

协同肿瘤研究临床试验信息系统 SOA 建模

孙明瑞 臧天仪 Nadir K.Salih

(哈尔滨工业大学计算机科学与技术学院 哈尔滨 150001)

摘 要 开放标准的肿瘤信息系统在不断发展, 以此应对现代肿瘤临床试验带来的挑战。本文提出面向服务的软件范例, 用于衍生临床试验信息管理系统, 支持多机构协同肿瘤研究。提出的方案将临床试验(元)模型和基于WSRF(Web Services Resource Framework)的面向服务体系结构(Service-Oriented Architecture, SOA)相结合, 并且应用早期临床试验进行评估。尽管主要目的是针对肿瘤研究, 但也适用于具有相似信息模型的其它领域。

关键词 面向服务的体系结构; Web服务; 元模型; Web服务资源框架

SOA Modeling of Clinical Trial Information for Collaborative Cancer Research

SUN Ming-rui ZANG Tian-yi Nadir K.Salih

(School of Computer Science and Engineering, Harbin Institute of Technology, Harbin 150001)

Abstract The open-standards cancer informatics is being developed to address the challenges posed by modern cancer clinical trials. In this paper, a service-oriented software paradigm is presented to derive clinical trial information management systems for collaborative cancer research across multiple institutions. Our proposal is founded on a combination of a clinical trial (Meta) model and WSRF (Web Services Resource Framework)-based SOA (Service-Oriented Architecture), and is evaluated for use in early phase trials. Although primarily targeted at cancer research, our approach is readily applicable to other areas for which a similar information model is available.

Keywords SOA; web services; meta-model; WSRF

1 引 言

面向服务技术、元数据驱动技术用于解决肿瘤信息学带来的挑战, 已经成为发达国家研究重点^[1]。通用的肿瘤临床试验元模型^[2], 可控、可重用词汇术语和元数据元素^[3], 支持软件系统自动生成、数据共享以及互操作。

面向服务的体系架构(Service Oriented Architecture, SOA)是由Gartner在1996年提出的。它是一种模块化、分布式和松耦合风格的软件体系结构。SOA风格的应用程序使用可重用的跨应用程序和跨企业边界的业务构件。这些构件通过基于定义明确的接口来调用, 独立于底层的硬件和软件平台以及开

发语言^[4]。Web服务资源框架(Web Services Resource Framework, WSRF)在某种程度上是由Globus联盟推动的^[5], 它是一种使用Web服务对持久资源进行建模和访问的开放式架构^[6]。WSRF的技术规范依赖当前的Web服务标准, 通过Web服务资源的概念以及状态管理规范的定义, 解决了普通Web服务固有的无状态性限制。

本文对临床试验数据的建模提出了一种基于WSRF的面向服务软件范例。根据临床试验元模型的语义, 提出了一种基于WSRF的面向服务体系结构, 在协同肿瘤研究中使用WSRF对临床试验数据收集、管理和分析进行建模。余下部分组织如下: 第2部分是综述临床试验模型; 第3部分提出隐含的试验数据模型语义; 第4部分描述如果用WSRF对临床试验数据建模; 第5部分给出案例研究, 举例说明两个肿瘤初级临床试验系

统的使用，在第6部分进行总结。

2 临床试验元模型

临床试验元模型的理念是在试验的设计阶段，确保包括临床试验报告在内的所有明确指定信息有效的最好方式。临床试验元模型反映的理念如图1所示^[2]。

试验设计群组集合的一般部分包括试验的名称、概要、注册码、以及资金的来源和联系方式。获取的

项目中，模型的方法和随机化部分包括对每个患者群体的干预（即治疗），患者的资格标准，数据收集的设置和位置以及随机分配的技术规范。模型的部分定义病例报告表(Case Report Forms, CRFs)在临床试验的每个阶段支持患者的 workflow。通过临床试验执行期间匹配所有收集到的数据，使元模型能够在临床试验之间数据共享。定义明确的通用数据元素(Common Data Elements, CDEs)^[3]，即肿瘤概念和测量值的控制集。

