Sala de Situação aberta com dados administrativos para gestão de Protocolos Clínicos e Diretrizes Terapêuticas de tecnologias providas pelo SUS

Felipe Ferré[1], Gustavo Laine Araújo de Oliveira[2], Mariana Jorge de Queiroz[2], Flávio Gonçalves[3]

[1]Programa de Pós-graduação em Saúde Pública, Universidade Federal de Minas Gerais, Av. Prof. Alfredo Balena, 190 – sala 503, Belo Horizonte – Minas Gerais, Brasil

[2]Departamento de Gestão e Incorporação de Tecnologias e Inovação em Saúde, Ministério da Saúde, Brasil

[3]Departamento de Economia, Universidade Federal do Paraná

labxss@gmail.com, {gustavo.araujo,mariana.queiroz}@saude.gov.br, flaviogonsalves@hotmail.com

Abstract. *Public stakeholders and specialists must deal with 4 billion outpatient records for Health Technology Assessment (HTA) and Monitoring, often without the apparatus and technical knowledge in Big Data, which demands transparent knowledge management tools for decision making and statistical and predictive reproducibility. The objective was to implement an open-source platform for knowledge management, analysis and visualization of SUS data. The interactive extraction and visualization, documented in wiki format, of 60 clinical protocols for 6,505,722 users with 254,433,772 dispensing records between 2008 and 06/2019 was automated, available at https://labxss.shinyapps.io/sabeis_pcdt3/.*

Resumo. *Gestores e especialistas devem lidar com 4 bilhões de registros ambulatoriais para Monitoramento e Avaliação de Tecnologias em Saúde (ATS) do SUS, muitas vezes sem o aparato e conhecimento técnico em Big Data, o que demanda ferramentas de gestão do conhecimento transparente para tomada de decisão e reprodutibilidade estatística e preditiva. Objetivou-se implementar plataforma livre para gestão do conhecimento, análise e visualização de dados do SUS. Foi automatizada a extração e visualização interativa, documentada em formato wiki, de 60 protocolos clínicos em 254.433.772 registros de dispensação para 6.505.722 de usuários, entre 2008 e 06/2019, disponível em https://labxss.shinyapps.io/sabeis_pcdt3/.*

1. Introdução

Desde 1996, o Ministério da Saúde (MS) dissemina dados abertos passíveis de consumo por máquina e tabuláveis pelo público amplo por meio de aplicativos (e.g., Tabwin e Tabnet) desenvolvidos pelo Departamento de Informática – Datasus. Dados abertos são utilizados para controle social e fazem parte da tomada de decisão por setores governamentais não controladores dos dados de produção (Silva 2009).

O auxílio analítico à gestão pública consiste, invariavelmente, na coleta de dados e transposição em informação e conhecimento com auxílio de plataformas comerciais ou de desenvolvimento próprio, seja para verificação de estatísticas básicas, seja para análises complexas, como a análise de situação ante a multiplicidade de atores em dada questão. A gestão situacional de informação em saúde é mais do que a mera implantação de "software" ou plataforma de inteligência de negócios (do inglês, business intelligence - BI) com robustos recursos estatísticos e computacionais (Moya et al. 2010; Lichtenstein et al. 2011), uma vez que organizam o conhecimento normativo, analítico e preditivo conjuntamente com as decisões negociais tomadas. A Sala de Situação visa o planejamento estratégico situacional em saúde pública, com gestão de conhecimento, a partir de evidências derivadas de conjuntos de dados transpostos em indicadores retroalimentados pelas decisões políticas. (Matus 1989; Moya et al. 2010). Ao contrário da busca reprodutível, a busca sob demanda de dados e informações leva ao cálculo situacional imediatista ou puramente técnico, no lugar do cálculo situacional sistemático com processos documentados e baseados em procedimentos transparentes ao leitor (Moya et al. 2010; Matus 1996). Na contramão da coleta e processamento de dados *ad hoc*, a sala de situação em saúde sistematiza a extração, transformação e carga em repositórios, com a apresentação de informações históricas de uma situação conforme o interesse tecnológico, econômico, político e social para produção e gestão de conhecimento (Moya et al. 2010).

Existem diretrizes para que indicadores e metadados correlatos à saúde sejam disponibilizados (OPAS-OMS 2008; Pereira e Tomasi 2016; WHO 2015) e algumas iniciativas contemplam a boa prática de disseminar dados abertos (Smith e Sandberg 2018; O'Neill et al. 2017). Porém, ainda que parte dos esforços recentes das determinações em saúde venham, invariavelmente, das ciências da computação, é incipiente o ensejo de apresentar conjuntamente, de forma documentada e reprodutível, códigos-fonte e material explicativo da transposição, análise e visualização de dados (Ross, Lehman, e Gross 2012) como parte do conhecimento que compõe a tomada de decisão via salas de situação.

Apesar dos esforços empreendidos, as ferramentas oficiais (Tabwin e Tabnet) possuem capacidade de processamento limitada, obrigando o analista a restringir o período ou a abrangência territorial. Ainda, o código-fonte dos consolidadores não são disponibilizados pelo governo, bem como os repositórios tratados, de modo que a comunidade possa validar e evoluir o percurso analítico que respaldou a decisão, o que fere os princípios da publicidade e impessoalidade. A não abertura dos processos que levaram à consolidação dos dados para dada decisão não garante que os atos administrativos estejam livres de viés. (Araújo 2007). Desta forma, diversos esforços independentes foram realizados para consolidar dados do Sistema Único de Saúde - SUS (Santos e Gutierrez 2011; Câmara e Others 2011; Deininger et al. 2014; Moya et al. 2010; Sellera et al. 2019), ainda que não tenham sido iniciativas de ciência aberta definida, segundo a taxonomia FOSTER, como o movimento de disponibilização da pesquisa científica, dados e disseminação de modo acessível a todos os níveis de uma sociedade investigadora(O'Neill et al. 2017).

