

Implementação Computacional de Fusão Automática de Dados Distribuídos em Apoio à Gestão de Saúde

Renato Cerceau^{1,3}, Elaine Sigette^{1,4}, Ricardo Pires Mesquita¹, Jorge Zavaleta¹,
Ricardo Cerceau², Laci Mary Barbosa Manhães¹, Raimundo José Macário Costa⁵,
Luís Alfredo Vidal de Carvalho¹

¹ Programa de Engenharia de Sistemas e Computação (PESC/COPPE), ² Pesquisador
Universidade Federal do Rio de Janeiro (UFRJ) – Rio de Janeiro, RJ – Brasil

³ Agência Nacional de Saúde Suplementar (ANS) – Rio de Janeiro, RJ – Brasil

⁴ Universidade Federal Fluminense (UFF) – Niterói, RJ – Brasil

⁵ Universidade Federal Rural do Rio de Janeiro (UFRRJ) – Rio de Janeiro, RJ – Brasil

{cerceau,elaine,mesquita,mymanhaes, zavaleta}@cos.ufrj.br,
macario@ufrj.br, ricardo.cerceau@gmail.com, luisalfredo@ufrj.br

Abstract. *This paper presents the computational implementation of a data distributed fusion and consolidation model in support of health management. It is intended that this aggregation model work as a support and source of information to health care researchers aiming to planning health actions, especially in municipal instances. For this purpose, is implemented a process of obtaining and automated data preparation. We present the process of defining this model of obtaining the information, including the type of support, the process of obtaining the production registers, the connection model to social networks and academic production. From the model presented here were generated 91 e-books that are available in open access.*

Resumo. *Este trabalho apresenta os resultados obtidos na implementação computacional de um modelo de agregação e consolidação de dados distribuídos em apoio à gestão de saúde. Pretende-se que este modelo de agregação sirva de suporte e fonte de dados aos pesquisadores da área de saúde com vistas ao planejamento de ações de saúde, em especial em instâncias municipais. Com este propósito, é implementado um processo de obtenção e elaboração automática de dados. Apresentamos o processo de definição deste modelo de obtenção das informações, incluindo o tipo de suporte, o processo da obtenção dos registros das produções, o modelo de vinculação às redes sociais e a produção acadêmica. A partir do modelo aqui apresentado foram gerados 91 e-books que estão disponíveis, em acesso aberto.*

1. Introdução

As instituições médicas e governamentais da área da Saúde podem analisar uma grande e variada massa de dados multiestruturados para diagnosticar perfis ou grupos de pacientes. Esses dados são produzidos localmente, mas se encontram geograficamente distribuídas. A fusão e o armazenamento destes dados constituem-se como um grande desafio, uma vez que demanda largos recursos computacionais no gerenciamento dos mesmos. Como exemplo disso tem-se a grande quantidade de dados diferentes e não uniformizados, gerados por cada município que dificulta o agrupamento e o estudo comparativo.

No Brasil, o sistema nacional de saúde está organizado pelo Sistema Único de Saúde (SUS), onde a gestão e o planejamento das ações acontecem de forma regionalizada e hierarquizada em níveis de complexidade crescente (BRASIL, 1990). O documento ‘Plano de Saúde’ elaborado pela gestão (municipal, estadual e federal) serve como base para as atividades e programações. Este Plano é realizado de acordo com as características estabelecidas pelo Conselho Nacional de Saúde (BRASIL, 2002).

Nos municípios, no mínimo a cada quatro anos deve ser elaborado o Plano Municipal de Saúde (PMS). Este consiste em etapa diagnóstica e proposição de planejamento para ações de saúde (BRASIL, 2002; BRASIL, 2011). No planejamento de saúde devem ser considerados os serviços e as ações prestados pela iniciativa privada, de forma complementar ou não ao SUS (BRASIL, 2011).

Os pesquisadores (e gestores) municipais têm dificuldades na obtenção e identificação das informações para elaboração de seus PMS, especialmente nos municípios menores. As principais dificuldades estão relacionadas à identificação das fontes de dados e à qualificação profissional para se fazer e apresentar as devidas análises (CERCEAU *et al.*, 2001).

