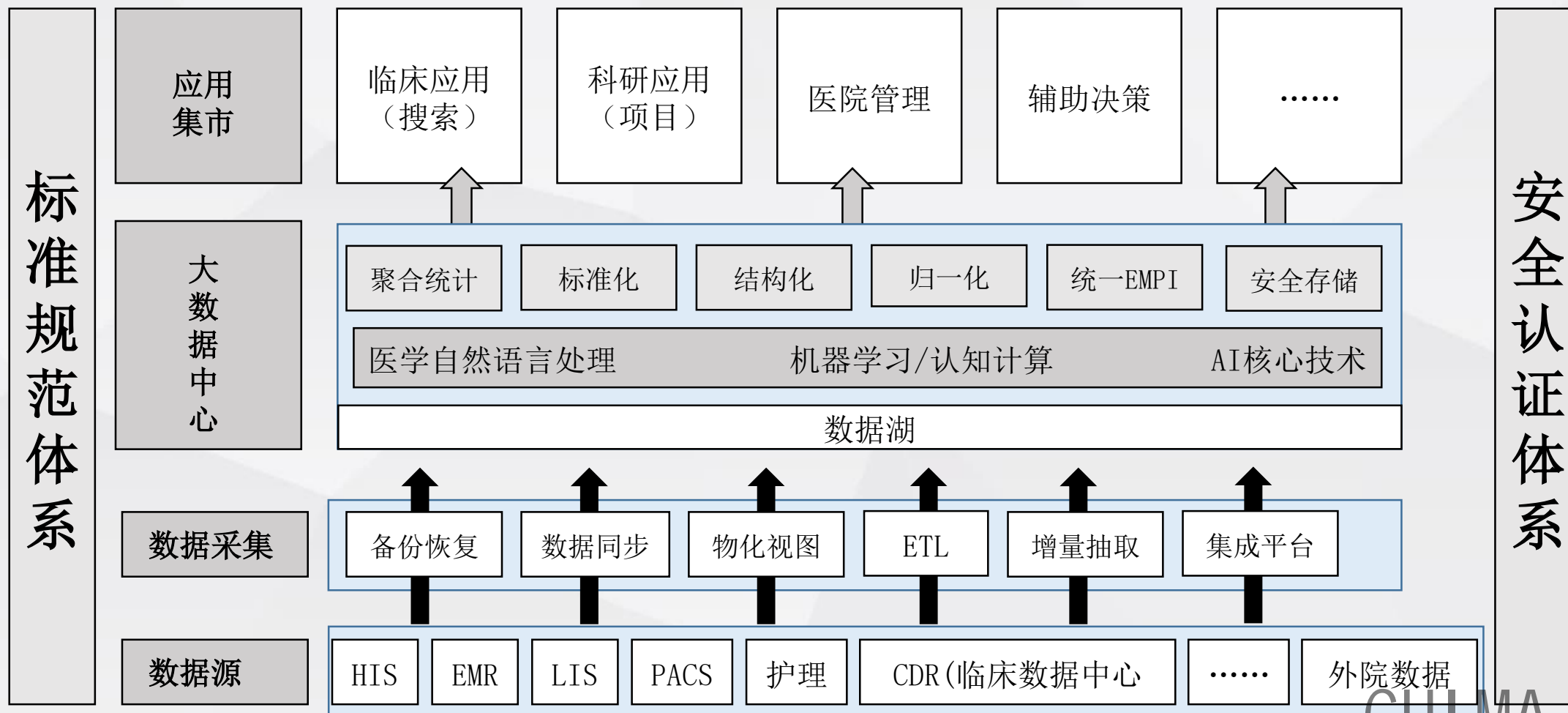
2、医疗机构采购市面成熟产品院内部署建设

- 对医院信息人员要求比较低，一般甲方承担项目经理职责
- 面向科研和临床、患者服务等
- 成效较快，利于普及并推进发展

本指南在内容编写时以第二种建设模式为主，在第二、三章节详细介绍了基于第二种模式下医疗机构在建设医疗大数据平台过程中的规范性、指导性内容

二、指南内容解读

医疗大数据平台架构



医疗大数据平台技术路线

当前医疗行业大数据平台技术多种多样，既有采用传统数据库（oracle、sqlserver、mysql等）建设大数据平台的，也有采用以hadoop等为代表的分布式计算的大数据平台。

本指南技术路线部分采用hadoop技术，主要原因为：

- 1、**高可靠性**：分布式存储和冗余，默认三副本存储
- 2、**高扩展性**：不停机线性扩容、横向扩展架构
- 3、**高效性**：分布式运算框架，海量数据处理
- 4、**高容错性**：可容忍某些部件故障和磁盘失效
- 5、**低成本**：普通X86服务器、硬件要求低

基于hadoop上述特点，所以本指南介绍的医疗大数据平台是采用上述技术的平台

医疗大数据平台安全体系

基于医疗大数据的4V特征（规模性、多样性、高速性、价值性——即患者医疗信息的高度聚合性）决定了医疗大数据平台建设的安全体系至关重要。

即最基础安全要满足医院安全等级保护的要求。

同时应从以下方面做好安全防护：

- 1、**硬件**安全：设备管理、巡检、维护、硬件访问权限
- 2、**网络**安全：私密性和安全性，如：防火线、流量控制器、堡垒机、审计
- 3、**数据**安全：数据脱敏、加密
- 4、**应用**安全：访问权限控制

