

# Sistema Distribuído para Intercâmbio de Dados de Saúde - SIDI

## **José Palazzo M. de Oliveira<sup>1</sup>**

*Engenheiro Eletricista (UFRGS-1968)*

*Mestre em Ciência da Computação (UFRGS-1976)*

*Doutor em Informática (INPG, França, 1984)*

*Professor Titular do Instituto de Informática da UFRGS*

*Vice-presidente da Câmara de PG da UFRGS*

*Pesquisador I do CNPq*

*Áreas de interesse: Modelagem Conceitual e Banco de Dados, Modelagem de Empresas e Workflow, Sistemas de Informação, Ensino a Distância*

## **Nina Edelweiss<sup>2</sup>**

*Engenheira Eletricista (UFRGS-1968)*

*Mestre em Ciência da Computação (UFRGS-1975)*

*Doutora em Ciência da Computação (UFRGS-1994)*

*Professor-Adjunto do Departamento de Informática Aplicada do Instituto de Informática da UFRGS*

*Professora e Pesquisadora do Programa de Pós-Graduação em Computação da UFRGS*

*Pesquisadora do CNPq*

*Áreas de interesse: Banco de Dados Temporais, Modelagem de Empresas*

## **Cora H. F. Pinto Ribeiro<sup>3</sup>**

*Tecnóloga em Processamento de Dados (UFRGS – 1975)*

*Mestre em Ciência da Computação (UFRGS – 1980)*

*Doutora em Ciência da Computação (UFRGS – 1995)*

*Professor-Adjunto do Instituto de Informática*

*Professora e Pesquisadora do Programa de Pós-Graduação em Computação da UFRGS*

*Áreas de Interesse: Modelagem Conceitual e Bancos de Dados Distribuídos*

## **PALAVRAS-CHAVE**

Sistema de informação da saúde – Dados heterogêneos – Operação de sistemas heterogêneos

### **RESUMO**

No Brasil, o acesso a dados relativos à saúde pública é limitado, devido à variedade de soluções adotadas pelas diferentes instituições de saúde para armazenar e recuperar as informações. Este artigo apre-

---

<sup>1</sup> E-mail: palazzo@inf.ufrgs.br

<sup>2</sup> E-mail: nina@inf.ufrgs.br

<sup>3</sup> E-mail: cora@inf.ufrgs.br

senta uma proposta para acesso integrado a informações do Sistema Único de Saúde Pública Brasileiro – SUS. A possibilidade de acesso integrado às informações clínicas e administrativas do SUS resulta em um conjunto relevante de benefícios sociais, desde o uso mais racional dos recursos disponíveis até a melhoria do atendimento médico. No sistema integrado, as informações fornecidas pelas instituições locais devem estar de acordo com um conjunto mínimo de dados requeridos pelo governo federal. Informações adicionais podem ser também fornecidas, a partir de decisões locais. O artigo apresenta os conceitos utilizados para a definição da arquitetura de um sistema integrado, de modo a proporcionar acesso heterogêneo e autônomo aos sistemas de banco de dados locais das instituições de saúde pública. A proposta aqui apresentada foi validada através da construção de um protótipo. Os resultados dessa proposta foram incorporados pela Companhia de Processamento de Dados do Município de Porto Alegre – PROCEMPA em sistemas utilizados pelo SUS.

## **1. INTRODUÇÃO**

A constante introdução de novos procedimentos médicos e testes para diagnósticos, complexos e dispendiosos, tem provocado um aumento significativo de custos nos sistemas de saúde pública no mundo inteiro [Donabedian 95]. Como resultado, políticas de reembolso estão sendo substituídas, em muitos países, pelo pagamento previamente estabelecido de despesas relacionadas a diagnósticos e/ou doenças específicas. Essa nova política leva a uma melhor utilização de recursos médicos. A utilização de recursos computacionais pode auxiliar a prática médica, tanto em diagnósticos como na análise da adequação de procedimentos, além de servir como um repositório integrado de diferentes fontes de informação médica. Previsões de resultados, bem como acesso a informações dispersas, podem dar origem a serviços de saúde de melhor qualidade e menor custo.

No Sistema de Saúde Pública Brasileiro, hospitais credenciados, públicos e privados, são reembolsados por procedimentos médicos e exames realizados em pacientes, de acordo com taxas estabelecidas pelo Sistema Único de Saúde – SUS. Entretanto, não existe uma preocupação com a adequação dos custos da medicina pública. Atualmente os recursos governamentais, centralizados no SUS, estão sendo transferidos para os municípios. Como resultado desta nova política, os municípios estão começando a buscar alternativas para enfrentar o problema de oferecer melhores serviços aos seus pacientes, sem ultrapassar os orçamentos locais. Buscando soluções para esse problema, a Companhia de Processamento de Dados do Município de Porto Alegre – PROCEMPA, juntamente com o Instituto de Informática da Universidade Federal do Rio Grande do Sul – II/UFRGS, desenvol-

veu um projeto de integração de dados de hospitais locais do Serviço de Saúde Pública – o Projeto SIDI.