Repositórios e código-fonte sem sistematização ou abertura do tratamento dificultam a governança do processo analítico e decisório e levam à fragmentação do apoio à

decisão(Arruda et al. 2015). A gestão de atendimentos de saúde especializados e estratégicos requer ferramentas capazes de lidar com mais de quatro bilhões de registros ambulatoriais disseminados. Na aplicação da Lei nº 12.401/2011 que explicita critérios para incorporação de tecnologias no Sistema Único de Saúde (SUS), devem ser integrados aos dados ambulatoriais de quase duzentos milhões de registros abertos de laudos individualizados por paciente para viabilizar Estudos de Utilização de Medicamentos (EUM). A decisão política acerca da composição do elenco disponível no SUS deve ser apoiada pela área do conhecimento Avaliação de Tecnologias de Saúde (ATS) com evidência clínica, avaliação econômica e impacto orçamentário (BRASIL 2014a). As avaliações culminam nos Protocolos Clínicos e Diretrizes Terapêuticas (PCDT), instrumentos que normatizam a decisão colegiada acerca do uso de tecnologias de saúde pelo SUS. Parte da aplicação do PCDT é registrada nos Laudos de Medicamentos Especializados - LME disseminados. Apesar de existirem mecanismos transparentes de determinações das decisões em saúde, não foi identificada na literatura propostas igualmente transparentes de extração, processamento, análise e visualização com dados abertos do SUS que são insumos para ATS. Visando suprir a lacuna de ferramentas analíticas abertas de Tecnologia da Informação em Saúde foi implementada a Sala Aberta de Inteligência em Saúde - SABEIS.

O objetivo do presente trabalho é apresentar plataforma tecnológica para sistematizar dados administrativos do SUS e gerir conhecimento da saúde integrado ao da informática, com código-fonte aberto, documentação para amplo público e painéis interativos acerca de medicamentos, referenciados por protocolo clínico.

2. Métodos

Na construção do SABEIS foram utilizados dados disseminados pelo Datasus e dados pelo e Instituto Brasileiro de Geografia e Estatística (IBGE). A ETL (do inglês, "Extract, Transform, Load" ou extração, transformação e carga) automatizada foi codificada em linguagem GNU bash, versão 4.4.20 e no Sistema de Gerenciamento de Banco de Dados (SGBD) PostgreSQL 10.8. A representação gráfica dos fluxos de coleta (extração e engenharia), processamento e análise dos dados são apresentados na figura 1

A ETL (figura 2) coleta os arquivos disponibilizados por sistema SUS, subsistema, estado, mês e ano e transforma em arquivo texto tabulado (*flat file*) CSV para ser incorporado no repositório com linguagem estruturada de consulta SQL (*Structured Query Language*).

O mesmo algoritmo pode ser processado na primeira carga e nas cargas subsequentes, bastando destinar os diretórios de estadiamento (*stage*), ainda que vazios, dos arquivos DBC e CSV. Os DBC são arquivos DBF compactados com o dbf2dbc.exe, os quais contém dados referentes a óbito, procedimentos ambulatoriais ou hospitalares, informações sobre estabelecimentos de saúde, equipes, equipamentos e, em alguns casos, tabelas de domínio. Os arquivos inexistentes ou com número de bytes diferentes do armazenado no estadiamento são listados e coletados a partir do diretório FTP (do inglês, *File Transfer Protocol*) do DATASUS em ftp://ftp.datasus.gov.br/. A verificação periódica do número de bytes do DCB e do número de registros do CSV ou no SGBD é necessária uma vez que o mesmo arquivo pode ser reprocessado pelo DATASUS.

