

医疗应用集成中基于语义的信息交换与集成^{*}

张抗抗 李庆忠

(山东大学计算机科学与技术学院 济南250100)

摘要 异构系统的应用集成和信息集成瓶颈已经成为企业信息化进程的严重障碍。为实现有效的企业应用与信息集成,我们以医疗行业为应用背景,建立了应用集成的语义模型,该模型包括资源语义模型、计算语义模型以及二者之间的绑定三部分。基于此模型,提出了一个平台结构的应用集成方案,将应用系统发布为面向平台的服务,并在平台中建立对服务和领域资源的语义描述,实现服务请求和服务提供的基于语义的映射。这种方式使得应用系统可以以一种统一的方式通过平台与其他系统进行信息的交互,增强了系统的独立性、灵活性和可维护性。

关键词 应用集成,本体,Web Service,集成语义模型,描述逻辑

Semantic-Based Information Interchange and Integration in Healthcare Application Integration

ZHANG Kang-Kang LI Qing-Zhong

(School of Computer Science and Technology, Shandong University, Jinan 250100)

Abstract The application and information integration bottleneck of heterogeneous systems has been the great obstacle of the enterprise informatization process. In order to settle this problem, we build a semantic model of application integration based on the background of healthcare application. The model consists of three parts: semantic resource model, semantic computation model and their binding. Based on the semantic model, we introduce an platform-based application integration approach and in the platform in which all application systems are published as services facing to the platform and semantic description of services and domain resource are built too. With the help of semantic information, the binding of service requester and supplier is built. Our approach makes application systems interchange with each other in a uniform mode and enhances their independency, flexibility and maintainability.

Keywords Application integration, Ontology, Web service, Integration semantic model, Description logic

1 引言

社会的信息化进程使得各种应用系统之间的联系越来越紧密,需要进行相互之间信息的交互和共享。但是,不同的系统往往都是异构的,它们采用不同的实现平台,不同的数据结构和实现方式,因此,如何提供一个有效的机制,实现异构系统之间的应用集成和信息集成,已经成为当前信息化进程中一个急需解决的问题。

在医疗应用领域,这个问题尤其突出。由于行业的复杂性和特殊性,医疗应用领域的信息化进程相对滞后,即使建立了统一的信息管理系统(HIS),一些特殊的应用系统如检验信息系统(LIS)、影像检查系统(RIS、PACS)等仍然由不同的专业厂商开发实施。另外,随着医疗保险和电子商务的发展,医院信息系统还要与医保系统、供应链系统等外部系统建立连接。这些系统之间的信息交互瓶颈已经成为

医院信息系统建设和应用的严重障碍。

目前异构系统之间的信息交换和集成已经引起业界的高度重视,并成为企业应用集成(EAI)研究领域的一个热点课题,XML、Web Service、工作流以及中间件等许多技术已经在处理异构系统的集成问题上发挥了极大的作用。在医疗应用领域也制定了HL7等标准用来规范系统间的信息交互。对于医疗领域的应用集成,国外的学术界和产业界已经做了大量的工作,但由于国内没有相关的标准和规范,HL7等标准在中国的应用和实施还刚刚起步,因此国内众多医院信息系统厂商的产品的系统间交互的接口标准也不统一,各家采用自己的方式进行系统直连,使整个医院系统的灵活性和可扩展性大大降低。虽然有些厂商利用中间件或其它技术实现了简单的、有限的医疗集成,但对于诸如医疗应用集成的标准、技术实现方式、业务流程的配置和管理、对遗留系统的集成等关键问题,还缺乏深入的探讨和研

^{*} 本课题得到国家“863”高技术研究发展计划基金(2002AA4Z3240)资助。张抗抗 博士研究生,主要研究领域为企业应用集成与信息集成。李庆忠 博士,教授,博士生导师,主要研究领域为应用集成与信息集成,数据库,数据库技术。

究,至今也难以形成较为理想的解决方案。

目前应用集成的主要研究方向是为应用系统的交互引入良好定义的语义,这种方向也覆盖了当前应用集成所使用的两种主要技术 Web Service 和工作流。Web Service 的提出为解决异构系统的互操作和集成提供了良好的机制^[1,2],但传统的 Web Service 技术缺乏语义的良好支持,目前 Web Service 发展的趋势是将语义网络的研究成果引入到 Web Service 中,形成语义 Web Service,通过定义良好的语义信息,使得可以自动地发现、执行和组装 Web Service,实现互操作^[2,3]。为了实现跨系统和跨企业边界的业务流程集成,流程管理技术逐渐从单纯的工作流向业务流程管理(BPM)的方向发展,并且也与 Web Service 技术相结合,利用 Web Service 的独立于具体应用的松散耦合特性和语义 Web Service 提供的语义信息,实现业务流程的动态生成和执行,达到系统间流程集成的目的。

