

INTEGRAÇÃO DE DADOS AMBULATORIAIS DE QUIMIOTERAPIA E RADIOTERAPIA REGISTRADOS NAS BASES DE DADOS DO SUS

R. C. Souza* e S. M. Freire*

*Departamento de Tecnologias da Informação e Educação em Saúde, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Brasil
e-mail: romulo@lampada.uerj.br

Resumo: *Dados relativos aos procedimentos de quimioterapia e radioterapia realizados no âmbito do Sistema Único de Saúde (SUS) são coletados pelo módulo de oncologia do subsistema de Procedimentos de Alta Complexidade do Sistema de Informações Ambulatoriais do SUS (APACSIA) e disponibilizados em arquivos mensais pelo Departamento de Informática do SUS (DATASUS). O presente trabalho desenvolveu uma rotina que integra os arquivos mensais da APAC a partir de 2008 utilizando uma plataforma para integração de dados, denominada PENTAHO e o banco de dados PostgreSQL. Um esquema de banco de dados, dados_apac, foi montado no PostgreSQL contendo as tabelas quimio, radio, casos e pacientes, que armazenam os dados de quimioterapia e radioterapia, casos de câncer e pacientes, respectivamente. Um conjunto de procedimentos foi criado no PENTAHO para extrair os dados dos arquivos mensais e armazená-los no esquema dados_apac. Estas rotinas estão automatizadas e podem ser utilizadas por um profissional de saúde que não tenha conhecimentos de informática. Com a integração destes arquivos, é possível identificarem-se casos de câncer, permitindo a produção de informações para o planejamento dos serviços, e não somente aquelas ligadas à gestão financeira.*

Palavras-chave: *sistema de informação ambulatorial, câncer, assistência ambulatorial, integração de dados.*

Abstract: *Data relating to chemotherapy and radiotherapy procedures performed within the Unified Health System (SUS) is collected by the oncology module of the subsystem of the High Complexity Procedures of the Outpatient Information System (APACSIA) and made available in monthly files by the SUS Department of Informatics (DATASUS). This study developed a routine that integrates the monthly archives of the APAC from 2008 onwards using a platform for data integration, called Pentaho and the PostgreSQL database data. A database schema, dados_apac, was created in PostgreSQL, containing the apac, cases of cancer and patient tables, which store data about chemotherapy and radiotherapy, cases of cancer and patients, respectively. A set of procedures was created in Pentaho to extract data from the monthly files and store them in the dados_apac scheme. These routines are automated and can be used by a health professional that does not have computer skills. With the integration of*

these files, it is possible to identify cancer cases, allowing the production of information for the planning of services, not just those related to financial management.

Keywords: *Care Information Systems, ambulatory care, cancer, data integration.*

Introdução

Dados relativos aos procedimentos de quimioterapia e radioterapia realizados em pacientes tratados no Sistema Único de Saúde (SUS) são registrados pelas unidades prestadoras de serviços (UPS) no módulo de oncologia do subsistema de Procedimentos de Alta Complexidade do Sistema de Informações Ambulatoriais do SUS (APACSIA). O objetivo principal desse sistema é administrativo, visando ao processo de faturamento pela UPS. Depois de processados pelos órgãos gestores de saúde, arquivos mensais contendo esses registros são disponibilizados para acesso público pelo Departamento de Informática do Sistema Único de Saúde (DATASUS) [1,2].

Gomes Jr et al desenvolveram uma metodologia para integrar os arquivos de dados do APACSIA [3], utilizando a linguagem SQL [4]. Souza et al. [5] criaram um aplicativo que automatizou o algoritmo de Gomes Jr et al. [3] e ofereceu ao usuário uma interface gráfica por meio da qual é possível gerar um conjunto de consultas. Essa proposta de integração se baseou no conceito de caso de câncer, proposto por Gomes Jr e Almeida [6], definido por meio de uma chave composta pelos campos CPF do paciente (cujo valor na base de dados disponibilizada é um *proxy* para o verdadeiro CPF do indivíduo), tipo do tumor primário, sexo e data de nascimento.

Essa integração permite a produção de informações que podem ser utilizadas para o monitoramento e planejamento da assistência oncológica no SUS, que não podem ser obtidas por meio dos arquivos analisados de forma isolada. Assim torna-se possível, por exemplo conhecer quantos novos casos de câncer e as características desses casos ao darem entrada no sistema, bem como obter o histórico desses casos. Costa [7] utilizou essa metodologia de integração para estudar as condutas terapêuticas adotadas por diferentes prestadores de serviços no tratamento a mulheres portadoras de neoplasia maligna de mama.

