

Uma Arquitetura para *Linked Data* para Armazenamento e Disponibilização de Dados de Saúde Pública

Romulo Eduardo Garcia Moraes¹, Flavia Cristina Bernardini¹, José Viterbo Filho²

¹Instituto de Ciência e Tecnologia
Universidade Federal Fluminense (UFF) – Rio das Ostras, RJ, Brasil

²Instituto de Computação
Universidade Federal Fluminense (UFF) – Niterói, RJ, Brasil

{romuloeduardo, fcbernardini}@id.uff.br, viterbo@ic.uff.br

Resumo. *O principal ator no cenário de Cidades Inteligentes é o cidadão, para quem, no Brasil, os serviços públicos oferecidos são precários. Um dos caminhos para esse problema é a construção de sistemas computacionais para facilitar a gestão das secretarias de saúde municipais, e oferecer aos cidadãos mecanismos para acesso à informação. O SINAN é o mais importante sistema de vigilância epidemiológica do Ministério da Saúde no Brasil. Em muitos municípios brasileiros, os dados são coletados e submetidos ao Ministério da Saúde, não sendo utilizados para auxiliar a gestão municipal da saúde. Neste trabalho, propomos uma arquitetura para disponibilização de dados do SINAN no padrão Linked Data. Também, apresentamos uma ferramenta web que faz uso desses dados para visualização de diferentes agravos e notificações. Tal ferramenta pode permitir melhor análise desses dados.*

1. Introdução

Segundo [LEE et al. 2013], o conceito de Cidade Inteligente está embasado em seis principais dimensões: economia, mobilidade, ambiente, pessoas, vida e governança inteligentes. Os autores de [GIFFINGER and GUDRUN 2010] ainda definem que “uma cidade é inteligente quando seus investimentos em capital humano e social, em transporte urbano e infraestrutura de TIC alimentam o desenvolvimento econômico sustentável e uma melhor qualidade de vida, com sábio gerenciamento de recursos naturais, através do governo participativo”. Segundo [COCCHIA 2014], a eficácia dos serviços é um aspecto necessário para as cidades serem mais inteligentes.

O Sistema Único de Saúde (SUS) constitui um dos maiores e mais modernos sistemas de saúde pública do mundo, beneficiando cerca de 180 milhões de brasileiros através do fornecimento de acesso integral, universal e gratuito a diversos serviços de saúde [FIOCRUZ — Fundação Oswaldo Cruz 2016]. Os princípios e objetivos do SUS são excelentes, mas não tem sido possível implantá-los da maneira desejada, pois há sérias dificuldades relacionadas tanto com seu financiamento quanto com a eficiência administrativa de sua operação. Um reflexo dessa ineficiência e precarização é mostrada numa pesquisa realizada pelo Instituto Data Folha em Outubro de 2015, na qual se verifica que aproximadamente 54% dos brasileiros avaliam o SUS como ruim ou péssimo. Discute-se, assim, a necessidade da melhoria na gestão dos limitados recursos financeiros disponíveis ao SUS.

Uma das necessidades para auxiliar o gestor de saúde é oferecer dados com qualidade, assim como oferecer ferramentas computacionais que possam identificar características da população em questão. Daí, o gestor poderá saber com grande precisão o estado atual de saúde da população e suas necessidades. Por outro lado, também é importante que a população tenha acesso a tais informações, para que tenha conhecimento da situação da saúde em seu município como um todo, além disso, é importante permitir que a população seja alertada quanto aos cuidados necessários na prevenção de diversos tipos de doenças/agravos e na contenção e prevenção do surgimento de epidemias, permitindo assim uma maior transparência nesses processos.