本文针对大型医院环境内各应用系统的集成,提出了一种平台结构的集成体系和框架,建立了一个应用集成语义模型,并着重讨论了其中的基于语义的信息交互和集成。基于这种模型,集成平台实现了各应用系统资源的统一表达,使得通过医疗应用集成平台连接的应用系统可以不用考虑所连接的系统的差异性,以一种统一的方式通过平台与其他系统进行信息的交互。

2 应用集成语义模型

有效的应用集成要求通信的系统双方明确地了解交互的信息的含义,即具有明确的语义。因此,建立应用集成语义模型是实现有效的应用集成的关键。语义模型建立的基础是本体^[4]技术,在此先给出本体的定义。

2.1 本体及其描述语言

定义1(本体) 是共享概念模型的明确的形式化规范说明,描述了一定领域的概念以及概念之间的关系。一个本体可以用四元组 $O(C, R, A, I)$ 表示,其中 C 表示类,也称为概念, R 表示类之间的关系, A 表示公理,是针对概念的永真断言, I 表示 C 的实例。

由于本体具有良好的概念层次结构和知识及语义表达能力,因此在知识的共享和重用方面发挥了重要的作用,并且随着语义网络的研究的深入,本体在面向语义的信息集成方面的应用正逐渐得到越来越多的研究者的重视。语义网络研究的一个主要目标是使可以通过内容而不仅仅是关键字访问网络资源成为可能。这方面的主要工作包括资源描述框架 RDF、Web 本体语言 OWL 以及 DAML+OIL 等,这些语言的主要理论基础是描述逻辑^[5]。描述逻辑

是一阶谓词逻辑的子集,是一种结合了逻辑和面向对象特征的形式化的知识表示工具。描述逻辑的基本元素包括类(也称为概念)和角色,概念是一组对象的集合,角色代表对象之间的二元关系。概念表达式和角色表达式分别由原子概念和原子角色通过构建符连接所得。

OWL 语言中的类和属性分别对应描述逻辑的概念和角色,并且 OWL 中也提供了丰富的构建符来创建类(概念)和属性(角色)表达式,这些构建符都可以在描述逻辑中找到它们的对应,当然,为了使用的方便,OWL 中包含了更多的构建符,但它们都可从构建符的完备集中导出。文[6]对 OWL 和描述逻辑作了比较深入的对比探讨。

将语义网络与 Web Service 结合起来的工作,即语义 Web Service^[3]产生了 DAML-s 以及 OWL-s^[7]等利用本体描述 Web Service 的标记语言,其中 DAML-s 建立在 DAML+OIL 的基础上,而 OWL-s 是 DAML-s 最新的版本,建立在 OWL 基础之上。语义 Web Services 工作组的目的是建立一个基于 OWL 的 Web Service 本体,以支持智能代理技术,实现语义网络上 Web Service 的自动化操作。

在 OWL-s 中,Web Service 通过本体来表示。在本体层次结构的最上层为 Service 类,该类的三个主要的属性为 presents、describedBy 和 supports,三个次高层的类 ServiceProfile、ServiceModel 和 ServiceGrounding 分别对应着三个属性。ServiceProfile 表示了“Web Service 做什么”,ServiceModel 表示了“Web Service 怎么做”,ServiceGrounding 表示了“如何访问 Web Service”,即“怎么用”。这样,从描述逻辑的角度看,定义了四个基本的概念 Service、ServiceProfile、ServiceModel 和 ServiceGrounding 和三个基本的角色 presents、describedBy 和 supports。

2.2 集成语义模型

应用集成语义模型即建立在本体及其描述语言 OWL 和 OWL-s 的基础之上。语义模型包括三个部分,资源语义模型、计算语义模型以及计算模型与资源模型的绑定。

资源语义模型定义的是应用领域范畴,是对业务流程的形式化的描述和表示。资源语义模型建立一个领域内共享的资源本体,对应用系统的流程信息、数据信息建模,是业务流程相关组件和过程控制以及所有的工作单元和数据的概念的集合。

