

# Identificação e classificação de casos de violência com o uso de *Data Science*

Eduardo R. V. Duarte<sup>1</sup>, Anilton S. Garcia<sup>1</sup>, Jair A. L. Silva<sup>1</sup>, Edleusa G. F. Cupertino<sup>2</sup>

<sup>1</sup>Programa de Pós-Graduação em Engenharia Elétrica  
Universidade Federal do Espírito Santo - Vitória - ES, Brasil

<sup>2</sup>Secretaria de Saúde do Espírito Santo  
Vitória - ES, Brasil

{eduardorvduarte@gmail.com, anilton@inf.ufes.br, jair.silva@ufes.br,  
edleusagfcupertino@gmail.com}

**Abstract.** *This paper proposes the creation of a database, semantically structured, where the data of the victims, made available in textual records, are organized with the purpose of enabling the use of tools to generate intelligence related to the problem of violence. From the database, the objective is to develop a set of tools using Data Science techniques in order to identify, after integration of the information, the possible causes and consequences of these cases of violence and, based on intelligent reasoning techniques, propose actions and measures that can prevent the most common cases of violence and / or mitigate its consequences.*

**Resumo.** *Este trabalho propõe a construção de um banco de dados, semanticamente estruturado, onde os dados dos registros das vítimas, disponibilizados em fichas textuais sejam organizados com a finalidade de possibilitar o uso de ferramentas para geração de inteligência em relação ao problema da violência. A partir da base de dados, o objetivo é desenvolver um conjunto de ferramentas com o uso de técnicas de Ciência de Dados de modo a possibilitar a identificação, após a integração das informações, das possíveis causas e consequências desses casos de violência e, com base em técnicas de raciocínio inteligente, propor ações e medidas que possam prevenir os casos mais comuns de violência e/ou amenizar suas consequências.*

## 1. Introdução

No Brasil e no mundo, o impacto da morbimortalidade por causas externas (violências e acidentes) constitui uma das maiores preocupações para chefes de Estado e dirigentes do setor Saúde. No Brasil, as violências e os acidentes representam a terceira causa de morte na população geral e a primeira na população de 1 a 49 anos.

Dados do Ministério da Saúde (MS) registraram, no período de 2000 a 2013, um total de 1.874.508 óbitos por causas externas. Entre o início e o fim desse período, houve um aumento de 28,1%, passando de 118.397 óbitos por causas externas, em 2000, para 151.683, em 2013. Em 2013, as causas externas representaram 12,5% do total de óbitos no País. Segundo o MS, em 2013, nos hospitais que integram o Sistema Único de Saúde (SUS), ocorreram 1.056.372 internações por causas externas, perfazendo 9,5% do total

de internações, enquanto em 2014, esse percentual foi de 9,9% [Brasil, da Saúde, M. 2016].

Preocupado com esses dados alarmantes, o Ministério da Saúde lançou em 2001 a Política Nacional de Redução da Morbimortalidade por Acidentes e Violências. Faz parte dessa política a Ficha de Notificação de Violência Interpessoal e Autoprovocada (vide Figura 1), que é a ficha de notificação compulsória de todos os casos suspeitos e ou confirmados de violência atendidos em todos os serviços de saúde do país, de forma a padronizar a coleta de dados, a partir de um conjunto de variáveis e categorias, que retratam as violências perpetradas contra grupos populacionais.

República Federativa do Brasil  
Ministério da Saúde

SINAN  
SISTEMA DE INFORMAÇÃO DE AGRAVOS DE NOTIFICAÇÃO  
FICHA DE NOTIFICAÇÃO INDIVIDUAL

Nº

Caso suspeito ou confirmado de violência doméstica/intrafamiliar, sexual, autoprovocada, tráfico de pessoas, trabalho escravo, trabalho infantil, tortura, intervenção legal e violências homofóbicas contra mulheres e homens em todas as idades. No caso de violência extrafamiliar/comunitária, somente serão objetos de notificação as violências contra crianças, adolescentes, mulheres, pessoas idosas, pessoa com deficiência, indígenas e população LGBT.