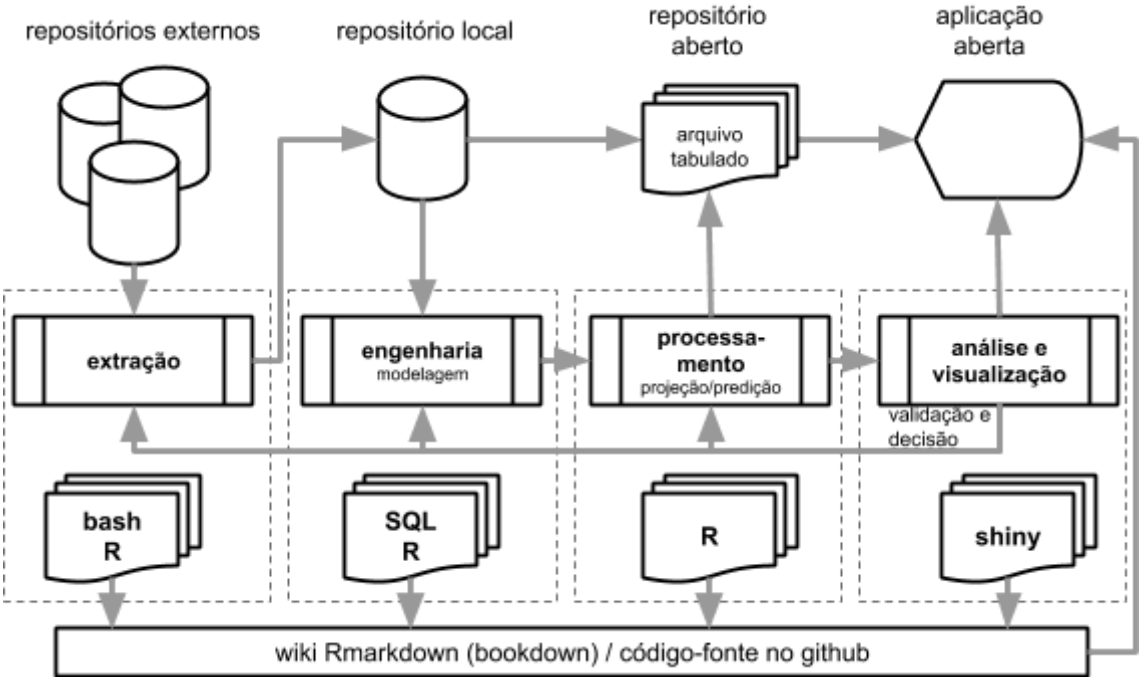

Figura 1: Sala Aberta de Situação em Saúde - SABEIS. Fluxo de processamento, documentação e apresentação. Fonte: elaboração própria com adaptação (Ferré 2017; Zaki e Wagner Meira 2014).

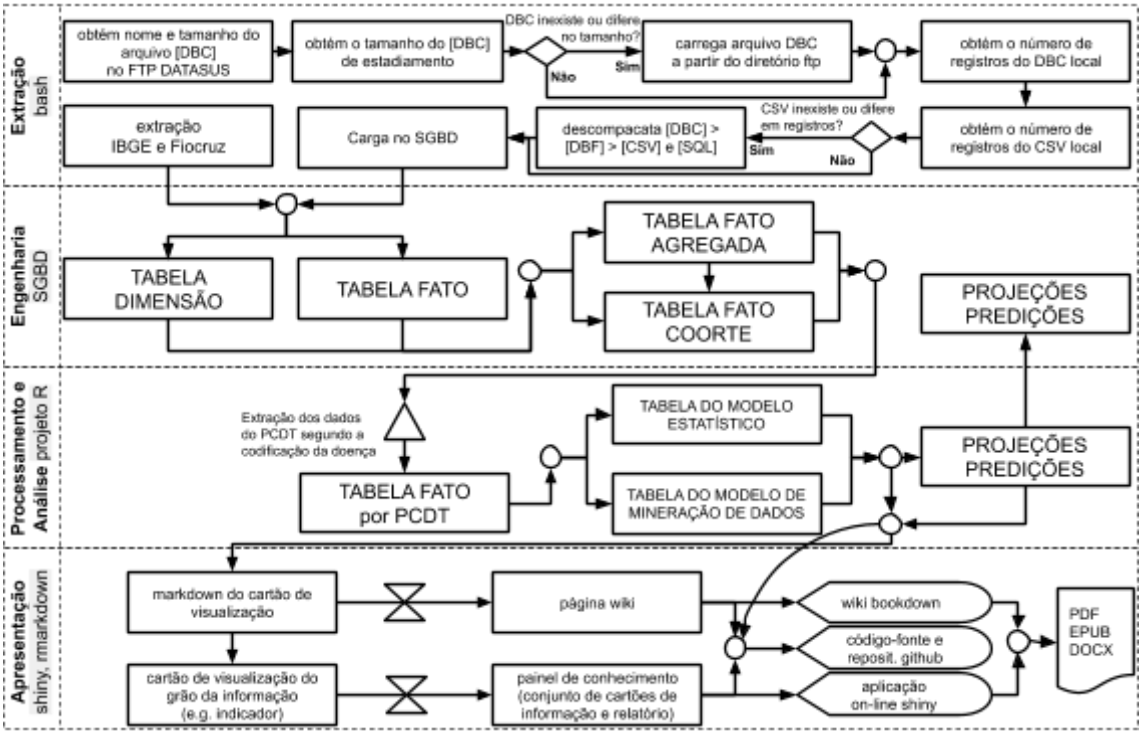

Figura 2: Fluxo de extração, transformação e carga. Fonte: elaboração própria.

O algoritmo proposto não trata volatilidade, isto é, o repositório local representa um espelho da versão mais recente do conteúdo disponível no FTP. Os dados de medicamentos do Componente Especializado da Assistência Farmacêutica (CEAF) originam-se do Sistema de Informações Ambulatoriais (SIA) e Cadastro Nacional de Estabelecimentos de Saúde (CNES) de gestão da Secretaria Especializada de Atenção à Saúde (SAES/MS). Apesar de tratar-se de dados administrativos, foram tomados como dados observacionais, do inglês (OMOP CDM *Observational Medical Outcomes Partnership - Common Data Model*), uma vez que trata-se de contato assistencial com data da dispensação, medicamento e diagnóstico, cujo registro é disseminado de forma individualizada com o Cartão Nacional de Saúde (CNS) encriptado, contendo sexo, idade e status de óbito, o que viabiliza a formação de coortes. O ano de nascimento, quando não disponível, foi obtido com base na diferença entre a idade e a competência do procedimento.

Os procedimentos de saúde do grupo "medicamentos" foram categorizados, respectivamente como, subgrupo e forma de organização, como mantidos pelo Sistema de Gerenciamento da Tabela de Procedimentos, Medicamentos e OPM do SUS (SIGTAP), Secretaria de Atenção Especializada à Saúde (SAES/MS), apenas acrescentando abreviatura com os quatro primeiros caracteres e todos os números da respectiva descrição, adicionando-se caracteres para desambiguação. Os diagnósticos foram registrados conforme a Classificação Internacional de Doenças e Problemas Relacionados à Saúde (CID-10) e hierarquizados conforme capítulos, grupos, categorias e subcategorias. O território foi caracterizado segundo o código IBGE do município e do gestor estadual. O estabelecimento foi identificado segundo o Cadastro Nacional de Estabelecimentos de Saúde (CNES, SAES/MS). A população residente segundo o código IBGE do município foi caracterizada por o sexo e faixa etária.