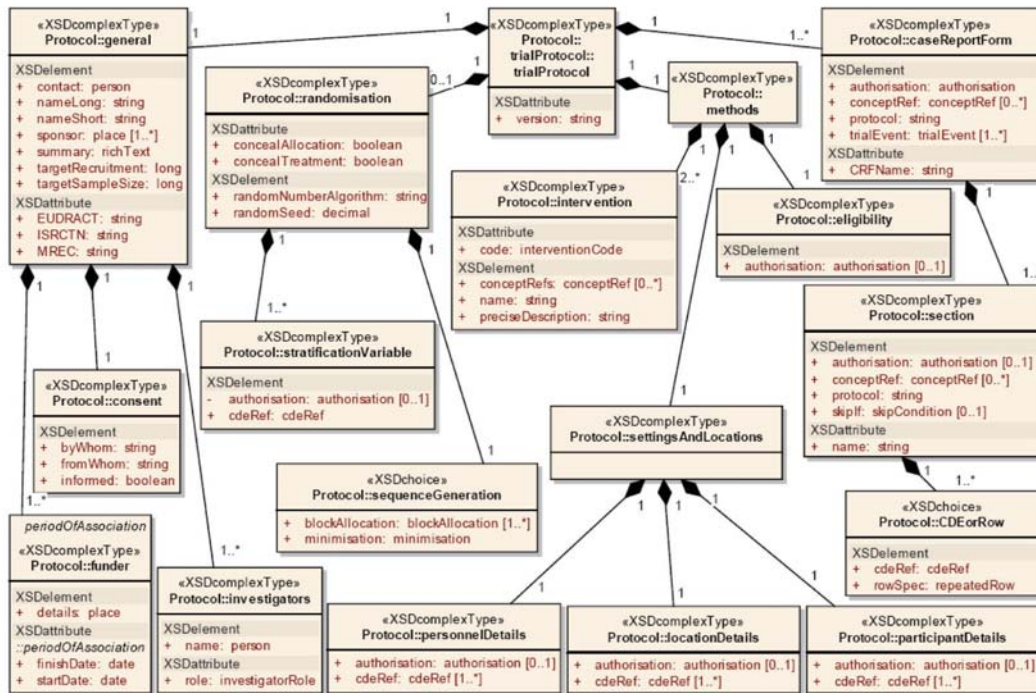


图1 临床试验元模型的高阶类图

3 隐含的试验数据模型语义

虽然UML类图是一个强大的设计工具，但是临床试验元模型只覆盖了临床试验数据建模所需要的一部分。因此，临床试验系统的设计和开发同样依赖肿瘤临床试验语义的隐性知识。

3.1 通用数据元素

根据一些基本类型定义CDE:

(1) CdeID, CDE标识符集，用于指定唯一的CDEs;

(2) CdeType, CDE值具有的类型集。考虑到不同信息系统之间数据的移动，支持语法的互操作性;

(3) CdeInfo, 完整定义CDE语义的元数据。考虑到不同数据集被比作鉴定覆盖或相关的内容，它支持语义的互操作性。

这些基本类型使用Z符号，总结如下^[7]:

[CdeID, CdeType, CdeInfo]

CDE类型详细说明:

```

Cde
-----
id: CdeID
valueDomain: CdeType
info: CdeInfo

```

在特定的研究领域中，CDEs用于对维持在这个研究领域中CDE（元数据）知识库的数据进行建模:

```

CdeRepository
-----
cdeSet: P Cde
-----
∀x, y. cdeSet • x.id = y.id ⇒ x = y

```

3.2 病例报告表

临床试验数据作为若干试验事件的结果，在试验的执行期间产生，每个结果都相当于临床试验执行的

一个阶段。例如，在注册登记阶段收集临床和个人病例数据，在随机化阶段分配治疗，对定期收集的随访资料进行治疗效果的评估。临床试验模型中试验项目的完全集如下：

$$\begin{aligned} \text{TrialEvent} ::= & \text{registration} \mid \text{eligibility} \mid \\ & \text{randomization} \mid \text{onStudy} \mid \text{treatment} \mid \text{offStudy} \mid \\ & \text{response} \mid \text{followUp} \mid \text{adverseEvent} \end{aligned}$$

通过填写从肿瘤CDE知识库中提取的包括CDEs在内的病例报告表，临床医师收集与试验事件相一致的数据。

| *cancerCdeRep*. *CdeRepository*

CRF通过一系列与它相一致的试验事件来完整的定义：

$$\begin{array}{|l} \text{CaseReportForm} \\ \text{events: seq TrialEvent} \end{array}$$

