

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA
DEPARTAMENTO DE COMPUTAÇÃO

Guilherme Servidoni da Silva

**Sobre a qualidade das bases de dados
públicas brasileiras: Estudo de caso na saúde**

São Carlos/SP

Setembro de 2022

Guilherme Servidoni da Silva

**Sobre a qualidade das bases de dados públicas
brasileiras: Estudo de caso na saúde**

Trabalho de Conclusão de Curso apresentado
ao Departamento de Computação como parte
dos requisitos para a conclusão da graduação
em Engenharia de Computação.

Orientação: Profa. Dra. Marilde Terezinha
Prado Santos

São Carlos/SP
Setembro de 2022

Agradecimentos

Primeiramente, agradeço à profa. Marilde, minha orientadora, por aceitar o desafio de me ajudar ao longo deste trabalho.

Agradeço também àqueles que me acompanharam ao longo da graduação, especialmente os membros dos grupos de extensão que participei - SECOMP, CATI Jr. e Enactus, que me mostraram todo o universo de possibilidades dentro e fora da universidade.

Esse trabalho também não teria sido possível sem os grandes amigos que fiz na Kearney, especialmente os membros do projeto que embasou esse trabalho inteiro - Franco, Maria, Daniel, Lucas e Danilo.

Por fim, a base de toda a minha vida, minha família - meus pais e minha irmã, que fizeram o possível e o impossível para que eu chegasse aqui.

*“Às vezes, é preciso escolher entre o que é certo, e o que é fácil.
(Albus Dumbledore)”*

Resumo

Com a crescente digitalização dos governos ao redor do mundo, esforços têm sido empenhados em prol da transparência de dados. Um dos principais marcos desse movimento é o workshop Open Government Data, realizado em 2007 em Sebastopol, na Califórnia, onde definiram-se oito princípios dos Dados Governamentais Abertos - DGAs (TAUBERER, 2007). Desde então, grandes iniciativas de instituições governamentais ao redor do mundo fazem-se presentes para tornarem seus dados mais transparentes, e dentre estas, instituições do Governo Federal brasileiro.

Por meio de iniciativas como a Lei de Acesso à Informação e o Portal Brasileiro de Dados Abertos, dados de diversas esferas dos governos municipais, estaduais e federal brasileiro tornam-se acessíveis a qualquer cidadão interessado. Todavia, o preço pela qualidade destes dados é tão primordial quanto a disponibilização em si, para que possam ser utilizados de maneira clara, ótima e granular. A partir de um recorte dos dados da saúde brasileira disponibilizados pelo Sistema Único de Saúde - SUS, via Datasus Tabnet, e da Agência Nacional de Saúde Suplementar - ANS, esse trabalho visa analisar a qualidade desses dados, propondo melhorias recicláveis a qualquer entidade governamental que vise a abertura de suas informações, prezando pela sua qualidade.

Alavancando conhecimentos adquiridos em um projeto executado pela multinacional Kearney Consultoria em Gestão Empresarial para um hospital de São Paulo, em que várias análises de mercado basearam-se em dados abertos da ANS e do SUS, analisou-se as dimensões para a qualidade dos dados propostas por Parker *et al.* (PARKER M.B., 2006). Para exemplificar onde a qualidade interfere de modo concreto na facilidade ou dificuldade do trabalho, análises sanitizadas desse projeto foram empregadas, mostrando o valor da abertura de dados governamentais, e como eles podem ser continuamente melhorados.

Palavras-chave: Dados públicos, ANS, Datasus, Tabnet, saúde, qualidade de dados.

Abstract

With the ever-growing digitalization of governments worldwide, great effort has been put towards data transparency. One important milestone of this trend is the Open Government Data workshop, which took place in 2007 in Sebastopol, California, where the eight Open Government Data principles (TAUBERER, 2007) were defined. Since then, large scale initiatives from government institutions worldwide were put in place to make their datasets more transparent, and among them, Brazilian institutions.

With initiatives such as the Information Access Law and the Brazilian Portal of Open Data (*Lei de Acesso à Informação* and *Portal Brasileiro de Dados Abertos*, respectively, unofficially translated), data from municipal, state and federal Brazilian government institutions are accessible to any interested citizen. However, effort towards data quality is as important as the openness of the data itself, so it can be used clearly, optimally, and as granularly as wished. Picking up from selected datasets made available by the Brazilian unified healthcare system (*Sistema Único de Saúde - SUS*), through the Datasus Tabnet platform, and the national agency for health insurance providers (*Agência Nacional de Saúde Suplementar - ANS*), this paper aims to analyze the quality of this data, proposing meaningful and reusable improvements to any government entity that wishes to open its data, with great care for quality.

Leveraging acquired knowledge from a project executed by the Chicago-based multinational consulting firm Kearney for a hospital based in São Paulo, where numerous market analyses were based on open data provided by SUS and the ANS, the dimensions for data quality proposed by Parker *et al.* (PARKER M.B., 2006) were used to analyze the available sets. In order to exemplify where such quality effectively interferes in the easiness of the work being done, sanitized analyses were brought up, showcasing the value of opening government data, and how they can be continuously improved.

Keywords: Public data, ANS, Datasus, Tabnet, health, data quality.

Lista de ilustrações

Figura 1 – Acesso ao painel de receitas e despesas da ANS no site de acesso à informação do Governo Federal.	23
Figura 2 – Definição dos dados de receitas e despesas da ANS a serem extraídos.	24
Figura 3 – Resultado da consulta realizada na figura 2.	24
Figura 4 – Página inicial da ANS via portal gov.br.	28
Figura 5 – Página da ANS no Portal Brasileiro de Dados Abertos.	29
Figura 6 – Análise Kearney da estratégia de operadoras de saúde selecionadas. Eixo X: Foco em qualidade. Eixo Y: Foco em custos.	30
Figura 7 – Análise Kearney de sinistralidade X ticket médio de operadoras selecionadas, feita inteiramente com dados da ANS. Eixo X: Ticket médio em R\$. Eixo Y: Sinistralidade da operadora em 2021 em %.	31
Figura 8 – Dados da Sompo Saúde Seguros, em uma visão por operadora do Mapa Assistencial da Saúde Suplementar da ANS.	32
Figura 9 – Amostra de colunas das características de produtos de saúde suplementar cadastrados na ANS.	33
Figura 10 – Amostra do dicionário das características de produtos de saúde suplementar cadastrados na ANS.	34
Figura 11 – Quantidade de operadoras ativas e <i>share</i> das classificadas como verticalizadas, de 2012 a 2021, e CAGRs respectivos (a definição de CAGR encontra-se no apêndice A.2).	35
Figura 12 – Página inicial do Datasus Tabnet.	36
Figura 13 – Página do Ministério da Saúde no Portal Brasileiro de Dados Abertos.	37
Figura 14 – Página inicial do openDatasus.	37
Figura 15 – Percentual de causas de morte no Brasil por grupo de causa, de 1996 a 2019, e CAGRs selecionados (definição de CAGR no apêndice A.2).	38
Figura 16 – Nota exibida ao se tentar baixar o conjunto mínimo de dados.	39
Figura 17 – Filtro intermediário do Datasus ao se extrair os dados de recursos físicos. Na abrangência geográfica, pode-se selecionar "Brasil por região, UF e município", ou um estado específico.	39
Figura 18 – Seleção de dados ao se extrair os recursos de consultório em ambulatórios do estado de São Paulo.	40
Figura 19 – Resultado da extração configurada nas figuras 17 e 18.	40
Figura 20 – Seleção de colunas para a extração de estabelecimentos cadastrados no CNES por tipo.	42
Figura 21 – Despesas assistenciais de 2020 por nível de atenção.	44

Figura 22 – Explicação superficial da metodologia empregada no projeto da Kearney para estimativa de tamanho de mercado por especialidade médica. . . .	45
Figura 23 – Visualização do conjunto de dados do IHME do gasto em saúde por planos privados por condição de saúde em 2016.	45
Figura 24 – Receitas e despesas extraídas do ANS Tabnet, com todas as colunas selecionadas, entre 2020 e 2022.	47

Lista de tabelas

Tabela 1 – Receita de contraprestações de operadoras de saúde selecionadas entre 2016 e 2020, em R\$ milhões.	46
Tabela 2 – Quebras e subquebras da base da ANS.	53
Tabela 3 – Quebras e subquebras da base do Datasus.	54

Sumário

1	INTRODUÇÃO	17
1.1	Objetivos	18
1.2	Justificativa e inspiração	18
2	METODOLOGIA	21
2.1	Crerios de qualidade	22
2.2	Comparação	25
2.3	Leis de proteoão de dados	25
3	RESULTADOS	27
3.1	Os 8 princípios	27
3.2	Dimensões da qualidade da informao	28
3.2.1	ANS Tabnet	28
3.2.2	Datusus Tabnet	36
3.2.3	Aplicaões reais	43
3.2.3.1	Nível de atenoão - ANS	43
3.2.3.2	Receita por operadora - ANS	46
3.2.3.3	Separao de períodos - Tabnet	46
4	CONCLUSÃO E CONSIDERAÇÕES FINAIS	49
	REFERÊNCIAS	51
	APÊNDICE A – APÊNDICE	53
A.1	Bases Datusus e ANS	53
A.1.1	Base ANS Tabnet	53
A.1.2	Base Datusus Tabnet	53
A.2	Definiao de CAGR	55

1 Introdução

A transparência governamental é assunto proeminente em várias esferas da atualidade: fóruns econômicos, critérios de investimento, cúpulas governamentais internacionais etc. Muito se fala de ESG - do inglês **E**nvironmental (ambiental), **S**ocial, e **G**overnance (governança). Por mais que essas discussões geralmente referenciem empresas privadas e suas responsabilidades para com a sociedade, o mesmo pode ser discutido para governos. Como posa Pelissero (PELISSERO, 2022), "Políticas ESG têm sido adotadas por governos de estado e locais em ritmo constante"(em tradução livre). Ao se falar de governança, um dos pilares fundamentais é a transparência: a própria S&P, referência em análise de crédito público e privado global, tem como um de seus quatro fatores fundamentais ao analisar a governança de uma instituição *transparência e relatórios* (GLOBAL, 2020).

Naturalmente, ao se falar de governança em governos, uma maior transparência vai ao encontro das melhores práticas. Segundo a Organização para a Cooperação e Desenvolvimento Econômico - OCDE (OCDE, s.d.), "Ao tornar suas bases de dados disponíveis, instituições públicas se tornam mais transparentes e responsáveis para cidadãos"(em tradução livre). Em países com significativa liberdade de mídia e de Internet, leis pró-transparência de dados são associadas a uma percepção de um nível inferior de corrupção (ŽUFFOVÁ, 2020). Em jurisdições com Índice de Desenvolvimento Humano mais elevado, há maior tendência ao cumprimento de normas governamentais de transparência (MICHENER G., 2017).

Todavia, para que políticas de transparência de dados sejam eficazes, os dados disponibilizados devem atender certos padrões de qualidade. Não existem métricas universais do que torna uma base de dados "boa"ou "ruim", mas existem estudos que buscam defini-las de acordo com o propósito utilizado. Quando se fala de transparência governamental, os *8 princípios para dados governamentais abertos* são comumente citados: tratam-se de princípios definidos por 30 acadêmicos pró-abertura governamental que reuniram-se em Sebastopol, CA, EUA, em dezembro de 2007 (TAUBERER, 2007). Os princípios são resumidamente listados abaixo.

1. **Completos:** Todos os dados públicos são disponibilizados.
2. **Primários:** Dados são coletados na fonte, com o maior nível possível de granularidade.
3. **Oportunos:** Disponibilizados o quanto antes possível.
4. **Acessíveis:** Disponíveis à maior gama de usuários para a maior gama de interesses.

5. **Processáveis por máquina:** Estruturados de modo razoável para processamento automático.
6. **Não-discriminatórios:** Disponíveis a qualquer um, sem necessidade de registro.
7. **Não-proprietários:** Disponíveis em formato sem controle exclusivo por qualquer entidade.
8. **Sem licença:** Não aplicável a copyrights, patentes, marcas registradas ou regulação de segredos.

Tratam-se de princípios fundamentais e básicos para um desenho eficiente de política de transparência de dados. Além disso, eles permitem a execução de estudos de caso de iniciativas de transparência governamentais, e esse é o foco desse trabalho.