Dados complementares foram coletados a partir do sítio do IBGE, relativo ao Índice de Desenvolvimento Humano Municipal (IDHM), população e Índice Nacional de Preços ao Consumidor Amplo (IPCA), para correção dos valores para agosto de 2019 segundo a respectiva competência (mês de realização do procedimento); e do projeto Redesbr, o qual caracteriza o grau de desenvolvimento socioeconômico em saúde tipificada por região de saúde(Albuquerque et al. 2017; Viana 2019).

As agregações foram realizadas com Linguagem de Consulta Estruturada (do inglês, *Structured Query Language* SQL) com modelagem de processamento analítico online (do inglês, Online Analytical Processing - OLAP) e normas de nomenclatura do Departamento de Informática do SUS (DATASUS) (BRASIL 2019). Os dados de domínio foram transpostos em tabelas dimensão e os oriundos de dados transacionais foram transpostos em tabelas fato de modo que cada agregação constituiu a soma simples da frequência e valores em reais relativos eventos discretos (procedimento, diagnóstico, território, características sociodemográficas). A performance de consultas foi otimizada com o recurso da desnormalização das tabelas fato, de modo a evitar o cruzamento com tabelas de dimensão. A ETL e as análises foram realizadas em computador doméstico com processador AMD Phenon II X6, com 16GB de memória RAM DDR3, SSD de 256GB com SGBD instalado em HD de 4TB com 7200 rotações.

Na caracterização das informações e visualização foi adotada a metodologia da RIPSA - Rede Intergerencial de Informações para a Saúde (BRASIL 2014b). Cada indicador,

usualmente expresso com numerador e denominador, foi descrito com as categorias explicativas conceituação, interpretação, usos, limitações, fontes (instituições responsáveis pela produção dos dados utilizados no cálculo do indicador e pelos sistemas de informação a que correspondem), método de cálculo, categorias sugeridas para análise e dados estatísticos e comentários. A wiki da extração, visualização e análise foi documentada em três níveis: i) origem e transposição dos dados, ii) visualização, iii) técnica estatística ou de aprendizado de máquina. Os indicadores foram tipificados em A - demográficos, B - socioeconômicos, C - mortalidade, D - morbidade, G - fatores de risco e de proteção, E - recursos e F - cobertura.

Na wiki foi disponibilizado o endereço para o código-fonte e repositório de dados, respectivamente, sob a licença *General Public License* - GPL 3.0 e *Open Database Licence* (ODbL). Conforme mostrado na raia "shiny/rmarkdown" da figura 2, cada indicador registrado na wiki foi transposto como cartão shiny, cujo conjunto compõe o painel de PCDT. Foram escritos painéis com a ferramenta Shiny, do projeto R, a qual adota conceitos do tema Bootstrap de visualização, o que possibilita recursos para a usabilidade como responsividade e harmonização visual. Cada visualização de indicador foi encapsulada em uma função *box*, cujo título foi precedido de "O que" o qual demarca o tipo de objeto, medicamento ou produto para saúde, "Onde" para o território, "Quem" para a população, "Quando" para demarcar a faixa de tempo e "Quanto" para valores monetários. Valores síntese foram assinalados com a função *infoBox*, de modo a destacar total de quantidades e valores. A interação do usuário foi orientada sob o fluxo "Escolha > Explore > Extrapole" onde cada sessão foi demarcada acima com *infoBox* contendo o nome do respectivo PCDT e do medicamento, quando aplicáveis. Após o usuário filtrar pelo PCDT e o período desejado, na etapa "Explore" obtém acesso ao painel analítico dotado de indicadores na forma de cartões de visualização com tabela, histograma, série temporal. Na seção "Extrapole" foram disponibilizadas visualização das predições baseadas em técnicas estatísticas e de aprendizado de máquina.

3. Resultados

Foram descritos 60 Protocolos Clínicos e Diretrizes Terapêuticas (PCDT) originados do processamento de 7.496 arquivos DBC com 77,5 gigabytes e 3.136.181.932 registros AM (medicamentos) e PA (guia de Autorização de Procedimentos de Alta Complexidade/Custo - APAC), referentes ao período de janeiro de 2008 a junho de 2019 gerados pelo Sistema de Informações Ambulatoriais SIA, Ministério da Saúde. A junção das tabelas AM e PA resultou na coleta de 254.433.772 registros sendo 213.175.500 (83,7%) com quantidade aprovada e número CNS encriptado para 6.505.722 usuários do SUS, referentes a 537 medicamentos conforme o código SIGTAP e 714 doenças segundo a subcategoria da CID-10 registradas no Sistema de Informações Ambulatorial SIA. Os PCDT com maior número de usuários são mostrado na tabela 1.

Na figura 3 é mostrada a página incial da Sala Aberta de Inteligência em Saúde (SABEIS), disponível no link https://labxss.shinyapps.io/sabeis_pcdt3/. Na página wiki são documentadas a ETL e apresentado código-fonte com doze mínimos exemplos viáveis (MEV), exemplificado na figura 4. MEV são demonstrações de cartões

(indicadores) contendo a explicação do ponto de vista da saúde (aqui aplicou-se a metodologia RIPSA) atrelado à implementação (código-fonte, entrada e dados e proposta de visualização). Foram propostos MEV em mapa de árvore, pirâmide etária, linha de tendência, barras horizontais empilhadas, diagrama de pareto, dispersão, mapa geográfico de calor fluxo, série temporal e curva de sobrevivência.