A metodologia proposta por Gomes Jr et al [3] pode

ser utilizada para arquivos do APAC/SIA até dezembro de 2007. A partir de 2008, a estrutura dos arquivos de dados da APAC sofreu alterações, com a inclusão de novos campos, como por exemplo, o cartão nacional de saúde, e exclusão de outros, como o CPF. Essa mudança implica que a metodologia proposta acima precisa ser adaptada para trabalhar com os dados a partir de 2008.

Assim, o presente trabalho teve por objetivo adaptar a rotina proposta por Gomes Jr et al [3] para a integração dos arquivos mensais da APAC a partir de 2008 e implementá-la em um ambiente que facilite a sua utilização por profissionais que não tenham conhecimento de informática.

Materiais e métodos

Foram utilizados nesse trabalho os arquivos mensais contendo os registros de produção de quimioterapia e radioterapia referentes ao Estado do Rio de Janeiro nos anos de 2012 e 2013, disponibilizados para acesso pelo DATASUS [8].

Entre as variáveis que compõem as estruturas desses arquivos encontram-se aquelas relativas a: 1) APAC, como datas de solicitação, autorização, apresentação, início e fim (validade), se é inicial ou de continuidade; 2) paciente, como idade, sexo, raça e cor, município de residência; 3) aspectos ligados à patologia, como tipo do tumor e estadiamento; 4) produção e faturamento, como código e valor do procedimento realizado; 5) UPS – código da unidade [9].

Os seguintes *softwares* foram empregados: manipulação dos arquivos de dados do DATASUS – Tabwin [10]; sistema de gerenciamento de banco de dados – postgresql [11] e o Pentaho [12]. O Pentaho é um *software* de código aberto, composto, entre outras, das seguintes ferramentas: 1) Pentaho Data Integration (PDI) – usado para a extração, transformação e carga do banco de dados; 2) Pentaho Schema Workbench (PSW) – usado para criar as visões analíticas (cubos); 3) Pentaho Report Designer (PRD) – usado para a criação de relatórios pré-definidos; 4) Pentaho Metadata Editor (PME) – usado para gerar os modelos de dados utilizados para a criação de relatórios *ad hoc*.

Inicialmente, os arquivos de dados foram baixados do *sítio* do DATASUS em forma compactada – formato DBC - e extraídos para o padrão xBase (DBF) para serem lidos pelo PDI.

No postgresql, criou-se um *esquema* chamado *dados_sus*. Com um *script* SQL [4], geraram-se duas tabelas com o objetivo de receberem os registros dos arquivos originais de quimioterapia e radioterapia, respectivamente. As estruturas dessas tabelas são idênticas às dos arquivos de dados correspondentes,

baixados do DATASUS.

Utilizando a interface do Pentaho Data Integration, foram configurados conjuntos de procedimentos que executaram as seguintes tarefas:

1) importação dos registros de quimioterapia e radioterapia dos arquivos originais;

2) geração de uma tabela unificada, chamada de *apac*, a partir da união de *quimio* e *radio*;

3) geração de uma tabela, chamada *caso*, a partir de *apac*, com base em uma chave composta pelas colunas documento (Cartão SUS - CNS) do paciente (cujo valor nos arquivos disponibilizados pelo DATASUS é apresentado de forma criptografada), sexo e os três primeiros dígitos do código de diagnóstico. Como a nova estrutura dos arquivos da APAC disponibilizados para *download* não inclui a data de nascimento (apenas a idade), essa definição de caso difere daquela proposta por Gomes Jr et al [3]. Para cada caso de câncer identificado, foi extraído da tabela *apac* apenas o primeiro registro, o mais antigo, segundo a data de apresentação do faturamento;

4) geração de uma tabela, chamada de *paciente*, a partir de *caso*, com base na coluna CNS.