A dispersão de dados médicos é uma fonte importante de desperdício de recursos públicos e de mau uso de recursos médicos. Com o objetivo de evitar a redundância nas solicitações de exames para diagnósticos médicos, todos os exames previamente realizados por um paciente, além de seus dados médicos, deveriam estar integrados e sempre disponíveis.

Embora a maioria dos procedimentos médicos siga um conjunto de passos e protocolos preestabelecidos, os diagnósticos ainda se baseiam na experiência médica e no julgamento pessoal do médico. A definição de alguns padrões na solicitação de testes de rotina, definidos por um médico especialista, pode reduzir substancialmente os custos médicos.

A otimização dos procedimentos médicos e a redução dos custos com o tratamento do paciente apresentam os seguintes requisitos básicos:

- integração de bancos de dados autônomos existentes, sem perda da autonomia local nem das informações disponíveis. Esta alternativa elimina a duplicação desnecessária de pedidos de exames e diminui o tempo necessário aos processos de diagnósticos, dando origem a um tratamento médico de melhor qualidade e com menor custo;
- fluxos de processamento críticos para protocolos de procedimentos terapêuticos e de diagnósticos. Ferramentas apropriadas deveriam apoiar os médicos especialistas na análise e definição desses fluxos de processamento, fornecendo aos estudantes de medicina e suas equipes novas ferramentas de apoio para os procedimentos médicos;
- integração de procedimentos médicos com análise e correlação de dados, permitindo recuperação automática de informações relacionadas relevantes.
- Este artigo apresenta os resultados obtidos durante o Projeto SIDI. O objetivo básico buscado nesse projeto foi o de disponibilizar para os usuários finais o acesso integrado a informações distribuídas através de bancos de dados autônomos e heterogêneos, pertencentes a diferentes serviços de saúde pública. Além de melhorar os cuidados médicos, o acesso integrado a bancos de dados médicos proporciona uma importante fonte para pesquisas clínicas, além de servir como base para estatísticas sobre saúde pública e para controle epidemiológico. Na segunda seção é apresentado o projeto tanto nos aspectos organizacionais quanto em sua estrutura geral. A seguir, na seção 3, é apresentada a base conceitual utilizada para o desenvolvimento do protótipo, na seção 4 são descritos os resultados obtidos e, finalmente, na seção 5 é apresentada uma avaliação geral sobre os resultados do projeto.

## 2. O PROJETO SIDI

O principal objetivo do Projeto SIDI foi a geração de um sistema distribuído de informação de saúde pública, a partir da integração de sistemas autônomos legados existentes. Apesar da integração, todos os sistemas participantes deveriam permanecer autônomos e inalterados. O sistema de informação resultante deveria proporcionar ao Sistema Público de Saúde – SUS, não somente a recuperação de informações médicas, mas também mecanismos para alertas de saúde pública, gerenciamento de informações médicas, controle e auditoria.

### 2.1 *Parceiros e Colaboradores do Projeto SIDI*

O projeto foi coordenado pelo II/UFRGS, que tinha como atribuição básica o suporte técnico nas áreas de Banco de Dados e de Sistemas de Bases de Conhecimento, de tecnologia de Internet e de modelagem de aplicações médicas.

O principal parceiro externo foi a PROCempa. Esta é responsável pela supervisão de onze companhias nacionais de processamento de dados, que atuam no desenvolvimento de software para o Sistema de Saúde Pública. Essa posição de liderança representa uma perspectiva promissora da aplicação do Projeto SIDI em nível nacional, com apoio do Ministério da Saúde.

Também colaboraram no projeto outras instituições acadêmicas e de pesquisa, tais como a UCPel (Universidade Católica de Pelotas – RS), InCor/USP (Instituto do Coração da Universidade Federal do São Paulo), Instituto de Cardiologia do Rio Grande do Sul, FUC–UFC (Fundação Universitária de Cardiologia da Universidade Federal do Ceará) e Centro de Informática em Saúde da Universidade Federal de São Paulo.

### 2.2 *Arquitetura SIDI*

A Internet é uma rede mundial de computadores que compreende milhares de redes regionais e locais, conectadas através de canais públicos de comunicação. Intranets proporcionam troca de informações locais e privadas, através de Redes Locais (*Local Area Networks* – LAN) e com a utilização de serviços de rede tradicionais. Essa utilização de soluções Internet em redes locais reduz as despesas com tecnologias e software adicionais, as quais são de alto custo. A Extranet é uma Intranet estendida, de modo a permitir acesso integrado, seguro e privado a sítios dispersos geograficamente, através de uma Rede Pública (*Wide Area Networks* – WAN). A Extranet inclui diferentes camadas de aplicações. Essa arquitetura está representada na Figura 1.