医疗大数据平台部署模式

在当前公有云技术发展情况下，当前医疗机构软件系统部署存在两种模式：

1、公有云模式：来自临床研究服务商、多中心研究

- 将软件或系统直接部署在公有云或混合云
- 安全体系一般由公有云服务商负责
- 安全具有挑战性

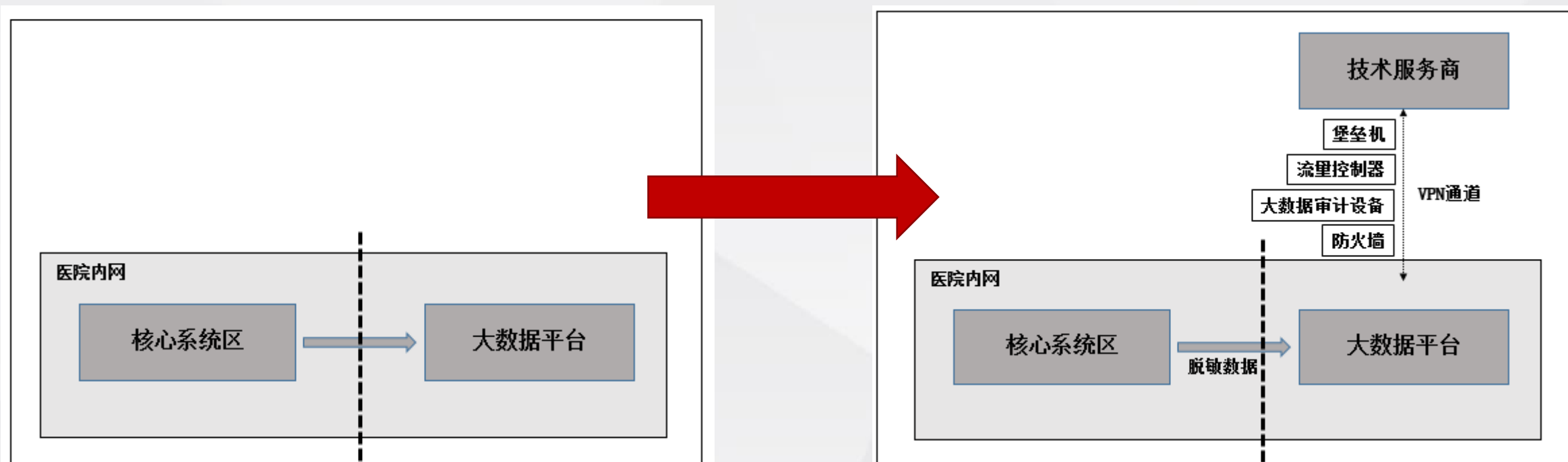
2、院内部署模式

- 安全保障措施可控
- 数据交互方便，不受网络、设备等限制

本指南介绍的医疗大数据平台建设以院内部署为核心内容讲解，同时考虑到当前平台建设后技术服务商维护特点，在部署模式中加入VPN维护模式部署。

二、指南内容解读

医疗大数据平台部署模式



基于院内网部署的医疗大数据平台所增加的远程维护模式VPN管理应由医疗机构把握

数据平台数据接入范围

数据接入范围

以患者为中心，以满足医疗大数据平台为临床、科研、管理、教学、患者服务为目的来确定医疗大数据平台数据接入范围。原则上包括：

- 以HIS为中心的就诊、医嘱、费用等
- 以EMR为中心的门诊病历、住院病历、护理相关等
- 以LIS为中心的检验数据
- 以PACS为中心的检查系统数据，如：超声、心电、病理、PECT、内镜等。
- 其它系统：患者移动应用、随访数据

医疗大数据平台数据接入范围是为普通医疗需求服务，考虑基因数据、样本数据、组学数据等特殊性的，所以指南数据接入范围未纳入上述数据（可作为扩展数据分析利用）。

二、指南内容解读

数据平台数据接入方式

| 接入方式 | 需要业务系统配合程度 | 数据实时性支持程度 | 跨平台兼容性 | 业务系统变更后带来影响 |
|-----------|------------|-----------|--------|-------------|
| 备份恢复 | 低 | 低 | 低 | 低 |
| 物化视图 | 低 | 中 | 低 | 高 |
| ETL抽取工具 | 高 | 中 | 中 | 低 |
| 集成平台 | 高 | 高 | 中 | 中 |
| 数据同步(OGG) | 低 | 高 | 高 | 低 |

指南基于医疗大数据平台数据接入方式未做特殊要求，平台建设机构可以依据本单位实际情况及上述接入方式优缺点等选择合适的接入方式

关于数据脱敏与加密

为什么要做数据脱敏与加密

基于数据安全的要求，同时由于大数据平台集成了医院所有医疗信息系统患者信息，数据体量大，敏感性高、数据泄漏后带来的风险比较高，同时大数据平台科研多中心研究等服务需要。

数据脱敏和加密应该怎么做

医疗大数据平台的数据本质是为临床所用，所以在做数据脱敏与加密时既要考虑安全要求，还要考虑实际应用需求。即在数据处理的过程中实现数据脱敏与加密，在数据应用过程中根据特殊要求实现数据反脱敏与加密。

数据脱敏与加密方案

在没有明确的法规参考前建议借鉴美国HIPAA法案，对患者部分敏感数据进行脱敏处理

数据脱敏与加密主体

医疗机构技术人员

数据处理

医疗大数据平台核心就是数据处理，数据处理实现了临床不可使用数据到可利用数据转换。采购现成平台开展医疗大数据平台的数据处理过程一般都是由技术服务商完成，在这个环节应关注以下问题：

1、提交处理前的数据验收

无论采用何种数据接入方式，数据在第一次到达平台时都应该做好数据验收工作，即通过接入方式到达平台的数据是否有缺失，内容是否正确等。

2、数据处理过程应遵循的标准

数据处理过程是一个分词、归一、标准化的过程，这个过程涉及标准化内容应严格遵循国家相关标准。

数据验证

数据验证指医疗大数据平台处理、生产后的数据与原业务系统数据的一致性、完整性和正确性。数据验收应由医疗机构信息部门完成，即核对平台处理后数据是否有误差、有缺失，建议以处理后数据对照原数据进行详细核对。

数据验证方法：

- 1、数据报告：初期平台建设方提供；
- 2、数据质量抽样检查：抽样比照HIS、EMR等业务系统单个患者数据核验；
- 3、病案数据抽样检查：随机抽取一定量打印病历，平台前台核对数据；
- 4、隐私数据检查：查看是否按要求去隐私化处理。

平台验收

平台验收是信息部门为代表的收货确认，是平台交付临床、科研和管理等科室客户使用的前提，主要包括功能验收、速度和性能验收。

1、速度和性能

- 1.响应时间：秒级查询；
- 2.系统容量：满足实际工作和未来五年业务需求；
- 3.系统并发用户数：满足未来5年以上发展需求；
- 4.安全性验收：VPN断开测试、数据监控和审计测试。

2、功能验收

- 1.高效搜索、相似疑难病例搜索；
- 2.患者全息视图、全量数据时间轴；
- 3.统计分析、可视化、多维分析；
- 4.权限、导出管理等。

数据授权与管理

医疗大数据平台建设部署完成是以平台授权开展使用为结束的，平台授权作为指南一部分重点关注以下内容：

- 1、**人员授权**：单独权限授权、HIS系统集成授权；
- 2、**数据授权**：根据用户身份、使用数据用途分配可查阅数据范围；
- 3、**日志记录**：用户使用情况追踪、可回溯；
- 4、**操作审计**：导出、隐私数据查询等敏感操作审批流程设置、操作审计监控。

二、指南内容解读

医疗大数据管理制度管理样例

1、依据：

《信息安全登记保护管理办法》、
《人口健康信息管理办法（试行）》、
《人类遗传资源管理暂行办法》、
《中华人民共和国网络安全法》。

2、审批：

数据的申请与审批、申请资格和范围、
数据权限分类、授权原则、使用期限、
授权记录等。

附录 B 医疗大数据管理制度示例

为更好地运用本书来指导医院医疗大数据平台的建设，正确、合理地利用医疗大数据，发挥医疗大数据平台建设的作用，本书收集、整理了部分医疗大数据平台数据利用、管理的相关制度，供读者参考。