计算语义模型定义的是数据操作,即对服务建模,建立了一个系统中发布的所有服务的本体,通过定义服务的结构和分类,将所有的服务用统一的模型组织起来,便于发现和调用。计算模型这些数据操作可以以任何应用程序或组件的形式出现,由于

Web Service 在解决异构系统集成中的巨大优势,本文中操作主要以 Web Service 的形式出现。

资源模型与计算模型的分离增强了系统的灵活性和扩展性,但因为资源模型所表示的数据是计算模型所表示的操作的工作对象,因此,必须建立有效的机制实现资源模型和计算模型的联系。资源语义模型与计算语义模型的绑定是将一个包含一组功能的计算模型和相应的资源模型相关联的过程,目的是为特定的资源发现并匹配最适合的操作,并为建立这种关联为用户提供指导。

下面分别给出集成语义模型及其组成部分的形式化定义。

定义2(集成语义模型) 集成语义模型 SM 表示为三元组 (RM, CM, B) , 其中 RM 表示资源语义模型, CM 表示计算语义模型, B 表示资源语义模型和计算语义模型的绑定。

定义3(资源语义模型) 面向应用领域的共享资源概念的本体, $RM = O(C, R, A, I)$, 其中 C 为应用领域概念集合。

定义4(计算语义模型) 系统中发布的服务的本体, $CM = O(C_s, R_s, A_s, I_s)$, 其中 C_s 为服务概念的集合, 每个服务表示为 $S(F, PL, IN, OUT)$, 四元组的子项分别表示服务的功能、处理逻辑、输入和输出, 其中 F, IN 和 OUT 均为资源语义模型本体中的概念。

定义5(绑定) 绑定 B 表示为 (S, f, C) , f 为映射机制, 在服务 S 和资源概念 C 之间建立映射。

3 医疗应用集成平台系统架构

本文提出一个解决大型综合性医院业务系统应用集成的解决方案, 称为医疗应用集成平台 (Healthcare Application Integration Platform, HAIP)。

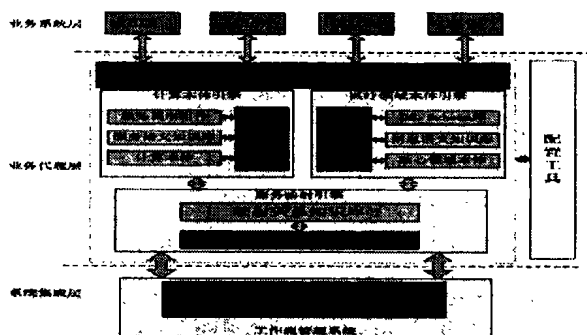


FIG. 1 医疗应用集成系统架构

图1 医疗应用集成系统架构

整个集成的医院业务系统架构包括三个层次, 如图1所示, 最外层为业务系统层, 中间为业务代理层, 最内层为系统集成层, 后两个层次构成医疗应用集成平台。业务系统层由实现医院业务的各个具体

的信息系统组成, 是应用集成的对象。系统集成层是整个应用集成平台的核心, 运行一个 workflow 管理系统, 主要包括两个部分: 一个 workflow 引擎和一个消息服务引擎, 主要进行业务流程的定义、解释和执行以及各互连系统间消息的传递。业务代理层是集成语义模型的具体实现, 应用系统通过它与集成核心连接。业务代理层的主要组成部分包括:

1. 业务系统接口封装: 将业务系统的对外接口封装为业务系统对外提供的服务以及向其他系统发出的服务请求;
2. 计算本体引擎: 负责对经过封装的业务系统服务进行描述和发布;
3. 医疗领域本体引擎: 接收和处理业务系统的服务请求, 并进行数据格式的转换和请求的转发;
4. 本体映射引擎: 进行应用系统间服务的请求与服务提供者之间的绑定(映射);
5. 配置工具: 此为平台管理工具, 可以进行平台基本参数及相关文档和组件的配置和定义。

