

## Geoprocessamento dos dados da saúde: o tratamento dos endereços

Geoprocessing of health data:  
treatment of information on addresses

Daniel Albert Skaba <sup>1</sup>  
Marília Sá Carvalho <sup>2</sup>  
Christovam Barcellos <sup>3</sup>  
Paulo Cesar Martins <sup>1</sup>  
Sonia Luiza Terron <sup>1</sup>

### Abstract

*This paper analyzes the current status of address data in the Brazilian Health Information System (SIS), with a view towards mapping large-city health events in geographic information systems (GIS) for risk analysis and evaluation. It is thus necessary to geocode these events to small geographic areas inside city limits. This study used a sample from the Reportable Health Events Information System (SINAN) database and also proposes alternatives to work with this large amount of events. Approximately 50% of addresses were referenced automatically and another 19% through conventional search. Data losses occurred mainly due to incomplete or inconsistent address data.*

*Geographic Information Systems; Risk Assessment; Communicable Diseases*

### Introdução

O georreferenciamento dos eventos de saúde é importante na análise e avaliação de riscos à saúde coletiva, particularmente as relacionadas com o meio ambiente e com o perfil sócio-econômico da população. Os Sistemas de Informações Geográficas (SIG), conjunto de ferramentas utilizadas para a manipulação de informações espacialmente apresentadas <sup>1</sup>, permitem o mapeamento das doenças e contribuem na estruturação e análise de riscos sócio-ambientais <sup>2,3</sup>. Para estas análises é necessária a localização geográfica dos eventos, associando informações gráficas (mapas) a bases de dados de saúde, alfanuméricas. O georreferenciamento de um endereço, definido como o processo de associação deste a um mapa terrestre, pode ser efetuado de três formas básicas: associação a um ponto, a uma linha ou a uma área <sup>4</sup>. O elemento geométrico resultante, associado a uma base de dados, é a unidade utilizada nos SIG.

Diversos trabalhos utilizam como fonte de dados os registros dos vários Sistemas de Informação em Saúde (SIS – <http://www.funasa.gov.br>), que contam com séries históricas de até vinte anos, em alguns casos. Nestas análises utiliza-se, na maioria dos casos, o município ou o bairro informado, para análise da distribuição espacial dos eventos a serem estudados. Nos municípios mais densamente povoados esta escala geográfica já é insuficiente, sen-

<sup>1</sup> Fundação Instituto Brasileiro de Geografia e Estatística, Rio de Janeiro, Brasil.

<sup>2</sup> Escola Nacional de Saúde Pública, Fundação Oswaldo Cruz, Rio de Janeiro, Brasil.

<sup>3</sup> Centro de Informação Científica e Tecnológica, Fundação Oswaldo Cruz, Rio de Janeiro, Brasil.

#### Correspondência

D. A. Skaba  
Fundação Instituto Brasileiro de Geografia e Estatística.  
Av. Brasil 15671,  
Rio de Janeiro, RJ  
21241-051, Brasil.  
dskaba@superig.com.br

do necessária a localização em áreas menores, principalmente nas áreas urbanas. Para aumentar a escala da análise, a utilização dos setores censitários vem sendo proposta em diversos trabalhos <sup>5</sup>, utilizando como fonte de informação o endereço dos registros dos SIS, georreferenciando-os ao setor censitário.

O objetivo deste trabalho é avaliar a qualidade dos endereços armazenados nos Sistemas de Informações de Agravos de Notificação (SINAN) e, com base na análise de uma amostra destes, propor alternativas para sua utilização, em grandes volumes, em SIGs, de forma a contribuir para viabilização, no país, de análise espacial em microáreas.

## Material e metodologia

### Bases de dados de referência

Para servir de base para localização dos endereços foram utilizados o *Cadastro de Segmentos de Logradouros* (Cadlog) e o arquivo preliminar das *Folhas de Coleta*, ambos criados a partir do *Censo 2000* da Fundação Instituto Brasileiro de Geografia e Estatística (IBGE). O Cadlog apresenta uma relação de logradouros por Setor Censitário <sup>6</sup>, com numeração de início e final do segmento, enquanto as *Folhas de Coleta* apresentam os endereços das unidades visitadas no censo. Para a localização manual, além dos cadastros, foram utilizados os mapas de localidades e de setores do *Censo 2000* <sup>7</sup>.