B.1 医院 A 大数据管理制度

关于使用临床数据资源开展科研工作的管理办法

第一章 总 则

第一条 为规范医院临床数据资源管理，推动临床科学研究工作的开展，根据《信息安全等级保护管理办法》（公通字〔2007〕43号）、《人口健康信息管理办法（试行）》（国卫规划发〔2014〕24号）、《人类遗传资源管理暂行办法》（国办发〔1998〕36号）、《中华人民共和国网络安全法》等法律法规，制定本办法。

第二条 临床数据资源是指医院临床信息系统所存储的涉及患者健康相关的数据、资料，即患者在医院就诊过程中记录的所有数据，包括门诊、住院、急诊、医技等系统采集的数据，并经患者唯一索引进行关联后产生的高质量数据。数据范围包括就诊信息、诊断信息、医嘱信息、用药信息、诊疗信息、会诊信息、护理信息、转科/转床信息、手术麻醉信息、操作信息、输血信息、检查信息、检验信息、放化疗记录、病历信息等。

第三条 本办法适用于使用临床数据资源用于开展科学研究相关工作，如开展临床课题研究、撰写研究论文等。医院相关职能科室根据主管部门要求上报医院数据的工作，不在本办法的适用范围内。

第四条 医院信息部门和病案室均可检索到临床数据资源。用于科学研究目的的临床数据资源限定由医院信息部门提供，病案室原则上不提供科研

用途的数据查询服务。

第二章 数据的申请与审批

第五条 申请资格和范围

(1) 数据使用申请人应为医院人力资源管理系统中正式登记的本院的在职工，如研究生申请使用，应由导师作为申请人发起申请；如进修医生申请撰写论文使用，则应由论文通讯作者发起申请，并经所在科室负责人审批通过。

(2) 原则上数据所涉及患者的就诊或住院科室应与申请人所在科室一致；如涉及其他科室的患者数据，需要获得相关的科室或医院管理部门的同意。

第六条 数据权限分类

(1) 经过脱敏处理后的临床数据查询与浏览（初级权限），初级权限可以供申请人查询患者的科室分布情况，为进一步的数据检索和导出提供支持。申请人获得初级权限后，可以使用基于大数据平台的临床科研平台进行患者数据检索，检索到的信息不包括患者的个人隐私基本信息。

(2) 授权范围内的临床数据导出（高级权限）：申请人提交高级权限申请材料，获批准后，交由信息部门导出申请项目的详细信息。

第七条 初级权限的申请

申请人需通过医院科研数据管理系统、OA等主管部门认可的其他方式申请临床科研平台账户。由医院信息部门建立使用者账号，账户对应权限的配置也根据具体数据使用需求的申请，经过申请人所在科室主任或医院管理部门批准后，予以配置。

第八条 高级权限申请

(1) 所需材料

除需要使用初级权限的申请流程外，高级权限申请还需要提供以下材料：研究方案；伦理批件；数据使用申请表（见附件）；其他相关材料。如果与其他单位合作使用，应提供双方合作协议。

(2) 审批流程

- ① 申请人准备以上所需材料后提出使用申请。
- ② 申请经申请人所在科室主任同意。
- ③ 申请人将所有材料交医院科研部门审核。
- ④ 科研部门通知涉及的其他科室主任审核。

二、指南内容解读

医疗大数据管理制度管理样例

3、权利和义务：

只能将临床数据资源用于数据申请的范围，申请人对数据负有保密责任。

4、知识产权：

平等互利、诚实信用、共同参与、共享成果，明确各方应享有的权利和承担的义务。

5、惩罚：

对于不规范使用数据的申请人，限制或拒绝提供数据申请服务。

⑤ 申请人将科研部门审核通过后的数据申请表和研究方案提交医院信息部门审核和备案。

⑥ 医院信息部门为申请人配置账户权限。

第九条 授权原则

(1) 初级权限查阅到的数据均经过脱敏处理。

(2) 初级查询权限有效期默认为“长期”，如人员调离，需通知信息部门停用账号。

(3) 高级权限有效期为“天”，批准权限到达截至期限后，系统自动终止其权限。

第十条 记录

医院信息部门根据基于大数据平台的临床科研平台的日志可回溯查看账号的查询或导出记录。

第三章 使用人的权利和义务

第十一条 凡需使用平台数据者，必须严格遵守国家的有关法律、法规和本办法的相关规定。

第十二条 使用人只能将临床数据资源用于数据申请的范围，不得用于其他用途，如更改数据用途，必须重新申请。

第十三条 使用人在数据使用结束后，应向医院科研部门提交数据使用报告备案。

第十四条 使用人对于申请获得的数据负有保密责任。

第十五条 数据的使用人在任何场合公开数据使用结果或发表署名文章、专著和论文时，须注明数据来源。

第十六条 数据的使用安全受法律保护，任何部门、机构或个人不得利用数据从事危害国家安全、社会公共利益和他人合法权益的活动。

第四章 知识产权

第十七条 对于单独使用医院的临床数据资源开展科学研究的，申请课题应以医院为第一承担单位，发表论文应以医院为第一作者单位。

第十八条 对于多家单位利用共同数据资源进行整合研究的，其知识产权按下列原则处理：

(1) 合作申请课题，课题承担单位和人员排序由双方通过协议约定。

(2) 合作发表论文，论文作者排序由双方根据贡献大小协商确定。

(3) 合作研究产生的其他科技成果，其使用权、转让权和利益分享办法由双方通过合作协议约定。协议没有约定好的，双方都有使用的权利，但向第三方转让须经双方同意，所获利益按双方的贡献大小分享。