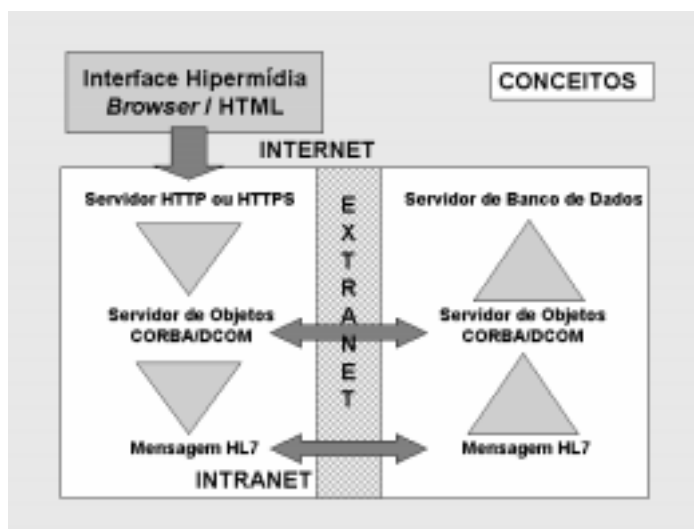


Figura 1 – Camadas da Intranet/Internet/Extranet

Uma das premissas do Projeto SIDI foi a aplicação de soluções padronizadas para a interoperabilidade e troca de dados fornecidos pela Internet. Sistemas baseados em HTML são utilizados para a integração de informações. O mecanismo de comunicação CORBA/DCOM [Özu 01] é utilizado para a recuperação, de forma orientada a objetos, de informações distribuídas. O padrão HL7 [HL7 01] é utilizado para a troca de informações médicas. O acesso integrado aos dados distribuídos, localizados em Intranets particulares dos hospitais e instituições de saúde participantes, é realizado através do conceito de *framework* estrutural de uma Extranet, controlada pela PROCEMPA (Figura 2).

O padrão HL7 (*Health Level Seven*), um padrão ANSI aprovado pela *American National Standard* para troca de documentos eletrônicos na área da saúde, permite que diferentes aplicações computacionais troquem conjuntos relevantes de informações médicas, clínicas e administrativas. O HL7 é formado por formatos padronizados – os protocolos HL7, os quais especificam a implementação de interfaces entre diferentes aplicações computacionais. Estes protocolos proporcionam a flexibilidade necessária para permitir a compatibilidade de conjuntos de dados distintos, que apresentam necessidades específicas. O HL7 engloba dados de admissão e alta de pacientes, observações clínicas, pedidos de exames, resultados e cobranças, além de outros dados de sistemas médicos. Em conformidade com o Sistema Legal Brasileiro, todos os usuários e produtores de sistemas de informação de saúde deveriam produzir um conjunto mínimo requerido de informações.

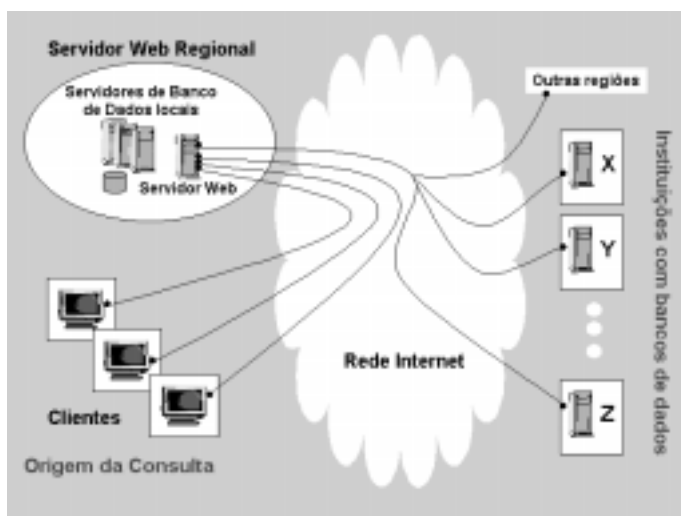


Figura 2 – Extranet SIDI

O padrão HL7 apresenta suporte a comunicações para diversas linguagens de programação, sistemas operacionais e ambientes de comunicação, abrangendo desde aplicações Internet até sistemas legados. Cada resultado de uma consulta é uma mensagem, adequada ao ponto de vista do usuário em questão. Cada mensagem é constituída de segmentos de dados compostos de um ou mais itens. Mensagens podem ser disparadas através de gatilhos. O conteúdo das mensagens, itens específicos de dados e condições de erros independentes da implementação física são definidos através da Definição de Mensagens Abstratas.