Utilizando o Tabwin, criou-se um arquivo DBF, chamado de *apac*, contendo todos os registros mensais de quimioterapia e radioterapia dos arquivos originais. Com esse DBF foram executados *scripts* em linguagem SQL com o objetivo de se contarem: 1) as quantidades mensais de registros de quimioterapia e radioterapia; 2) a quantidade de casos de câncer, de acordo com a definição proposta neste artigo; 3) as quantidades de tipos casos; 4) a quantidade de pacientes, considerando apenas o campo CNS. Procedimento semelhante foi executado a partir do esquema *dados_sus* com as tabelas *apac*, *caso* e *paciente*. Os resultados, então, foram comparados para verificar a correção do processo utilizado no PDI.

As rotinas desenvolvidas e instruções de como configurar o ambiente do Pentaho para realizar a integração das bases propostas nesse trabalho podem ser obtidas por meio de uma consulta aos autores.

Resultados

A Figura 1 mostra, parcialmente, o modelo do esquema *dados_sus*. A tabela *apac* é resultado da integração dos registros de quimioterapia e radioterapia. Dessa tabela são extraídos os dados para as tabelas casos de câncer e pacientes. Como se observa, um paciente pode estar relacionado a um caso ou mais de câncer e é possível recuperar todos os registros de procedimentos realizados para cada paciente.

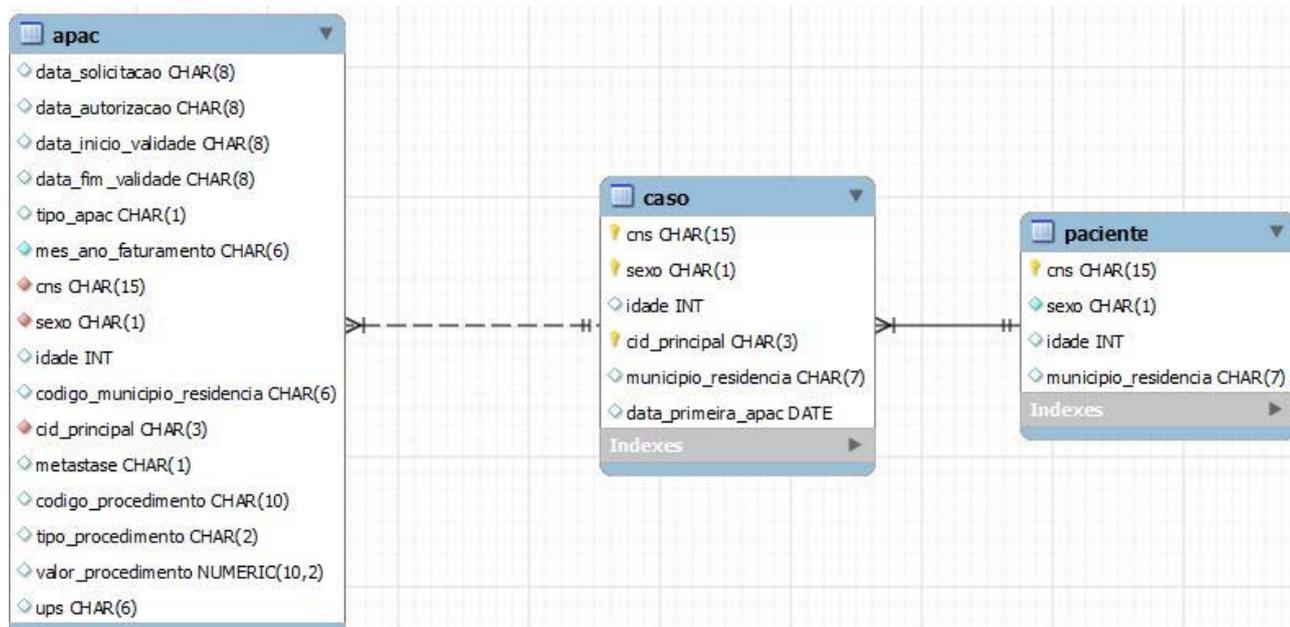


Figura 1 – Representação parcial do modelo de dados do esquema *dados_sus*.

Os resultados obtidos a partir da execução dos *scripts* SQL utilizando o Tabwin e o arquivo *apac.dbf* foram os mesmos daqueles conseguidos com as tabelas do esquema *dados_sus* do PDI.

Discussão

O Brasil conta com vários sistemas públicos de informação que visam a coleta e o armazenamento de dados sobre eventos ligados à saúde da população durante o ciclo de vida do indivíduo – do nascimento ao óbito. Esses sistemas são de reconhecida importância para a produção de informações de interesse epidemiológico, vigilância em saúde, entre outros [13][14].