第十九条 有关人类遗传资源的国际合作项目应当遵循平等互利、诚实信用、共同参与、共享成果的原则，明确各方应享有的权利和承担的义务，充分、有效地保护知识产权。

第五章 惩罚

第二十条 违反本办法规定，有下列行为之一者，将其列入医院临床数据资源限制使用名单，限制或拒绝提供数据申请服务。

(1) 擅自更改数据用途。

(2) 擅自传播或转让数据。

(3) 公开数据使用结果或者发表署名文章、专著和论文时未声明数据来源。

(4) 歪曲使用数据结果。

(5) 使用后未向科研部门备案。

第二十一条 对违反以上条款者，医院有权追究其法律责任。

第六章 附 则

第二十二条 本办法由医院信息部门和科研部门负责解释。

第二十三条 本办法自颁布之日起施行。

二、指南内容解读

搜索技术

结构化归一

数据挖掘

机器学习

分布式计算

知识推理

知识图谱

应用场景

临床应用：大数据搜索、MDT、患者全息视图和临床决策支持

科研应用：科研思路探索与发现、基于时间模型的科研分析和专病库

管理应用：精细化管理、病案管理、病历评分体系、VTE风险评估、ICD辅助编码

患者服务：智能导诊、候诊

药物研究：受试者智能招募、RBM质控核查、AE/SAE自动报警、试验数据辅助采集

教学应用：疾病图谱、临床数据与知识库关联应用

二、指南内容解读

应用场景-大数据全量搜索

病历搜索 > 高级搜索

患者维度 某患者的全部就诊的合集满足搜索条件, 该患者符合搜索条件的就诊能被搜索到

就诊维度 单次就诊满足搜索条件, 则该次就诊能被搜索到

编辑关系

| 逻辑关系 | 搜索主题 | 搜索条件 | 值域范围 |
|------|------|------|--------|
| | 诊断名称 | 包含 | 胃恶性肿瘤 |
| | 诊断类型 | 等于 | 出院诊断 |
| AND | 出院科室 | 包含 | 消化肿瘤病房 |
| AND | 年龄 | > | 60 |

筛出患者 2,392 清空条件 搜索

搜索结果 ☆收藏 导入到项目

搜索结果 统计分析 [病历相关统计数据转移到这里啦](#)

您的权限内, 相关病人2392个, 病历67110份。 [查看权限](#)

*** 男 出生年月 1953-**-** 病人标识 *** 相关病历 8份 (住院: 3份) [患者全景](#)

住院日期 2012-07-20 08:43:26 就诊医院 北京大学肿瘤医院(本院) 住院科室 消化肿瘤病房 出院科室: 消化肿瘤病房 住院天数 3天 出院科室 消化肿瘤病房

住院日期 2012-06-13 09:05:49 就诊医院 北京大学肿瘤医院(本院) 住院科室 消化肿瘤病房 出院科室: 消化肿瘤病房 住院天数 8天 出院科室 消化肿瘤病房

住院日期 2012-07-11 09:27:24 就诊医院 北京大学肿瘤医院(本院) 住院科室 消化肿瘤病房 出院科室: 消化肿瘤病房 住院天数 1天 出院科室 消化肿瘤病房

纳入标准 3 排除标准 1

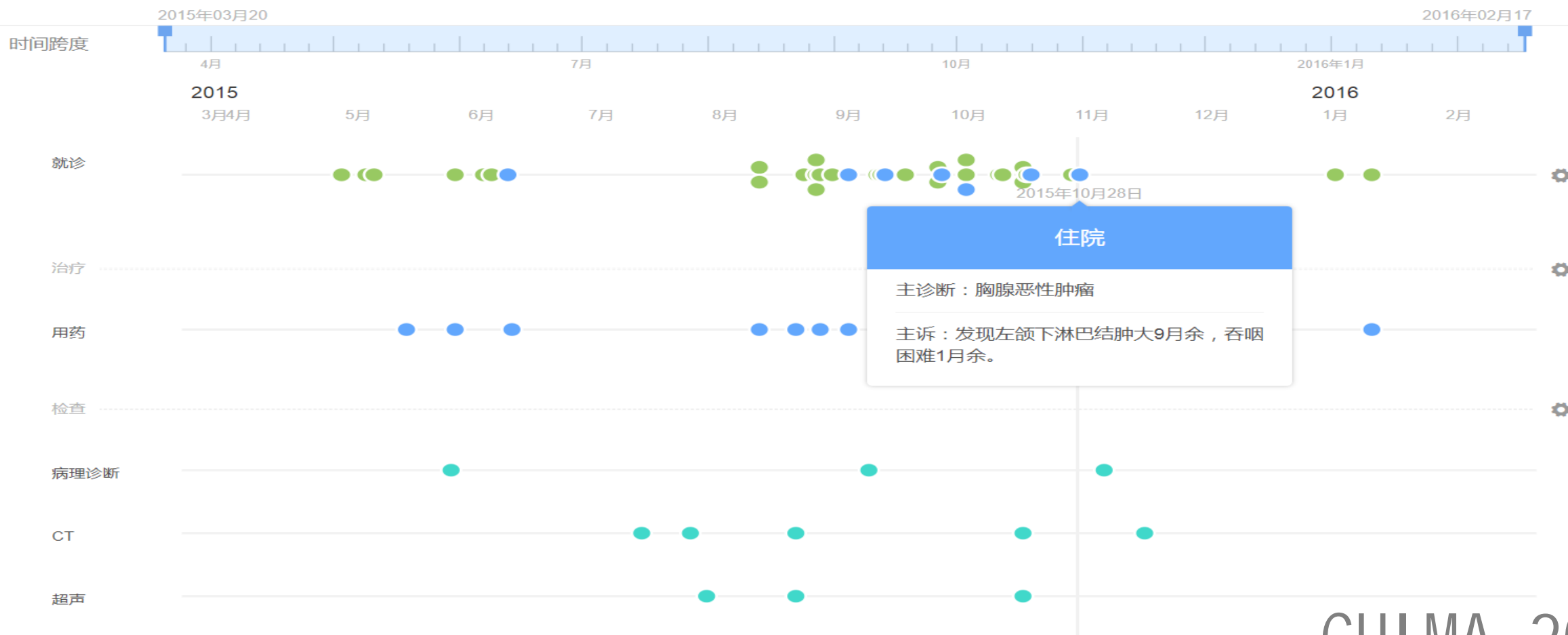
```
graph LR; A[+] --- B[并且]; A --- C[或者]; B --- D[手术名称 包含 乳房病损切除术]; B --- E[药品名称 包含 紫杉醇注射液]; C --- F[药品名称 包含 阿帕替尼];
```