Buscando auxiliar o agrupamento e fusão dos dados, em sua maioria disponibilizada de forma pulverizada na Internet, foi criada uma metodologia que automatiza a obtenção e a consequente geração de informações empregando técnicas computacionais. Procura-se aperfeiçoar os métodos empregados na coleta de dados, além de proporcionar uma maior disponibilidade de informações devidamente qualificadas, com o intuito de tornar o processo decisório dos gestores mais rápido – além de popularizar o acesso aos dados de saúde municipais, sejam eles de organizações públicas ou privadas.

Nesse sentido, este trabalho se apoia em técnicas do domínio da *e-science* que objetivam preservar dados públicos da Saúde, ele tem como objetivo auxiliar pesquisadores de formação multidisciplinar que necessitem organizar dados científicos para disponibilizá-los em acesso aberto (*Open Access*) na forma de livros digitais (*e-books*). Os dados são obtidos por meio de grande coleta de dados em bases distribuídas (e multiestruturadas) para serem consolidados na busca de evitarmos que esses dados se percam em um grande volume de dados, por falta de tratamento técnico adequado, como alerta Costa & Cunha (2014). Buscamos preservar os dados das séries temporais divulgadas na rede mundial de computadores pelas instituições governamentais da área de saúde. Diversas instituições têm optado por apresentar apenas a informação mais recente, usualmente excluindo do acesso *online* as bases históricas. O resultado da pesquisa é de cunho prático e possui impacto social tendo grande valor na área da gestão da saúde pública.

Este trabalho está organizado da seguinte forma. Na seção 2 apresentamos os trabalhos relacionados. Na seção 3 apresentamos os Materiais e Métodos. Na seção 4 apresentamos os principais resultados. Na seção 5 apresentamos as conclusões.

2. Trabalhos Relacionados

Foram propostos guias de orientação para elaboração de PMS, por diferentes instâncias governamentais, com obtenção manual de informações a partir de diversas fontes de dados (BAHIA, 2009; BRASIL, 2009).

Uma das estratégias para disseminação de informações sobre os municípios consiste na geração dos “Cadernos de Informação de Saúde”. Com a utilização do programa IDBMUN (DATASUS, 2015a), são geradas planilhas Excel®, uma para cada

área geográfica solicitada com indicadores obtidos das diversas bases de dados do Ministério da Saúde. Os dados disponíveis provêm basicamente do sistema TABNET (DATASUS, 2015), que possibilita a pesquisa por Informações de Saúde. Trata-se de um tabulador que torna disponíveis informações sobre o SUS. A consolidação de dados não se processa de maneira automatizada. CERCEAU *et al.* (2001) demonstrou a possibilidade de obtenção de dados descentralizados, do MS e outras fontes de dados externas ao DATASUS, a partir de fontes dispersas na Internet e de forma automática.

A importância na descrição dos experimentos científicos tem aperfeiçoado no meio acadêmico o desenvolvimento da proveniência de dados (CRUZ *et al.*, 2009; GUIMARÃES & CAVALCANTI, 2009).

Estudos de geração automática de documentos XML para representação de bancos de dados na Web foram propostos a partir de esquemas conceituais. MARTINS (2004) empregou *templates* buscando a conversão automática de informações parametrizadas de banco de dados para texto descritivo com informações curriculares baseadas na Plataforma Lattes. Para aplicações multimídia com informação semântica foram implantadas soluções utilizando DTDs XML como base para manipulação e troca de informações (TEIXEIRA, 2005). A tecnologia RDF (*Resource Description Framework*) tem sido empregada neste contexto.

A literatura recente aponta para utilização de geração automática de metadados baseados em contexto para suporte a Web Semântica (FERREIRA, 2006). O conceito de agentes tem sido mencionado ao serem empregados programas de computador que coletam conteúdo da Web de diversas fontes, processam estas informações e compartilham os resultados com outros programas.

O presente estudo contribui com a captura automática de dados de informação de saúde de proveniência difusa, implementando um pipeline para compilação/fusão de dados públicos (multiestruturados) da área da saúde.