1.1 Objetivos

Ao tomar como base os dados disponibilizados pela Agência Nacional da Saúde Suplementar - ANS e pelo Datasus (iniciativa de disponibilização de dados do Sistema Único de Saúde), esse trabalho visa realizar a avaliação qualitativa de bases representativas disponibilizadas por essas entidades. Com isso, espera-se:

- Valer-se de metodologias academicamente validadas para traçar paralelos entre as bases, sugerindo melhorias e reconhecendo esforços potencialmente replicáveis por outras agências governamentais.
- Mostrar como dados de boa qualidade fazem significativa diferença a partir de um trabalho real desenvolvido a partir de dados públicos por uma consultoria multinacional de gestão empresarial.

Com isso, almeja-se que uma realidade real de mercado possa servir de exemplo para políticas cada vez mais transparentes e eficazes de dados governamentais abertos.

1.2 Justificativa e inspiração

Como mencionado anteriormente, esse trabalho busca mostrar com exemplos concretos a diferença que dados de qualidade podem fazer. Entre novembro de 2021 e abril de 2022, o autor deste trabalho integrou um time da Kearney, consultoria de gestão multinacional baseada em Chicago, IL, EUA, no desenho do planejamento estratégico para 2030 de um grande hospital filantrópico da cidade de São Paulo¹. Um dos principais

¹ O nome do hospital não pode ser revelado por questões contratuais.

entregáveis do trabalho foi um conjunto de análises de macro tendências na saúde brasileira, como panoramas da saúde suplementar, evolução de diagnósticos etc. Salvo pesquisas pontuais proprietárias, a vasta maioria das análises valeu-se de dados públicos, de fontes como a ANS, o Datasus, a Associação Nacional dos Hospitais Privados - ANAHP, a Pesquisa Nacional de Saúde - PNS, o Conselho Federal de Medicina - CFM, entre outros.

Justamente por grande parte do trabalho ter envolvido análises aprofundadas desses dados abertos, várias oportunidades de possíveis melhorias foram debatidas pela equipe, e como poderiam-se contornar dificuldades advindas da ausência de dados, ou de dados não ideais. Valendo-se do conhecimento prévio adquirido e análises executadas para apresentação ao hospital, trazê-las aqui configura, idealmente, um esforço para que agentes governamentais pensem em como podem melhorar suas políticas de dados abertos, e que o país usufrua de suas várias consequências positivas. Por exemplo, como concluem Machado, Martins e Leite, "As melhorias em bases de dados são necessárias, [...] contribuindo para estudos estratégicos que subsidiem a tomada de decisões no planejamento de hospitais e redes de atenção à saúde"(MACHADO J., 2016).

Também, no que tange o projeto executado pela Kearney, alavancado neste trabalho, pode-se fazer um exercício de pensamento: dados de maior qualidade resultariam em um planejamento estratégico mais preciso, o que tornaria a estratégia do hospital filantrópico em questão mais eficaz, e poderia resultar em um incremento de qualidade aos pacientes atendidos, os quais, em grande parte, são advindos do Sistema Único de Saúde. Dados abertos, nesse caso, podem melhorar o panorama da saúde brasileira.

2 Metodologia

Para que seja possível analisar de modo comparativo bases de dados no que tange a qualidade destas, e para que a análise seja de fato relevante, alguns pontos intuitivos podem ser levantados:

1. **As bases, idealmente, tratam de temas similares.** Isso é útil para que possa-se ampliar a análise para o conteúdo provido pelas bases, e não somente aspectos organizacionais das mesmas.
2. **As bases são de amplo interesse público.** Para que cumpra o objetivo de servir como exemplo para outras agências governamentais e suas políticas de transparência, exemplos de destaque nacional devem ser procurados.
3. **As bases possuem vasta quantidade de dados, potencialmente relacionados.** Uma quantidade significativa de dados permite analisar sua consistência entre diferentes sub-bases, e a própria quantidade ampla oferece uma visão da organização dos dados e uniformidade deles.

Dentre as várias bases de dados de relevância nacional da saúde brasileira, optou-se por analisar os dados da ANS e do Datasus Tabnet. Em relação aos critérios previamente apresentados:

1. Ambas tratam da saúde brasileira - enquanto a ANS trata da saúde suplementar, o Datasus trata tanto da saúde pública, quanto de alguns usos da saúde privada pelo SUS, e de estatísticas gerais da saúde brasileira.
2. Ambas são de amplo interesse público. Em 2020, o planejamento estratégico da Kearney estimou o mercado de saúde brasileiro em 762 bilhões de reais, dos quais 312 são públicos, e 450, privados. De acordo com o IBGE ([NOTÍCIAS, 2021](#)), isso representa mais de 10% do PIB brasileiro do mesmo ano.
3. Ambas possuem vários subgrupos de dados, que podem ser checados entre si por consistência - por exemplo, doenças cujo volume diagnosticado aumenta nas bases do Datasus potencialmente refletem um maior volume de consultas na especialidade correspondente nas bases da ANS.

2.1 Critérios de qualidade

Ao realizar uma análise da qualidade dos dados, os 8 princípios dos dados governamentais abertos¹ são pilares fundamentais para uma estrutura funcional. Contudo, outras métricas podem ser buscadas, tanto para aprofundar as análises, quanto para torná-las mais adequadas aos objetivos do trabalho. Em 2020, foi conduzida uma análise da qualidade do Portal Brasileiro de Dados Abertos (SILVA AMBROZINA, 2020), que além de considerar os 8 princípios, também leva em consideração as **dimensões da qualidade da informação**, propostas em 2006 por Parker *et al.* (PARKER M.B., 2006), os quais são¹:

- **Acessibilidade:** Os dados são facilmente acessíveis e disponíveis para download.
- **Acuracidade:** Os dados são corretos e confiáveis.
- **Adequação:** Os dados são adequados de acordo com o que os interessados exigem.
- **Credibilidade:** Os dados são verdadeiros e credíveis.
- **Integridade:** Os dados não são faltantes, e são suficientes.
- **Consistência:** Todos os dados são apresentados no mesmo formato.
- **Facilidade de manipulação:** Os dados são facilmente manipuláveis.
- **Livre de erro:** Os dados são corretos, confiáveis, e livres de erro.
- **Objetividade:** Os dados são imparciais.
- **Relevância:** Os dados são aplicáveis, úteis e relevantes.
- **Representação:** Os dados são legíveis, consistentes e possuem uma estrutura formal.
- **Reputação:** Os dados são altamente considerados no que diz respeito ao seu conteúdo.
- **Segurança:** Os dados são apropriadamente restritos, onde necessário, para manter sua segurança.
- **Velocidade:** Os dados são baixados e/ou recuperados em velocidade adequada.
- **Atualização:** Os dados devem ser atualizados.
- **Entendimento:** Os dados devem ser facilmente entendidos/compreendidos.

¹ Os princípios foram ligeiramente modificados para se adequarem a uma realidade mais atual do que quando foram originalmente escritos, sem alteração semântica.

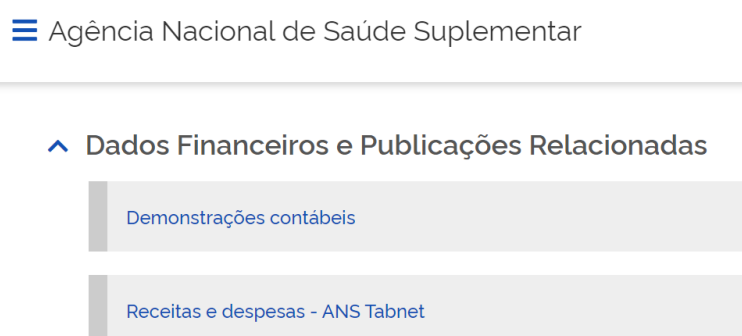
- **Valor agregado:** Os dados devem ser benéficos e proporcionar vantagens pelo seu uso.

Ao levar em consideração a análise realizada por Ambrozina do Portal Brasileiro de Dados Abertos, é importante ressaltar que o Portal em si não é uma base de dados. De acordo com sua seção de perguntas frequentes, "O portal funciona como um catálogo federado que facilita a busca e uso de dados publicados pelos órgãos do governo" (ABERTOS, s.d.). Não obstante, investigar a qualidade dos dados do portal pode ser algo feito de modo similar à uma análise dos dados em si das bases por ele catalogadas. Sendo assim, utilizando a avaliação qualitativa do Portal como base, isso foi replicado para os estudos de caso desse trabalho - a ANS e o Datasus Tabnet.

As bases da ANS e do Datasus são apresentadas por meio de uma plataforma chamada **Tabnet**, o que também torna sua comparação facilitada. De acordo com o Datasus, trata-se de um tabulador genérico de domínio público que permite organizar dados de forma rápida, conforme a consulta que se deseja tabular (SAÚDE, s.d.). É possível baixá-las em um formato próprio (.dbc), para que se tenha acesso local à base completa por meio de um software desenvolvido para sua visualização, chamado TabWIN. Porém, como uma das dimensões da qualidade da informação é a **acessibilidade**, faz sentido analisá-las em extração direta via interface web.

De fato, todas as bases seguem um processo razoavelmente padronizado de organização em colunas, exibição, e possibilidade de download em .csv, que pode ser manipulado por softwares mais tradicionais, como Microsoft Excel, scripts em Python etc. Por exemplo, caso se queira obter no ANS Tabnet a receita de contraprestações de cada operadora médico-hospitalar ativa no país nos últimos 5 anos, o processo é razoavelmente simples: a partir da seção da ANS do site de acesso à informação do Governo Federal, acessam-se as receitas e despesas, como mostra a figura 1.

Figura 1 – Acesso ao painel de receitas e despesas da ANS no site de acesso à informação do Governo Federal.



A partir disso, selecionam-se os valores de linha, coluna, conteúdo e períodos, além de filtros caso se queira utilizá-los, como mostra a figura 2.

Figura 2 – Definição dos dados de receitas e despesas da ANS a serem extraídos.

Receitas e Despesas (Dados do 1º trimestre de 2022 – Data de corte 24/06/2022)

Linha:
 Coluna:
 Conteúdo:

Períodos Disponíveis

Seleções Disponíveis

Modalidade:

Operadora:

Grupo Modalidade:

Ordenar pelos valores da coluna

Fonte: ans.gov.br/anstabnet/cgi-bin/dh?dados/tabnet_rc.def. Acesso em: 30/08/2022.

Por fim, os dados são exibidos dentro do próprio site, com opções de download em .csv e .dbc ao final da tabela exibida na figura

Figura 3 – Resultado da consulta realizada na figura 2.

