

# 网格技术在医学研究中的应用实例—— caBig 简介

张 燕, 李 鸷 (北京大学临床肿瘤学院, 北京 100036)

An Example of caBig: Application of Grid in Medical Research ZHANG Yan, LI Zhi

摘 要: 全文介绍美国国立卫生研究院(NIH)通向医学生物学未来研究路线图计划的试点项目**癌症生物医学信息网(caBig)**的主要特点、目标和研究内容。NCI认为caBig的出现将会改变癌症研究进行的传统方式,全面推进恶性肿瘤医学领域各方面的研究。

关键词: caBig; 网格; 互联网; 肿瘤

中图分类号: R730.1; TP393 文献标识码: C 文章编号: 1004-0242(2008)05-0354-02

2007年9月,美国国立癌症研究所(NCI)副所长、**癌症生物医学信息网(the Cancer Biomedical Informatics Grid)**,以下简称**caBig**的创始人 Kenneth Buetow教授等在北京大学临床肿瘤学院,“国际医学信息学在临床肿瘤药物试验中的应用”研讨会上,详细介绍了NCI的caBig的现状与目标。**caBig在我国的应用将使广大医务人员更加高效快捷地共享世界各国先进的肿瘤治疗信息和资源,全面提升临床科研和治疗的信息化水平**。我们认为其应用和发展应该引起我国医学和IT专业人员的高度重视。

## 1 网格的现状和特点

**caBig是2003年美国国立卫生研究院(NIH)制定的通向医学生物学未来研究路线图计划(NIH Roadmap)的一个试点项目**。NCI先后还启动了其它一系列研究项目,包括癌症基因组解剖工程(CGAP)、癌症分子工程(AMPAP)、临床试验数据库、癌症模型数据库、癌症实验室信息管理系统(caLINS)等。caBig是一个开放源码,开放式接入,自愿性的信息网络,使癌症研究者根据商定的共同标准和需求,共享工具、标准、数据、应用和技术,它建立一个基础信息框架,连接癌症医学研究团队的每一部分,形成一个协作网络,或称网格。

网格(Grid)是利用高速国际互联网或专用网络把地球上广泛分布的计算资源、存储资源、通信资

收稿日期: 2007-10-12

源、网络资源、软件资源、数据资源、信息资源、知识资源等连成一个逻辑整体,采用开放标准,实现网络虚拟环境上的资源共享和协同工作,最终实现用户在网格这个虚拟组织环境上进行资源共享和协同工作,消除信息孤岛和资源孤岛。

我们以一家新建工厂进行的可行性分析为例。企业将自身的历史数据交付共享,财务顾问公司将自己的财务分析/预测系统交付共享,决策咨询公司也将自己的假设性分析系统交付共享。如此一来,参加可行性分析的各方形成了一个虚拟组织,所有资源由各方共享,大家可在全球不同的地点共同工作,最后完成任务。按照这种理念,虚拟组织需要的是一个非常富有弹性的共享关系。共享的时间、共享的权限、资源的数量都会不断发生变化。

## 2 caBig的目标与期望

近年来,癌症研究方法和技术的飞速进展,在基因和蛋白质的研究、临床试验和其他众多研究领域带来了知识和信息的“爆炸”<sup>[1]</sup>。不过从海量信息中获取多少利益要受到几个因素的制约。具体地说,没有一个为个体研究者和研究机构提供方便的共享数据的公用机制,并且没有一个给癌症研究者使用的统一的基础或通用标准,致使他们经常在得不到外界帮助的条件下孤立地工作,不能方便地共享数据、工具和他人开发的新技术。

作为一个试验项目,caBig的具体目标是: 有

着不同需求和能力的各种癌症中心是可以加入同一个网络或网格,共享数据、应用和技术;展示与NCI合作的癌症中心是如何获得开发新的工具和系统的能力;各癌症中心可以积极利用caBig使其研究工作实现更大的价值;建立一个不断扩展的、涵盖全球的癌症研究节点的通用构架。