Dados Gerais	1	Tipo de Notificação		2 - Individual	
	2	Agravado/ença		VIOLENCIA INTERPESSOAL/AUTOPROVOCADA	Código (CID10) Y09
	3	Data da notificação			
	4	UF	5	Município de notificação	
Dados do Paciente	6	Unidade Notificadora		<input type="checkbox"/> 1- Unidade de Saúde <input type="checkbox"/> 2- Unidade de Assistência Social <input type="checkbox"/> 3- Estabelecimento de Ensino <input type="checkbox"/> 4- Conselho Tutelar <input type="checkbox"/> 5- Unidade de Saúde Indígena <input type="checkbox"/> 6- Centro Especializado de Atendimento à Mulher <input type="checkbox"/> 7- Outros	
	7	Nome da Unidade Notificadora		Código Unidade	
	8	Unidade de Saúde		Código (CNES)	
	9	Data da ocorrência da violência			
Notificação Individual	10	Nome do paciente		11	
	12	(ou) idade		11	
	13	Sexo		15	
	14	Gestante		15	
	16	Escolaridade		15	
	17	Número do Cartão SUS		18	

Figura 1. Primeira parte da Ficha de Notificação de Violência Interpessoal

Apesar de haver instruções específicas para o preenchimento dessa ficha, muitas vezes esse preenchimento não é feito de maneira adequada. Existe a possibilidade de a vítima preencher informações antagônicas em campos que se complementam ou mesmo inserir informações importantes para o entendimento do tipo de violência em um campo de ‘Observações Adicionais’, que pode ficar esquecido na hora de analisar os dados.

Com este trabalho busca-se desenvolver uma modelagem conceitual das informações e estruturar semanticamente a base de dados, a partir das fichas e documentos de referência, para que seja possível integrar essas informações preenchidas nas fichas de notificação em meios digitais e alimentar os bancos secundários, de forma a auxiliar a análise dos gestores e, por consequência, às vítimas de violências.

## 2. Objetivos

O objetivo geral deste projeto é propor um modelo conceitual baseado em ontologias para a gestão da informação e do conhecimento na área de vigilância epidemiológica, com foco inicial nas questões relativas à violência.

Os objetivos específicos do projeto são:

- Estruturar semanticamente o banco de dados existente, modelado com base na ontologia definida para o domínio;
- Uma vez estruturado o banco, apontar possíveis incorreções nos registros, buscando excelência de informação;

- Utilizar esse banco de dados semanticamente estruturado para o desenvolvimento de diversos painéis (*dashboards*), atualizados à medida em que os novos dados são introduzidos, de forma a auxiliar na melhor compreensão dos cenários de violência presentes nas Fichas de Notificação;
- Auxiliar no processo de localização e exploração de novos conhecimentos que o Sistema de Saúde, apesar de possuir, não consegue identificar claramente, enriquecendo a criatividade e gerando inteligência competitiva e, dessa forma, embasando o processo de tomada de decisão;
- Desenvolver ferramentas de Ciência de Dados com o objetivo de apontar a possibilidade de novos casos mais graves de violência (ex: tentativa de suicídio ou feminicídio) baseado em informações já disponíveis nos bancos de dados referentes às vítimas ou aos agressores;

### 3. Justificativa

O elevado índice de violência faz com que o Brasil seja considerado um dos países mais perigosos do mundo. Isso traz inúmeros problemas para o país que incluem altas taxas de homicídio, evasão escolar, superlotação da população carcerária, baixo aproveitamento da capacidade turística, além de muitas outras desvantagens.