No caso das APACs de radioterapia e quimioterapia, para melhor explorar o potencial das mesmas em gerar informações de cunho assistencial, duas questões se colocam: 1) integração das bases, já que normalmente os dados são mantidos em arquivos distintos e mensais; 2) oferecimento ao usuário/analista meios para produzir/visualizar as informações da maneira mais fácil possível, considerando o fato de que, geralmente, quem utiliza tais dados são profissionais que podem não possuir habilidades em informática.

A abordagem mostrada no presente trabalho lida com os dois problemas acima e poderia ser adaptada para ser utilizada nas bases da APAC anteriores a 2007. Quando comparada à rotina de Gomes Jr et al.[3], esta abordagem possui as seguintes vantagens: 1) dispensa o usuário de conhecimento de linguagem SQL para executar o processo de integração dos arquivos; 2) diminui o tempo de processamento nas atualizações dos dados, uma vez que, a cada novo arquivo disponibilizado, apenas seus registros são adicionados à

tabela previamente unificada; 3) gera automaticamente as tabelas de casos de câncer e paciente; 4) exporta as tabelas resultantes para formato que pode ser lido por *softwares* como planilhas eletrônicas e pacotes estatísticos. Em relação à aplicação desenvolvida por Souza et al [5], essa proposta: 1) é mais facilmente mantida e atualizada; 2) permite uma melhor disponibilização e distribuição para usuários que tenham interesse em sua utilização; 3) pode ser integrada a outras aplicações/bancos de dados.

Apesar de resolver o problema da integração dos arquivos APAC oncologia a partir do ano de 2008, a proposta aqui apresentada não consegue realizar a integração dos dados desse período com os de anos anteriores, por duas razões: 1) nos arquivos disponibilizados pelo DATASUS não há dados que permitem a identificação do paciente, seja por método determinístico ou probabilístico; 2) o campo CPF, que poderia ser utilizado como variável *proxy* para identificar o paciente, está presente na estrutura dos arquivos até o ano de 2007, sendo substituída a partir de 2008 pelo campo CNS.

A mudança da definição de caso de câncer nesse trabalho em relação à proposta de Gomes Jr et al [3] deve-se à ausência das variáveis data de nascimento e CPF no dicionário de dados da APAC a partir de 2008. Na nova definição, o CPF foi substituído pelo CNS e a chave não inclui a data de nascimento. Acreditamos que a data de nascimento tenha um papel marginal na definição de um caso de câncer, servindo para distinguir casos quando duas pessoas diferentes possuem o mesmo CPF. Por outro lado, ela pode introduzir um erro de classificação quando a mesma pessoa tiver registros com datas de nascimento diferentes.

Um caso de câncer não pode entendido como um

paciente, já que um indivíduo está sujeito a apresentar mais de um tipo de neoplasia. Trabalhando com uma base de dados identificada proveniente da APAC, e aplicando a metodologia probabilística de vinculação de registros, Souza [15] concluiu que o campo CPF, ainda que disponibilizado de forma “mascarada”, seria uma variável *proxy* para paciente. Nesse trabalho, estamos supondo que o CNS teria a mesma função que o CPF, mas isso precisa ser melhor investigado.

O Pentaho Data Integration possui as seguintes características: 1) possibilidade de conexão com os principais sistemas gerenciadores de bancos de dados disponíveis; 2) permite a exportação de resultados para arquivos de diferentes formatos, como textos e planilhas eletrônicas; 3) uso livre e código aberto; 4) ampla comunidade de usuários, o que tem promovido uma permanente evolução do *software*, além do suporte aos usuários; 5) pode ser utilizado de forma independente dos demais aplicativos da plataforma Pentaho.

Nesse trabalho foram utilizados dados dos anos de 2012 e 2013 para o estado do Rio de Janeiro. Entretanto, a metodologia aqui utilizada pode ser aplicada para todos os anos a partir de 2008 e para os demais estados da federação.

Como evolução do presente trabalho pretende-se: 1) implementar funções no PDI para executar o *download* dos arquivos diretamente do DATASUS e realizar a descompactação dos mesmos; 2) utilizar outros módulos do Pentaho, como o Pentaho Schema Workbench, Pentaho Report Designer, e o Pentaho Metadata Editor para criar visões analíticas e permitir a realização de processamento analítico online, mineração de dados e criação de relatórios pré-definidos e *ad hoc*.