Após o estudo de diferentes softwares que poderiam ser utilizados, o componente *OCX Symphonia* foi selecionado como uma alternativa eficiente para a codificação e decodificação das mensagens HL7 enviadas pela Web. No Projeto SIDI, todas as consultas recebidas por um servidor regional são transmitidas para cada nodo dos bancos de dados, no formato HL7, utilizando o software *Symphonia*. Estas consultas devem ser então traduzidas para a Linguagem de Manipulação de Dados (DML) compatível com a base local. As respostas às consultas seguem o caminho inverso, da DML para o formato HL7, sendo enviadas de volta ao servidor. O servidor junta as respostas parciais, recebidas a partir dos diferentes bancos de dados, de uma forma bastante simples (concatenação), reunindo os componentes de uma página HTML completa ou do documento de resposta HL7. Todas as respostas solicitadas e serviços são direcionados a um único servidor Web, responsável pelo processamento de consultas (através de uma página Web), e serviços (falhas em nodos, problemas de *timing*, mapeamentos, etc.), como pode ser observado na Figura 2.

### **3. INTEROPERABILIDADE ENTRE BANCOS DE DADOS ATRAVÉS DO MAPEAMENTO DO ESQUEMA CONCEITUAL**

O Projeto SIDI integra sistemas autônomos e heterogêneos. Os sistemas de informação médica participantes armazenam informações colhidas ao longo de vários anos e são independentes uns dos outros. Cada um dos hospitais participantes possui tecnologia e procedimentos operacionais próprios, incluindo o processamento específico dos dados locais. Como consequência, os bancos de dados distribuídos e as aplicações correspondentes apresentam elementos distintos. Devido a essa diversidade, o Projeto SIDI adotou a premissa de manter a autonomia de cada um dos hospitais participantes. Para alcançar esse objetivo, o problema da heterogeneidade foi resolvido, pela equipe da UFRGS, no nível conceitual dos bancos de dados.

O mapeamento entre esquemas conceituais [Yan 01, Miller 00, Ribeiro 96] permite uma visão global e complementar dos diferentes aspectos de uma mesma entidade real, representados em diferentes bancos de dados, sem que as modificações locais requeridas em metodologias de integração de banco de dados heterogêneos [Miller 01] se façam necessárias. A comparação entre os diferentes esquemas conceituais envolvidos leva à identificação das diferentes representações locais de uma mesma entidade do mundo real, como por exemplo, das diferentes informações, clínicas e administrativas, de um paciente. Nessas representações equivalentes, adotadas nos diferentes bancos de dados, podem existir semelhanças e conflitos de representação. Nesse contexto, a resolução de conflitos e a unificação de esquemas conceituais, adotadas em metodologias de integração, são substituídas pela criação de funções de correlação entre estruturas de esquemas diferentes, porém equivalentes [Abiteboul 97, Kashyap 95, Shet 90]. A reconciliação de discrepâncias, associada à integração semântica [Papakonstantinou 96, Kent 91], é substituída pelo relativismo semântico [Saltor 91, Shet 91], permitindo diferentes interpretações da realidade. As modificações locais de esquemas não são consideradas para efeitos de mapeamento, de modo que a autonomia local não é ameaçada.

No SIDI, o mapeamento entre os esquemas conceituais dos diferentes bancos de dados foi utilizado como suporte para o acesso integrado às informações provenientes destes repositórios de informação.

A tradução de consultas para a linguagem de manipulação de dados específica também assegura a autonomia local do SGBD em uso.

#### ***3.1 Mapeamento do Esquema Conceitual***

O método de mapeamento utilizado [Ribeiro 95, Kantorski 2000] se baseia na idéia de bancos de dados corporativos compartilhando dados locais com todos os membros da federação, sem alterações nos programas de aplicação locais ou nos

métodos de acesso aos dados locais, como adotado em Sistemas de Bancos de Dados Federados – FDBS fracamente acoplados [90].

O processo de mapeamento de representações equivalentes é precedido pela disponibilização das informações locais, a partir da geração do esquema de exportação – EE, contendo a representação daqueles dados que estão sendo disponibilizados para acesso externo, e da tabela de objetos locais – TOL correspondente, onde todas as entidades representadas no banco de dados são especificadas e descritas.

A conversão dos esquemas locais para o esquema de exportação representado em um modelo canônico elimina a ocorrência de conflitos de modelagem de dados, tornando a comparação dos esquemas conceituais bem mais simples [Batini 86]. Adicionalmente, a proteção de dados locais pode ser assegurada através da exclusão, no esquema de exportação, de todos os dados privados. Informações semânticas adicionais podem ser facilmente acrescentadas ao esquema de exportação durante o processo de geração do esquema de exportação, por meio de especificações complementares de tipos e de domínios de atributos. O processo de enriquecimento semântico desta etapa deve incluir o assinalamento de todos os atributos que podem ser utilizados como identificador (como identificador de paciente, carteira de identidade e CPF), permitindo assim o mapeamento de representações equivalentes durante o acesso integrado às informações.