3. Materiais e Métodos

A etapa de diagnóstico consiste em levantamento das informações para a Análise da Situação de Saúde com identificação, formulação, priorização e análise dos problemas de saúde em um determinado território (BAHIA, 2009). O território é o local onde se estabelece o atendimento de saúde propriamente dito. Usualmente, a partir de uma unidade mínima de agregação estabelecida como território (por exemplo, o bairro) empregam-se agregações dos territórios de forma ascendente (ou seja, agregam-se bairros para formar o território do município, agregam-se municípios para formar microrregiões, regiões de saúde, macrorregiões, estados etc.).

A identificação dos problemas e necessidades de saúde da população de um município, o pesquisador da equipe municipal deve, basicamente, caracterizar a população, suas condições de vida, seu perfil epidemiológico e a constituição do sistema municipal de saúde (BAHIA, 2009). Importante destacar que cada município tem autonomia para definir as linhas gerais do processo de elaboração do seu PMS, consoante os princípios e diretrizes adotadas na legislação básica e normas complementares do SUS. Esta autonomia denota uma grande quantidade de dados diferentes e não uniformizados, dificultando o agrupamento e o estudo comparativo por município.

3.1. Metodologia de Coleta de Dados

Neste trabalho, usando a regulamentação estabelecida para o SUS (BRASIL, 1990; BRASIL, 2011), foi estabelecida como unidade mínima de agregação o território *municipal* e a agregação segundo a *região de saúde* (isto é, agregação de municípios limítrofes). Desta forma, em junho de 2013 foram obtidas as bases territoriais correspondentes a 5.625 municípios, agregados em 437 regiões de saúde (DATASUS, 2015; DATASUS, 2015a).

O conteúdo foi formado por dados das áreas de saúde pública, privada e de saúde suplementar (planos e seguros de saúde). Foram incluídos todos os municípios dos seguintes estados: Bahia, Alagoas, Minas Gerais, Rondônia, Roraima, São Paulo, Sergipe, Acre, Amapá, Espírito Santo, Rio de Janeiro e Rio Grande do Norte (Figura 1).



Figura 1. Estados selecionados para edição de livros digitais.

Foi realizada adaptação do software CADERNOS (DATASUS, 2015a) para obtenção de dados relacionados aos municípios que possibilita a inserção de tabelas nos modelos ‘municipais’ (planilhas Excel®). Foi realizada adaptação do *script* e da planilha município (por se tratar da unidade mínima de agregação estabelecida).

A partir dos dados obtidos foram gerados os argumentos, que são as frases que serão apresentadas no relatório final, obtidas a partir do tratamento dos dados pela concatenação dos elementos de informação (com flexão de número: singular/plural e heurísticas para formação das frases). Pelo agrupamento dos argumentos, foi estabelecido o texto padronizado com as informações relacionadas a cada município. Os dados empregados são secundários, obtidos de fontes públicas na Internet e não contém dados com variáveis identificadoras de indivíduos.

3.1.1. Fontes de Dados

A partir da proposta de 'DIAGNÓSTICO DE SAÚDE' (CERCEAU *et al.*, 2001) foram obtidos dados relacionados aos seguintes temas:

- I. **Informações Gerais** (Dados Municipais Básicos e Divisão Administrativa)
- II. **Demográfica** (População total, por sexo, e segundo extratos: Menores de 1 ano; Crianças - 1 a 9 anos; Jovens - 10 a 19 anos; Adultos - 20 a 59 anos; Idosos - 60 anos e mais; Mulheres em idade fértil - 10 a 49 anos; Proporção da população feminina em idade fértil).
- III. **Capacidade instalada** (Quantidade de estabelecimentos por tipo de convênio – SUS, ou Particular, ou em Plano de Saúde, Público, ou Privado - e, segundo tipo de atendimento prestado - Internação Ambulatorial, Urgência, Diagnose e terapia, Vigilância epidemiológica e sanitária, Farmácia ou cooperativa)
- IV. **Saúde suplementar** (ou planos e seguros de saúde) - quantidade de vínculos de pessoas aos planos e seguros de saúde.