Esse trabalho apresenta uma implementação no ambiente Pentaho de uma proposta para a integração das bases que compõem o módulo de oncologia do APAC/SIA, de modo a identificar casos de câncer. A partir desta integração, os recursos do Pentaho podem ser explorados para produzir informações úteis para o melhor planejamento dos serviços de assistência à saúde.

Referências

- [1] BRASIL Ministério da Saúde/Secretaria de Atenção à Saúde/Departamento de Regulação, Avaliação e Controle/Coordenação Geral de Sistemas de Informação. Manual Técnico Operacional SIA/SUS - Sistema de informações Ambulatoriais - Aplicativos de captação da produção ambulatorial APAC Magnético - BPA Magnético - VERSIA - DEPARA - FPO Magnético. março de 2010. Disponível em http://www.saude.am.gov.br/docs/programas/bucal/manual_sia/MANUAL_OPERACIONAL_SIA.pdf.
- [2] Secretaria de Atenção à Saúde Ministério da Saúde, Departamento de Atenção Especializada, Departamento de Regulação, Avaliação e Controle. SIA - Sistema de Informação Ambulatorial do SUS - Oncologia - Manual de Bases Técnicas - 17a. edição. 2014. Disponível em http://www1.inca.gov.br/inca/Arquivos/manual_oncologia_17_edicao.pdf.
- [3] Gomes Jr SC, Di Martino R, Almeida RT. Rotinas de integração das tabelas do Sistema de Autorização de Procedimentos de Alta Complexidade em Oncologia do Sistema Único de Saúde. Cadernos de Saúde Coletiva. 2003;11(2):231–54.
- [4] Kriegel A, Trukhnov BM. SQL Bible. Second Edition. Indianapolis, Indiana: Wiley Publishing; 2008.
- [5] Souza RC, Freire SM, Almeida RT. Sistema de informação para integrar os dados da assistência oncológica ambulatorial do Sistema Único de Saúde. Cadernos de Saúde Pública. 2010; 26(6):1131–40.
- [6] Gomes Jr SC, Almeida RT. Identificação de um caso novo de câncer no Sistema de Informação Ambulatorial do Sistema Único de Saúde. Cad Saúde Coletiva. 2004; 12(1):57–68.
- [7] Costa MR. Comparação das condutas terapêuticas no tratamento ambulatorial das mulheres com câncer de mama [Dissertação]. [Rio de Janeiro]: Universidade Federal do Rio de Janeiro; 2005.
- [8] Ministério da Saúde. Departamento de Informática do SUS. Disponível em <http://www2.datasus.gov.br/DATASUS/index.php?area=0901&item=1&acao=22&pad=31655>.
- [9] Ministério da Saúde, DATASUS. Disseminação de Informações do Sistema de Informações Ambulatoriais do SUS (SIASUS), Informe Técnico 2014-01, 2014. Disponível em: ftp://ftp.datasus.gov.br/dissemin/publicos/SIASUS/200801/Doc/IT_SIASUS_1401.pdf. Acesso em 10 mar 2014.
- [10] The PostgreSQL Global Development Group. The world's most advanced open source database [internet]. [citado 20 de julho de 2014]. Disponível em: www.postgresql.org.
- [11] Pentaho. A comprehensive platform for data integration & Business Analytics [internet]. [citado 20 de julho de 2014]. Disponível em: <http://community.pentaho.com/projects/data-integration/>.
- [12] Ministério da Saúde. Secretaria Executiva do DATASUS. Tabwin [internet]. [citado 20 de julho de 2014]. Disponível em: <http://www.datasus.gov.br>.
- [13] Ministério da Saúde/Organização Panamericana de Saúde/Fundação Oswaldo Cruz. A experiência brasileira em sistemas de informação em saúde Volume 1. 2009
- [14] Silva, JPL; Travasso, C; Vasconcellos, MM; Campos, LM. Revisão sistemática sobre encadeamento ou linkage de bases de dados secundários para uso em pesquisa em saúde no Brasil. Cad. Saúde Coletiva. 2006; 14(2):197-224.
- [15] Souza RC. Desenvolvimento de um armazém de dados a partir da integração de sistemas de informação em saúde para apoiar a gestão da assistência oncológica Rio de Janeiro 2012 [Tese]. Universidade do Estado do Rio de Janeiro; 2012.