筛出患者 185 清空条件 搜索

二、指南内容解读

应用场景-患者全息视图

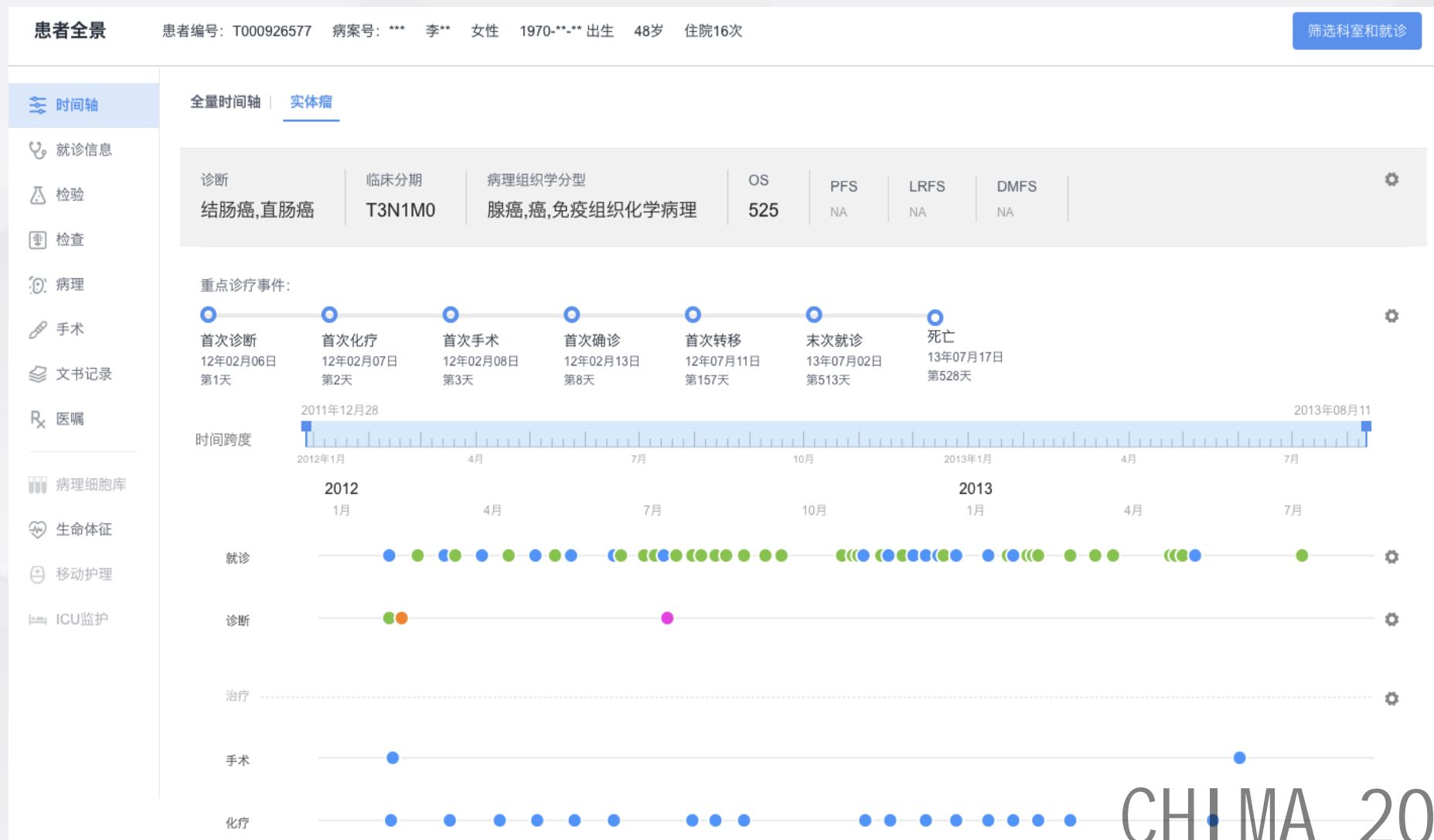
全量时间轴 | 实体瘤



二、指南内容解读

应用场景-全息视图/治疗时间轴

- 以**诊疗事件**、**时间**为主线，提供单个患者各种重要事件的时间节点信息。
- 设定关注指标功能，实现特定关注指标追踪



二、指南内容解读

应用场景-基于大数据平台MDT



张** 患者全景

年龄
54岁

就诊科室
胃肠中心一病区

目前诊断
1、食管胃结合部癌cT3N1
2、阑尾术后
3、糖尿病

会诊目的
下一步治疗方案

主诉

诊断胃癌4月，术前化疗3周期后

入院记录-现病史

2017年12月 即4个月前无明显诱因出现后背痛，间断性发作，无放射痛，不伴有腹痛腹胀、恶心呕吐、反酸烧心、腹泻便不适。曾口服耐信、谷氨酰胺片治疗，症状略有好转。

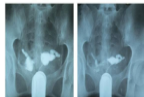
2018-01-08 胃镜：距门齿39cm食管下段、食管胃结合部大弯隆起型肿物，阻塞管腔，齿状结构消失，SIEWER

2018-01-09 超声胃镜：距门齿39-43cm隆起病变，累及腹段食管、胃底及贲门，齿线受累，分期为uT3N1。

2018-01-09 病理：中低分化腺癌，Lauren分型：肠型。免疫组化结果显示:HER2(2+)。

2018-01-04 CT：胃底贲门部胃壁增厚，考虑Bormann 3型胃癌（cT3）。胃周淋巴结，转移不排除。胸部CT：显示占位征象。颈部超声：双侧锁骨上低回声淋巴结，观察。

2018-01-16 全麻行腹腔镜探查术，探查阴性。



2018-01-18 行第1周期新辅助化疗，方案为SOX，具体用药剂量为奥沙利铂250mg 静滴 D1+替吉奥 140mg，分片晚4片，D1-14。后回当地医院行SOX方案2周期，末次化疗结束2018-03-16。治疗后评价：-复查EUS(本院)26)：食管胃结合部病变，病变呈低回声，侵及1-3层，局部与第4层关系密切，考虑受累可能；病变截面最大厚度约1.2cm，截面最大截面积181平方毫米。胃周小弯侧单发小淋巴结，直径约0.9cm。

2018-3-25 复查腹部增强CT：胃底贲门部胃壁增厚较前减轻，原较厚处约17mm (IM65)，现约10mm (IM10) 与腹段食管及正常胃壁分界不清，浆膜面较清晰。病变后方与胰腺脂肪间隙清晰。No.3组淋巴结稍缩小，现约7x5mm (IM69)，原约10x8mm (IM69)。腹膜后小淋巴结较前稍饱满 (IM94)。双侧肾上腺未见肿大淋巴结。未见腹水征象。考虑：胃底贲门部胃壁增厚较前减轻。胃周淋巴结稍缩小；腹膜后淋巴结稍饱满。

2018-3-19 复查胸部CT：胸部CT平扫未见明显占位征象。

S0001111 bjccancer_xxb_28
讨论组：2018-04-01 14:00 消化系统mdt讨论 | 创建科室：胃肠外科

病历摘要

基本信息

患者姓名 周** 性别 女 年龄 46

目前诊断 肺恶性肿瘤,慢性浅表性胃炎

Q 查找文献

会诊目的 后续治疗方案

病情介绍

B I U 14pt

主诉

发现血糖升高24年，血糖控制不佳伴双下肢酸胀3月。

病理

右肾下极外侧梭形细胞软组织肉瘤 (13x12.5x10cm)，免疫组化结果诊断纤维肉瘤

病历统计

相似病历

住院

入院诊断：肺结核

相似度：★★★★★

主诉：咳嗽、咳痰1月余，加重伴发热、喘憋5天

于 2012-07-07 08:14:00 就诊于呼吸内科 | *** 男性 78岁

病历详情 患者详情

住院

入院诊断：肺结核

相似度：★★★★★

主诉：咳嗽、咳痰2周，伴发热5天

于 2012-12-19 21:25:00 就诊于呼吸内科 | *** 女性 73岁

病历详情 患者详情

住院

出院诊断：肺结核

相似度：★★★★★

主诉：咳嗽、咳痰伴盗汗1月余

于 2014-10-27 14:41:00 就诊于呼吸内科 | *** 男性 46岁

病历详情 患者详情

住院

出院诊断：支气管扩张伴感染,肺间质纤维化,陈旧肺结核,冠状动脉粥样硬化性心脏...