这一部分也是本文讨论的重点, 将在第4节进行详细的讨论。有关业务流程的定义、执行和管理部分在本文中不作详细的论述。

4 基于语义的信息交换和信息集成

业务代理层是整合异构的业务系统, 使用统一的消息传递机制和业务流程进行应用集成的关键环节, 主要包括业务系统接口的封装、描述与发布、服务请求的处理和不同系统间消息格式的转换以及服务请求与服务提供之间的绑定(映射), 即建立应用集成平台的计算语义模型和资源语义模型, 并建立计算模型和资源模型之间的绑定。

4.1 业务系统接口的封装、描述与发布

由于 Web Service 的松散耦合特性和接口的统一性, 使用 Web Service 开发应用系统接口对于集成来说是一种理想的选择。但是, 由于一些客观因素的存在, 完全使用 Web Service 实现应用系统接口也是不现实的, 因此, 在 HAIP 中进行了一些扩展, 提供了包括 Web Service、COM 组件在内的几种接口方式。但由于 Web Service 技术在解决异构系统集成方面所表现出来的巨大的优越性, 这种方式在 HAIP 的设计和实现中占据了重要的位置。

对于 Web Service 方式, 首先在 HAIP 中提供了一个本地的 UDDI, 作为局域网络内 Web Service 的注册中心, 进行应用系统服务的注册和发布; 然后结合使用 OWL-s 和 WSDL 进行 Web Service 的描述^[8]。为了增强对 Web Service 语义的描述, OWL-s 以统一的计算机可操作的方式提供了一个核心的 Web Service 描述本体。在其三个基本组成部分中, ServiceProfile 和 ServiceModel 被认为是抽象的, 因

为它们没有提供 Web Service 的消息格式、协议和网络地址的细节指定,而 ServiceGrounding 提供了这些具体的细节。OWL-s 和 WSDL 的结合就在于 ServiceGrounding 可以通过 WSDL 来实现,在这一点上,它们是可以兼容的。每一个发布的服务 S 都可以看作是 Service 概念的一个实例,它可以简单地表示为如下三元组:

$S(P, M, G)$ P ServiceProfile, M ServiceModel, G ServiceGrounding

对于以其他方式提供的服务接口,如 COM 组件、CORBA 组件等,因为组件的接口元数据是以专用的二进制格式存储的,必须通过专用的应用程序接口(API)进行访问,因此,在 HAIP 中开发部署了一组组件,称为 AIC(Application Integration Component)组件,分别实现对不同的应用接口组件的调用。由于 WSDL 是可扩展的,它不仅限于描述基于 XML 和 SOAP 的服务,还可以用来描述使用其它格式的服务。因此,对于 AIC 组件,也通过扩展的 WSDL 进行统一的描述。所有服务的 WSDL 描述文档共同构成服务语义知识库。这样,从计算本体引擎向上可以提供一种统一的服务的注册与发布机制,向下则由计算本体处理器来统一进行不同方式接口的调用。

4.2 服务请求处理及消息格式转换

医疗领域本体引擎负责接收和处理来自服务请求者的服务请求,接收请求消息,并将消息转换为其目标系统所能识别的格式。在医疗应用领域,HIS 系统是医疗业务处理的核心,应用请求一般都由 HIS 系统发起,其他系统为其提供服务。在 HAIP 与各应用系统进行消息交互中,采用 XML 作为消息的标准格式,并且在平台中保存各类消息的 XML 文档的格式定义,使用 XML Schema 的形式保存。

XML 只是消息的实现形式,其内容的构成是由 HL7 标准^[9]定义的。由于 HL7 标准是针对西方特别是美国的医疗体制的特点而制订的,许多方面并不适合我国的实际情况,另外标准的本身还存在着许多不完善的地方,比如关于医疗保险方面的问题,HL7 中就没有涉及。因此,在实际应用中,对 HL7 标准进行了一些扩充和本地化工作,补充了一些原来标准中没有涉及到的业务内容的消息(目前主要是关于医保的),并且将所有涉及到系统间交互的消息类型使用 XML Schema 进行了格式定义,形成 XML Schema 库,保存在 HAIP 平台中,作为消息实例验证和语义描述的依据。

消息语义知识库中保存有对消息的 XML Schema 的语义描述,描述的信息包括消息的类型、消息的来源以及执行的功能等。当领域本体处理器接收到一条消息,可以根据消息语义知识库中对该