A tabela de objetos locais contém o nome e a descrição de todos as entidades do mundo real representados no esquema de exportação, além de componentes que permitem a inclusão de informações complementares que possam facilitar a identificação de diferentes representações, aparentemente não relacionadas, de um mesmo objeto real.

Inicia-se, então, a etapa de identificação de equivalência entre objetos locais e os representados nos demais bancos de dados, armazenada na tabela de equivalência de objetos – TEO, e das equivalências e funções de mapeamento entre atributos destas representações equivalentes, armazenados na tabela de equivalência de atributos – TEA. Todas estas informações, além de informações sobre o banco de dados propriamente dito, são armazenadas em um repositório comum, denominado Registro. Todos os componentes da metodologia estão representados na Figura 3.

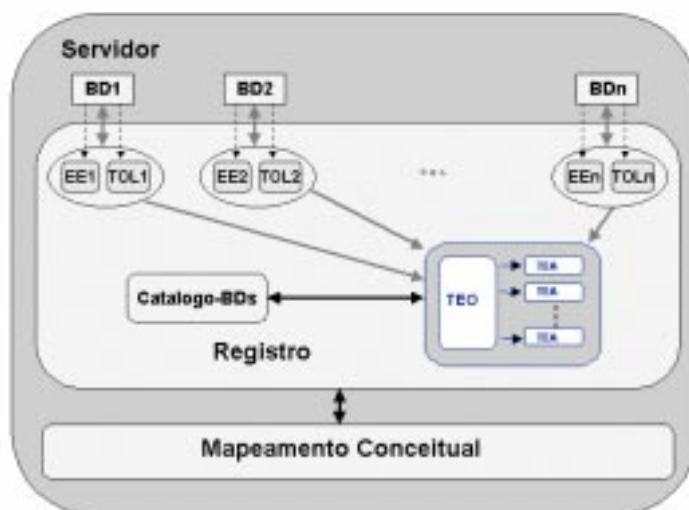


Figura 3 – Mapeamento de Esquemas Conceituais

O mapeamento conceitual inicia com a comparação dos esquemas de exportação e tabela de objetos locais. É importante enfatizar que informações semânticas adicionais podem ser incluídas durante a geração do esquema de exportação, facilitando esta etapa.

O processo divide-se em dois passos principais: (1) identificação de equivalência de representações de objetos, e (2) identificação de equivalência domínio de atributos de representações equivalentes.

A equivalência entre as representações de objetos, assim como restrições e conflitos identificados, é armazenada na tabela de equivalência de objetos – TEO.

Todos os relacionamentos existentes entre atributos de objetos equivalentes, bem como funções de mapeamento e procedimentos de registro de inconsistências de instâncias, a serem aplicados quando do acesso integrado às informações são armazenados na tabela de equivalência de atributos – TEA. A identificação da equivalência entre atributos que podem ser identificadores é fundamental nesta etapa.

Para que um atributo seja utilizado como o identificador de um objeto, ele deve ser associado a um, e somente um, objeto real [Jonscher 94]. Conseqüentemente, um atributo identificador deve satisfazer duas condições básicas: deve ser *único*, pertencendo a somente um objeto, e deve ser *simples*, ou seja, deve representar uma só propriedade associada a um objeto [Kent 92]. Os atributos que satisfazem essas condições são denominados neste artigo de *únicos*.

O objetivo da inclusão de funções de mapeamento de domínio de atributos é possibilitar a utilização de analogia semântica, de modo que diferenças de jargão possam ser informadas ao usuário, de acordo com seu próprio vocabulário. Como

resultado, um código externo para uma determinada droga pode ser traduzido para um código local. Essas funções de mapeamento são utilizadas para traduções de domínio, durante o acesso aos bancos de dados.

Além de manipular discrepância de dados e diferenças de domínios, o enfoque do mapeamento também faz a manipulação da inconsistência de valores. Sempre que atributos equivalentes apresentam inconsistência de informações no nível da banco de dados, esta inconsistência deve ser mostrada ao usuário. Registros de inconsistências podem ser implementados através da detecção de inconsistências e de procedimentos de manipulação, com base nas informações de equivalência.

### ***3.2 Mapeamento Conceitual no Projeto SIDI***

No SIDI, o processo de mapeamento e acesso a informações foi reduzido ao tratamento das informações essenciais. Essas informações essenciais compreendem um conjunto de dados médicos definido, pela comunidade médica, como sendo o conjunto de dados críticos para o tratamento médico efetivo, acrescido de informações médicas adicionais, legalmente necessárias para fins de saúde pública. Esse enfoque simplificou a implementação inicial do método, mas também diminuiu o escopo das informações acessadas.