Os temas I a III foram obtidos a partir das bases de dados do DATASUS (<http://www.datasus.gov.br>), por meio de alterações no *script* do modelo Município aplicado pelo software CADERNOS.

O tema IV foi obtido no TABNET ANS (<http://www.ans.gov.br/anstabnet/>). As informações de Beneficiários por Município foram obtidas para os planos de assistência médica e para aqueles exclusivamente odontológicos, segundo as variáveis: faixa etária, época de contratação, segmentação, sexo e tipo de contratação.

3.1.2. Argumentos

O processo de geração dos argumentos se estabelece por emprego de heurísticas. Foram estabelecidas frases *templates*, com elaboração de diferentes saídas de argumentos, segundo o(s) atributo(s) de entrada.

No Quadro 1, formalizam-se dois exemplos. O tratamento da primeira frase pode ser realizado por uma heurística simples que se caracteriza pela formação da frase usando uma estrutura de seleção simples. No segundo caso, usa-se concatenação, como uma heurística composta.

Quadro 1. Exemplo de heurística simples e composta

Exemplificando: Tendo sido estabelecido um município qualquer ('X'), pretendemos ter a informação sobre a quantidade de pessoas que moram neste município. Para este fim, devemos obter a tabela de quantidade de pessoas residentes, segundo o sexo. Como 'argumento-template' podemos definir como entradas ('Y') dois exemplos.

ARGUMENTO 1 ('Y' representa a quantidade de indivíduos):

- na ausência de dados (Y=0): "O município 'X' não possui habitantes".
- se caso unitário (Y=1): "O município 'X' possui 1 habitante."
- senão (Y>1): "O município 'X' possui 'Y' habitantes."

ARGUMENTO 2 (caso de concatenação)

('H' representa a quantidade de homens; 'M' representa a quantidade de mulheres):

- na ausência de dados (H=M=0): "" (nenhuma informação, pois o município não possui habitantes).

-- início default: "A população está composta por "

- caso somente apareça um sexo (não concatena):

se caso unitário:

* masculino (H=1; M=0): "1 homem."

* feminino (H=0; M=1): "1 mulher."

como caso maior que 1:

* masculino (H>1; M=0) " _____ homens."

* feminino (H=0; M>1) " _____ mulheres."

- caso apareça os dois sexos (então, concatena as informações):

se caso masculino (H=1) "1 homem."

* senão (H>1) " _____ homens."

se caso feminino (M=1) "e 1 mulher."

* senão (M>1) "e _____ mulheres."

3.1.3. Concatenando Informações

Os argumentos foram empregados para gerar as informações dos blocos, segundo a ordenação estabelecida na etapa de coleta de dados (Tabela 1). Os blocos foram organizados em sequência para geração do Diagnóstico de Saúde Municipal.

Tabela 1: Diagnóstico de Saúde Municipal

Informações Gerais	Estrutura Municipal	Saúde Suplementar - Vínculos
<input type="checkbox"/> Dados Básicos Brasil <input type="checkbox"/> Divisão Administrativa Demográfica <input type="checkbox"/> População	<input type="checkbox"/> Rede Ambulatorial <input type="checkbox"/> Produção Ambulatorial <input type="checkbox"/> Rede Hospitalar <input type="checkbox"/> Base Instalada	<input type="checkbox"/> Planos de Assistência médica <input type="checkbox"/> Planos exclusivamente Odontológicos

3.1.4. Tipo de Suporte do Conteúdo

Optou-se por manutenção do interesse primário pela difusão dos dados, com popularização das informações, e com foco no acesso às pessoas e instituições para uso prático dos relatórios. Os dados foram compilados em formato de livros digitais (e-book) com tamanho reduzido e fácil difusão (arquivos PDF, *portable document format*, com menos de 2 MB).

4. Resultados

No início do desenvolvimento, em julho de 2013, foi produzida a primeira versão dos livros incluindo as capitais do Brasil (Saúde Suplementar: Capitais do Brasil), com registros ISBN para versão PDF e para versão impressa. Os dados disponíveis para as 27 capitais estaduais foram: dados populacionais do Instituto Brasileiro de Geografia e Estatística – IBGE (dezembro/2012); quantidade de estabelecimentos por tipo de convênio segundo tipo de atendimento prestado, pelo Cadastro Nacional de Estabelecimentos de Saúde – CNES (junho/2013); e, dados sobre planos e seguros de saúde, pela Agência Nacional de Saúde Suplementar – ANS (dezembro/2012).