3.3 设置及地点

元模型的设置和地点构件指定了收集的包括试验地点（如医院或临床试验中心）、医护人员和患者的数据。这是由元数据知识库提取的所有CDEs组合：

$$\begin{array}{|l} \text{SettingsAndLocations} \\ \text{locationCdeSet: P cancerCdeRep.cdeSet} \\ \text{personnelCdeSet: P cancerCdeRep.cdeSet} \\ \text{patientCdeSet: P cancerCdeRep.cdeSet} \end{array}$$

3.4 试验设计

为了基于WSRF的试验系统开发的目的，临床试验由Settings and Locations定义，所有患者的CRFs，以及与试验地点、医护人员、患者和每个患者表单的项目相一致的CDEs所构成：

$$\begin{array}{|l} \text{TrialDesign} \\ \text{forms: P CaseReportForm} \\ \text{eventCdeSet: TrialEvent} \rightarrow \text{P cancerCdeRep.cdeSet} \\ \text{SettingsAndLocations} \\ \text{dom eventCdeSet} = \cup \{f.\text{forms} \bullet (\text{ran } f.\text{events})\} \\ \forall f_1, f_2. \text{forms} \bullet f_1 \neq f_2 \Rightarrow \text{ran } f_1.\text{events} \cap \text{ran } f_2.\text{events} = \emptyset \end{array}$$

针对临床试验的地点、医护人员和患者实例，定义为：

$$\begin{array}{|l} \text{LocationInstance} \\ \text{id: N} \\ \text{locationDetails: P (Cde} \times \text{CdeType)} \end{array}$$

$$\begin{array}{|l} \text{PersonnelInstance} \\ \text{id: N} \\ \text{locationIds: P N} \\ \text{personnelDetails: P (Cde} \times \text{CdeType)} \end{array}$$

$$\begin{array}{|l} \text{PatientInstance} \\ \text{id: N} \\ \text{locationId: N} \\ \text{patientDetails: P (Cde} \times \text{CdeType)} \end{array}$$

临床试验中，地点、医护人员和患者实例之间的关系详细说明如下：

$$\begin{array}{|l} \text{trialLocations: TrialDesign} \rightarrow \text{P LocationInstance} \\ \text{trialPersonnel: TrialDesign} \rightarrow \text{P PersonnelInstance} \\ \text{trialPatients: TrialDesign} \rightarrow \text{P PatientInstance} \\ \text{trialPatientCrfs: TrialDesign} \times \text{PatientInstance} \rightarrow \\ \text{P CaseReportForm} \\ \forall t. \text{TrialDesign}; c. \text{Cde}; v. \text{CdeType}; \\ l. \text{trialLocations } t; p. \text{trialPersonnel } t; \\ pt. \text{trialPatients } t; l_1, l_2. \text{trialLocations } t \\ ((c, v) \in l.\text{locationDetails} \Rightarrow c \in t.\text{locationCdeSet}) \wedge \\ ((c, v) \in p.\text{personnelDetails} \Rightarrow c \in t.\text{personnelCdeSet}) \wedge \\ ((c, v) \in pt.\text{patientDetails} \Rightarrow c \in t.\text{patientCdeSet}) \wedge \\ (l_1.\text{id} = l_2.\text{id} \Rightarrow l_1 = l_2) \wedge p.\text{locationIds} \subseteq \{l.\text{id}\} \wedge \\ pt.\text{locationId} \in \{l.\text{id}\} \wedge \text{trialPatientCrfs}(t, pt) \subseteq t.\text{forms} \end{array}$$

3.5 授权

基于角色的访问控制，通过授权元素融入进指定敏感临床试验数据模型的每个部分，由临床试验模型强制执行。试验医护人员和患者所起的作用集为：

$$\begin{aligned} \text{Role} ::= & \text{patient} \mid \text{coordinator} \mid \text{clinician} \mid \\ & \text{research_nurse} \mid \text{statistician} \end{aligned}$$

合理的操作集对于特殊的角色是可用的：

$$\text{AccessType} ::= \text{creation} \mid \text{modification} \mid \text{querying} \mid \text{retrieval}$$