Receitas e Despesas (Dados do 1º trimestre de 2022 – Data de corte 24/06/2022)

Receita de contraprestações por Ano segundo Operadora
Grupo Modalidade: Operadoras médico-hospitalares
Período: 2018-2022

Operadora	2018	2019	2020	2021	2022	Total
TOTAL	192.070.062.637	207.583.098.415	217.477.021.944	239.204.543.067	57.876.338.923	914.211.064.985
005711-BRADESCO SAUDE S.A.	21.821.023.153	24.047.350.176	24.438.499.057	26.836.625.786	6.399.529.537	103.543.027.709
326305-AMIL ASSISTENCIA MEDICA INTERNACIONAL S.A.	20.210.210.423	20.932.205.303	19.278.921.500	19.484.933.662	3.825.351.100	83.731.621.987
006246-SUL AMERICA COMPANHIA DE SEGURO SAUDE	15.466.130.544	17.172.488.412	18.191.455.825	19.096.375.853	4.980.028.831	74.906.479.466
359017-NOTRE DAME INTERMEDICA SAUDE S.A.	5.519.052.424	6.527.394.142	8.835.850.986	9.662.274.653	2.496.762.206	33.041.334.410
368253-HAPVIDA ASSISTENCIA MEDICA S.A.	4.685.774.204	5.383.304.930	6.074.614.150	7.477.092.979	2.302.502.573	25.923.288.837
346659-CAIXA DE ASSISTENCIA DOS FUNCIONARIOS DO BA	4.885.781.364	5.632.025.940	6.086.469.514	6.335.004.577	1.573.893.749	24.513.175.145
393321-UNIMED-RIO COOPERATIVA DE TRABALHO MEDICO D	4.286.140.864	4.501.592.080	4.816.744.877	4.999.650.888	1.263.367.943	19.867.496.650
339679-CENTRAL NACIONAL UNIMED - COOPERATIVA CENTR	2.891.215.151	4.078.715.813	5.064.097.850	6.032.404.197	1.480.617.148	19.547.050.159
343889-UNIMED BELO HORIZONTE COOPERATIVA DE TRABAL	3.917.146.007	4.215.048.143	4.524.351.113	4.867.570.654	1.283.439.580	18.807.555.498
323080-GEAP AUTOGESTAO EM SAUDE	4.065.181.462	4.144.592.478	4.058.847.478	3.919.774.350	936.815.489	17.125.211.258
302147-PREVENT SENIOR PRIVATE OPERADORA DE SAUDE L	2.865.332.797	3.487.199.584	4.096.432.474	4.927.513.530	1.212.966.758	16.589.445.144
352501-UNIMED PORTO ALEGRE - COOPERATIVA MEDICA LT	2.589.839.749	2.919.967.472	3.033.825.414	3.349.016.722	740.933.771	12.633.583.128
304701-UNIMED CURITIBA - SOCIEDADE COOPERATIVA DE	2.463.821.622	2.691.256.032	2.776.985.722	2.999.585.252	642.273.525	11.573.922.154
335690-UNIMED CAMPINAS - COOPERATIVA DE TRABALHO M	2.077.397.638	2.263.211.198	2.525.520.948	2.685.260.473	620.444.063	10.171.834.320
000701-UNIMED SEGUROS SAUDE S/A	2.428.548.103	1.884.028.171	2.236.797.017	2.402.070.281	883.842.956	9.835.286.527
416428-SUL AMERICA SERVICOS DE SAUDE S.A.	1.881.592.274	2.240.130.370	2.272.086.068	2.827.440.835	16.699.955	9.237.949.502
317144-UNIMED DE FORTALEZA SOCIEDADE COOPERATIVA M	1.672.285.576	2.188.855.209	2.250.049.269	2.514.864.236	557.359.386	9.183.413.676
000582-PORTO SEGURO - SEGURO SAUDE S/A	1.426.853.751	1.613.904.109	1.823.357.416	2.109.455.135	673.566.760	7.647.137.172

Fonte: ans.gov.br/anstabnet/cgi-bin/tabnet?dados/tabnet_rc.def. Acesso em: 30/08/2022.

2.2 Comparação

Durante a análise qualitativa das bases, paralelos subjetivos podem ser traçados entre elas, os quais podem ilustrar como uma agência pode melhorar a partir de estratégias seguidas por outras, e idealmente isso pode ser escalado para diferentes agências governamentais. Além disso, como exercício de pensamento, alguns paralelos podem ser traçados até com bases de outros países, que podem oferecer vantagens e desvantagens em comparação às bases brasileiras de saúde.

Por fim, para a ilustração concreta da avaliação de qualidade, exemplos foram selecionados a partir do planejamento estratégico executado pela Kearney para um hospital filantrópico de São Paulo. Além de funcionarem como exemplos de uso real das bases analisadas, eles permitem demonstrar como as bases podem melhorar sua arquitetura de dados, em relação aos critérios de qualidade utilizados, e com relação à outros problemas enfrentados na utilização das bases no projeto.

2.3 Leis de proteção de dados

Enquanto analisar o cumprimento dos conjuntos à Lei Geral de Proteção de Dados - LGPD (e outras leis de outros países para esse fim), foge ao escopo desse trabalho, é importante verificar minimamente o cumprimento dela por parte dos objetos de estudo, especialmente pela ampla importância dada ao tema nos últimos anos, como a emblemática *General Data Protection Regulation* - GDPR (Proteção Regulatória de Dados Gerais, em tradução livre), vigente para os países da União Europeia.

3 Resultados

Todas as análises puras das bases foram feitas desconsiderando quaisquer constatações previamente feitas no projeto da Kearney que foi usado ao longo do trabalho, para que não houvessem vieses. Elas seguem a seguinte estrutura:

1. Análise geral das bases da ANS e do Datasus segundo os 8 princípios dos DGAs;
2. Análise geral das bases da ANS segundo as dimensões da qualidade da informação de Parker *et al.*;
3. Análise geral das bases do Datasus segundo as dimensões da qualidade da informação de Parker *et al.*;

Ao longo das análises e posteriormente, para que os questionamentos de qualidade não permaneçam abstratos, análises selecionadas do planejamento estratégico da Kearney para o hospital de São Paulo são trazidas, posando questionamentos e constatações sobre a qualidade dos dados dispostos e seu uso.

Adicionalmente, como citado no item 2.3, vale ressaltar que todos os dados dos conjuntos analisados encontravam-se anonimizados por IDs não vinculáveis às pessoas específicas que poderiam dizer respeito, tanto na ANS, quanto no Datasus.

3.1 Os 8 princípios

Como listados anteriormente¹, os 8 princípios caracterizam se os dados em questão são "abertos" ou não. Nesse aspecto, ambas as bases se saem bem, por ser um pilar fundamental e não negociável.

1. **Completos:** Ambas as bases dispõem de todos os dados que se comprometeram a inserir. Quando algo passa a não ser mais atualizado, isso está descrito.
2. **Primários:** Por se tratarem das geradoras de grande parte dos dados, por consequência são primários. Quando isso não ocorre, como é o caso da população residente no Datasus, por exemplo, isso é descrito logo no título (nesse caso, que a fonte primária é o IBGE).
3. **Oportunos:** Após o período de processamento do dado em questão, ele é inserido no arquivo público. Por exemplo, no caso da figura 3, o acesso foi realizado em agosto e os dados do 2º trimestre de 2022 já haviam sido inseridos.

4. **Acessíveis:** Não há qualquer restrição de acesso aos dados, ou de como eles podem ser exibidos, a não ser por restrições lógicas contornáveis. Por exemplo, na figura 2, não se pode escolher múltiplos anos com uma coluna selecionada, mas isso é contornável obtendo-se os dados ano a ano.
5. **Processáveis por máquinas:** Sempre há uma visualização web e uma opção de download em .csv. Nem sempre ela é padronizada entre bases diferentes, mas ela sempre é processável por máquina.
6. **Não discriminatórios:** Em nenhum momento, se é pedido um registro.
7. **Não proprietários:** Enquanto o formato .dbc disponibilizado é proprietário (DBC..., 2017), sempre há a disponibilização em .csv, que é um formato não proprietário por se tratar simplesmente de valores separados por vírgula.
8. **Livres de licenças:** Os dados não possuem licenças aplicáveis.

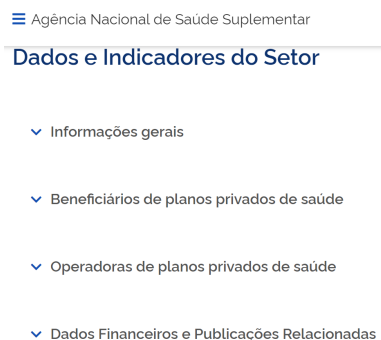
3.2 Dimensões da qualidade da informação

Ao se analisar as dimensões da qualidade da informação, há maior espaço para não cumprimento pleno dos requisitos, e uma investigação mais detalhada das bases é requerida. Por englobar uma quantidade mais ampla de bases, naturalmente há maior espaço para cumprimento parcial das dimensões nas bases do Datasus do que da ANS. Uma lista exaustiva dos níveis e subníveis das bases analisadas encontra-se no apêndice A.1 do trabalho.

3.2.1 ANS Tabnet

Por ser uma base com menor quantidade de subquebras de dados possíveis, o ambiente do ANS Tabnet é mais controlado, totalmente formatado na estrutura do Tabnet.

Figura 4 – Página inicial da ANS via portal gov.br.



Fonte: <https://www.gov.br/ans/pt-br/acesso-a-informacao/perfil-do-setor/dados-e-indicadores-do-setor>.
Acesso em: 31/08/2022.

Também há um link direto para o ANS Tabnet acessível via Datasus e o portal gov.br, que deveria levar à página inicial própria do ANS Tabnet (e não via gov.br). Todavia, o link em questão (<ans.gov.br/anstabnet>) encontrava-se somente com texto, e sem acesso às bases, pelo menos, durante a totalidade do mês de agosto de 2022, quando essa seção do trabalho foi fechada. De fato, esse foi o endereço utilizado ao longo de todo o projeto da Kearney, mas ele parece ter sido descontinuado, e o link foi mantido. Todos os dados puderam ser acessados via gov.br, e por isso não houve prejuízo ao trabalho. Porém, esse descasamento de links prejudica a acessibilidade, pois torna confuso o acesso à página.

Abaixo, encontra-se a análise das bases de dados da ANS de acordo com as dimensões da qualidade da informação.

1. Acessibilidade:

Todos os dados propostos estão disponíveis via web, oferecendo várias quebras administrativas e filtros, e disponibilizando o download em .csv ou .dbc. Além disso, encontram-se listados no Portal Brasileiro de Dados Abertos, com 57 conjuntos de dados catalogados, o que oferece mais uma possibilidade de acesso aos dados, como mostra a figura 5. O Portal é um indexador de dados de várias entidades governamentais, como explicado na seção 2.1.

Figura 5 – Página da ANS no Portal Brasileiro de Dados Abertos.



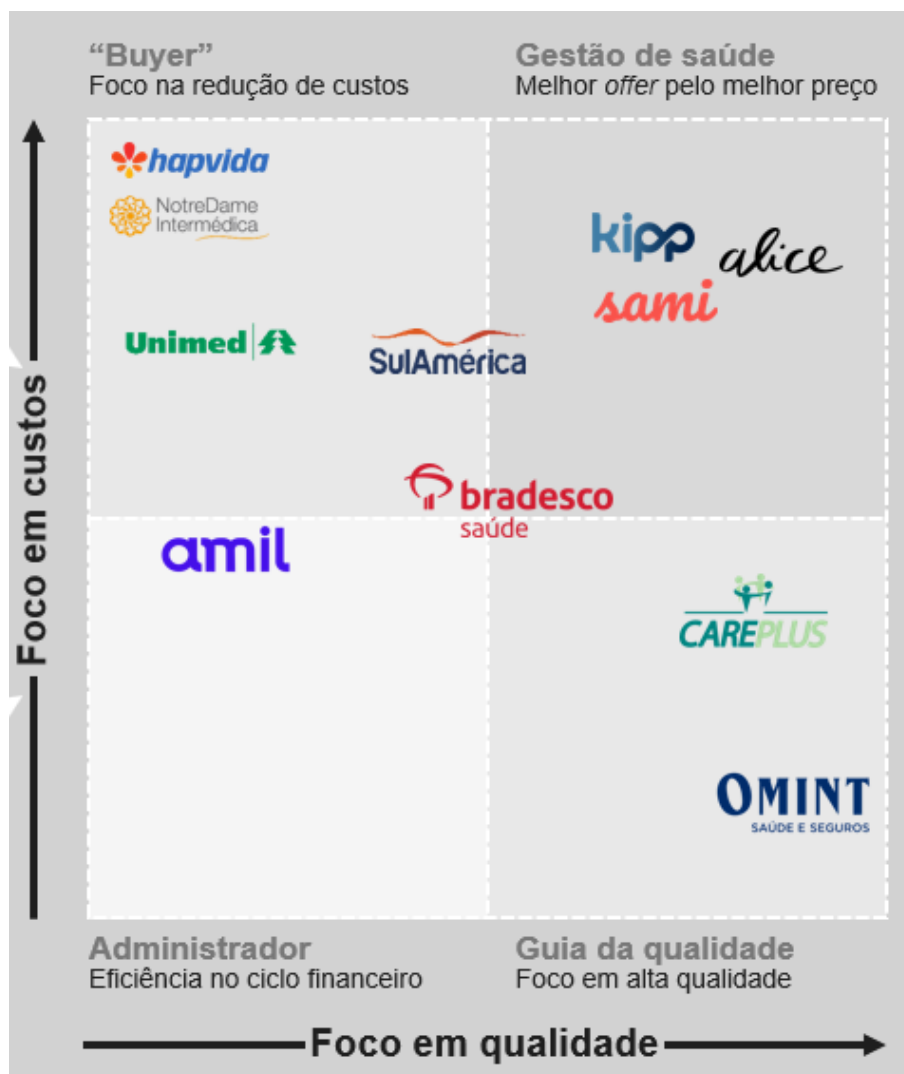
Fonte: <https://dados.gov.br/organization/agencia-nacional-de-saude-suplementar-ans>. Acesso em: 31/08/2022.