A atualização dos dados deveria ser realizada pela ANS a cada três meses. Na ausência de atualização, foi enviada solicitação de informação e busca pela atualização dos dados, usando a Lei de Acesso à Informação (LAI), através do *e-SIC* – Sistema Eletrônico do Serviço de Informação ao Cidadão (BRASIL, 2011a). Finalmente, em agosto de 2013 as informações foram atualizadas para os períodos de março de 2013 e junho de 2013. Por conta disso, houve correção de alguns dados da competência de dezembro de 2012. Assim, foram elaborados e consolidados contendo os dados atualizados para junho de 2013.

Em agosto de 2013 foi publicada a segunda edição do livro digital (*e-book*), contendo as 27 capitais estaduais: SAUDE SUPLEMENTAR: CAPITAIS DO BRASIL (CERCEAU & CERCEAU, 2013). Em seguida, 12 estados foram contemplados (Figura 1), totalizando 2.549 municípios analisados (45,3% do total). Foram publicados 91 livros digitais contendo 123 regiões de saúde (28,1% do total). A Tabela 2 apresenta os quantitativos de regiões de saúde, municípios e livros digitais produzidos.

Desde dezembro de 2013 foram tornados disponíveis 69 e-books através do RESEARCHGATE® (<http://www.researchgate.net/>), totalizando 3.615 visualizações e 2.581 downloads (Tabela 3).

5. Conclusão

A fusão automatizada possibilitou a consolidação da informação a partir de fontes distribuídas e multiestruturas de dados. Através de uma abordagem prática foi possível fazer a implementação computacional de um pipeline de modelo de agregação e fusão de dados na área da gestão da saúde pública. A fusão de dados municipais obtida na

forma de regiões de saúde possibilita utilização das informações em apoio ao gestor de saúde municipal e regional, fortalecendo as práticas do SUS ao promover o reconhecimento das demandas e divulgar o planejamento na saúde pública, privadas e na saúde suplementar (isto é, planos e seguros de saúde). A oportunidade de geração de livros digitais populariza as informações e introduz inovação na área de saúde.

Tabela 2. Quantitativo de regiões de saúde, municípios e livros digitais, segundo Brasil Regional e Estados.

BRASIL REGIONAL ESTADOS	QUANTIDADE DE REGIOES DE SAUDE	QUANTIDADE DE MUNICIPIOS	TIPO DE SUPORTE PDF (e-book)
Capitais*	-	27	2
Nordeste			33
Alagoas	10	102	5
Bahia	31	417	19
Rio Grande do Norte	7	167	6
Sergipe	7	75	3
Norte			5
Acre	3	22	1
Amapá	2	16	1
Rondônia	4	52	2
Roraima	1	15	1
Sudeste			51
Espírito Santo	4	78	3
Minas Gerais	28	853	26
Rio de Janeiro	9	92	4
São Paulo	17	645	18
TOTAL GERAL	123	2.561**	91

* publicadas 1ª e 2ª versões; ** capitais contempladas estão duplicadas (2549 municípios exclusivos).

Tabela 3. Quantitativo de visualizações e downloads de livros digitais, segundo Brasil Regional e Estados.

BRASIL REGIONAL ESTADOS	QUANTIDADE DE VISUALIZAÇÕES	QUANTIDADE DE DOWNLOADS
Capitais	150	78
Nordeste	845	560
Alagoas*	-	-
Bahia	845	560
Rio Grande do Norte	-	-
Sergipe		
Norte	43	21
Acre	43	21
Amapá	-	-
Rondônia	-	-
Roraima	-	-
Sudeste	2.577	1.922
Espírito Santo	371	142
Minas Gerais	1.350	880
Rio de Janeiro	-	-
São Paulo	856	900
TOTAL GERAL	3.615	2.581

* livro(s) do Estado não se encontra(m) depositado(s) no Researchgate®.