消息类型(Schema)的描述,查找到该消息所对应的语义信息。

在平台中还建立了一个医疗领域的核心领域本体,该本体建立了一个集成系统共享的概念词汇表,通过该本体,可以实现不同系统间消息的语义翻译。

4.3 服务绑定

服务绑定是指服务的请求与提供之间的映射,具体工作包括服务的发现、定位和调用。目前采用的方案是建立一个服务映射关系知识库,知识库中记录了消息类型与处理该类型消息的服务的映射,通过这种映射,实现操作和数据的绑定。

当集成平台接收到服务请求者发出的消息后,首先对该消息进行分析,确认消息的类型以及消息中包含的数据信息。根据消息类型,通过服务映射引擎定位服务,然后根据注册中心发布的关于服务的信息,将原消息的数据内容重新组织成服务可以识别的消息,并实现对服务的调用,将新生成的消息传递给服务,最后将服务的结果返回给请求者,这样就完成了一次系统间的服务请求交互。

结束语 本文论述了企业应用集成的语义模型,基于此模型,提出了一个医疗应用集成平台的总体系统架构,着重讨论了其中基于语义的信息交换和集成。基于这种总体结构,在山东省立医院信息化建设中,我们以 Microsoft 公司的 BizTalk Server 为支撑平台,实现了一个医疗应用集成平台,为解决医疗信息领域的应用集成提供了一个很好的解决方案。该系统的实际运行情况表明,这种方式使得应用系统可以以一种统一的方式通过平台与其他系统进行信息的交互,增强了系统的独立性、灵活性和可维护性。另外,虽然目前应用集成平台是依托医疗行业的应用背景设计开发的,但只要在平台中配置不同的业务逻辑,整个系统架构可以非常方便地应用到其它行业的企业应用集成。

目前实现的服务绑定方法是采用预先定义的映射关系知识库的形式,但是这种方式缺乏灵活性,无法实现智能的服务发现与映射,并且对处理涉及多个服务的复杂业务流程也缺乏足够的支持。因此,下一步的工作重点是充分利用已经建立的服务和资源的语义描述,建立语义推理机制,实现服务的动态的智能的发现和绑定。

另外,必须针对不同的应用系统接口开发相应的 AIC 组件,才能在平台中实现对服务的统一描述。因此,下一步的另外一个是探讨接口的封装机制,将不同的应用系统接口组件封装为 Web Service,从而可以简化和统一它们在集成平台中的注册和发布。从长远来看,这也是一种趋势。

参考文献

- 1 Hu Zaijun, Kruse E, Draws L. Intelligent binding in the engi-

- neering of automation systems using ontology and Web services. Systems, Man and Cybernetics, Part C, IEEE Transactions on, 2003, 33(3): 403~412
- 2 Tsai T-M, Yu Han-Kuan, Liao Ping-Yao, Shih H-T. Semantic modeling among Web services interfaces for services integration - SOTA (smart office task automation) platform. In: Proc. 14th Intl. Workshop on Database and Expert Systems Applications, Sept. 2003. 579~583
- 3 McIlraith S A, Son T C, Zeng Honglei. Semantic Web services. Intelligent Systems, IEEE: [see also IEEE Expert]. 2001, 16(2): 46~53
- 4 邓志鸿,唐世渭,张铭,杨冬青,陈捷. Ontology 研究综述. 北京大学学报(自然科学版), 2002, 38(5): 730~738

- 5 Calvanese D, Giacomo G D, Lenzerini M. Description logics: foundations for class-based knowledge representation. Logic in Computer Science. In: Proc. 17th Annual IEEE Symposium on, July 2002. 359~370
- 6 Zuo zhihong, Zhou mingtian. Web Ontology Language OWL and its description logic foundation. Parallel and Distributed Computing, Applications and Technologies. PDCAT'2003. In: Proc. of the Fourth Intl. Conf. on, Aug. 2003. 157~160
- 7 <http://www.daml.org/services/>
- 8 DAML-s Coalition working document. Describing Web Services using OWL-s and WSDL. Oct. 2003. <http://www.daml.org/services/owl-s/1.0/owl-s-wsdl.html>
- 9 <http://www.hl7.org/Library/standards.cfm>

(上接第70页)