Este trabalho está dividido como segue: Na Seção 2 é apresentado o objetivo deste trabalho. Na Seção 3, são apresentados conceitos de Web Semântica e dados ligados, conceitos sobre o SINAN (Sistema de Informação de Agravos de Notificação) e outros conceitos necessários para a compreensão deste trabalho

2. Objetivos da Pesquisa

Atualmente, no Brasil, uma das ferramentas utilizadas pelos gestores de saúde para coletar e analisar dados sobre a saúde da população brasileira, é o SINAN que, segundo [LAGUARDIA et al. 2004], foi construído para ser operado nas Unidades de Saúde, em nível municipal, para coletar e processar dados sobre agravos de notificação no território nacional. O processamento eletrônico é feito nos níveis centrais das secretarias municipais de saúde. É alimentado pela notificação e investigação de casos de doenças e agravos que constam da Lista Nacional de Doenças de Notificação Compulsória (LNDC), e todos os dados são centralizados no Ministério da Saúde (MS). Os dados são preenchidos em formulários, e a secretaria de saúde regional é responsável por transferir esses dados para formato eletrônico. No entanto, o SINAN foi construído para centralização dos dados no MS, e não para auxiliar a gestão da saúde em nível municipal. Por outro lado, esses dados podem ser utilizados pela Secretaria da Saúde para auxiliar na melhoria da prestação de serviço de saúde ao cidadão. Para isso, disponibilizar os dados em um formato que possa ser utilizado como *Linked Data* e/ou na Web Semântica pode facilitar o processo de gestão [VITERBO et al. 2012]. O objetivo deste trabalho é propor um formato para disponibilização de dados do SINAN em RDF. Para isso, foi utilizado o Esquema RDF para definição do modelo de dados. Foi desenvolvida uma ferramenta Web para processar os dados nesse formato e visualização por parte de gestores na área de saúde.

3. Contribuições Esperadas

O principal acesso da população às informações governamentais abertas deverá ser via Web. A Web Semântica [BERNERS-LEE et al. 2001] possibilita serviços inteligentes, tais como agentes intermediadores de informações, agentes de busca, filtros de informações, etc. A Web Semântica somente é possível quando níveis de interoperabilidade são estabelecidos, facilitada pelos esforços de padronização da W3C, que inclui o RDF (*Resource Description Framework*) e o Esquema RDF. Um Esquema RDF provê um modelo de dados do tipo OAV (Objeto-Atributo-Valor) para meta-dados. Assim, o Esquema RDF permite a definição de um vocabulário particular que pode ser usado pelos atributos RDF. Uma maneira de disponibilizar os dados representados em RDF é através

da notação JSON (*JavaScript Object Notation*), um formato de fácil escrita e leitura para seres humanos.

O SINAN [LAGUARDIA et al. 2004] foi projetado para ser operado pelas unidades de saúde em todo o território nacional. As doenças e agravos sujeitos a notificação são os que constam na LNDC, e é facultado a estados e municípios incluírem outros problemas de saúde. Para alimentar o SINAM, são utilizados dois tipos de formulários padronizados. A **Ficha de Notificação** (FN) é utilizada para comunicar uma ocorrência de determinada doença ou agravo à saúde que conste na LNDC, que pode ser (i) Notificação Individual, relacionada à suspeita da ocorrência de problema de saúde de notificação compulsória de interesse nacional, estadual ou municipal; (ii) Notificação Negativa, relacionada à não ocorrência de doenças de notificação compulsória, utilizada para indicar que os profissionais de saúde e o sistema de vigilância da área estão alertas para a ocorrência de tais eventos; e (iii) Notificação de Surto, relacionada a agravos de origem desconhecida, situações epidêmicas de doenças que não constam na LDNC, e surto de doenças da LDNC com alteração no padrão epidemiológico e casos agregados das doenças que constam na LDNC, mas cuja taxa de notificações seja elevada. O formulário para o preenchimento das fichas é único para os três tipos de notificação. A **Ficha de Investigação** (FI) é um roteiro de investigação, utilizada desde a suspeita de infecção até a conclusão do caso, e permite a inserção de dados que possibilitam identificação da fonte de infecção e mecanismos de transmissão da doença/agravo. A FI permite a realização de uma análise mais detalhada do caso investigado, e, conseqüentemente, a geração de dados mais precisos em relação a doença/agravo investigado. Para cada doença/agravo que consta na LNDC, há uma Ficha de Investigação específica¹.