Na realização deste projeto foi registrado que podemos obter o aperfeiçoamento dos mecanismos de atualização de informações governamentais empregando a Lei de Acesso a Informações, pressionando o ente público para transparência e revisão de suas bases de dados.

6. Agradecimentos

Agradecemos ao CNPq e a CAPES pelas bolsas concedidas e ao professor Sergio Manuel Serra da Cruz/UFRRJ pela revisão do manuscrito.

Bibliografia

- BAHIA. (2009). **Manual Prático De Apoio À Elaboração De Planos Municipais De Saúde**. Acesso em 03/05/2015, disponível em <http://bit.ly/1dm1OOy>
- BRASIL. (1990). **Lei N° 8.080**. Acesso em 03/05/2015, disponível em http://www.planalto.gov.br/ccivil_03/leis/l8080.htm
- BRASIL. (2002). **Norma Operacional Da Assistência À Saúde / Sus - Noas-Sus 01/02**. Acesso em 15/04/2015, disponível em <http://bit.ly/1JKKela>
- BRASIL. (2009) **Sistema de planejamento do SUS : uma construção coletiva : instrumentos básicos**– 2. ed. – Brasília : Ministério da Saúde, 2009
- BRASIL. (2011). **Decreto N° 7.508**. Acesso em 03/05/2015, disponível em https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/decreto/d7508.htm
- BRASIL. (2011a). **LEI N° 12.527**. Acesso em 03/05/2015, disponível em http://www.planalto.gov.br/ccivil_03/_Ato2011-2014/2011/Lei/L12527.htm
- BRASIL. (2015). e-SIC. **Sistema eletrônico do serviço de informação ao cidadão**. Acesso em 03/05/2015, disponível em <http://www.acesoainformacao.gov.br/>
- CERCEAU, R., SILVA, A., ANDRADE Jr., F., & ALMEIDA, R. (2001). **O uso da internet no apoio à gerência dos serviços de saúde no brasil**. Memórias II Cong. Latinoamericano Ing.Biomédica <http://dx.doi.org/10.13140/2.1.3825.8560>
- CERCEAU, R.; CERCEAU, R. (2013) **Saúde suplementar : capitais do brasil**. 2. ed. Rio de Janeiro: Cerceau. <http://dx.doi.org/10.17655/9788567211657>
- COSTA, M.M.; CUNHA, M.B (2014). **O bibliotecário no tratamento de dados oriundos da e-science: considerações iniciais**. *Perspect.ciênc.inf.*, Belo Horizonte, acesso 12/06/2015. <http://dx.doi.org/10.1590/1981-5344/1900>.
- CRUZ, S.M.S.; CAMPOS, M.L.M.; MATTOSO, M. **Towards a Taxonomy of Provenance in Scientific Workflow Management Systems**. *In: IEEE 2009* <http://dx.doi.org/10.1109/SERVICES-I.2009.18>
- DATASUS. (2015). **TABNET**. Acesso em 03/05/2015, disponível em <http://datasus.saude.gov.br/informacoes-de-saude/tabnet>
- DATASUS. (2015a). **Cadernos de Informação de Saúde**. Acesso em 03/05/2015, disponível em <http://tabnet.datasus.gov.br/tabdata/cadernos/cadernosmap.htm>
- FERREIRA, E. C. (2006). **Geração Automática de Metadados: uma contribuição a Web Semântica**. Tese (Doutorado), USP, Escola Politécnica.
- GUIMARÃES, M.P.; CAVALCANTI, M.C.R. **Proveniência de dados na área de Bioinformática**. IME. Seção de Engenharia de Computação. 2009
- MARTINS, S. R. (2004). **Geração automática de textos em plataformas de governo eletrônico: um estudo de caso na Plataforma Lattes**. Dissertação, UFSC, Centro Tecnológico. Programa Pós-Graduação Engenharia de Produção.
- TEIXEIRA, E. C. (2005). **Utilizando XML para publicação de dados multimídia na web**. Dissertação, UFSC, Programa Pós-Graduação Ciência da Computação.