As informações essenciais representadas no formato HL7 são compostas pelos seguintes grupos de informação:

- Identificação do paciente
- Diagnóstico
- Condições críticas (para procedimentos de alerta)
- Retornos programados e tratamentos futuros
- Procedimentos médicos
- Cada um destes componentes contém atributos específicos. Com isso, o processo de geração do esquema de exportação e da tabela de objetos locais ficou reduzido apenas à representação local dos elementos, e atributos correspondentes, referenciados no HL7. Contudo, expandir o escopo da representação não é um processo complexo.

## **4. ATIVIDADES E RESULTADOS**

A arquitetura do Projeto SIDI foi desenvolvida através de um esforço conjunto entre o II e a PROCEMPA. A definição do modelo de dados resultante, representando as informações do sistema de saúde, foi o resultado de diversas reuniões e discussões a respeito da utilização de diferentes alternativas.

A análise e a escolha de ferramentas de software, o desenvolvimento do protótipo e a implementação final foram realizados pela PROCEMPA, com base nas seguintes requisitos:

- utilização de um servidor Web central para cada região geográfica brasileira;
- submissão de consultas através de um *browser* Internet ou de um programa de aplicação;
- utilização do padrão HL7 para a troca de mensagens entre os usuários e os servidores Web;
- desenvolvimento de todos os programas de aplicação em Java.

O Centro de Informática na Saúde da Universidade de São Paulo (CIS–UNIFESP) tem uma grande experiência em aplicações computadorizadas na área da saúde, desenvolvendo atividades em diversas áreas, tais como apoio médico remoto, sistemas de informação de hospitais, sistemas de apoio ao diagnóstico e objetos distribuídos. Sua equipe, que possui conhecimentos médicos aprofundados, juntou-se ao Projeto SIDI com a finalidade de fornecer os conhecimentos da área médica. Essa parceria foi fundamental para o andamento do projeto, principalmente na fase de definição da arquitetura, com importante contribuição no conhecimento das tecnologias e dos padrões utilizados em saúde pública nos principais países e instituições do mundo. Um destes, o padrão HL7 (*Healthcare Level 7*), foi adotado no protótipo construído no Projeto SIDI. Outra importante contribuição do CIS–UNIFESP esteve ligada à definição do conjunto mínimo de dados dos pacientes a serem adotados como obrigatórios. A interação do CIS com os outros participantes do projeto deu origem a importantes resultados, entre os quais pode ser ressaltado o treinamento de recursos humanos nas aplicações HL7 e no uso de tecnologias de suporte.

O Consórcio Brasileiro de Componentes de Software foi criado como uma parceria entre as instituições do Projeto SIDI, a DATASUS e outras organizações brasileiras de saúde pública, com o objetivo de distribuir, em nível nacional, componentes de software específicos para a área da saúde, de modo que companhias de menor porte, com pouca experiência em novas tecnologias, possam assimilar facilmente esse software.

Em parceria com as instituições do Projeto SIDI, o CIS–UNIFESP submeteu a especificação do Interpretador de Dados Médicos (*Healthcare Data Interpretation Facility*) ao CORBAMED, um comitê internacional responsável pela criação de padrões de especificação para sistemas de saúde. Essa submissão foi a primeira participação brasileira no CORBAMED, estabelecendo a base para a padronização desta especificação, atualmente em desenvolvimento.

A contribuição da UCPEL no projeto foi na área de Descoberta de Conhecimento. A meta buscada nesse contexto foi a extração de conhecimento de documentos

textuais (não estruturados), além da organização dessas informações em um formato padronizado. Esse conhecimento é buscado em fichários médicos e inclui o reconhecimento implícito de padrões. Foi implementada uma ferramenta, denominada Hiperdicionário, para armazenar *links* (relacionamentos) entre palavras. Esses *links* são utilizados para relacionar palavras, registrar sinônimos, correlacionar afinidades léxicas, etc. O Hiperdicionário apóia todas as ferramentas de consultas, na busca pelo contexto ou através de afinidades. As ferramentas de extração de informações também utilizam esse Hiperdicionário na identificação de afinidades léxicas e na análise sintática. Além disso, o Hiperdicionário auxilia, durante o processo de descoberta de conhecimento, a detecção de palavras correlacionadas de um texto.

Outra ferramenta implementada faz o agrupamento de documentos textuais. Nela, um determinado conjunto de documentos é analisado, buscando similaridades entre os textos e partições em geral. Após o agrupamento, é feita uma análise, com o objetivo de extrair um centróide para cada *cluster*. O centróide apresenta a característica média dos elementos do *cluster*, sendo representado como uma lista de palavras, de acordo com a frequência de cada palavra no *cluster*. Análises posteriores podem ser feitas sobre o centróide, de forma a extrair padrões para cada *cluster*, como por exemplo, pacientes com a mesma enfermidade. É realizada a correção ortográfica nos textos médicos para evitar enganos que podem introduzir erros durante os processos de descoberta de conhecimento. No processo de correção desses textos são realizadas análises morfológica, sintática e semântica.