Tabela 1: Dados coletados em outubro de 2019 dos subsistemas AM (medicamentos) e PA (guia de Autorização de Procedimentos de Alta Complexidade/Custo - APAC) mantidos pelo Sistema de Informações Ambulatoriais SIA, Ministério da Saúde referentes às dez doenças mais frequentes regidas por Protocolos Clínicos e Diretrizes Terapêuticas.

PCDT	início	procedimento SIGTAP	diagnóstico principal	usuários	registros	municípios de residência
Acne	200801	10	3	416.586	2.405.373	4.555
Artrite Reumatoide	200801	81	8	315.765	13.405.350	5.202
Asma	200801	112	3	637.492	17.755.014	5.093
Dislipidemia	200801	174	8	1.175.804	28.286.888	4.623
Doença de Alzheimer	200801	74	3	354.377	10.005.982	4.784
Dor Crônica	200801	39	2	273.496	2.303.467	4.130
Esquizofrenia	200801	109	8	756.797	26.372.810	5.353
Glaucoma	201308	9	8	274.914	10.911.866	2.474
Osteodistrofia Renal	200801	122	4	550.852	30.338.645	5.491
Osteoporose	200801	113	18	360.143	8.510.700	4.293
outros	200801	433	262	2.042.075	62.879.405	5.570

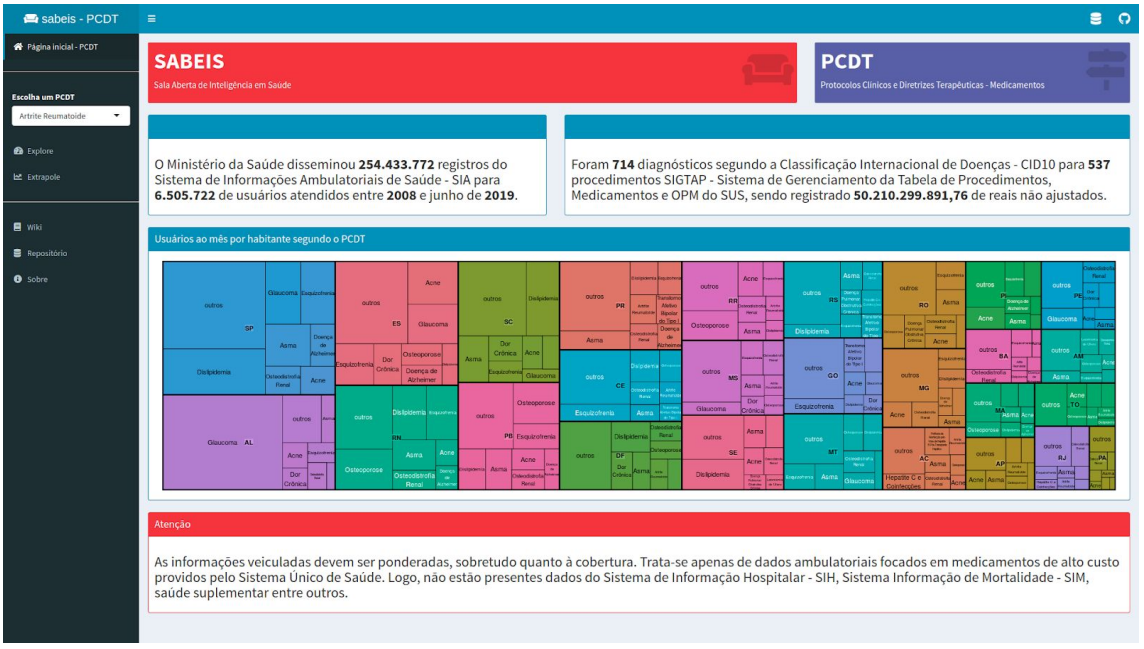

Figura 3: Tela inicial da Sala Aberta de Inteligência em Saúde (SABEIS).

3.2 Explore - Quem - usuários do PCDT - D

3.2.1 Conceituação

Proporção de usuários de Protocolo Clínico e Diretrizes Terapêuticas - PCDT por faixa-etária e sexo.

3.2.2 Interpretação

- Expressa uma relação quantitativa entre sexo e faixa-etária em anos.
- As faixa-etária de cada usuário é obtida a partir do ano de nascimento mais frequente registrado no Sistema de Informação Ambulatorial.
- Se igual a 100, todos os usuários são homens ou mulheres de uma dada faixa-etária.

3.2.3 Usos

- Analisar variações na distribuição da população por sexo e faixa-etária.
- Subsidiar processos de planejamento, gestão e avaliação de políticas públicas de medicamentos.
- Identificar necessidades de estudos sobre os fatores condicionantes das variações encontradas.

3.2.4 Limitações

- Trata-se de registros administrativos de informações ambulatoriais, logo, não é carantida a informação do ponto de vista do registro civil de cada paciente.
- A idade é calculada em anos com base no ano de nascimento mais frequente em relação ao número de registros no Sistema de Informações Ambulatoriais. O cálculo desejável decorreira da data de nascimento.