授权感知的试验设计定义了一种规则，指定了在每个角色上允许执行的CDEs操作：

$$\begin{array}{|l} \text{AuthorizedTrialDesign} \\ \text{trial: TrialDesign} \\ \text{accessRules: Role} \times \text{AccessType} \rightarrow \text{P Cde} \\ \forall \text{roleCdeSet: ran accessRules} \bullet \text{roleCdeSet} \\ \subseteq \cup \{(\cup \{e. \text{TrialEvent} \bullet (\text{trial}.\text{eventCdeSet } e)\}), \\ (\cup \{\text{trial}.\text{locationCdeSet}\}), (\cup \{\text{trial}.\text{patientCdeSet}\}), \\ (\cup \{\text{trial}.\text{personnelCdeSet}\})\} \end{array}$$

4 临床试验资源建模

4.1 面向虚拟组织的SOA模型

在上一节所描述的基于临床试验语义，对肿瘤研究虚拟组织进行的临床试验研究，提出了服从WSRF的SOA模型。如图2所示，分层的系统结构自然地映射到一个肿瘤研究虚拟组织中。

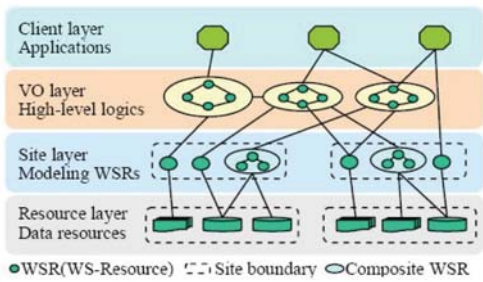


图2 面向虚拟组织的SOA模型

由下而上各层次为数据资源层（资源层），Web服务资源建模层（位置层）和高级肿瘤研究逻辑层（虚拟组织层）。在不同的契约和监管环境下，肿瘤研究特征多样化，在不同的体系中存储和管理重要数据资源。资源层由许多重要的基于临床和组织的、包含在肿瘤研究中的数据资源构成，比如临床试验CDE知识库，临床试验和试验数据模型仓库等。

Web服务资源建模层在每个位置都包含单一复合的Web服务资源。Web服务资源模拟潜在的数据资源状态，提供一个服从WSRF标准的接口与它们交互作用。一般来说，这些部署在各个站点的Web服务资源模型，用来管理和操纵特殊位置的信息和数据。

基于Web服务资源基础架构模型，使用Web服务组合和编排技术，在虚拟组织层中为协同肿瘤研究设计了高层的业务逻辑。此外，除了根据虚拟组织规则管理和维护成员资格，虚拟组织层通过利用基于Web服务资源的访问控制机制，为肿瘤研究实现了安全策略。虚拟组织层的业务逻辑执行的越多，留给客户层的工作量就越少。

4.2 临床试验Web服务资源

我们使用在WSRF规范中定义的Web服务接口来隐藏资源属性，实现临床试验Web服务资源。在CDE层和资源层，临床试验Web服务资源确保计算的互操作性以及语法和语义的互操作性。在Web服务资源属性中定义的6个端口类型^[8]，可以满足在3.5节中指定的访问试验数据的需求。每个端口类型都显示一个与端口类型名称相同的单一操作。使用这些接口，我们可以在运行时动态地检索、查询、更新、插入和删除Web服务资源的资源属性。详细说明如下：

- (1) GetResourcePropertyDocument 允许用户检索与Web服务资源相联系的所有资源属性值；
- (2) GetResourceProperty 允许用户通过它的限定名 (Qualified Name, QName)，包括命名空间和内部名的名称，来访问任何资源属性的值；
- (3) GetMultipleResourceProperties 允许用户

通过它们的限定名立刻访问一些资源属性的值；

(4) QueryResourceProperties 允许用户在资源属性文档中执行复杂的查询。XPath可作为查询语言；

(5) PutResourcePropertyDocument 允许用户使用新的资源属性文档，完整地取代Web服务资源的属性值；

(6) SetResourceProperties 允许用户在服务的资源属性中请求一个或几个修改。通过参数可以指定三种行为：更新、插入和删除。

4.3 临床试验资源共享

为了支持资源共享，依据访问类型，Web服务资源属性端口类型可以分为两种：读Web服务资源属性端口类型和写Web服务资源属性端口类型。前者提供了阅读权限的操作，后者提供了写权限的操作。因此我们为资源提供了两类Web服务资源：只读的Web服务资源为Web服务实现只读的Web服务资源属性，读写的Web服务资源对Web服务实现所有的Web服务资源属性，如图3所示。