Todavia, a página principal do ANS Tabnet, como descrita anteriormente, encontra-se incompleta, o que pode causar confusão inicial e possível desistência da procura pelos dados. A página e seus vínculos em outros sites deveriam ser desativados, ou a página retomada.

2. Acuracidade:

Por mais que não seja possível verificar exatamente a acurácia de todos os dados pelo fato da ANS se tratar da fonte deles, análises de especialistas da Kearney indicam que eles, de fato, fazem sentido, e portanto são acurados. Por exemplo: com base em relacionamentos construídos com clientes e outros estudos feitos no passado, a Kearney possuía uma análise qualitativa dos planos de saúde no país, plotados em uma matriz de foco em qualidade X foco em custos, como ilustra a figura 6.

Figura 6 – Análise Kearney da estratégia de operadoras de saúde selecionadas. Eixo X: Foco em qualidade. Eixo Y: Foco em custos.



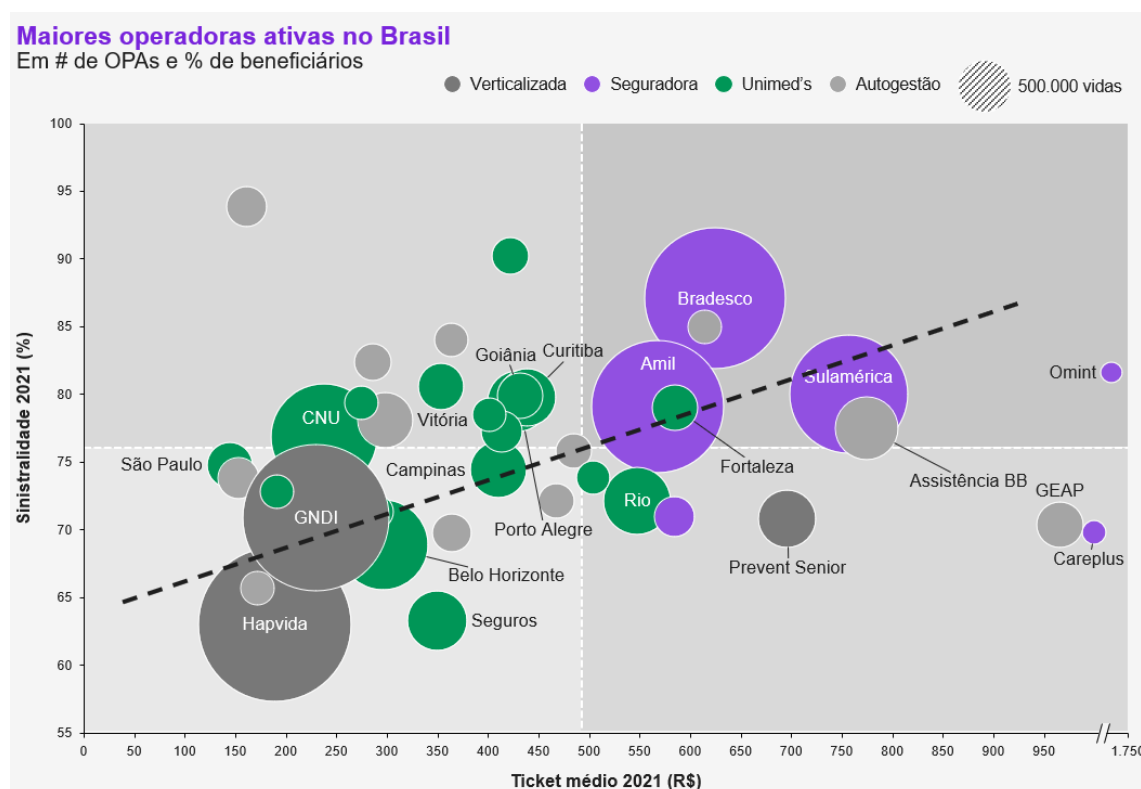
Fonte: Kearney.

Ao comparar as operadoras¹ previamente mapeadas com um gráfico de sinistralidade (% de receita de contraprestações das operadoras dividido pelas suas despesas operacionais/com pacientes - quanto menor a sinistralidade, maior a margem de lucro) por ticket médio (valor médio pago por beneficiário da operadora), oriundo

¹ Unimed's agrupadas só facilitam a visualização, pois atuam de modo independente.

totalmente de dados da ANS, há consonância dos dados - operadoras com maior ticket médio possuem maior foco na qualidade do serviço, enquanto operadoras mais focadas no custo priorizam uma sinistralidade baixa, por vezes resultando em um ticket médio relativamente inferior. Essa matriz encontra-se abaixo, na figura 7.

Figura 7 – Análise Kearney de sinistralidade X ticket médio de operadoras selecionadas, feita inteiramente com dados da ANS. Eixo X: Ticket médio em R\$. Eixo Y: Sinistralidade da operadora em 2021 em %.



Fonte: Kearney.

Nessa análise, destaca-se principalmente o posicionamento de Omint e Careplus, que possuem alto ticket médio e alto foco na qualidade.

Com isso, a acuracidade dos dados parece adequada.

3. Adequação:

Os conjuntos estão de acordo com as exigências dos interessados, porém nem sempre no formato mais adequado. Por exemplo: vários dados encontram-se em PDFs, como o Mapa Assistencial da Saúde Suplementar e o Prisma Econômico da Saúde Suplementar. Esses exemplos, particularmente, tornaram-se um *dashboard* em Power BI a partir de 2019, o que melhora a situação, mas ainda assim não é ideal, pois não existe uma maneira simples de extraí-los. É disponibilizada uma planilha, porém a única quebra se dá por modalidade de operadora (autogestão, cooperativa médica,

filantropia etc), o que compromete a granularidade da informação. No *dashboard*, por exemplo, há granularidade por operadora, como mostra a figura 8.

Figura 8 – Dados da Sompso Saúde Seguros, em uma visão por operadora do Mapa Assistencial da Saúde Suplementar da ANS.



Fonte: Mapa Assistencial da Saúde Suplementar, disponível em

<<https://www.gov.br/ans/pt-br/aceso-a-informacao/perfil-do-setor/dados-e-indicadores-do-setor>>.

Acesso em: 31/08/2022.

Isso implicou no projeto da Kearney, por exemplo, em alguns comandos de cópia e cola em planilhas para a manipulação. Isso mostra que alguns dados, possivelmente, não estão de acordo com as exigências dos interessados, e também não são tão facilmente manipuláveis (problema explorado no item 7 abaixo).

Nessa frente, a unificação e disposição adequada dos dados em alto nível de granularidade é deficitária.

- Credibilidade:** A ANS é tida como uma agência reguladora reputável. Em 2022, por exemplo, recebeu destaque midiático ao suspender 70 planos de saúde de 8 operadoras (AQUINO, 2022).
- Integridade:** Os dados não são ordinariamente trocados, e quando são colhidos de outras fontes, são adequadamente citadas. Por exemplo, o Mapa Assistencial da Saúde Suplementar, citado no item 3, ao se valer do Sistema de Informações de Produtos (SIP), cita-o diretamente.
- Consistência:** O formato do Tabnet, presente na vasta maioria dos dados, mostra alto grau de consistência. Só não é completa devido aos itens anteriormente citados, em PDFs e *dashboards*. Todavia, a vasta maioria, como os 57 dados disponibilizados

no Portal Brasileiro de Dados Abertos (ver figura 5), encontram-se no mesmo padrão, em formato .csv e .dbc.

7. Facilidade de manipulação:

Como abordado no item 3, não há padronização total dos dados da ANS. Sempre que um arquivo é disponibilizado na maior granularidade somente em *dashboards* ou em PDFs, há dificuldade de manipulação, visto que nem sempre um programa de reconhecimento óptico de caracteres extrai dados da maneira almejada. Porém, como abordado no item 6, a vasta maioria dos dados encontra-se unificada, o que facilita a extração e manipulação dessa maioria.

8. **Livre de erro:** A reputação da agência, conjuntamente à aparente confirmação dos dados com conhecimentos de mercado, como abordado no item 2, configuram evidências de dados livres de erro. Não é possível garantir esse item, porém é plausível tê-lo como assegurado.
9. **Objetividade:** Os dados são imparciais e objetivos.
10. **Relevância:** Os dados são de absoluta relevância para obter-se visões da saúde suplementar no país, visto que é a única agência reguladora nacional do tema.
11. **Representação:** A representação dos dados segue um padrão tabular, presente em todos os dados que se valem da plataforma Tabnet. Quando há dados dispostos de modo mais técnico, há um dicionário acompanhando os dados, como ilustram as figuras 9 e 10 para as características dos produtos da saúde suplementar cadastrados na ANS.

Figura 9 – Amostra de colunas das características de produtos de saúde suplementar cadastrados na ANS.

ID_PLANO	CD_PLANO	NM_PLANO	CD_OPERADORA	RAZAO_SOCIAL	GR_MODALIDADE
18414598	490006215	POTENCIAL	422762	Sociedade Portu	Filantropia
18500088	492526222	POTENCIAL	422762	Sociedade Portu	Filantropia
18500108	492528229	CONFIANÇA	422762	Sociedade Portu	Filantropia
18443929	490721213	DENTAL IND	422975	Dentalmaster Lt	Odontologia De Gr
18381488	489143211	COLETIVO EI	422771	Plan Saúde Ltda	Medicina De Grup

Fonte: Características dos Produtos da Saúde Suplementar, disponível em

<<https://www.gov.br/ans/pt-br/aceso-a-informacao/perfil-do-setor/dados-e-indicadores-do-setor>>.

Acesso em: 31/08/2022.

Figura 10 – Amostra do dicionário das características de produtos de saúde suplementar cadastrados na ANS.

Nome do Campo	Tipo	Tamanho	Descrição
ID_PLANO	NUM	8	Identificação única do plano. <i>Nota:</i> Este campo corresponde à chave primária "ID_PLANO" do conjunto de dados "Abrangência Geográfica dos Planos de Saúde".
CD_PLANO	CHAR	30	Código de registro do produto na ANS. Para os planos contratados após 1º de janeiro de 1999 trata-se do número de registro concedido para o plano de acordo com a RN nº 100/05. Para os planos contratados até 1º de janeiro de 1999 e os planos adaptados à Lei 9.656/98, refere-se ao código de identificação do plano no Sistema de Cadastro de Planos Privados de Assistência à Saúde (SCPA), conforme a RN nº 56/03.
NM_PLANO	CHAR	140	Nome Comercial do Plano registrado na ANS.
CD_OPERADORA	NUM	6	Registro de operadora de plano privado de assistência à saúde concedido pela ANS a pessoa jurídica para operação no setor de saúde suplementar. <i>Nota:</i> Este campo corresponde à chave primária "Registro ANS" dos conjuntos de dados abertos "Operadoras de planos de saúde Ativas" e "Operadoras de planos de saúde Canceladas".
RAZAO_SOCIAL	NUM	140	Razão Social da Operadora

Fonte: Mapa Assistencial da Saúde Suplementar, disponível em

<<https://www.gov.br/ans/pt-br/aceso-a-informacao/perfil-do-setor/dados-e-indicadores-do-setor>>.

Acesso em: 31/08/2022.

As bases da ANS não mostram nenhum problema aparente ao se analisar a representação dos dados.

12. Reputação:

Ao encontro da credibilidade (item 4), além de fontes reputáveis, os dados apresentam consistência histórica, o que assegura segurança nesse aspecto.