Para este trabalho, foram abstraídos conceitos das FNs e FIs para permitir uma disponibilização de dados mais eficaz. As FN e FI possuem seções em comum: (i) **Dados Gerais** — dados sobre o tipo de notificação, que, na FN, pode ser preenchido com um dos três tipos de notificação, e, na FI, pode conter o agravo/doença, a data de notificação, UF, município e unidade de saúde onde foi feita a notificação, e data dos primeiros sintomas; (ii) **Dados de Identificação Individual**, contendo dados pessoais do paciente; e (iii) **Dados Residenciais**, contendo dados de endereço e contato do paciente. As FIs e FNs possuem seções para a adição de dados complementares. A FI exige o preenchimento de uma quantidade maior de campos, que variam de ficha para ficha, de acordo com a doença/agravo investigado. Já as seções **Conclusão** (dados sobre a conclusão da investigação) e **Dados Laboratoriais** (dados sobre exames realizados no paciente) são comuns a todas as FIs.

Na Figura 1 apresentamos uma arquitetura para coleta, distribuição e análise dos dados. FNs e FIs devem ser importados do SINAN. Utilizando o esquema RDF que denominamos *finsinan*, os formulários são convertidos para JSON e podem ser armazenados em uma base de dados. Deve ser observado que, devido aos dados de saúde serem sigilosos, a implementação de tal arquitetura em uma secretaria de saúde deve levar em conta a questão da segurança no acesso a tais dados. Uma versão compacta do

¹Modelos das FIs e FNs podem ser encontrados em <http://www.saude.sp.gov.br/cve-centro-de-vigilancia-epidemiologica-prof.-alexandre-vranjac/homepage/downloads/fichas-de-investigacao> e <http://www.visa.goias.gov.br/post/ver/165260/fichas-de-notific>

Esquema RDF finsinan é ilustrado na Figura 2². Temos cinco classes principais: Ficha (ficha), Unidade de Saúde (unidadeDeSaude), Paciente (paciente), Funcionário (funcionario), e Agravado/Doença (agravoDoenca). além da classe Ficha de Investigação (fichaDeInvestigacao), que é uma subclasse de Ficha (rdfs:subClassOf). A classe agravoDoenca representa um dos agravos ou doenças que constam na LNDC, e sua importância é ter um modelo de representação de dados detalhados sobre o agravo ou doença (agente patogênico, sintomas, método de contração, etc.). A classe unidadeDeSaude foi criada com o objetivo de reunir informações importantes sobre as unidades de saúde onde são realizadas as notificações e investigações. A classe funcionario representa um funcionário da área de saúde e que trabalha (trabalhaEm) em uma unidade de saúde. Caso um funcionário faça uma investigação, ele será representado como um investigador (investigador), que é uma subclasse de Funcionário, e caso ele faça uma notificação, será representado como um notificante (notificante). A classe ficha reúne as seções que são comuns as FIs e FNs — numeração da ficha (numero), dados gerais (notificacao) e o agravo/doença. Os dados das seções de identificação individual (dadosPessoais) e dados residenciais (dadosResidenciais) foram movidos para a classe paciente, que por sua vez, passou a ser uma propriedade da classe ficha. Os dados comuns a todas as FIs e FNs são representados no esquema pela classe ficha. A subclasse fichaDeInvestigacao contém os dados complementares das FIs — dados laboratoriais (dadosLaboratoriais), de conclusão (conclusao), e de hospitalização (hospitalizacao). Os dados que não são comuns a todas as FIs, mas específicas para cada agravo/doença, devem ser representados como uma subclasse de fichaDeInvestigacao, que devem ter a propriedade agravoDoenca preenchida com os dados do agravo/doença a qual se referem.