Adicionalmente, uma ferramenta para extrair informações dos textos médicos foi também implementada. Nesta, as informações codificadas em textos livres são extraídas automaticamente e colocadas em um formato estruturado para análise posterior. A ferramenta utiliza *tags* como indicadores de informações nos textos. Estes *tags* são criados manualmente, após uma análise exaustiva de textos da área e da linguagem técnica utilizada.

Outro objetivo do Projeto SIDI é o disparo de gatilhos (*triggers*) de alerta com base na análise integrada de dados médicos. Um serviço de alertas tem por objetivo detectar situações especiais em gerar avisos que podem ser úteis durante um processo de tomada de decisão. Por exemplo, um serviço de alerta pode prevenir um médico que o resultado do exame de potássio está abaixo do normal. Para gerar esses alertas, é necessário definir uma boa base de conhecimentos, na qual são descritas todas as condições que podem produzir esses alertas. Como uma só instituição de saúde não seria capaz de criar uma base grande o suficiente para cobrir todas as áreas da medicina, é necessário padronizar esse conhecimento para permitir o compartilhamento de informações entre diferentes centros.

O padrão utilizado na área médica é a Sintaxe de Arden, através do qual pode ser criado um Módulo Médico Lógico – MML, que descreve a situação na qual o alerta pode ser gerado. O MML apresenta *frames* que descrevem a condição de invocação, a regra a ser testada e a ação a ser tomada caso a regra seja verdadeira. O mapeamento do MML para o banco de dados não é uma tarefa trivial. Como cada banco de dados apresenta uma

estrutura particular, é necessário prover ferramentas para auxiliar esse mapeamento. Há, ainda, a necessidade de ferramentas para ativar automaticamente as regras do MML quando necessário e de ferramentas para executar a ação quando a regra for verdadeira.

Uma série de componentes de software foi implementada para representar os módulos orientados a objetos. Esses componentes são capazes de testar as regras, executar as consultas necessárias ao banco de dados e executar as respectivas ações. Foi também implementado um monitor de eventos, o qual detecta situações especiais nos bancos de dados (tais como a entrada de resultados de um novo teste de laboratório), com o objetivo de disparar módulos específicos sempre que determinadas condições forem verdadeiras.

## **5. CONCLUSÃO**

Os tópicos principais tratados nesse projeto foram a Manipulação de Dados Heterogêneos e a Interoperabilidade de Sistemas Heterogêneos. Um protótipo operacional para a troca de informações sobre pacientes entre o Hospital de Pronto Socorro – HPS (o principal hospital de emergências médicas de Porto Alegre) e o centro regional comunitário de saúde Bom Jesus foi produzido. Logo após a chegada e a identificação de um paciente nesse centro de saúde, o diagnóstico clínico preliminar é registrado no sistema. Sempre que as condições clínicas do paciente requeiram sua transferência para o HPS, os dados desse paciente são também transferidos para o banco de dados do HPS. Todos os registros médicos desse paciente (atuais e passados, tanto do HPS como do centro de saúde Bom Jesus) estão disponíveis para consulta no banco de dados do HPS. Quando o paciente recebe alta do HPS, as informações médicas relativas ao período de internação são enviadas novamente ao centro de saúde Bom Jesus. Esse protótipo deverá ser aplicado a um conjunto maior de instituições médicas em um futuro próximo.

O Projeto SIDI apresenta um enfoque inovador para o acesso a um Sistema de Saúde Pública. Baseia-se na tecnologia atual de Bancos de Dados e de Sistemas Baseados em Conhecimento, incluindo as potencialidades de navegadores Internet e de sistemas cliente–servidor. Conceitos adotados no projeto e no desenvolvimento de Sistemas de Informação Inteligentes e em Sistemas de Bancos de Dados Especialistas deverão ser incorporados ao projeto, desse modo satisfazendo os requisitos de monitoramento de procedimentos médicos, alertas, disparo de ações corretivas, mineração de dados, além de outros aspectos a serem analisados. Diversas cooperações bilaterais continuam em andamento, consolidadas através desse projeto.