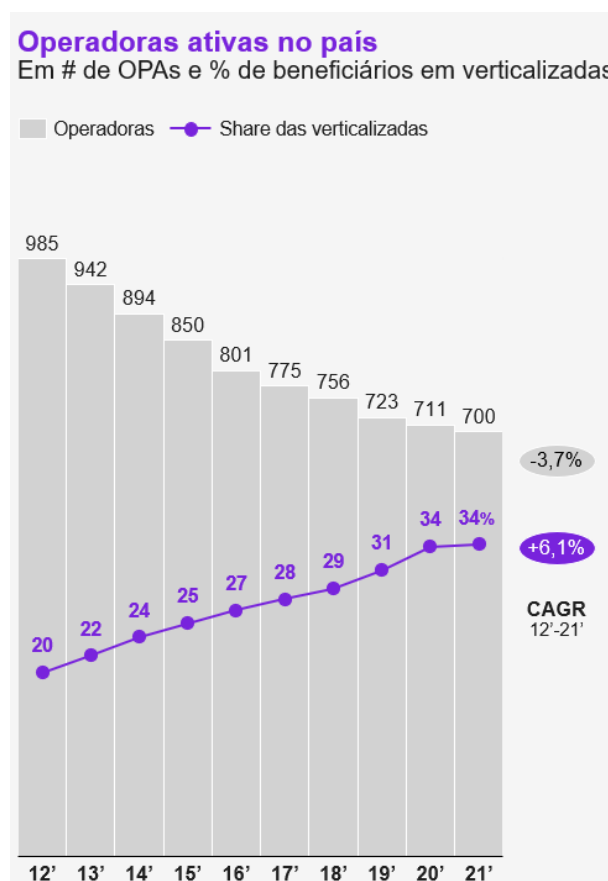
Existe uma análise feita pela Kearney que ilustra esse aspecto da reputação justificada pela consistência histórica, que também argumenta pela credibilidade das bases: nos últimos anos, tem-se observado uma tendência de verticalização no mercado de operadoras de saúde do país. De acordo com Tiago Reis, "Verticalização é uma metodologia empresarial que centraliza a cadeia de produção"(REIS, 2019). Dois exemplos notórios de verticalização desse mercado são, por exemplo:

- A compra da SulAmérica (seguradora que opera planos de saúde) pela Rede D'Or (rede de hospitais) (ALVES, 2022).
- A compra de dois hospitais da marca Leforte pelo DASA (Diagnósticos da América) (ANDRADE, 2021).

Ou seja, a expansão de um negócio na cadeia produtiva: uma rede de diagnósticos comprando hospitais, uma rede de hospitais comprando uma operadora de saúde etc. De fato, isso pode ser observado nos dados da ANS ao analisar-se a quantidade de operadoras ativas no país, e o *share* (fatia de mercado) das operadoras verticalizadas

em receita bruta (classificadas por município com metodologia proprietária), como ilustra a figura 11.

Figura 11 – Quantidade de operadoras ativas e *share* das classificadas como verticalizadas, de 2012 a 2021, e CAGRs respectivos (a definição de CAGR encontra-se no apêndice A.2).



Fonte: Kearney.

Conclui-se que a fonte mostra ser reputável.

- Segurança:** Não há restrições de informação por motivo de segurança. Ademais, a própria árvore FTP dos dados exibidos é acessível em <http://ftp.dadosabertos.ans.gov.br/FTP/>.
- Velocidade:** Ao longo dos usos da base, não houveram problemas relacionados à velocidade de download.
- Atualização:** A ANS não discorre sobre o processo de atualização em si, porém, os dados disponibilizados sempre são oportunos: dados trimestrais são disponibilizados no trimestre seguinte, dados anuais no ano seguinte, mensais no mês seguinte etc. Ou seja, a avaliação é positiva.
- Entendimento:** Não existe um glossário de siglas quando estas são utilizadas, o que pode prejudicar um pouco o uso por parte de um leigo. Todavia, são siglas facilmente

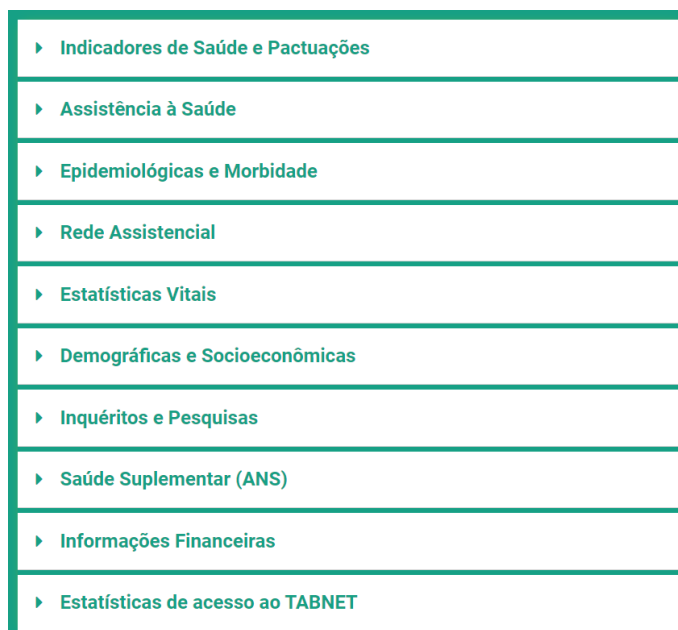
buscáveis online, como por exemplo, nos dados de ressarcimento ao SUS, o que é um "capítulo CID-10", o que é básico para quem já é familiar no assunto, mas não é de conhecimento popular (CID é uma **C**lassificação **I**nternacional de **D**oenças).

17. **Valor agregado:** Naturalmente, é lógico concluir que há alto valor agregado nessas bases, visto, por exemplo, as análises do projeto real da Kearney que populam esse trabalho.

3.2.2 Datasus Tabnet

Enquanto conjuntos de dados mais amplos, o Datasus Tabnet requer, naturalmente, maior coordenação por parte dos envolvidos na construção e agregação dos dados. Nota-se um esforço de padronização ao navegar pelos conjuntos oferecidos, porém isso nem sempre ocorre. Ao entrar no Datasus, depara-se com quebras intuitivas, como mostra a figura 12.

Figura 12 – Página inicial do Datasus Tabnet.



Fonte: <https://datasus.saude.gov.br/informacoes-de-saude-tabnet/>. Acesso em: 01/09/2022.

Uma lista completa com as quebras e subquebras disponíveis no momento da escrita desse trabalho encontra-se no apêndice A.1.2. Uma análise superficial das quebras não encontrou sobreposições, o que demonstra boa arquitetura das bases, de sistemas não concomitantes. Abaixo, encontra-se a análise das bases de dados do Datasus Tabnet de acordo com as dimensões da qualidade da informação.

1. **Acessibilidade:**

Assim como no ANS Tabnet, todos os dados propostos estão disponíveis via web, oferecendo várias quebras administrativas e filtros, e disponibilizando o download em

.csv ou .dbc. Porém, ao contrário da ANS, não há listagem no Portal Brasileiro de Dados Abertos, o que pode prejudicar a acessibilidade do dado procurado. Existe uma página do Ministério da Saúde no Portal, mas os conjuntos não são correspondentes (figura 13).

Figura 13 – Página do Ministério da Saúde no Portal Brasileiro de Dados Abertos.

The screenshot shows the 'Ministério da Saúde - MS' page on the Brazilian Open Data Portal. The page features a search bar with the text 'Pesquisar conjuntos de dados...' and a magnifying glass icon. Below the search bar, it indicates '34 resultado(s) encontrado(s)' and a dropdown menu for 'Ordenar por: Relevância'. The main content area lists two datasets: 'Unidades Básicas de Saúde - UBS' and 'Febre Amarela em humanos e primatas não-humanos - 1994 a 2021'. The 'Unidades Básicas de Saúde - UBS' dataset has a 'CSV' download option. The 'Febre Amarela...' dataset has 'PDF' and 'CSV' download options. On the left side, there is a profile section for 'Ministério da Saúde - MS' with a description and statistics showing '0 Seguidores' and '34 Conjuntos de dados'.

Fonte: <https://dados.gov.br/organization/ministerio-da-saude-ms>. Acesso em: 01/09/2022.

Além dessas duas fontes de dados vinculadas ao Ministério da Saúde, há uma 3ª: o openDatasus, que disponibilizava 16 conjuntos de dados quando este trabalho foi escrito, principalmente relacionados à pandemia de Covid-19, como mostra a figura 14.

Figura 14 – Página inicial do openDatasus.

The screenshot shows the 'openDataSUS' website. The page features a search bar with the text 'Pesquisar' and a magnifying glass icon. Below the search bar, it indicates '16 conjuntos de dados encontrados' and a dropdown menu for 'Ordenar por: Relevância'. The main content area lists two datasets: 'Campanha Nacional de Vacinação contra Covid-19' and 'SRAG 2021 e 2022 - Banco de Dados de Síndrome Respiratória Aguda Grave - incl...'. The 'Campanha Nacional de Vacinação contra Covid-19' dataset has 'PDF', 'API', and 'CSV' download options. The 'SRAG 2021 e 2022...' dataset has a 'PDF' download option. On the left side, there is a navigation menu with categories: 'Organizações' (16), 'Grupos' (2), and 'Etiquetas' (5). The 'Organizações' category is expanded, showing 'Ministério da Saúde' with 16 datasets.

Fonte: <https://opendatasus.saude.gov.br/dataset>. Acesso em: 01/09/2022.

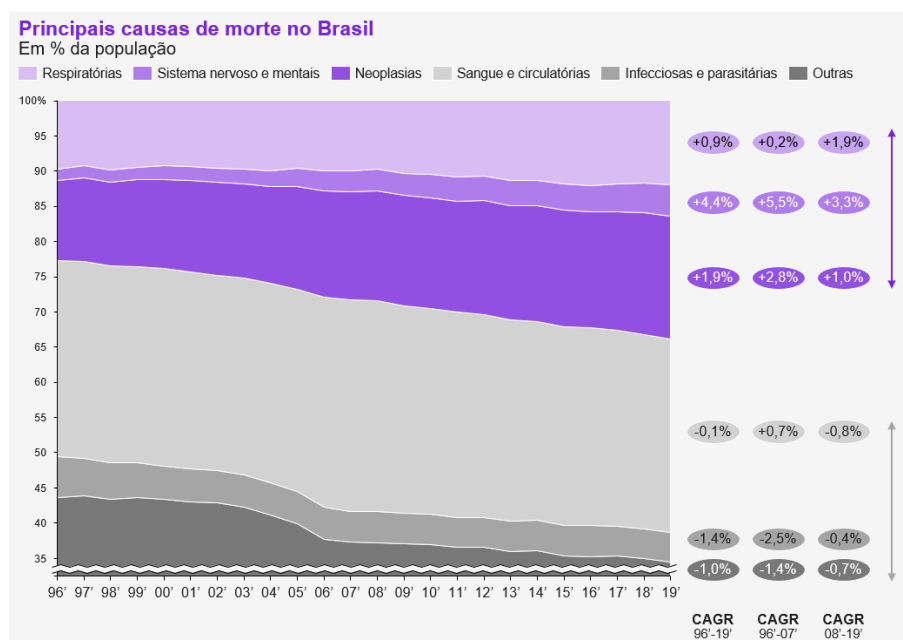
Durante a investigação dessas bases, os conjuntos mapeados na página do Portal Brasileiro de Dados Abertos se sobrepunham parcialmente com os do openDatusus. Porém, nenhum deles possuía sobreposição com aqueles do Datusus Tabnet. Devido a essa dispersão, a acessibilidade pode ficar prejudicada ao interessado que ainda não é familiarizado com as bases em questão.

2. Acuracidade:

Assim como no caso da ANS, não há como especificamente validar a acurácia dos dados. Todavia, ao longo do projeto executado pela Kearney, os dados utilizados mostraram-se consistentes com resultados previstos, o que leva a crer que são, de fato, acurados.

Uma análise que foi executada pelo projeto e comprova uma tendência esperada foi a relação entre o envelhecimento populacional e o perfil de causas de morte no país. Sabe-se que a população brasileira está em processo de envelhecimento - a CNN noticiou, em 2021, com base em um estudo do IPEA, que espera-se que aproximadamente 40% da população brasileira seja idosa em 2100, contra 7,3% em 2010 (JANONE, 2021). Com isso, esperava-se que o projeto constatasse que o perfil de causas de morte no país evoluiria para um perfil mais relacionado à doenças crônicas e neoplasias. De fato, a plotagem desses dados obtidos via Datusus explicita essa realidade, como mostra a figura 15.

Figura 15 – Percentual de causas de morte no Brasil por grupo de causa, de 1996 a 2019, e CAGRs selecionados (definição de CAGR no apêndice A.2).