A Figura 2 mostra uma espécie de ‘Atlas de violência’ do Brasil em 2017, destrinchando o cenário da violência ao longo do território brasileiro. Um dado que se destaca neste infográfico é a elevada proporção de mortes violentas não esclarecidas em relação ao total de mortes violentas. Altas proporções de mortes violentas não esclarecidas quando comparadas às mortes por causas externas indicam problemas na qualidade do sistema de informação da saúde [Cerqueira et al. 2016].

Enquanto em países desenvolvidos as mortes violentas indeterminadas representam um resíduo inferior a 1% do total de mortes violentas, em alguns estados do Brasil esse número foi superior a 10%. Isto ocorre, pois nesses países se reconhece a importância de se descobrir as causas que levaram o indivíduo a óbito, como elemento fundamental para evitar mortes futuras.



Figura 2. Atlas da violência 2017

Esse conjunto de fatores ajudam a reforçar o fato que se deve fazer um estudo mais aprofundado do cenário da violência, identificando suas possíveis causas de modo a tentar agir de maneira preventiva para evitar possíveis novos casos.

#### 4. Metodologia

As etapas de desenvolvimento deste trabalho estão divididas da seguinte forma:

1. Pesquisa bibliográfica sobre o tema incluindo-se nessa etapa as bases de divulgação científica, a documentação existente na Organização Mundial da Saúde (OMS) e no MS;
2. Modelagem Conceitual baseada em ontologias das informações referentes aos diversos tipos de notificações;
3. Utilização dos dados do VIVA/SINAN para realizar a estruturação de uma base de dados semanticamente estruturada com base na ontologia definida para o domínio.
4. Desenvolvimento de painéis (*dashboards*) que possam ajudar a compreender melhor o cenário de violência presentes nas Fichas de Notificação;
5. Utilização de técnicas de *Data Science/Machine Learning* para identificar correlações entre os dados e tentar traçar perfis de possíveis novos casos de agressão baseado nas informações já disponíveis das Fichas de Notificação.

#### 5. Resultados Obtidos e Esperados

Neste momento da pesquisa, a modelagem conceitual baseada em ontologia das informações constantes das fichas de notificações já está desenvolvida. Essa modelagem foi feita utilizando o software *Protégé*, baseado nas informações presentes nas fichas de notificação. Uma parte da ontologia desenvolvida é mostrada na figura 3, onde é possível ver alguns campos da subclasse de “Identificadores” utilizados.