### Bases da saúde

Neste trabalho foram utilizados três arquivos de doenças transmissíveis, provenientes do SINAN, oriundos de: Campinas, com 410 registros de dengue (janeiro a junho/2003); Macapá, com 1.140 registros de hepatite (1999-2003) e Rio de Janeiro com 2.544 casos de leptospirose (1990-1996).

### Metodologia

Este trabalho foi desenvolvido em quatro fases. A primeira fase consistiu na compatibilização das tabelas dos SIS com as tabelas bases de consultas (cadastro de segmentos de logradouros e folhas de coleta do IBGE), separando tipo, título e nome do logradouro e número da unidade de residência em subcampos. Neste procedimento são utilizadas tabelas de tipos de logradouros (rua, avenida, travessa etc.), e de títulos (doutor, governador, presidente etc.), além das abreviações conhecidas destes, geradas pelo IBGE para

o *Censo 2000*. Para o processo foi desenvolvido um aplicativo específico, de conversão de *strings*, em VisualBasic 6<sup>®</sup> adaptado às características de cada tabela fonte da informação.

Na segunda fase foi feita a busca automática dos endereços, que se deu em vários níveis, a partir da combinação entre os subcampos (tipo, título, nome e número) da informação de endereço. O resultado desta busca leva à localização de apenas um setor, mais de um setor ou nenhum setor. Nos dois últimos casos, passa-se às próximas fases de busca manual.

A terceira fase foi uma pesquisa manual dos endereços que apontaram para mais de um setor censitário. Foi uma operação relativamente rápida, facilitada pela relação de setores selecionados pela fase anterior e utilizando outras informações contidas na tabela, como os pontos de referência e complementos. A quarta e última fase consistiu na pesquisa manual dos endereços em que não há setores indicados automaticamente.

## Resultados

Na Tabela 1 são apresentados os resultados da pesquisa com a amostra selecionada. Na coluna de busca automática é apresentado o número de casos em que foi encontrado apenas um setor automaticamente. Na seguinte, de busca manual, é apresentado o número de casos em que foi possível especificar o setor censitário a que pertence o endereço em uma pesquisa manual aos cadastros e mapas. Na penúltima coluna (não encontrado) estão os casos em que os endereços estão incompatíveis com os cadastros e mapas ou poderiam estar em mais de um setor.

## Discussão

A escolha dessas amostras não pretendeu fazer uma abordagem sistemática de todas as possibilidades de registro de endereço. A atual estrutura dos SIS indica que existem pelo menos cinco formatos de registro de endereços, onde o único campo de fato padronizado é o município. Além disso, buscamos nestes exemplos mostrar os problemas em situações bem variadas, mesmo que alguns destes sistemas já tenham alterado a máscara de entrada de dados. Inclusive porque o uso de georreferenciamento pode permitir recuperar séries históricas de importância para a saúde pública.

A primeira grande diferença, que justifica em parte o grau de acerto, diz respeito às dife-

Tabela 1

Distribuição dos resultados do georreferenciamento e por arquivo.

Arquivo	Em branco	Busca automática	Busca manual	Não encontrado	Total
Campinas (dengue)	19 (5%)	202 (49%)	139 (34%)	50 (12%)	410
Macapá (hepatite)	1 (0%)	821 (72%)	228 (20%)	90 (8%)	1.140
Rio de Janeiro (leptospirose)	95 (4%)	1.233 (48%)	421 (17%)	795 (31%)	2.544
<b>Total</b>	115 (3%)	2.256 (55%)	788 (19%)	935 (23%)	4.094

renças no modo de armazenamento do endereço. No arquivo de dengue de Campinas, o endereço está todo contido em um único campo, inclusive as informações complementares. No de leptospirose do Rio de Janeiro, o ponto de referência está em um campo separado e há um outro campo com a informação de bairro. Por último, no arquivo de hepatite de Macapá, o endereço está composto por quatro campos: nome do logradouro, número de porta, complemento e referência. Estas diferenças implicam procedimentos específicos na primeira fase do desenvolvimento (compatibilização das tabelas). Além disso, a qualidade dos dados armazenados é também fator determinante. Em duas amostras há cerca de 5% dos endereços em branco. Nestes não há qualquer possibilidade de se mapear o evento dentro do município.

Outro fator importante é a natureza do evento. Algumas doenças, como é o caso, por exemplo, da leptospirose, têm frequência maior em áreas sem um bom saneamento básico, como periferias e favelas, locais onde os endereços não apresentam regularidade, havendo grande perda de informação. O tratamento de endereços em favelas, acampamentos e invasões é um problema que merece estudo especial.