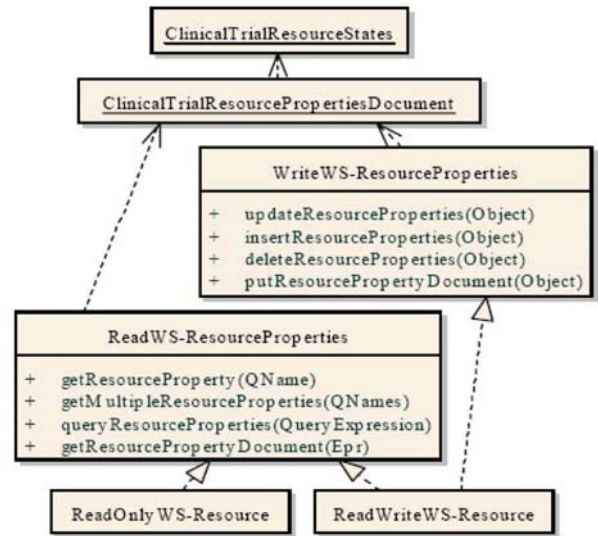


图3 只读和读写的Web服务资源

此外，我们还对数据共享和访问控制目的资源的部分资源属性，提供只读的和读写的Web服务资源。一个CRF由一系列的事件所组成。使用授权元素，表单中的每个事件都会覆盖表格层的针对于这个部分的访问规则。可以在每个部分以及整个表单实施只读的和读写的Web服务资源。

4.4 临床试验资源持久性

WSRF不指定一个范例来永久的存储资源属性。为了处理这类问题，在我们的Web服务资源中，实现了load()和store()操作。它们确保负载的资源属性从持续到永久储存。按需调用这两个方法，我们可以确信

内存中资源属性（资源状态）的值与磁盘中的值是同步的。

为了在网格环境中与远程数据资源进行交互，采用OGSA-DAI作为数据中间件^[9]。在跨地域的分布式数据资源中，OGSA-DAI协助我们处理数据的访问和整合，通过Web服务接口进行数据的查询、更新、转换和交付。

4.5 基于角色的Web服务资源组

在Web服务资源组规范的基础上，我们提出了基于角色的Web服务资源组来解决以上提出的问题。基于角色的Web服务资源组作为一个Web服务资源来实现，结合其它Web服务资源的端点引用一起来满足成员规则标准。Web服务资源的端点引用是动态地生成的，并且可以动态地发现和检查。对加入群组的角色所授权的访问权限（只读或读写），成员被限制只允许访问Web服务资源。作为一个示例，图4描述了协调者和临床医师的Web服务资源组。在某种程度上，基于角色的Web服务资源组影响了4.3节所描述的资源属性访问控制机制。

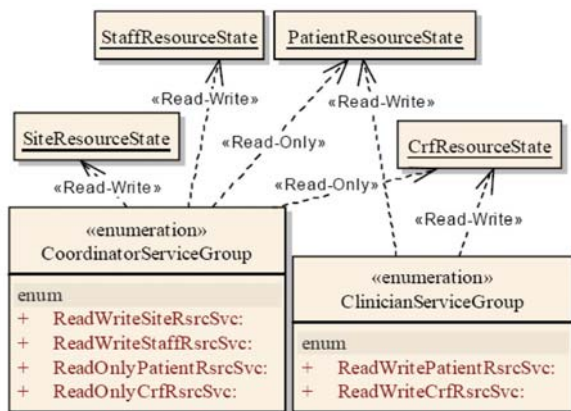


图4 协调者和临床医师Web服务资源组

作为一个Web服务，基于角色的Web服务资源群组也可以与Web服务安全和其它授权和身份认证技术一起，将系统的安全最大化。因此，基于角色的Web服务资源组不仅是简化Web服务资源集的发现和管理的自组织分类机制，同样也是能够有效地支持基于角色访问控制的基本安全机制。