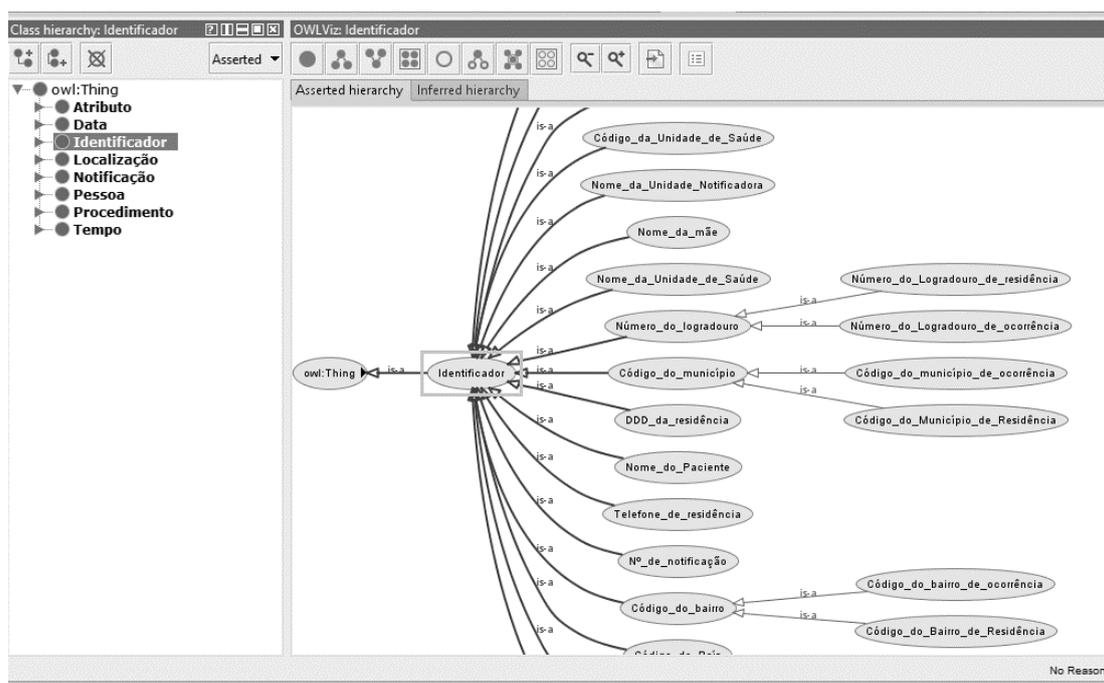


Figura 3. Parte da ontologia desenvolvida no Protégé

Outra parte do trabalho já bem desenvolvida é a elaboração de uma base de dados semanticamente estruturada contendo os dados das vítimas de agressão de maneira estruturada. Utilizando o software *Tableau Desktop* realizou-se essa formulação da base de dados. A figura 4 mostra o *preview* das primeiras linhas de informação da base de dados com as primeiras colunas de informações presentes na ficha de notificação sendo exibidas no software *Tableau Desktop*.

Violência ES 2011 a 2018 (Violência ES 2011 a 2018\_...)

Connection: Live (selected), Extract

Filters: 0 | Add

Violência ES 2011 a 2018

Sort fields: Data source order

Show aliases:  Show hidden fields:  1,000 rows

| #                        | Abc                      |
|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| Violência ES 2011 a 2018 |
Número da notifi...	Tipo de notific...	Id do agravo	Data da notificação	Semana epidemiol...	Ano da notificação	UF da notificação	Município da notifi...	
2939166	2	Y09	4/22/2013	201317	2013	32	320530	
2939155	2	Y09	3/27/2013	201313	2013	32	320530	
2962373	2	Y09	5/2/2013	201318	2013	32	320470	
2967452	2	Y09	4/19/2013	201316	2013	32	320150	
2962374	2	Y09	5/2/2013	201318	2013	32	320470	
3023236	2	Y09	4/30/2013	201318	2013	32	320500	
2967450	2	Y09	4/19/2013	201316	2013	32	320150	
2967293	2	Y09	5/9/2013	201319	2013	32	320190	
2939165	2	Y09	4/23/2013	201317	2013	32	320530	
2967292	2	Y09	5/9/2013	201319	2013	32	320190	
3022913	2	Y09	5/5/2013	201319	2013	32	320500	
3022914	2	Y09	4/28/2013	201318	2013	32	320500	

Figura 4. Registros iniciais da fonte de dados desenvolvida no *Tableau Desktop*

A etapa de confecção de *dashboards* que mostrem a situação da violência Interpessoal/Autoprovocada no Brasil já está em andamento.