3.2.5 Fonte

- DATASUS/MS - Ministério da Saúde, Departamento de Informática do SUS [ftp://ftp.datasus.gov.br/dissemin/publicos/SIASUS/200801_/Dados/].
- IBGE - Instituto Brasileiro de Geografia e Estatística [https://educa.ibge.gov.br/jovens/conheca-o-brasil/populacao/18318-piramide-etaria.html]
- RIPSA - Rede Intergerencial de Informações em Saúde. Razão de sexos [http://fichas.ripsa.org.br/2012/a-2/?l=pt_BR]

3.2.6 Métodos de Cálculo

Fórmula geral:

$$P_{SF}\frac{n_{SF}}{T} \times 100 \tag{2.1}$$

onde P_{SF} é a proporção da respectiva combinação de sexo S e faixa-etária F, cujos domínios são, respectivamente

$$S = \{feminino, masculino\}$$

$$F = \left\{ \begin{array}{ccccc} 00-04 & 05-09 & 10-14 & 15-19 & 20-24 \\ 25-29 & 30-34 & 35-39 & 40-44 & 45-49 \\ 50-54 & 55-59 & 60-64 & 65-69 & 70-74 \\ 75-79 & 80-84 & 85-89 & 90+ & \end{array} \right\}$$

n_{SF} é o número de usuários do sexo e da faixa etária e T é o total de usuários do PCDT, obtido por $T = \sum n_{SF}$.

3.2.7 Categorias Sugeridas para Análise

- Unidade geográfica: Brasil, Grandes Regiões, Estados, Distrito Federal e Regiões Metropolitanas.
- Município de residência
- Medicamento

3.2.8 Dados Estatísticos e Comentários

Visualizar no painel SABEIS > Explore.

3.2.9 Notas

Realizado conforme metodologia RIPSA. Vide a seção *fonte*.

3.2.10 Visualização

Pirâmide etária.

3.2.11 Conjunto de dados e Código-fonte

Pré-requisitos

Baixe o arquivo **db_sabeis.tf_pcdt_paciente.csv** na pasta dataset.

Certifique-se de que os pacotes R abaixo estejam instalados.

```
library(pyramid)
```

R script

Lê arquivo csv e ajusta para o formato percentual.

Ajuste o caminho do arquivo para o diretório onde baixou.

Plota a pirâmide etária com o pacote pyramid.

Observe o filtro **co_pcdt**, o qual extrai dados apenas para um PCDT.

```
tf_pcdt_paciente = read.csv(file = 'db_sabeis.tf_pcdt_paciente.

# ajusta para o formato percentual
tf_pcdt_paciente$qt_paciente_m =
  tf_pcdt_paciente$qt_paciente_m/tf_pcdt_paciente$qt_registros_

tf_pcdt_paciente$qt_paciente_f =
  tf_pcdt_paciente$qt_paciente_f/tf_pcdt_paciente$qt_registros_

pyramid(
      subset(tf_pcdt_paciente, co_pcdt == 10)[,c(2,3,6)],
      Llab="Masculino",Rlab="Feminino",
      Lcol="navy",
      Ldens=5,
      Rcol="red",
      Rdens=10, GL=FALSE, Clab = "Idade", main = "usuários (%)"
    )
```

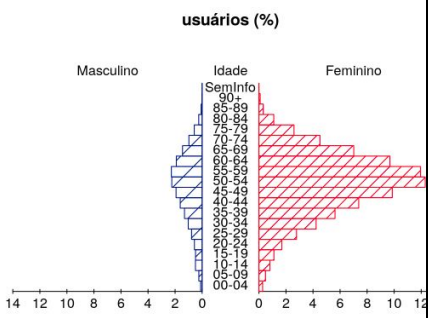

Tabela 2.2: Amostra da tabela de entrada.

	qt_paciente_m	qt_paciente_f	sg_fa
9	3.754098	4.491803	40-44
10	4.114754	4.901639	45-49
11	4.754098	6.557377	50-54
12	5.655738	6.393443	55-59
13	4.836066	6.704918	60-64

Figura 4: Exemplo mínimo viável documentando um cartão (indicador) na wiki.

4. Discussão

A principal contribuição do presente trabalho é apresentar-se como esforço de ciência aberta, publicizando as informações do repositório (Wazlawick 2020), com o uso de software livre (Sousa 2020), diferententemente do objetivado por ferramentas análogas identificadas (Câmara e Others 2011; Deininger et al. 2014; Moya et al. 2010; Santos e Gutierrez 2011; Assunção et al., [s.d.]; Sellera et al. 2019), em particular, integrando os conhecimentos de ATS com as ciências da computação para atender equipes multidisciplinares e com profissionais dotados de dupla formação. Quanto a taxonomia FOSTER para ciência aberta (O'Neill et al. 2017) foram contemplados quatro níveis de funcionalidades e documentação para extração, engenharia, processamento, análise e visualização (figura 1): "acesso aberto" (definição e iniciativas), "dado aberto" (uso e reuso, big data, definição de dados, padrões de dados, dados abertos governamentais), ciência aberta reprodutível (definições, laboratório aberto e anotações, fluxo aberto de dados, código-fonte aberto, guias reprodutivos, testes reprodutivos) e "ferramentas de ciência aberta" (repositórios abertos e serviços abertos).

A abordagem proposta ofertou tripla governança para estruturação do esforço técnico em constituir salas de situação: i) geracional, ii) serviço de software, iii) serviço de hardware. i) A transparência na coleta, transformação e apresentação de dados é necessária para a continuidade geracional da avaliação e monitoramento, onde o baixo grau de documentação e rastreabilidade dos procedimentos demanda engenharia reversa ou novo desenvolvimento a cada composição da equipe, algo comum na gestão da saúde pública com cenários de maioria sem estabilidade na carreira (Druck 2016). ii) A adoção de software livre, por sua vez, é essencial para a governança das operações na gestão pública (SciELO 2018), uma vez que a não renovação de contratos para suporte e atualização de softwares fechados pode implicar no risco crítico da inoperação. iii) Foi utilizada tecnologia leve de hardware, cuja aplicação também funciona em ambiente local sem conexão à internet.