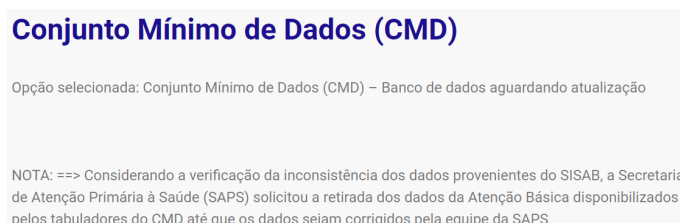


Fonte: Kearney.

Além disso, outra possível evidência de acuracidade encontrada ao longo da pesquisa

desse trabalho é o Conjunto Mínimo de Dados, disponível na quebra de *assistência à saúde*: Ao ser verificada uma inconsistência, os dados foram removidos até que sejam corrigidos, conforme nota do site (figura 16).

Figura 16 – Nota exibida ao se tentar baixar o conjunto mínimo de dados.



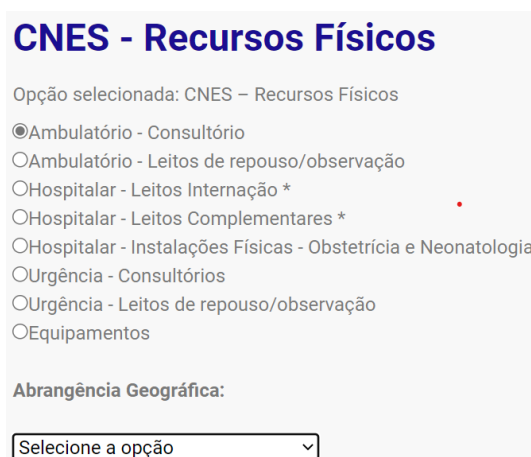
Fonte: <https://datasus.saude.gov.br/conjunto-minimo-de-dados-cmd/>. Acesso em: 01/09/2022.

Assim, há evidências de que os dados são acurados.

3. Adequação:

No que tange a adequação, os dados são mais próximos do que se espera de uma padronização do que os dados disponibilizados pela ANS. Da totalidade presente e verificada durante a escrita desse trabalho, todos levavam a uma página uniforme de geração tabular de dados, que é a própria visualização do Tabnet. Em algumas extrações existe um filtro intermediário de preenchimento necessário, porém este também é padronizado. O filtro intermediário e a visualização Tabnet são exemplificadas para os recursos físicos do Cadastro Nacional de Estabelecimentos de Saúde - CNES, disponíveis na quebra de rede assistencial do Datasus Tabnet, nas figuras 17 (filtro), 18 (seleção de dados) e 19 (resultado da seleção).

Figura 17 – Filtro intermediário do Datasus ao se extrair os dados de recursos físicos. Na abrangência geográfica, pode-se selecionar "Brasil por região, UF e município", ou um estado específico.



Fonte: <https://datasus.saude.gov.br/cnes-recursos-fisicos>. Acesso em: 01/09/2022.

Figura 18 – Seleção de dados ao se extrair os recursos de consultório em ambulatórios do estado de São Paulo.

➤ CNES - RECURSOS FÍSICOS - AMBULATÓRIO - CONSULTÓRIOS - SÃO PAULO

Linha	Coluna	Conteúdo
Município	Não ativa	Clínica Básica
Município gestor	Região de Saúde (CIR)	Clín. Especializada
Região de Saúde (CIR)	Macrorregião de Saúde	Clínico Indiferente
Região de Saúde/Município	Divisão administ estadual	Odontologicos

➤ PERÍODOS DISPONÍVEIS

Jul/2022
Jun/2022
Mai/2022
Abr/2022
Mar/2022
Esv/2022

➤ SELEÇÕES DISPONÍVEIS

- Município
- Município gestor
- Região de Saúde (CIR)
- Macrorregião de Saúde
- Divisão administ estadual
- Microrregião IBGE
- Região Metropolitana - RIDE
- Ensino/Pesquisa

Fonte: <https://datasus.saude.gov.br/cnes-recursos-fisicos>. Acesso em: 01/09/2022.

Figura 19 – Resultado da extração configurada nas figuras 17 e 18.

➤ CNES - RECURSOS FÍSICOS - AMBULATÓRIO - CONSULTÓRIOS - SÃO PAULO

Clínica Básica segundo Município
Período: Jul/2022

Município	Clínica Básica
TOTAL	43.527
350010 Adamantina	14
350020 Adolfo	2
350030 Aguaí	37
350050 Águas de Lindóia	27
350055 Águas de Santa Bárbara	3
350070 Agudos	33

Fonte: <https://datasus.saude.gov.br/cnes-recursos-fisicos>. Acesso em: 01/09/2022.

Os dados são razoavelmente adequados.

4. **Credibilidade:** No que tange o conteúdo exibido ser verdadeiro e credível, não existem grandes razões para duvidar. Todavia, seria incompleto não ressaltar os ataques hackers e brechas de segurança já encontradas no Datasus. Em dezembro de 2021, por exemplo, foi-se noticiado que o Ministério da Saúde havia sido notificado duas vezes sobre um risco de ataque hacker até que, eventualmente, ele ocorreu (NEIVA, 2021). Assim, a credibilidade não é total.
5. **Integridade:** As seleções executadas ao longo do projeto da Kearney e testadas para esse trabalho pareceram íntegras, sem elementos faltantes. Adicionalmente,

fontes externas, assim como na ANS, são citadas.

6. **Consistência:** A maneira de exibição dos dados mostra-se altamente consistente. No Datasus Tabnet, o problema de documentos em PDF ou visualizações alternativas, como acontece na ANS, não foi detectado. A avaliação desse aspecto é positiva.
7. **Facilidade de manipulação:** Alinhado com a dimensão de consistência citada (item 6), dados consistentes ao longo dos conjuntos tornam a manipulação significativamente facilitada.
8. **Livre de erro:** Assim como na ANS, a reputação da instituição, conjuntamente à aparente confirmação dos dados com conhecimentos de mercado, como ilustrado no item 2, configuram evidências de dados livres de erro. Não é possível garantir esse item, porém é plausível tê-lo como assegurado.
9. **Objetividade:** Os dados, novamente, são imparciais e objetivos.
10. **Relevância:** A relevância dos dados disponibilizados é inquestionável. Por lidar com métricas de morbidade, diagnósticos e outros temas relevantes à políticas de saúde nacionais, esses dados logicamente ajudam a guiar decisões de saúde públicas e privadas ao redor do país.
11. **Representação:** Não há problemas na representação dos dados, o que também é, em grande parte, fruto da notação tabular do Tabnet propagada em todos os conjuntos de dados disponíveis.
12. **Reputação:** Assim como no caso da ANS, a reputação da entidade, conjuntamente à aparente precisão dos dados, como abordado no item 2, configuram evidências de dados livres de erro. Também é possível garantir esse item, porém é plausível tê-lo como assegurado.
13. **Segurança:** Enquanto não há maneira explícita de burlar a segurança dos dados, como citado no item 4, a segurança da plataforma já foi repetidamente posta em voga por hackers. Em fevereiro de 2021, o "hacker sincero" invadiu a plataforma, alegando que sua segurança continuava questionável (VENTURA, 2021).
14. **Velocidade:** A velocidade sustentou-se adequada durante o projeto da Kearney e os testes realizados para esse trabalho.
15. **Atualização:**

Conjuntos de dados que ainda são atualizados apresentam-se em janelas temporais satisfatórias, assim como na ANS: dados mensais no mês seguinte, trimestrais no trimestre seguinte, etc. Por exemplo, ao extrair informações advindas do CNES,

como quantidade de estabelecimentos por tipo, no início de setembro de 2022 já haviam valores de julho de 2022, como mostra a figura 20.

Figura 20 – Seleção de colunas para a extração de estabelecimentos cadastrados no CNES por tipo.

Fonte: <https://datasus.saude.gov.br/cnes-estabelecimentos>. Acesso em: 01/09/2022.

A frequência de atualização, por esse aspecto, parece satisfatória. Todavia, existe uma grande gama de conjuntos que não são mais coletados. Isso é, principalmente, um questão de políticas públicas e foge ao escopo desse trabalho, porém prejudica análises temporais executadas por vários anos, o que é útil para avaliar evoluções de tendências na saúde. Dados do Datasus Tabnet que não são mais atualizados (não-exaustivos), por exemplo, são:

- Indicadores de saúde e pactuações/Indicadores e Dados Básicos - somente valores para 2012.
- Indicadores de saúde e pactuações/Pacto pela saúde - somente valores para 2010 e 2011.
- Indicadores de saúde e pactuações/Pactos de atenção básica - somente valores entre 2001 e 2006.
- Assistência à Saúde/Atenção básica de saúde da família - somente valores entre 1998 e 2015.
- Epidemiológicas e morbidade/Estado nutricional (SISVAN) - somente valores entre 1996 e 2017.

16. **Entendimento:** Assim como no caso da ANS, não existe um glossário de siglas para que pessoas leigas entendam com mais facilidade alguns termos técnicos, como o caso da sigla "CID" explicado no item 16, que se repete aqui.

17. **Valor agregado:** O valor dos conjuntos de dados do Datasus é significativo, por balizar grandes decisões de saúde pública e privada no país.

3.2.3 Aplicações reais

Além da análise pura das dimensões da qualidade da informação, também existem ponderações subjetivas que podem ser feitas a respeito das bases. Nesse aspecto, o projeto executado pela Kearney é bem ilustrativo, pois demonstra a diferença real que uma boa qualidade de dados públicos pode fazer, mesmo que em um contexto privado de mercado.

Enquanto análises interessantes para exemplificação de como a qualidade de bases de dados pode interferir em uma construção, o conteúdo desta seção atua como ilustrativo, dando exemplos do dia a dia de conteúdos que foram feitos para o projeto da Kearney usando os conjuntos retratados até o momento.

3.2.3.1 Nível de atenção - ANS

Ao observar-se a cadeia produtiva da saúde, existe uma quebra possível em atenção **primária**, atenção **secundária**, e atenção **terciária**. Essa quebra se refere ao nível de profundidade do paciente no ecossistema de saúde, seja ele público ou privado. De acordo com a Sami, *healthtech* de planos de saúde nacional (O... , 2021):

1. **Atenção primária:** "É a porta de entrada do sistema de saúde, composto por médicos de família e comunidade e equipes de enfermagem que prestam o primeiro atendimento, resolvem a maioria das necessidades de saúde e coordenam a utilização dos níveis mais complexos de atenção". Para a Kearney², são especialidades introdutórias, como clínica médica, geriatria e pediatria, por exemplo.
2. **Atenção secundária:** "[Visa] receber pacientes encaminhados do setor primário que necessitam de atendimentos mais especializados, com médicos voltados exclusivamente para a solução de problemas específicos e [...] que precisam de um suporte maior e mais adequado para manter a estabilidade do indivíduo". Para a Kearney², são especialidades de nicho, como cardiologia, hematologia e neurologia, por exemplo.
3. **Atenção terciária:** "Compreende os hospitais de grande porte, sendo privados ou públicos. Eles são aptos para receberem casos de alta complexidade e que necessitam de atendimento especializado, invasivo, rápido e de urgência." Para a Kearney², compreende principalmente procedimentos hospitalares que envolvem internação.

Nesse sentido, uma das análises centrais do projeto era a estimação do valor de mercado do setor de saúde no país, onde a saúde suplementar (advinda de planos de saúde)

² Opinião dos especialistas em saúde, internos e externos, consultados ao longo do projeto.

seria dividida entre atenção primária, secundária, terciária e pós-tratamento (reabilitação e cuidados pós-cirúrgicos), e também dividida por especialidade médica.

É possível estimar o valor de mercado da saúde suplementar por nível de atenção pelo Mapa Assistencial da Saúde Suplementar da ANS, classificando suas linhas entre atenção primária, secundária, terciária e outros (a classificação executada pela Kearney nesse projeto não pôde ser divulgada a pedido do cliente), como mostra a figura 21.