5 案例研究

通过本文提出的基于WSRF的面向服务软件规范，开发实现了两个临床试验信息管理系统原型，基于Globus Toolkit 4^[10]（网格中间件）的Neat和tAnGo临床试验，用来评估提出范例的有效性。

Neat和tAnGo是真实的肿瘤临床试验^[11, 12]，在分析阶段就已经完成了数据的采集。我们能够基于人工的但是真实的患者数据进行试验的模拟执行，并对临床试验医护人员示范，使他们直接参与到不同肿瘤研究中心的操作试验中去。从试验的协调者、临床医师以及参与试验执行的IT人员那里所获得的反馈信息，用来改进临床试验数据模型的范例。

服从WSRF的临床试验服务的开发基于Neat和tAnGo模型，该模型是临床试验元模型的实例。对于这两个试验，实现了多种Web服务资源和不同的Web服务接口，对设置、地点、CRFs以及试验的试验数据建模。对模型中指定的角色，建立基于角色的Web服务资源组。

6 总结

本文描述的面向虚拟组织的SOA模型和建模临床试验数据方法，代表了基于WSRF的面向服务软件范例，用以开发支持协同肿瘤研究的临床试验管理系统。这表明了服从WSRF的Web服务如何用于支持肿瘤临床试验的语义。将我们的方案应用于真实的临床试验，证明了它的有效性。

开发的Web服务资源和Web服务资源组提供了一个临床试验Web服务资源基础设施，使语法、语义和计算的互操作性能够支持肿瘤研究的实施。其它的高级服务是我们未来要考虑的研究内容，如患者工作流管理系统和跨临床试验的数据分析服务。虽然建模的数据资源具有领域特殊性，但是也可以将我们的范例很容易地应用到其它领域，基于CDE的信息模型对临床试验的其它类型也是有效的。特别地，安全感知Web服务资源共享机制能够直接适用于其它的基于WSRF的SOA应用。

参考文献

- [1] CancerGrid Project [EB/OL]. 2012, <http://www.cancergrid.org>.
- [2] Harris S, Calinescu R. "CancerGrid clinical trials model 1.0," Oxford University Computing Laboratory, Oxford, UK, CancerGrid Tech. Rep. MRC/1.4.1.1 [EB/OL]. 2012, https://cancergrid.org/index.php?option=com_remository&Itemid=26&func=fileinfo&id=20.
- [3] Toujilov I V, Maccallum P. "Common data element management architecture," Oxford University Computing Laboratory, Oxford, UK, CancerGrid Tech. Rep. MRC-1.1.2 [EB/OL]. 2012,

- https://cancergrid.org/index.php?option=com_remository&Itemid=26&func=fileinfo&id=15.
- [4] Gartner. Service-Oriented Architecture (SOA) [EB/OL]. 2012, <http://www.gartner.com/it-glossary/service-oriented-architecture-soa/>.
- [5] Alliance G. The WS-Resource Framework [EB/OL]. 2004. <http://www.globus.org/wsrf>.
- [6] OASIS WSRF v1.2 standard [EB/OL]. http://www.Oasis-open.org/committees/tc_home.php?wg_abbrev=wsrfWeb.
- [7] Woodcock J, Davies J. Using Z: specification, refinement, and proof [J]. Prentice Hall. 1996, Vol. 39.
- [8] Ferraiolo D F, Kuhn D R, Chandramouli R. Role-Based Access Control [M]. Norwood: Artech House. Inc., 2003.
- [9] The University of Edinburgh. OGSA-DAI. [EB/OL]. 2012, <http://www.ogsadai.org.uk>.
- [10] Globus Toolkit 4 [EB/OL]. 2012, <http://www.globus.org/toolkit>
- [11] Poole C, Earl H. NEAT: National breast cancer study of epirubicin plus CMF versus classical CMF adjuvant therapy [EB/OL]. 2012, <http://public.ukcrn.org.uk/Search/StudyDetail.aspx?StudyID=643>.
- [12] Poole C, Howard H, Dunn J. tAnGo: A phase III randomised trial of gemcitabine in paclitaxelcontaining, epirubicin based adjuvant chemotherapy for women with early stage breast cancer, 2003.