相似度：★★★★★

主诉：咳嗽咳痰3年，伴发热1天

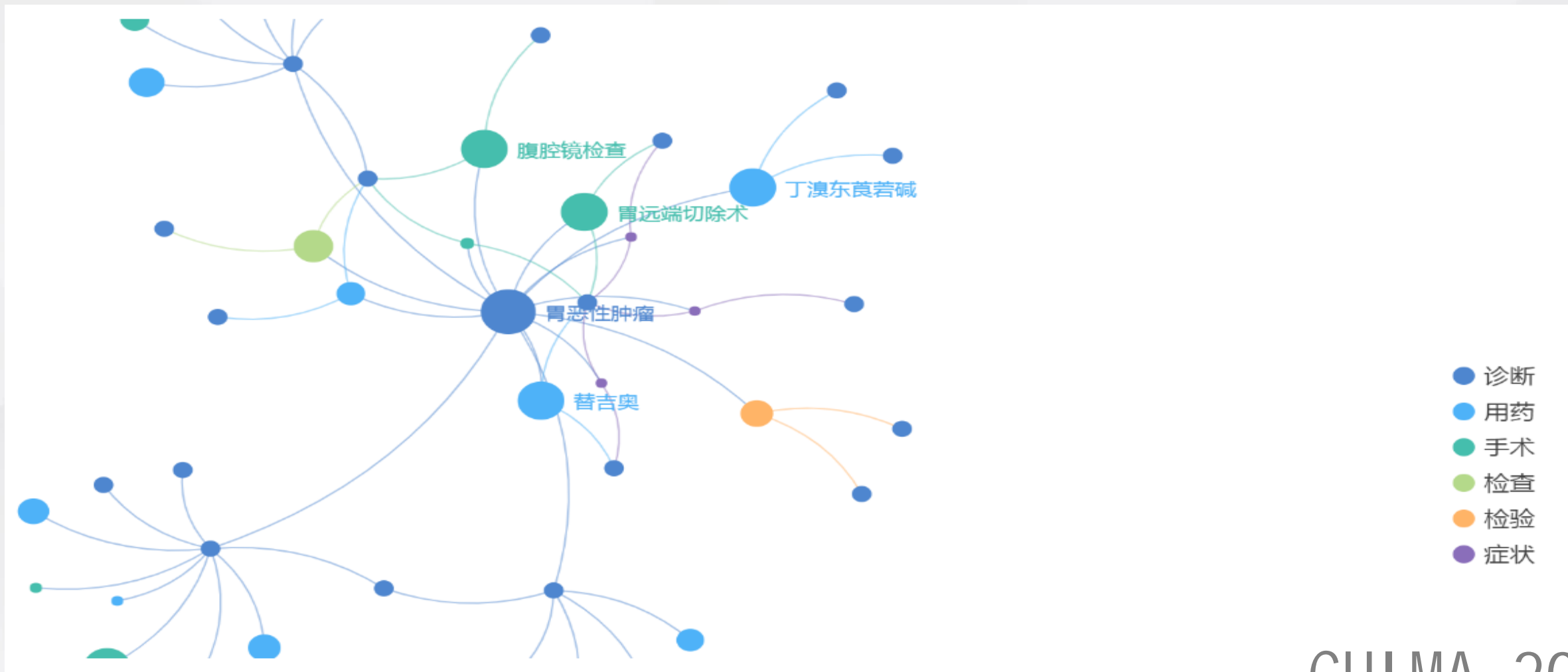
于 2013-04-30 23:20:00 就诊于呼吸内科 | *** 男性 80岁