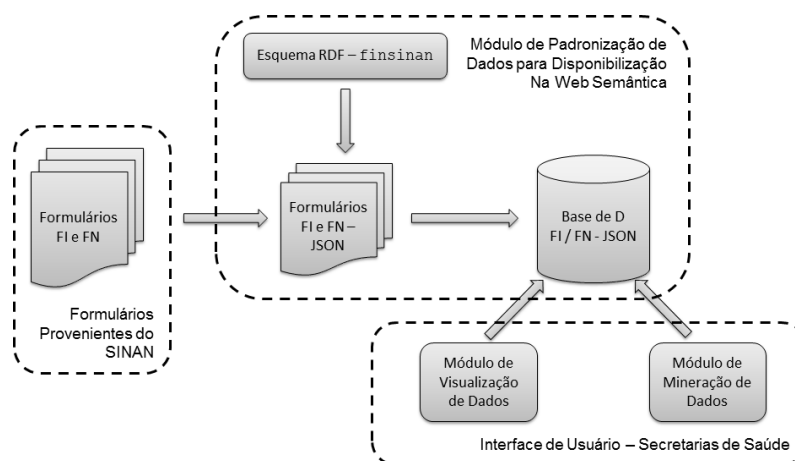


Figura 1. Processo de distribuição e processamento dos dados para auxiliar a gestão da saúde nas Secretarias Municipais de Saúde

4. Resultados já Alcançados

Para avaliar a arquitetura proposta, construímos uma ferramenta Web para visualização de dados, mostrando uma das potencialidades de uso da arquitetura³. Na Figura 3, é mos-

²O esquema completo pode ser acessado em <https://github.com/regmoraes/sis/blob/master/SisViDA, arquivo Finsinan RDF Schema>

³O código da ferramenta pode ser obtido em <https://github.com/regmoraes/sis/tree/master/SisViDA>.

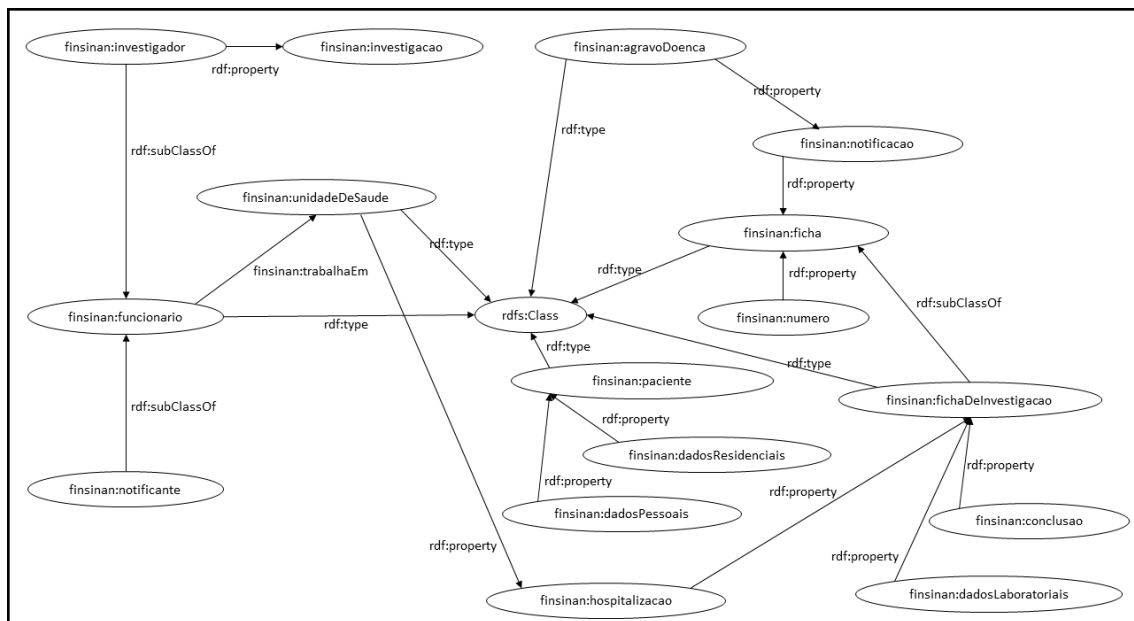


Figura 2. Versão compacta do Esquema RDF para Fichas de Investigação

trada a interface para visualização dos dados, onde é possível visualizar as notificações e investigações baseadas na sua geolocalização, extraída de finsinan:dadosResidenciais. Ilustramos ocorrências de casos fictícios de dengue que ocorreram na cidade de Rio das Ostras. Essa interface foi apresentada para pessoas do governo local de Rio das Ostras, que acharam a proposta bastante interessante.