Figura 21 – Despesas assistenciais de 2020 por nível de atenção.

id_item	Item	Despesa (R\$)	Despesa %a/a
A	Consultas Médicas	20,575,003,444.72	-20.2%
A.01	Consultas Médicas Ambulatoriais	15,676,259,738.30	-16.7%
A.02	Consultas Médicas em Pronto Socorro	4,809,860,175.09	-25.2%
B	Outros Atendimentos Ambulatoriais	14,226,129,733.81	-3.3%
C	Exames	32,093,990,740.81	-10.8%
D	Terapias	14,446,996,359.56	-0.9%
E	Internações	75,591,769,415.80	-5.9%
H	Demais despesas médico-hospitalares	7,945,636,724.94	-1.1%
I	Procedimentos Odontológicos	2,739,178,998.46	-18.5%
I.01	Consultas odontológicas iniciais	159,994,813.43	-16.5%
I.03	Procedimentos preventivos	407,411,728.22	-15.3%
I.07	Exodontias simples de permanentes (12 anos ou mais)	38,617,908.64	-9.8%
I.10	Próteses odontológicas	106,219,781.85	-12.6%
I.11	Próteses odontológicas unitárias (coroa total e restauração metálica fundida)	131,888,025.32	-17.2%

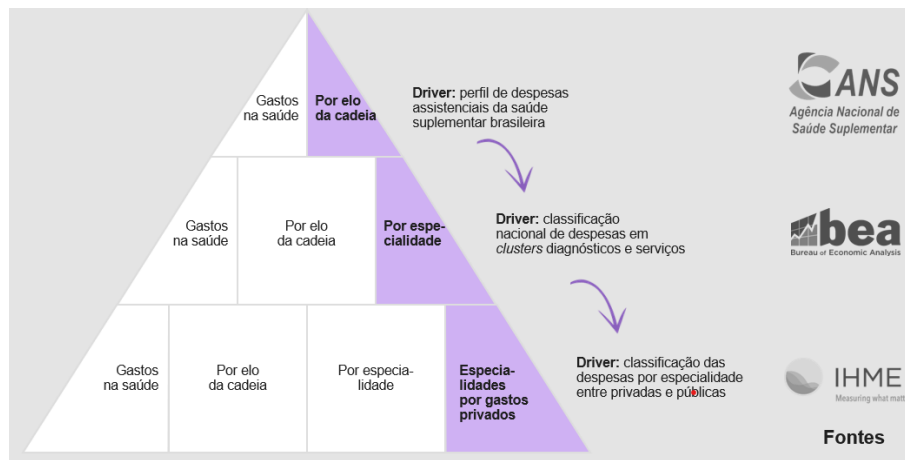
Fonte: <https://www.gov.br/ans/pt-br/aceso-a-informacao/perfil-do-setor/dados-e-indicadores-do-setor>, Mapa Assistencial da Saúde Suplementar de 2021. Acesso em: 02/09/2022.

Todavia, essa não é, idealmente, a maior granularidade possível, que é a granularidade por especialidade médica. Um dos 8 princípios fundamentais dos DGAs é que eles devem ser primários (item 2), e isso envolve a apresentação na maior granularidade possível. Isso inviabilizou o cálculo direto, no projeto, do tamanho de mercado por especialidade médica, essencial para que o hospital determinasse, por exemplo, grau de investimento em cada especialidade no hospital, para equilibrar as ofertas de serviço à demanda do mercado. A ANS disponibiliza o volume de consultas por especialidade médica, mas não o valor gasto, o que não é diretamente conectável com as despesas da figura 21.

Na ocasião, foi utilizada uma metodologia própria e restrita para essa estimação, que envolveu se basear em outros países cuja quebra por especialidade é disponibilizada, como os EUA. Por ser uma metodologia de estimação proprietária, ela não pôde ser exposta nesse trabalho, mas no caso dos EUA, utilizou conjuntos de dados como os do Bureau of Economic Analysis - BEA e do Institute for Health Metrics and Evaluation - IHME, que realizam esse trabalho de estimativa por especialidade.

A metodologia em questão é explicada superficialmente na figura 22.

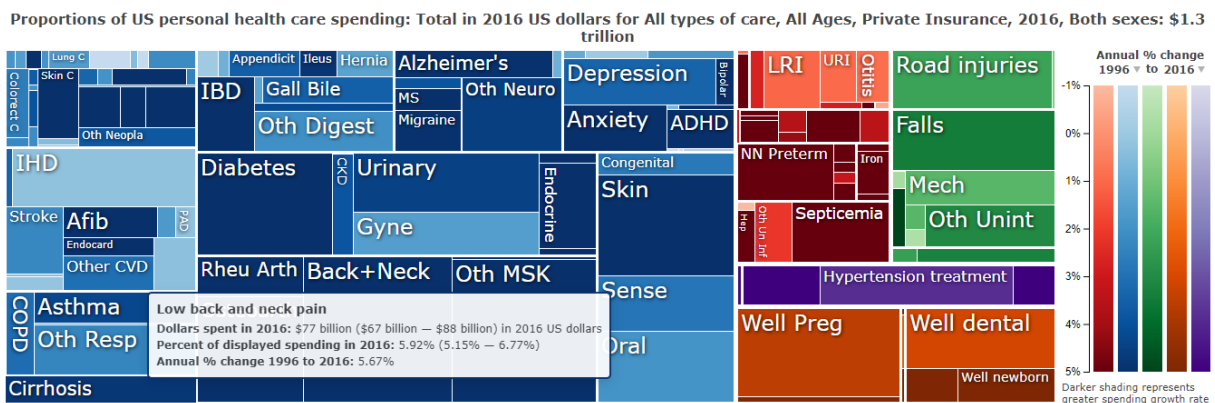
Figura 22 – Explicação superficial da metodologia empregada no projeto da Kearney para estimativa de tamanho de mercado por especialidade médica.



Fonte: Kearney.

É possível perceber na base gráfica do IHME, por exemplo, que a quebra em alto nível de especialidades médicas é possível, como mostra a figura 23. Nela, é filtrado o gasto nos EUA por condição de saúde (que pode ser atrelada a uma especialidade médica), somente para planos de saúde privados, que era o objetivo de estudo dessa análise no projeto da Kearney.

Figura 23 – Visualização do conjunto de dados do IHME do gasto em saúde por planos privados por condição de saúde em 2016.



Fonte: <https://vizhub.healthdata.org/dex/>. Acesso em: 02/09/2022.

Nesse aspecto, a base da ANS não oferece a maior granularidade, e prejudica a conexão entre as bases ofertadas: possui despesas assistenciais por nível de atenção, porém uma visualização por especialidade somente em volume de consultas, e não em gastos declarados. Para que bases como essas possam ser utilizadas de modo ótimo e máximo, deve haver consistência e possibilidade de conexão entre bases correlatas, como nesse caso.

3.2.3.2 Receita por operadora - ANS

No ANS Tabnet, há a disponibilização histórica de receitas e despesas das operadoras cadastradas, e uma análise do projeto da Kearney envolveu entender a fatia de mercado das principais operadoras do país. Todavia, duas observações foram verificadas nesse conjunto de dados que tiveram que ser contornadas, o que prejudica, por exemplo, na dimensão da qualidade da informação do conjunto ser livre de erros e de atualização (item 2.1):

- Existem valores claramente inconsistentes³ em algumas operadoras da base de dados. Por exemplo, a receita da Unimed de Rio Branco, da Associação Evangélica Beneficente de Londrina e da IBBCA 2008 Gestão em Saúde, apresentam informações zeradas sem explicação aparente:

Tabela 1 – Receita de contraprestações de operadoras de saúde selecionadas entre 2016 e 2020, em R\$ milhões.

Operadora	2016	2017	2018	2019	2020
Unimed Rio Branco	107,0	121,6	135,2	0	163,1
Associação Evangélica Beneficente de Londrina	85,4	100,0	0	122,4	133,3
IBBCA 2008 Gestão em Saúde	66,9	0	53,5	54,8	45,7

- A base de dados não possui um método nem observações para contabilizar fusões e aquisições. Por exemplo, em 2021 foi anunciada a maior fusão até o momento da saúde privada do país, entre as operadoras Hapvida e Notre Dame Intermédica (REUTERS, 2021), mas em todas as bases que envolvem operadoras, até mesmo em 2022, quando a fusão já havia transcorrido todos os trâmites formais (NETTO, 2022), as operadoras ainda eram exibidas de modo separado nas bases e precisavam ser unidas posteriormente ao longo do projeto, para análise precisa.

Nessa frente, um trabalho de normalização e/ou maior rigor na coleta dos dados contribuiria em conjuntos de dados mais confiáveis.

3.2.3.3 Separação de períodos - Tabnet

Tanto no ANS Tabnet, quanto no Datasus Tabnet, ao se selecionar mais de um período, os períodos são somados para as colunas desejadas, sem possibilidade de visualização de colunas e subcolunas, como ilustra a visualização de receitas e despesas no ANS Tabnet na figura 24.

³ A semântica de "consistência" aqui não deve ser confundida com a ilustrada no item 6 das dimensões da qualidade da informação. Aqui, se refere a dados inconsistentes no próprio conjunto, enquanto nas dimensões, se refere aos conjuntos serem apresentados no mesmo formato.

Figura 24 – Receitas e despesas extraídas do ANS Tabnet, com todas as colunas selecionadas, entre 2020 e 2022.

Receitas e Despesas (Dados do 1º trimestre de 2022 – Data de corte 24/06/2022)

Receita de contraprestações, Outras receitas operacionais, Despesa assistencial, Despesa administrativa, Despesa de comercialização, Outras despesas operacionais segundo Operadora
Período: 2020-2022

Operadora	Receita de contraprestações	Outras receitas operacionais	Despesa assistencial	Despesa administrativa	Despesa de comercialização	Outras despesas operacionais
TOTAL	528.202.681.942	27.345.999.540	424.794.589.450	52.239.013.944	17.234.391.827	40.864.879.807
005711-BRADESCO SAUDE S.A.	57.674.654.380	376.962.317	52.886.844.045	2.347.122.986	2.738.707.654	1.309.513.122
326305-AMIL ASSISTENCIA MEDICA INTERNACIONAL S.A.	42.589.206.261	253.862.893	36.168.965.448	4.248.588.209	1.854.007.958	706.492.183
006246-SUL AMERICA COMPANHIA DE SEGURO SAUDE	42.267.860.509	88.188.832	34.964.180.105	2.363.460.964	2.771.935.674	1.015.674.607
359017-NOTRE DAME INTERMEDICA SAUDE S.A.	20.994.887.844	1.034.824.732	16.552.341.550	1.856.186.779	1.320.842.965	802.844.015
368253-HAPVIDA ASSISTENCIA MEDICA S.A.	15.854.209.703	351.156.542	11.247.264.557	2.041.743.748	826.584.386	868.318.805
346659-CAIXA DE ASSISTENCIA DOS FUNCIONARIOS DO BA	13.995.367.841	576.753.291	11.977.142.052	819.848.063	3.404	541.010.344
339679-CENTRAL NACIONAL UNIMED - COOPERATIVA CENTR	12.577.119.195	298.448.976	10.246.230.839	1.067.850.358	590.940.520	656.519.081
393321-UNIMED-RIO COOPERATIVA DE TRABALHO MEDICO D	11.079.763.707	369.479.491	8.857.164.129	917.973.094	828.541.678	761.755.636
343889-UNIMED BELO HORIZONTE COOPERATIVA DE TRABAL	10.675.361.348	475.083.002	7.768.849.467	1.351.461.570	136.321.957	724.521.817

Fonte: https://www.ans.gov.br/anstabnet/cgi-bin/dh?dados/tabnet_rc.def. Acesso em: 02/09/2022.

Ao executar avaliações de séries históricas, é natural buscar a variação ano a ano, para entender-se dinâmicas de evolução de mercado. Ao mostrar a soma dos períodos selecionados, e não os valores ano a ano, há uma grande ineficiência da extração dos dados. Em última instância, é possível realizar a extração ano a ano, porém os dados não são apresentados conforme os usuários a esperam, prejudicando o critério de adequação (item 3).

Isso parece ser um problema da visualização web do Tabnet, visto que também ocorre na visualização do Datasus Tabnet. Uma grande melhoria ao sistema Tabnet, portanto, seria a possibilidade dessa visualização em colunas e subcolunas, especialmente para séries temporais.

4 Conclusão e considerações finais

Como plausível concluir, quanto maior a quantidade de conjuntos de dados a gerir sob uma única instituição, maior a dificuldade de manter a qualidade dos mesmos ao serem disponibilizados. Como discutido ao longo do trabalho, políticas de transparência de dados são fundamentais em instituições de governança sólidas, especialmente públicas. Quanto mais a transparência alia-se à qualidade, mais eficaz é a abertura dos dados.