病历详情 患者详情

退出全屏

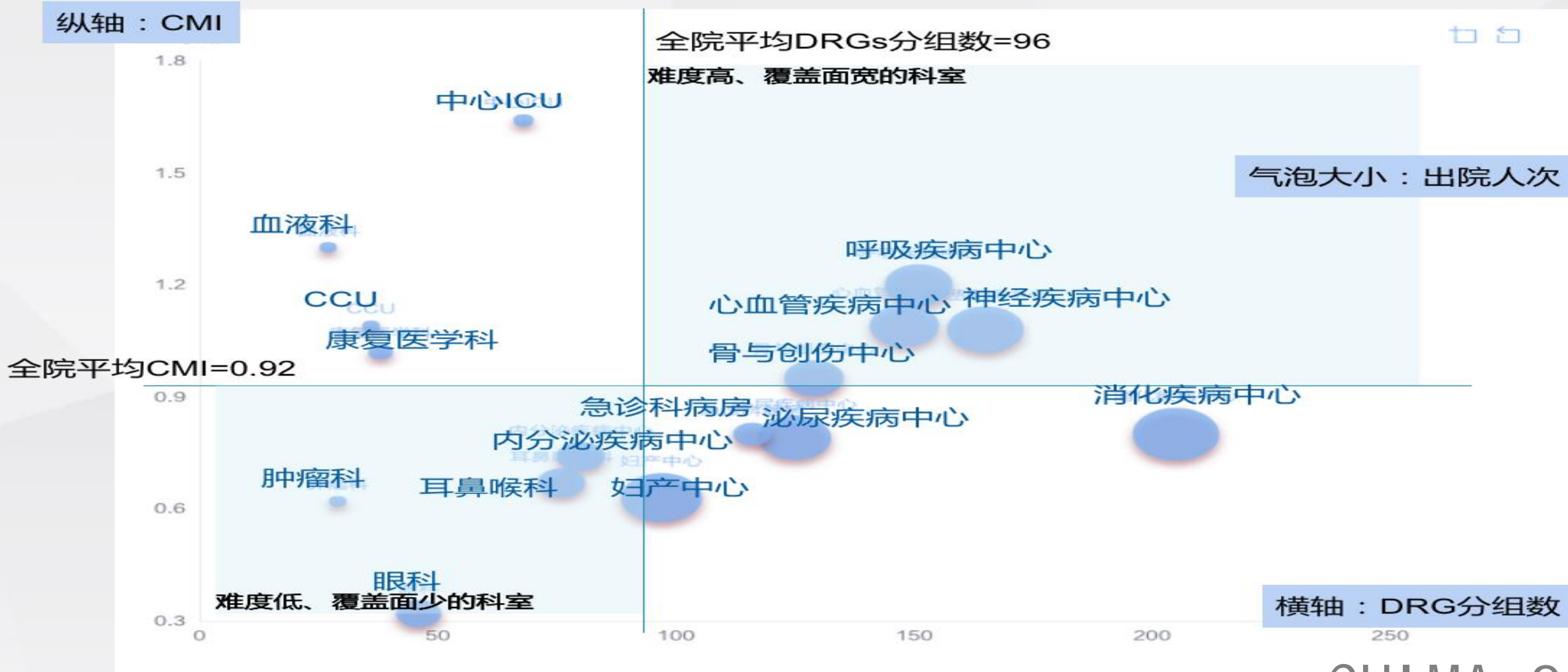
二、指南内容解读

应用场景-疾病图谱



二、指南内容解读

应用场景-医院管理



二、指南内容解读

- 医疗大数据平台如何在医院立项?
- 医院上了临床数据中心, 还需要上大数据平台吗?
- 如何判断大数据平台供应商, 需要更关注哪些方面? 原有做临床科室科研项目的厂商是否能选择?
- 为提高数据质量, 而修改业务系统, 如何操作?
- 大数据技术依靠医院自我学习还是和厂商合作? 如何保证持续性和自主性?



- 是否能让临床科室直接从大数据平台导出数据? 如果科研项目需要患者隐私信息, 有什么好的方法?
- 数据申请和使用的管理, 您认为主要由医院哪个职能科室管理?
- 大数据平台管理的难点有哪些? 为什么?

- 大数据平台如何在院内推广使用?
- 如何进行大数据平台效益评价, 大数据平台成果不显著, 可能的原因是什么?

二、指南内容解读

附录B 医疗大数据管理制度示例

附录B 医疗大数据管理制度示例

为更好地运用本书来指导医院医疗大数据平台的建设，正确、合理地利用医疗大数据，发挥医疗大数据平台建设的作用，本书收集、整理了部分医疗大数据平台数据利用、管理的相关制度，供读者参考。

B.1 医院A 大数据管理制度

关于使用临床数据资源开展科研工作的管理办法

第一章 总 则

第一条 为规范医院临床数据资源管理，推动临床科学研究工作的开展，根据《信息安全等级保护管理办法》（公通字〔2007〕43号）、《人口健康信息管理办法（试行）》（国卫规划发〔2014〕24号）、《人类遗传资源管理暂行办法》（国办发〔1998〕36号）、《中华人民共和国网络安全法》等法律法规，制定本办法。

第二条 临床数据资源是指医院临床信息系统所存储的涉及患者健康相关的数据、资料，即患者在医院就诊过程中记录的所有数据，包括门诊、住院、急诊、医技等系统采集的数据，并经患者唯一主索引进行关联后产生的高质量数据。数据范围包括就诊信息、诊断信息、医嘱信息、用药信息、诊疗信息、会诊信息、护理信息、转科/转床信息、手术麻醉信息、操作信息、用血信息、检查信息、检验信息、放疗记录、病历信息等。

第三条 本办法适用于使用临床数据资源用于开展科学研究相关工作，如开展临床课题研究、撰写研究论文等。医院相关职能科室根据主管部门要求上报医院数据的工作，不在本办法的适用范围内。

第四条 医院信息部门和病案室均可检索到临床数据资源。用于科学研究目的的临床数据资源限定由医院信息部门提供，病案室原则上不提供科研

附件 使用临床数据资源开展科研工作申请表

| | | |
|---|--|----------|
| 申请单编号（管理部门填写）： | | |
| 申请人姓名 | 工号 | |
| 科室 | 联系方式 | |
| 申请材料（请打“√”） | | |
| <input type="checkbox"/> 研究方案 | <input type="checkbox"/> 任务批件（指国家及省部级课题） | |
| <input type="checkbox"/> 伦理批件 | <input type="checkbox"/> 中国人类遗传资源管理办公室许可批件 | |
| <input type="checkbox"/> 其他材料 | | |
| 相关课题是否已经在科研部门备案 | <input type="checkbox"/> 是 | 备案课题名称： |
| | | 备案课题负责人： |
| | | 备案课题编号： |
| | | 备案课题来源： |
| | <input type="checkbox"/> 否 | |
| 申请数据范围（包括申请使用患者的疾病类型、涉及科室、起止时间等） | | |
| 疾病类型 | | |
| 涉及科室： 1. 2. 3. 4. | | |
| 时间范围： | | |
| 其他说明： | | |
| 申请人承诺： | | |
| 我将严格按照《××医院关于使用临床数据资源开展科研工作的暂行管理办法》相关规定执行。本人承诺绝不擅自以直接、间接、口头或书面等形式向其他任何个人、科室和单位泄露所获得的医院临床数据。若本人未按医院规定执行，所造成的一切不良后果，均由本人承担。 | | |
| 签字： 年 月 日 | | |
| 申请人所在科室主任意见： | | |
| 科室主任签字： 年 月 日 | | |

附件1 ××医院数据查询统计申请表

| | | | |
|--------------------|-------|------|--|
| | | 编号： | |
| 申请人姓名 | | 所在科室 | |
| 联系电话 | | 申请日期 | |
| 申请事由 | | | |
| 查询内容 | | | |
| 所在科室领导意见 | 年 月 日 | | |
| 机关领导意见 | 年 月 日 | | |
| 信息部门意见 | 年 月 日 | | |
| 相关工程师签字（请标注结果反馈日期） | 年 月 日 | | |

附件2 ××医院信息系统敏感数据查询统计申请表

| | | | |
|-------------------|------------------------------|------------------------------|---------------------------------|
| 制表： 编号： | | | |
| 申请人姓名 | | 申请科室 | |
| 联系电话 | | 申请日期 | |
| 数据类别 | <input type="checkbox"/> 药品类 | <input type="checkbox"/> 耗材类 | <input type="checkbox"/> 患者个人信息 |
| 申请事由 | | | |
| 查询内容 | | | |
| 申请科室领导意见 | 年 月 日 | | |
| 主管处室领导意见 | 年 月 日 | | |
| 主管院长意见 | 年 月 日 | | |
| 信息部门意见 | 年 月 日 | | |
| 信息工程师查询签字（标注反馈日期） | 年 月 日 | | |