A iniciativa aberta possibilita a reutilização ou comparação de implementações análogas de ETL sobre as mesmas fontes, ainda que os objetivos sejam distintos. Identificou-se ETL disponível em software livre para extração de dados do SUS hospitalares, de nascidos vivos e óbitos, porém, não ambulatoriais, escopo do presente trabalho (Sousa 2020). Um exemplo cujo código-fonte da ETL não foi identificado e que não utilizou *software* livre, o que inviabiliza a replicação sem aquisição de ferramentas proprietárias, é o MINERSUS (Santos e Gutierrez 2011), cujo *data warehouse* (repositório) coletou e estruturou dados ambulatoriais, hospitalares, de notificações e agravos, de estabelecimentos de saúde e sociodemográficos. Ainda, a presente abordagem diferenciou-se de estudos de monitoramento de indicadores, os quais usualmente apresentam apenas a descrição dos indicadores para profissionais da saúde, sem ilustrar a extração e apresentação do ponto de vista informacional (Pereira e Tomasi 2016; Câmara e Others 2011; Moya et al. 2010).

Outra contribuição do presente trabalho não observada na literatura foi estabelecer uma sala aberta dedicada para ATS do SUS, considerando dados administrativos como primários para a gestão orientada por protocolos clínicos. Espera-se que a utilização de processos extrativos transparentes possa beneficiar os esforços da transposição de dados administrativos em dados secundários para Estudos de Utilização de Medicamentos

(EUM), em especial, para Avaliação de Tecnologias de Saúde (ATS). Uma externalidade esperada é a adoção das premissas elencadas de ciência aberta pelos Núcleos de Avaliação de Tecnologia em Saúde (NATS), integrantes da REBRATS - Rede Brasileira de Avaliação de Tecnologias em Saúde (REBRATS 2020) de modo padronizar achados com base em dados populacionais.

Dentre as dificuldades, decorre a ausência de PCDT em formato consumível por máquina (e.g. ontologia), o que demanda transposição semântica manual do PDF disponível em portarias, cuja tarefa consiste em integrar linhas terapêuticas com SIGTAP e listas oficiais referentes a medicamentos e outras tecnologias de saúde. Outra tarefa é expandir a coleta para outros subsistemas do SUS disponíveis, referentes a fístula, laudos diversos, acompanhamento multiprofissional, nefrologia, quimioterapia, radioterapia, tratamento dialítico e boletim individual de modo a se aproximar do número de 134 PCDT existentes. As limitações, entretanto, não reduzem a importância em trabalhar com dados abertos existentes utilizando-se do massivo conteúdo disponível não apenas para achados administrativos, mas epidemiológicos seguindo exemplo do modo de operação da ciência colaborativa traçado pela genômica (Choudhury et al. 2014) ou pesquisa de ensaios clínicos (Ross, Lehman, e Gross 2012).

5. Conclusão

O presente trabalho mostrou a construção da SABEIS, plataforma aberta de informática em saúde com dados de dispensação de medicamentos unificados e sistematizados com software livre para avaliação e reavaliação de protocolos clínicos. A implementação proposta é a codificação funcional de uma tecnologia. Espera-se que a tecnologia apresentada seja instanciada em diversos ambientes de trabalho, inclusive com agregação estadual e municipal, para achados científicos convergentes, ao menos sob o aspecto da tecnologia da informação em saúde.

Referências

Albuquerque, Mariana Vercesi de, Ana Luiza D'ávila Viana, Luciana Dias de Lima, Maria Paula Ferreira, Edgard Rodrigues Fusaro, e Fabíola Lana Iozzi. 2017. "Desigualdades regionais na saúde: mudanças observadas no Brasil de 2000 a 2016". *Ciência & Saúde Coletiva* 22 (abril): 1055–64.

Araújo, Luiz César G. de. 2007. "Conselho Nacional de Secretários de Saúde: A gestão administrativa e financeira do SUS". *Brasília: CONASS*.

Arruda, C., S. G. R. Lopes, Mhal Koerich, e D. R. Winck. 2015. "Redes de atenção à saúde sob a luz da teoria da complexidade". *Esc. Anna Nery Rev. Enferm* 19: 169–73.

Assunção, Renato M., Osvaldo S. F. Carvalho, Marcos O. Prates, e Marcelo Almeida Campos. [s.d.]. "Detecção de anomalias nos pagamentos do Sistema Único de Saúde". In , 459–68. Sociedade Brasileira de Informática em Saúde.

BRASIL. 2014a. *Diretrizes metodológicas. Análise de impacto orçamentário. Manual para o Sistema de Saúde do Brasil*.