### ***KEYWORDS***

*Health information system – Data heterogeneity – Heterogeneous systems interoperation*

## **ABSTRACT**

*The access to healthcare information in Brazil is limited, due to the variety of information storage and retrieval solutions adopted by different healthcare institutions. This paper presents a proposal for integrated access to healthcare information in the Brazilian Public Health System – SUS. The integrated access to clinical and administrative information granted by the system results in a broad spectrum of social benefits, ranging from a more rational use of resources to better medical care. Information provided by local institutions should fulfill the minimal set of data required by the federal government. Additional information may also be released, according to local decision. Architectural concepts for accessing heterogeneous and autonomous local database systems are presented. The current proposal was validated by a prototype, and results were incorporated by PROCEMPA – the city of Porto Alegre Data Processing Company – in actual production systems currently in use at the SUS.*

## **REFERÊNCIAS BIBLIOGRÁFICAS**

- [Abit 97]. S. Abiteboul et al. “Correspondence and Translation for Heterogeneous Data”. Proceedings of the International Conference on Database Theory (ICDT-97), Delphi, Greece, p.352–363, 1997.
- [Bati 86]. C. Batini et al. “A Comparative Analysis of Methodologies for Database Schema Integration”. ACM Computing Surveys, v.18, n.4, p.323–364, dec. 1986.
- [Dona 95]. A. Donabedian. “The Quality of Care how can it be assessed?”. JAMA 1988, v. 260, n.12, sept. 1988.
- HL7 01 HL7. Data Model Development, HL7, nov, 2001. Página Web acessada em nov. 2001: [www.hl7.org](http://www.hl7.org)
- [Jons 94]. D. Jonscher and K.R. Dittrich. “An Approach for Building Secure Database Federation”. Proc. VLDB’94, p. 24–35, Santiago, Chile, sept. 12–15, 1994.
- [Kant 00]. G. Kantorski and C.P. Ribeiro. “Heterogeneous Database Interoperability through the WWW”. Anais do 15<sup>o</sup> Simpósio Brasileiro de Banco de Dados (SBBD-00), João Pessoa (PA), Brasil, p.79–88, 2000.
- [Kash 95]. V. Kashyap and A. Sheth, A. “Semantic and Schematic Similarities between Database Objects: A Context-based approach”. VLDB Journal, v. 5, n. 4, p.276–304, 1995.
- [Kent 91]. W. Kent, “The Breakdown of the Information Model in Multidatabase Systems”. SIGMOD Record, New York, v.20, n.4, p.10–15, dec. 1991.
- [Kent 92]. W. Kent et al., “Object Identification in Multidatabase Systems”, Proc. IFIP WG2.6 DS-5 Conf. on Semantics of Interoperable Databases, Austrália, nov., 1992.
- [Mill 00]. R.J. Miller, et al. “Schema Mapping as Query Discovery”. Proceedings of the 26<sup>th</sup> International Conference on Very Large Databases. Cairo, Egypt, sept 2000.

- [Mill 01]. Renée J. Miller, Mauricio A. Hernández, Laura M. Haas, Ling-Ling Yan, C. T. Howard Ho, Ronald Fagin, Lucian Popa: *The Clio Project: Managing Heterogeneity*. SIGMOD Record 30 (1): 78–83, 2001.
- [Özsu 01]. M.T. Özsu and P. Valduriez. “*Princípios de Sistemas de Banco de Dados Distribuídos*”. Tradução da segunda edição, Rio de Janeiro, Editora Campus, 711 p, ISBN 85–352–0713–9, 2001.
- [Papa 96]. Y. Papakonstantinou, et al. “Object Fusion in Mediator Systems”, Proceedings of the 22<sup>th</sup> International Conference on Very Large Databases, Mumbai (Bombay), Índia., 1996.
- [Ribe 95]. C.H.F.P. Ribeiro, “Banco de dados Heterogêneos: Mapeamento dos Esquemas Conceituais em um Modelo Orientado a Objetos”, Tese de doutorado, CPGCC, UFRGS, Brazil, dec, 1995.
- [Ribe 96]. C.H.F. Pinto Ribeiro and J. Palazzo M. de Oliveira. “Multidatabase Interoperability Through Conceptual Schema Mapping in an Object Oriented Model”. Proceedings PDCS’96, p.647–652, Dijon, France, oct 25–27, 1996.
- [Salt 91]. F. Saltor, et al. “Suitability of Data Models as Canonical Models for Federated Databases” SIGMOD Record, ACM Computing Surveys, v. 22, n. 4, p.44–48, ACM, dec 1991.
- [Shet 90]. A. Sheth and J. Larson. “Federated Database Systems for Managing Distributed, Heterogeneous, and Autonomous Databases”. ACM Computing Surveys, New York, v.22, n.3, p.183–236, sept. 1990.
- [Shet 91]. A. Sheth. “Semantic Issues in Multidatabase System”. Preface by the Special Issue Editor, SIGMOD Record, New York, v.20, n.4, p.5–9, dec. 1991.
- [Yan 01]. L.L. Yan, et al. “Data-Driven Understanding and Refinement of Schema Mappings”. Proceedings of the ACM SIGMOD, Santa Barbara, California, USA, may 21–24, 2001.