Figura 3. Processo de distribuição e processamento dos dados para auxiliar a gestão da saúde nas Secretarias Municipais de Saúde

Como trabalho relacionado, podemos citar o Sistema de Monitoramento Diário de Agravos (SIMDA), disponível em <http://tcl.sms.fortaleza.ce.gov.br/simda/login/auth>, utilizado pela Secretaria Municipal de Saúde da Prefeitura de Fortaleza, que tem como objetivo auxiliar no monitoramento, visualização e controle de casos de dengue, varicela e diarreia notificados e/ou investigados na cidade de Fortaleza. O SIMDA é alimentado diariamente com dados provenientes do SINAN, gerados pelas unidades locais de saúde. No entanto, o SIMDA é limitado na visualização de dados — apenas três tipos de agravos e doenças —, não está disponível para outras prefeituras,

e não propõe um modelo de unificação, para futuramente poderem ser utilizados como dados ligados. Com a utilização da arquitetura proposta, é possível armazenar os dados referentes a notificações e investigações de quaisquer tipo agravos/doenças que constam na LNDC, permitindo que, futuramente, os dados possam ser ligados a outras fontes de dados. Caso essa sistematização seja adotada em todos os municípios, extração de novos conhecimentos em nível municipal, estadual e até nacional pode ser facilitada.

5. Conclusões e Trabalhos Futuros

Neste trabalho, foi apresentada uma arquitetura para disponibilização e análise de dados do SINAN para *Linked Data*. Para ilustrar uma possibilidade de uso da arquitetura, foi desenvolvido um sistema web para visualizar os dados, que permite identificar os principais agravos ocorridos bem como a distribuição desses agravos no município. Tal sistema permite que tanto a gestão na saúde pública quanto o acesso a informações relevantes à população sejam mais transparentes. Atualmente, está em discussão no município de Rio das Ostras o uso desse sistema por parte de sua Secretaria de Saúde. Futuramente, com a distribuição dos dados em formato RDF/JSON, técnicas e métodos para ligação de dados e descoberta de novos conhecimentos podem ser propostos e avaliados junto a essas secretarias municipais.

Agradecimentos

Os autores gostariam de agradecer ao suporte financeiro do CNPq através do edital Universal - MCTI/CNPq Número 14/2013.

Referências

- BERNERS-LEE, T., HENDLER, J., and LASSILA, O. (2001). The semantic web. *Scientific American*.
- COCCHIA, A. (2014). *Smart City: How to Create Public and Economic Value with High Technology in Urban Space*, chapter Smart and Digital City: A Systematic Literature Review. Chapter 2. Progress in IS Series, Springer Verlag.
- FIOCRUZ — Fundação Oswaldo Cruz (2016). SUS. [online]. Disponível em <http://pensesus.fiocruz.br/sus>. Acessado em 12/05/2016.
- GIFFINGER, R. and GUDRUN, H. (2010). Smart cities ranking: an effective instrument for the positioning of cities? *ACE: Architecture, City & Environ.*, 4(12):7–25.
- LAGUARDIA, J., DOMINGUES, C., CARVALHO, C., LAUERMAN, C., MACÁRIO, E., and GLATT, R. (2004). Sistema de informação de agravos de notificação em saúde (sinan): desafios no desenvolvimento de um sistema de informação em saúde. *Epidemiologia e Serviços de Saúde*, 13(3).
- LEE, J. H., PHAALB, R., and LEE, S. (2013). An integrated service-device-technology roadmap for smart city development. *Technological Forecasting and Social Change*, 80(2):286–306.
- VITERBO, J., GAMA, V., SARAIVA, D., CASANOVA, M., SALAS, P., BREITMAN, K., MAGALHÃES, R., FRANZOSI, E., and CHAVES, M. (2012). Open government data in brazil. *IEEE Intelligent Systems*, 27:45–49.