———. 2014b. "Rede Interagencial de Informações para a Saúde. Fichas de qualificação da RIPSA-2010: mortalidade". RIPSA Brasília. 1º de setembro de

2014. http://www.ripsa.org.br/.
———. 2019. “Norma de Padronização de Nomenclatura - DATASUS”. 1º de outubro de 2019. http://datasus.saude.gov.br/mad-norma-de-padronizacao-de-nomenclatura/.
Câmara, Tarsila Myrtle Alves, e Others. 2011. “Plano de implantação da sala de situação na diretoria geral de gestão do trabalho da secretaria municipal de saúde do Recife”. https://www.arca.fiocruz.br/handle/icict/29224.
Choudhury, Suparna, Jennifer R. Fishman, Michelle L. McGowan, e Eric T. Juengst. 2014. “Big Data, Open Science and the Brain: Lessons Learned from Genomics”. *Frontiers in Human Neuroscience* 8 (maio): 239.
Deininger, Layza de Souza Chaves, Kerle Dayana Tavares de Lucena, Daniela Cristina Moreira Marculino de Figueiredo, César Cavalcanti da Silva, Ana Eloísa Cruz de Oliveira, e Ulisses Umbelino dos Anjos. 2014. “A sala de situação da dengue como ferramenta de gestão em saúde”. *Saúde em Debate* 38 (100): 50–56.
Druck, Graça. 2016. “A terceirização na saúde pública: formas diversas de precarização do trabalho”. *Trabalho, educação e saúde* 14 (1): 15–43.
Ferré, Felipe. 2017. *Descoberta de interações medicamentosas com mineração de dados: Modelo para extração, engenharia, processamento e análise*. Novas Edições Acadêmicas.
Lichtenstein, Flavio, Agostinho Tavares, Ivan Torres Pisa, e Daniel Sigulem. 2011. “Sistemas de apoio à decisão baseados em diretrizes interpretadas por computador: um breve histórico e outros tópicos”. *Journal of health informatics in developing countries* 3 (4). http://www.jhi-sbis.saude.ws/ojs-jhi/index.php/jhi-sbis/article/view/157.
Matus, Carlos. 1989. *Adeus senhor presidente: planejamento, antiplanejamento e governo*. Litteris.
———. 1996. *Política, planejamento & governo: graph*. 2º ed. Vol. 1. IPEA.
Moya, José, João Baptista Risi Junior, Ayrton Martinello, Ernani Bandarra, Helvécio Bueno, e Otaliba Libânio de Morais Neto. 2010. *Sala de Situação em Saúde: compartilhando as experiências do Brasil*. Organização Pan-Americana da Saúde, Ministério da Saúde.
O’Neill, Gareth, Lidia Borrell-Damian, Stephan Kuster, e Marcel Swart. 2017. “FOSTER Facilitate open science training for european research”. 1º de maio de 2017. https://www.fosteropenscience.eu/toolkit.
OPAS-OMS. 2008. *Indicadores básicos para a saúde no Brasil: conceitos e aplicações*. Organização Pan Americana da Saúde, Organização Mundial da Saúde.
Pereira, Bernadete dos Santos, e Elaine Tomasi. 2016. “Instrumento de apoio à gestão regional de saúde para monitoramento de indicadores de saúde”. *Epidemiologia e Serviços de Saúde* 25 (junho): 411–18.
REBRATS. 2020. “Rede Brasileira de Avaliação de Tecnologias em Saúde - REBRATS”. 18 de janeiro de 2020. http://rebrats.saude.gov.br/.
Ross, Joseph S., Richard Lehman, e Cary P. Gross. 2012. “The Importance of Clinical Trial Data Sharing: Toward More Open Science”. *Circulation. Cardiovascular Quality and Outcomes* 5 (2): 238–40.
Santos, Ricardo da Silva, e Marco Antonio Gutierrez. 2011. “MINERSUS--Ambiente computacional para extração de informações para a gestão da saúde pública por meio da mineração dos dados do SUS”. *Research on Biomedical Engineering* 24

(2): 77–90.
SciELO. 2018. “Fiocruz facing the challenge of Open Science for Development and Public Health | SciELO in Perspective”. SciELO in Perspective. 21 de maio de 2018. https://blog.scielo.org/en/2018/05/21/fiocruz-facing-the-challenge-of-open-science-for-development-and-public-health/.
Sellera, Paulo Eduardo Guedes, Christiane Braga Martins de Brito, Márcia Benévolo Jovanovic, Soraia Ofugi Rodrigues, Carlos Fernando Dal Sasso de Oliveira, Soraya Oliveira dos Santos, e Luci Fabiane Scheffer Moraes. 2019. “A Implantação do Sistema de Monitoramento e Avaliação da Secretaria Estadual de Saúde do Distrito Federal (SES/DF)”. *Ciência & Saúde Coletiva* 24 (6): 2085–94.
Silva, N. P. 2009. “A utilização dos programas TABWIN e TABNET como ferramentas de apoio à disseminação das informações em saúde [Dissertação de Mestrado]”. *Rio de Janeiro: Escola Nacional de Saúde Pública Sergio Arouca, Fundação Oswaldo Cruz*.
Smith, Göran, e Johan Sandberg. 2018. “Barriers to innovating with open government data: Exploring experiences across service phases and user types”. *Information Polity* 23 (3): 249–65.
Sousa, Marcos Vinicius do Carmo. 2020. “Data warehouse e ETL para dados da saúde pública. Uma aplicação prática”. Organizado por Kelly Rosa Braghetto. BSc, Universidade de São Paulo.
Viana, Ana Luiza D’ávila. 2019. “Região e Redes | Banco de Indicadores”. Redesbr. 2019. http://www.resbr.net.br/indicadores/view/.
Wazlawick, Raul Sidnei. 2020. “Data Warehouse Relatórios e-SUS AB PEC”. 2020. https://integracao.esusab.ufsc.br/dw/.
WHO. 2015. “Global reference list of 100 core health indicators”. World Health Organization. https://apps.who.int/iris/bitstream/handle/10665/173589/WHO_HIS_HSI_2015.3_eng.pdf.
Zaki, Mohammed J., e Jr Wagner Meira. 2014. *Data Mining and Analysis: Fundamental Concepts and Algorithms*. Cambridge University Press.