敏感数据定义：涉及统计医师个人或临床科室有关药品、高值耗材（和卫计委特殊管控的低值耗材）信息；涉及患者姓名、身份证、联系方式、家庭住址等个人信息。

二、指南内容解读

附录C 医疗大数据平台建设案例



附录D 术语定义

1. EMPI (Enterprise Master Patient Index)

患者主索引。20 世纪末出现的医疗信息化专业用语，简单来说，它是患者基本信息检索目录。

2. EMOI (Enterprise Master Object Index)

企业主对象索引。患者医疗就诊记录无法准确关联到某次就诊行为时，通过大数据技术手段，找到时间上最合适的就诊行为，挂载到具体的一次就诊上（门诊/住院）。

3. MDT (Multi-Disciplinary Treatment)

多学科会诊。由多学科资深专家以共同讨论的方式，为患者制定个性化诊疗方案的过程，尤其适用于肿瘤、肾衰、心衰等复杂疾病的诊疗。

4. CRF (Case Report Form)

病例报告表。在临床试验中用以记录每名受试者在试验过程中的症状、体征或实验室检查数据的文件。

5. SMO (Site Management Organization)

临床机构管理组织。

6. CRO (Contract Research Organization)

合同研究组织。20 世纪 80 年代初起源于美国，它是通过合同形式为制药企业、医疗机构、中小医药医疗器械研发企业甚至各种政府基金等机构在基础医学和临床医学研发过程中提供专业化服务的一种学术性或商业性的科学机构。

7. CRA (Cooperative Research Association)

临床检查员。由申办者任命并对申办者负责的具备相关知识的人员，其任务是监察和报告试验的进行情况和核实数据。

8. HIPAA (Health Insurance Portability and Accountability Act)

联邦健康保险携带与问责法案。

9. PHI (Protected Health Information)

受保护的健康信息。

20. Hbase

一个分布式的、面向列的开源数据库。

21. Hive

基于 Hadoop 的一个数据仓库工具，可将结构化的数据文件映射为一张数据库表，并提供简单的 SQL 查询功能，可将 SQL 语句转换为 MapReduce 任务进行运行。

22. Sqoop

一个开源的工具，主要在 Hadoop (Hive) 与传统数据库之间传递数据，可将一个关系型数据库(如 MySQL、Oracle、Postgres 等)中的数据导入 Hadoop 的 HDFS，也可将 HDFS 的数据导入关系型数据库。

23. Mahout

Apache Software Foundation (ASF) 旗下的一个开源项目，提供一些可扩展机器学习领域的经典算法的实现，旨在帮助开发人员更加方便快捷地创建智能应用程序。

24. ZooKeeper

一个开放源码的分布式应用程序协调服务，是 Google 的 Chubby 的一个开源实现，是 Hadoop 和 Hbase 的重要组件。

25. Ambari

一种基于 Web 的工具，支持 Apache Hadoop 集群的供应、管理和监控。

26. Spark

专为大规模数据处理而设计的、快速通用的计算引擎。

27. Tez

Apache 开源的支持 DAG 作业的计算框架。

28. ES (Elasticsearch)

一个基于 Lucene 的搜索服务器，它提供一个分布式多用户能力的全文搜索引擎。

29. VPN (Virtual Private Network)

虚拟专用网络。属于远程访问技术，在公用网络上建立专用网络，进行加密通信。

30. IPS (Intrusion Prevention System)

入侵防护系统。计算机网络安全设施，是能够监视网络或网络设备的网络资料传输行为的计算机网络安全设备，能够即时中断、调整或隔离一些不

二、关于指南征求意见稿



CHIMA《医疗大数据平台建设指南》（征求意见稿）发布会

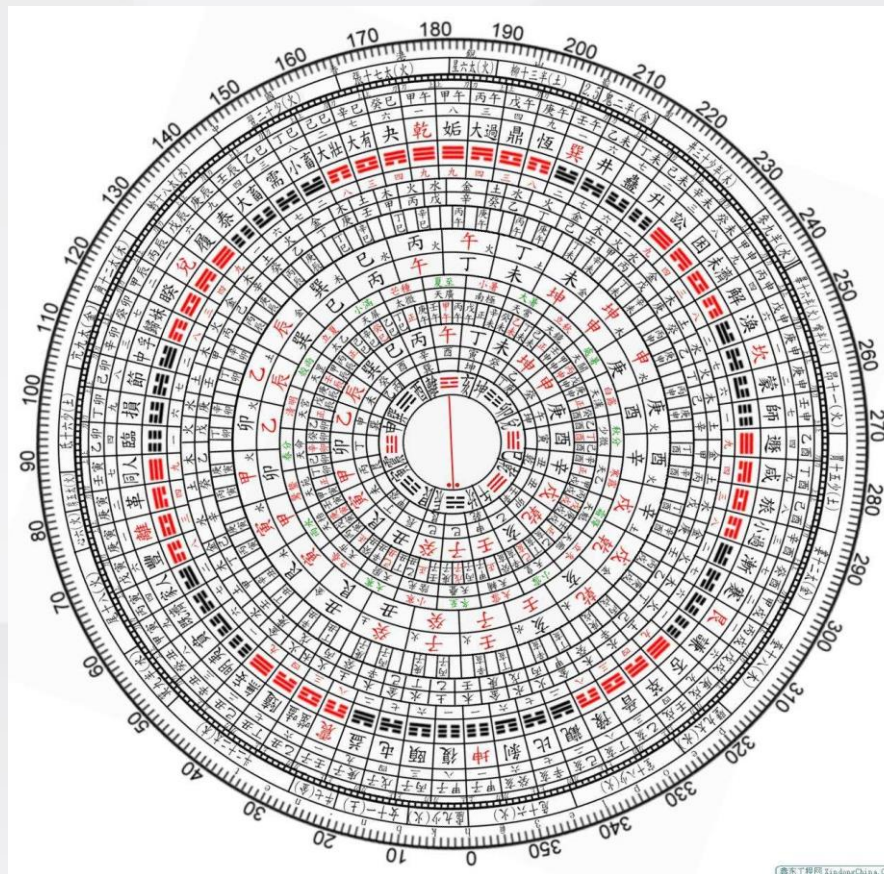


对关心指南、提出宝贵意见和建议的各位专家在此表示衷心感谢！

2019年5月6日，在中国医院协会信息专业委员会主持下，《医疗机构医疗大数据平台建设指南》（征求意见稿）召开了发布会

发布会后广大医疗信息同行对《医疗大数据平台建设指南》提出了许多宝贵意见和建议，同时也提出了很多尖锐性问题，本指南特在后记部分进行了回答说明。

中国医院协会信息专业委员会计划于明年发布《医疗机构医疗大数据平台建设指南》
第二版



THANKS

谢 谢 聆 听

欢迎行业同道继续支持和完善