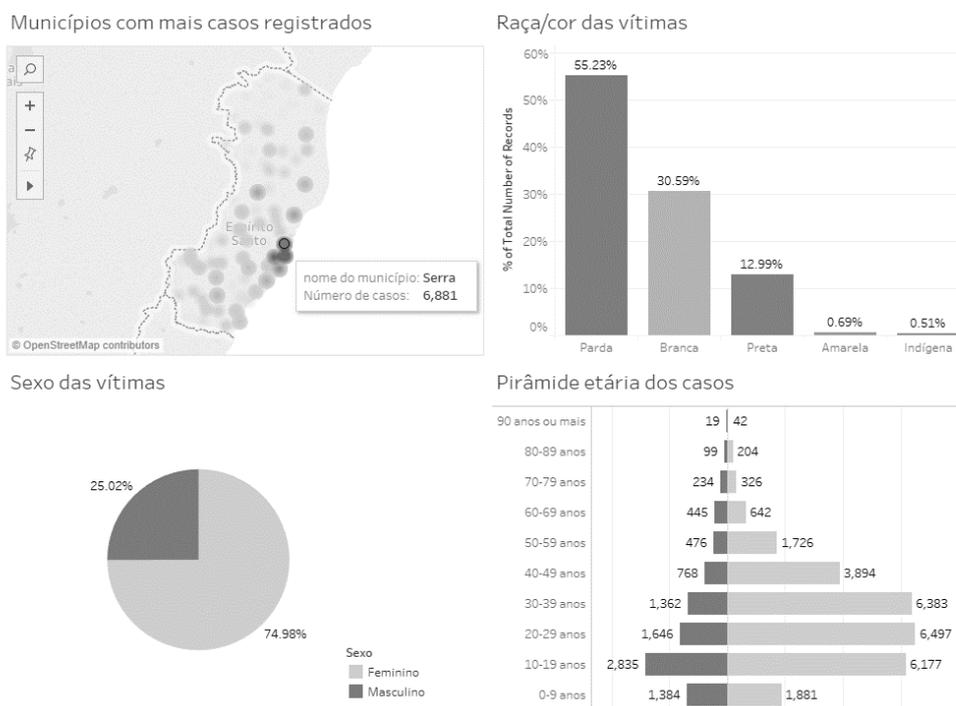


Figura 5. Exemplo de *dashboard* desenvolvido no *Tableau Desktop*

Inicialmente foram elaborados alguns painéis com os dados disponibilizados pela Secretaria Estadual da Saúde do Espírito Santo, mas a ideia é expandir esse raciocínio para todo o Brasil, já que a maneira de preencher as fichas é idêntica em todos os estados. A figura 5 mostra um dos *dashboards* desenvolvidos com o software *Tableau desktop*.

Ao fim desta pesquisa, espera-se obter dentro do conteúdo pesquisado e durante a elaboração da dissertação de mestrado, os seguintes resultados:

- Desenvolvimento de ferramentas de correlação e predição baseada em tecnologias *Data Science / Machine Learning*;
- Indicação de possíveis novos casos de violência e agressão baseados nos dados dos casos já existentes nas Fichas de Notificação.

Junto a estes resultados espera-se prover:

- Apresentação do trabalho aos gestores do SUS locais e de âmbito nacional;
- Participação em um *workshop* nacional;
- Participação em um evento em nível nacional;
- Submissão de um artigo em um periódico do Sistema Nacional de Saúde;
- Submissão de um artigo em um periódico internacionalmente qualificado.

## 6. Referências

- Corassa, R. (2018) “Violência por Parceiro Íntimo e Homicídios de Mulheres no de 2011 a 2016”. Dissertação de mestrado, Universidade Federal do Espírito Santo.
- Cerqueira, D., Lima, S., Bueno, S., Valencia, L., Hanashiro, O., Machado, P., Lima, A. (2017) “Atlas da Violência”. Online, [http://www.ipea.gov.br/portal/images/170602\\_atlas\\_da\\_violencia\\_2017.pdf](http://www.ipea.gov.br/portal/images/170602_atlas_da_violencia_2017.pdf).
- Chaiken, J., Rhodes, W. (1994) “Predicting Violent Behavior and Classifying Violent Offenders.” In: NATIONAL RESEARCH COUNCIL. Understanding and Preventing Violence, Volume 4, cap.
- Brasil, da Saúde, M. (2016) “VIVA: Instrutivo – Notificação de Violência Interpessoal e Autoprovocada”. Online, [http://bvsms.saude.gov.br/bvs/publicacoes/viva\\_instrutivo\\_violencia\\_interpessoal\\_autoprovocada\\_2ed.pdf](http://bvsms.saude.gov.br/bvs/publicacoes/viva_instrutivo_violencia_interpessoal_autoprovocada_2ed.pdf).