像caBig这样规模宏大、雄心勃勃的生物信息工程还没有先例,NCI将其设想为“癌症研究的万维网”;遍及世界各地的研究者将具有一个公共的caBig平台,能够使用通用工具、迅速地转换、连接和分析各种不同的来源的数据。癌症研究组织中的成员也能根据他们的需要和兴趣为caBig做出积极的贡献。

caBig的领导人Kenneth Buetow说:“我们的目标是创建一个能**加速发现过程的网格**,它能将癌症研究者团队、数据和工具连接为一体”。迄今为止,NCI及其指定的癌症中心已形成了最早的一批caBig的合作开发者,确保caBig能直接满足癌症研究界的需要。作为一个开放源码,开放式过程接入,其他成员的癌症研究机构以及私营机构任何时候都可以自由参加caBig。自2003年7月caBig试点启动以来,已有超过50个NCI指定的癌症中心参与了caBig的版本、方法论和结构的开发。参与的癌症中心对于如何依据现有的创新工具和可利用的数据集来测试caBig的可行性提出了设想,他们现正在积极地进行试点作业。随着试点阶段的进展,参与的癌症中心将继续主动指导着caBig的开发、评价和改进。开发者的最终希望是,caBig能够成为一个由国家的和国际的生物医学研究领域组成的的大型共同体,它们共同承认开放的和共享的生物信息学工具、标准、结构和数据的重要性。如果这一希望实现,将对所有生物医学研究和医疗保健团体产生重大影响。

caBig将影响整个癌症研究团体,包括研究癌症起源和机制的基础科学机构,研究预防、早期筛查和治疗的机构<sup>[2]</sup>,以及从事以病人为研究对象新的诊断或治疗方法的临床试验机构。**利用caBig,基础研究的科学家将能更好地整合与自己实验室或世界各地其它研究室不相同的数据表格。**caBig增强每一个中心的试验者的工作力度和范围,从而缩小试验范围,**更迅速地得出病人能否得到更佳结局的推论。**通过直接满足每一个研究人员和机构对于易用的、需要费用的一致生物信息学工具、标准和统一的架构,促进生物医学研究,改善病人结局。**caBig将帮助癌症研究者们汲取大量的现存数据,形成新的合作。**利用caBig,个体研究者能更迅速和有效地回答问题,在各个方面加速癌症研究的进程。因此NCI认为,

在其实现自己的挑战目标——“到2015年消除癌症的痛苦和死亡”的过程中,caBig将起重要作用。

NCI在caBig中的作用是提供协调监督和对公共的、私人的和学术机构的合作者的工作协调,但caBig的定义、开发者仍然是癌症研究团体,最终将涉及更多的国家和国际团体。**caBig保证多中心共享数据的安全和人类研究中的隐私保护。**

### 3 已经发布的 caBig 的产品

以下是caBig已经发布的产品,可以从caBig的Web网页<sup>[3]</sup>获取:CTMS——Clinical Trials Management System,临床试验管理系统(C3D C3PR 临床试验软件兼容性评价 报表工具 中间件等);I-CR——Integrative Cancer Research,一体化的癌症研究(基因、蛋白质、ca序列、临床路径等);TBPT——Tissue Banks and Pathology Tools,组织库和病理学工具(同盟组织数据集白皮书、调查和结果的详细需求书、caties等);ARCH——architecture,体系结构(Cabig兼容性指导原则、cagird 1.0、查询语言白皮书、用户管理白皮书、caBig安全白皮书);VCDE——Vocabularies and Common Data Elements,词汇和公用数据元素;SP——Strategic Planning,战略规划(caBig战略规划等);DSIC——Data Sharing and Intellectual Capital(数据共享、出版原则、安全管理推荐等);TRAINING——培训(cacore课程、中间件等)。

### 4 caBig 的未来

按照Kenneth Buetow的介绍,caBig未来的框架包含下几个模块:C3PR——Cancer Central Clinical Participant Registry(参与者登记制度),caAERS——Cancer Adverse Event Reporting System(癌症不良事件报告系统),PSC——Patient Study Calendar(病人研究日历),CTOM——Clinical Trials Object Model,C3D——Cancer Central Clinical Database(癌症临床中心数据库)。

### 参考文献:

- [1] [http://bahcabigcontracts.com/\[EB/OL\].](http://bahcabigcontracts.com/[EB/OL].) 2007-09-24.
- [2] [http://www.cancer.gov/newscenter/pressreleases/caBIGQandA\[DB/OL\].](http://www.cancer.gov/newscenter/pressreleases/caBIGQandA[DB/OL].) 2007-09-12.
- [3] [https://cabig.nci.nih.gov/\[DB/OL\].](https://cabig.nci.nih.gov/[DB/OL].) 2007-09-10/2007-09-24.