门户是基于 Jetspeed 门户框架实现的。Jetspeed 框架以 Portlet 构件为核心内容。基于 Portlet 构件的

门户在页面定制的方面具有独特的优势。门户开发者和用户都可以使用定制页面功能。区别只是根据权限不同,定制的自由度有所不同。

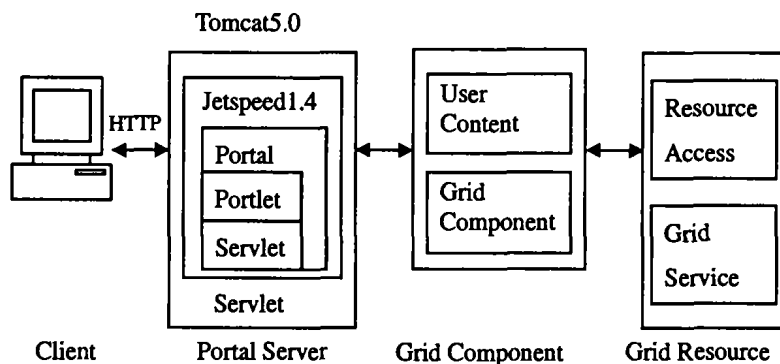


图6 服务网格门户的体系结构

在基于 Jetspeed 框架的门户中,提供三种定制功能:页面内容定制、页面布局定制、页面外观定制。门户开发者开发的所有 Portlet 和注册 Portlet 并控制 Portlet 显示模式的 PSML (Portal Structure Markup Language) 文件被事先部署在 Jetspeed 框架下,它们可以根据定制者的意愿来构成个性化的门户页面。图7为采用 Jetspeed 门户框架开发的服务网格门户的示例。用户登录到服务网格门户后,可以方便地选择门户提供的各项功能,根据用户意愿定制页面中显示的内容,添加或删除功能各异的 Portlet。

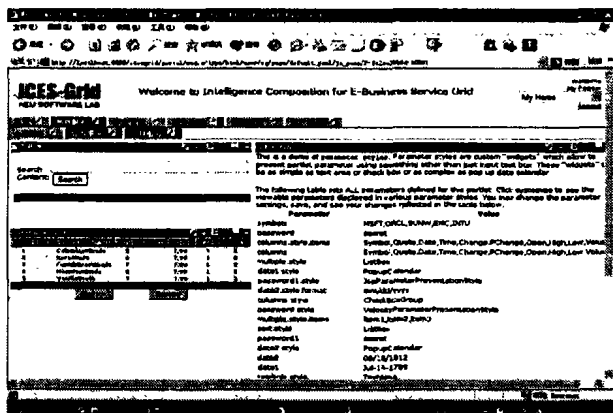


图7 服务网格门户示例

结束语 在网络越来越受到关注的今天,网络的设计与实施成为人们研究的重要课题。而网格门户是网格能否投入实际应用的重要一环。如何简化门户的开发过程,快速地建立起功能完善并且使用方便的服务网格门户同样是建立服务网格的重要课题。相对于一些现有网格门户系统庞大、结构复杂、开发困难、不易部署及更新等特点,本文中设计和实现的服务网格门户具备了层次清晰、功能完整、轻量开发、易于部署及更新等优点,使服务网格门户更加容易得到广泛实现。随着门户技术的成熟和网格计算的不断发展,网格门户的解决方案将会不断完善和发展。

参考文献

- 1 Unger J, Haynos M. A visual tour of Open Grid Services Architecture. <http://www-900.ibm.com/developerWorks/cn/grid/gr-visual/index-eng.shtml>. 2003, 10
- 2 Malhotra V. Jetspeed 第2部分:高级 portlet 技术. IBM developerWorks Java 技术专区. <http://www-900.ibm.com/developerWorks/cn/cnedu.nsf/java-onlinecourse-bytitle/>. 2003, 06
- 3 Laszewski G V, Alunkal B, Amin K. The Java CoG Kit User Manual Draft Version 1.1. Sandeep Nijure Argonne National Laboratory Mathematics and Computer Science Divisional. <http://www-unix.globus.org/cog/manual-user.pdf>. 2003, 10
- 4 Basney J. GSI Credential Management with MyProxy. GGF8 Production Grid Management RG Workshop. <http://www.ncsa.uiuc.edu/~jbasney/ggf8-myproxy.ppt>. 2003, 7
- 5 What is PSML? (c) 2000-2004, Apache Software Foundation. <http://portals.apache.org/jetspeed-1/psml.html>. 2004
- 6 Novotny J, Russell M, Wehrens O. GridLab Portal Design. <http://www.gridlab.org/WorkPackages/wp-4/Documents/GridSphere.pdf>. 2003, 9