No Brasil, nos municípios onde se investiu maciçamente na construção de SIGs, como é o caso, por exemplo, de Belo Horizonte (Minas Gerais) <sup>8</sup> e Porto Alegre (Rio Grande do Sul), o

georreferenciamento parte desta base cartográfica detalhada. Entretanto, o custo desta estratégia é muito alto e transcende o setor saúde. A estratégia alternativa, que também permitiria bons resultados na microlocalização de eventos de saúde, baseia-se na criação de bons cadastros alfanuméricos e na padronização da entrada de endereços. Na maioria das cidades, as regras básicas de padronização de endereços não são obedecidas, principalmente nas periferias e nas áreas de ocupação irregular, como as favelas. Algumas cidades apresentam alto índice de numeração irregular e outras (como Brasília e Palmas) utilizam endereços por quadra e não por logradouro. Além disso, os mapas urbanos digitais, quando disponíveis, não seguem um padrão único.

A alternativa utilizada foi a de se referenciar os endereços ao setor censitário, que está disponível para todo o país e tem referência aos dados sócio-econômicos coletados nos censos e pesquisas do IBGE.

O problema aqui analisado tem impacto substantivo em diversas questões, que vão de estudos e pesquisas onde o mote é a análise geográfica, às atividades de vigilância epidemiológica, sendo particularmente relevante na perspectiva da construção de uma vigilância em saúde de base territorial, integrando a ocorrência dos agravos registrados nos diferentes SIS com aspectos ambientais relevantes.

## Resumo

*Este trabalho faz uma análise da situação atual das informações de endereços nos Sistemas de Informações em Saúde (SIS) em alguns municípios, visando a sua utilização em Sistemas de Informações Geográficas (SIG), para a análise e avaliação de riscos dos eventos de saúde pública em grandes cidades, com localização destes eventos em áreas intra-urbanas. Utiliza como base de dados uma amostra dos cadastros do Sistema de Informações de Agravos de Notificação (SINAN) e tem como objetivo propor alternativas para aproveitamento de grandes volumes de dados já existentes. Nas amostras trabalhadas, cerca de metade dos endereços foi localizada automaticamente e aproximadamente 19%, em uma busca manual. As principais perdas ocorreram por falta de completitude ou de consistência nos dados de endereço.*

*Sistemas de Informação Geográfica; Avaliação de Risco; Doenças Transmissíveis*

## Colaboradores

D. A. Skaba, M. S. Carvalho e C. Barcellos participaram na organização e elaboração do texto e seleção dos dados do SINAN. D. A. Skaba, S. L. Terron e P. C. Martins participaram na definição e desenvolvimento do aplicativo utilizado e na estruturação dos cadastros e mapas do IBGE.

## Referências

1. Aronoff S. Geographic information systems: a management perspective. Ottawa: WDL Publications; 1990.
2. Barcellos C, Ramalho W. Situação atual do geoprocessamento e da análise de dados espaciais em saúde no Brasil. *Revista IP – Informática Pública* 2002; 4:221-30.
3. Richards TB, Rushton G, Brown CK, Fowler L. Geographic information and public health: mapping the future. *Public Health Rep* 1999; 114:359-73.
4. Eichelberger P. The importance of addresses – The locus of GIS. In: *Proceedings of the URISA 1993 Annual Conference*. Park Ridge: Urban and Regional Information Systems Association; 1993. p. 200-11.
5. Carvalho MS, Cruz OG. Análise espacial por microáreas: métodos e experiências. In: Veras RP, Barreto ML, Almeida Filho N, organizadores. *Epidemiologia: contextos e pluralidade*. Rio de Janeiro: Editora Fiocruz; 1998. p. 79-89.
6. Fundação Instituto Brasileiro de Geografia e Estatística. *XI Recenseamento Geral do Brasil: manual de delimitação de setores*. Rio de Janeiro: Fundação Instituto Brasileiro de Geografia e Estatística; 1997.
7. Skaba DA, Terron SL. Mapas urbanos digitais do censo 2000: uma abordagem tecnológica. *Revista IP – Informática Pública* 2003; 5:205-19.
8. Oliveira CM. Lançamento de endereços no geoprocessamento de Belo Horizonte. In: *Anais do XXI Congresso Brasileiro de Cartografia [CD-ROM]*. Sociedade Brasileira de Cartografia; 2003.

---

Recebido em 14/Jun/2004

Versão final reapresentada em 18/Ago/2004

Aprovado em 20/Ago/2004