Em relação estritamente aos conjuntos analisados, a plataforma de disponibilização Tabnet é muito útil, pois ajuda a uniformizar as informações em um formato que, quando aprendido, pode ser utilizado rapidamente para novos conjuntos. Ele, em si, cabe melhorias, como a possibilidade de seleção de subcolunas abordada na seção 3.2.3.3. A arquitetura do Tabnet em si foge ao escopo do trabalho, contudo, é razoável propor que, caso alguma melhoria seja implementada na plataforma, haja um mecanismo de atualização da arquitetura de dados de todas as instituições que o utilizam, para que sejam replicadas uniformemente em várias instituições governamentais.

A relevância dos conjuntos abordados é inquestionável, vide o próprio projeto da Kearney cujo conhecimento adquirido foi alavancado neste trabalho. Mesmo que seja um exemplo da esfera privada, a relevância do tema para a esfera pública é inquestionável, e pode basear várias decisões de políticas governamentais.

A Lei de Acesso à Informação, promulgada em 2012 no Brasil, permite que qualquer cidadão solicite dados ao governo ([DADOS, s.d.](#)), e a conscientização dessa possibilidade é essencial para que se construa um país mais transparente, com informações igualmente transparentes, nunca abrindo mão da qualidade dos dados. Como investigou Craveiro, Soletto e Machado, existe demanda considerável, especialmente em alguns setores governamentais específicos, como administração, saúde e educação ([CRAVEIRO G., 2020](#)). Frente essa demanda, espera-se que esse trabalho ajude interessados a entender, com casos de uso concretos, como vasculhar dados públicos, e que ajude instituições públicas a abrir, com prezo máximo pela qualidade, dados que são de direito dos cidadãos brasileiros.

Referências

- ABERTOS, P. B. de D. *Perguntas frequentes*. s.d. <<https://dados.gov.br/pagina/faq>>. Último acesso em: 30/08/2022. Citado na página 23.
- ALVES, A. *Rede D'Or compra SulAmérica e acelera consolidação no setor de saúde*. 2022. <<https://www.cnnbrasil.com.br/business/rede-dor-compra-sulamerica-2/>>. Último acesso em: 01/09/2022. Citado na página 34.
- ANDRADE, B. *Dasa conclui a compra de 3 hospitais por R\$ 1,7 bilhão*. 2021. <<https://www.moneytimes.com.br/dasa-conclui-a-compra-de-3-hospitais-por-r-17-bilhao/>>. Último acesso em: 01/09/2022. Citado na página 34.
- AQUINO, M. *Após reclamações, ANS suspende 70 planos de saúde de oito operadoras*. 2022. <<https://www.metropoles.com/brasil/saude-br/apos-reclamacoes-ans-suspende-70-planos-de-saude-de-oito-operadoras>>. Último acesso em: 31/08/2022. Citado na página 32.
- CRAVEIRO G., S. S. J. A. M. J. da S. Um balanço da demanda de dados abertos no Brasil. In: . [S.l.: s.n.], 2020. p. 273–296. Citado na página 49.
- DADOS, E. de. *Solicitando dados via Lei de Acesso à Informação*. s.d. <<https://escoladedados.org/tutoriais/solicitando-dados-via-lei-de-acesso-a-informacao/>>. Último acesso em: 03/09/2022. Citado na página 49.
- DBC Format. 2017. <http://socialledge.com/sjsu/index.php/DBC_Format>. Último acesso em: 30/08/2022. Citado na página 28.
- GLOBAL, S. *What is the “G” in ESG?* 2020. <<https://www.spglobal.com/en/research-insights/articles/what-is-the-g-in-esg>>. Último acesso em: 29/08/2022. Citado na página 17.
- JANONE, L. *Estudo aponta que idosos vão representar 40% da população brasileira em 2100*. 2021. <<https://www.cnnbrasil.com.br/nacional/estudo-aponta-que-idosos-va-representar-40-da-populacao-brasileira-em-2100/>>. Último acesso em: 01/09/2022. Citado na página 38.
- MACHADO J., M. M. d. C. L. I. P. Qualidade das bases de dados hospitalares no Brasil: alguns elementos. In: . [S.l.: s.n.], 2016. p. 567–581. Citado na página 19.
- MICHENER G., C. E. e. N. I. Da Opacidade à transparência? Avaliando os Cinco Anos da Lei de Acesso no Brasil. 2017. Citado na página 17.
- NEIVA, L. *Ministério da Saúde foi alertado duas vezes sobre risco de ataque hacker*. 2021. <<https://congressoemfoco.uol.com.br/projeto-bula/noticia/ministerio-da-saude-foi-alertado-duas-vezes-sobre-risco-de-ataque-hacker/>>. Último acesso em: 01/09/2022. Citado na página 40.
- NETTO, V. *Após fechar fusão com Hapvida, NotreDame deixará de ser negociada na B3*. 2022. <<https://www.cnnbrasil.com.br/business/>>

apos-fechar-fusao-com-hapvida-notredame-deixara-de-ser-negociada-na-b3-em-11-2/>. Último acesso em: 02/09/2022. Citado na página 46.

NOTÍCIAS, A. I. *PIB cai 4,1% em 2020 e fecha o ano em R\$ 7,4 trilhões*. 2021. <<https://agenciadenoticias.ibge.gov.br/agencia-sala-de-imprensa/2013-agencia-de-noticias/releases/30165-pib-cai-4-1-em-2020-e-fecha-o-ano-em-r-7-4-trilhoes>>. Último acesso em: 29/08/2022. Citado na página 21.

O que são os níveis de atenção primária, secundária e terciária? 2021. <<https://blog.samisaude.com.br/niveis-de-atencao-a-saude/>>. Último acesso em: 02/09/2022. Citado na página 43.

OCDE. *Open Government Data*. s.d. <<https://www.oecd.org/digital/digital-government/open-government-data.htm>>. Último acesso em: 29/08/2022. Citado na página 17.

PARKER M.B., M. V. D. I. H. R. e. W. G. An evaluation of Information quality frameworks for the World Wide Web. In: *8th Annual Conference on WWW Applications*. Bloemfontein: [s.n.], 2006. Citado 3 vezes nas páginas 7, 9 e 22.

PELISSERO, J. *Trends in the Implementation of ESG Policies in State and Local Governments*. 2022. <<https://www.scu.edu/ethics/all-about-ethics/trends-in-the-implementation-of-esg-policies-in-state-and-local-governments/>>. Último acesso em: 29/08/2022. Citado na página 17.

REIS, T. *Verticalização: entenda como funciona esse conceito produtivo*. 2019. <<https://www.suno.com.br/artigos/verticalizacao/>>. Último acesso em: 01/09/2022. Citado na página 34.

REUTERS. *Hapvida e Notre Dame Intermédica anunciam fusão*. 2021. <<https://g1.globo.com/economia/noticia/2021/03/01/hapvida-e-notre-dame-intermedica-anunciam-fusao.ghtml>>. Último acesso em: 02/09/2022. Citado na página 46.

SAÚDE, C.-G. de Disseminação de Informações em. *Tutorial TABNET*. s.d. <http://tabnet.datasus.gov.br/Tutorial/Tutorial_tabNet_FINAL.pdf>. Último acesso em: 30/08/2022. Citado na página 23.

SILVA AMBROZINA, A. A. M. D. e. d. O. R. A. de A. P. Qualidade da Informação dos Dados Governamentais Abertos: Análise do Portal de Dados Abertos Brasileiro. In: . Fortaleza: [s.n.], 2020. p. 31–47. ISSN 1984-7297. Citado na página 22.

TAUBERER, J. *8 Principles of Open Government Data*. 2007. <https://public.resource.org/8_principles.html>. Último acesso em: 28/08/2022. Citado 3 vezes nas páginas 7, 9 e 17.

VENTURA, F. *DataSUS é invadido de novo e hacker reclama: “continua uma b*****”*. 2021. <<https://tecnoblog.net/noticias/2021/02/18/datasus-e-invadido-de-novo-e-hacker-reclama-continua-uma-b/>>. Último acesso em: 01/09/2022. Citado na página 41.

ŽUFFOVÁ, M. Do FOI laws and open government data deliver as anti-corruption policies? Evidence from a cross-country study. In: . [S.l.: s.n.], 2020. p. 101480. Citado na página 17.

APÊNDICE A – Apêndice

A.1 Bases Datasus e ANS

A.1.1 Base ANS Tabnet

A tabela 2 ilustra as quebras e subquebras do ANS Tabnet no momento da escrita desse trabalho.

Tabela 2 – Quebras e subquebras da base da ANS.

Quebra	Subquebra
Beneficiários	Geografia
Beneficiários	Operadora
Beneficiários	Taxa de cobertura
Operadoras	Status do registro
Operadoras	Receitas e despesas
Planos de saúde	Tipo de contratação
Ressarcimento ao SUS	Tipo de atendimento
Demandas dos consumidores e fiscalização	Reclamações dos beneficiários
Demandas dos consumidores e fiscalização	Solicitação de informações
Demandas dos consumidores e fiscalização	Demandas NIP
População residente	População residente

A.1.2 Base Datasus Tabnet

A tabela 3 ilustra as quebras e subquebras do Datasus Tabnet no momento da escrita desse trabalho.

Tabela 3 – Quebras e subquebras da base do Datasus.

Quebra	Subquebra
Indicadores de saúde e pactuações	Indicadores e dados básicos
Indicadores de saúde e pactuações	Rol de diretrizes, objetivos, metas e indicadores
Indicadores de saúde e pactuações	Pacto pela saúde
Indicadores de saúde e pactuações	Pactos de atenção básica
Indicadores de saúde e pactuações	Indicadores municipais
Assistência à saúde	Produção hospitalar
Assistência à saúde	Produção ambulatorial
Assistência à saúde	Imunizações
Assistência à saúde	Atenção básica - saúde da família
Assistência à saúde	Vigilância alimentar e nutricional
Assistência à saúde	Conjunto mínimo de dados
Epidemiológicas e morbidade	Morbidade hospitalar no SUS
Epidemiológicas e morbidade	Casos de AIDS
Epidemiológicas e morbidade	Casos de hanseníase
Epidemiológicas e morbidade	Casos de tuberculose
Epidemiológicas e morbidade	Doenças e agravo de notificação
Epidemiológicas e morbidade	Casos suspeitos de SCZ
Epidemiológicas e morbidade	Programa de controle de esquistossomose
Epidemiológicas e morbidade	Estado nutricional
Epidemiológicas e morbidade	Hipertensão e diabetes
Epidemiológicas e morbidade	Câncer de colo de útero e de mama
Epidemiológicas e morbidade	Sistema de informação do câncer
Epidemiológicas e morbidade	Tempo até o início do tratamento oncológico
Rede assistencial	Estabelecimentos
Rede assistencial	Recursos físicos
Rede assistencial	Recursos humanos
Rede assistencial	Equipes de saúde
Estatísticas vitais	Nascidos vivos
Estatísticas vitais	Mortalidade
Estatísticas vitais	Painéis diversos de monitoramento
Estatísticas vitais	Correção e redistribuição de óbitos
Demográficas e socioeconômicas	População residente
Demográficas e socioeconômicas	Educação
Demográficas e socioeconômicas	Trabalho e renda
Demográficas e socioeconômicas	PIB
Demográficas e socioeconômicas	Saneamento
Inquéritos e pesquisas	PNAD – Pesquisa Nacional por Amostra de Domicílios
Inquéritos e pesquisas	VIGITEL – Vigilância de fatores de risco e proteção para doenças crônicas
Inquéritos e pesquisas	VIVA – Vigilância de violências e acidentes
Informações financeiras	Recursos Federais do SUS (por Município)
Informações financeiras	Valores aprovador da produção SUS (por Prestador)
Informações financeiras	Guia de autorização de pagamento
Estatísticas de acesso ao Tabnet	Estatísticas de acesso ao Tabnet (atualizado somente até 2018)

A.2 Definição de CAGR

Em tradução livre, de acordo com a Gartner, "CAGR é a taxa de crescimento de receita anualizada entre dois anos". Ou seja, uma porcentagem fixa de crescimento a partir do ano inicial analisado, a partir do qual o valor do ano final seria o valor esperado. Matematicamente, é dado por:

$$\left(\frac{V_{final}}{V_{inicial}}\right)^{\frac{1}{t}} - 1 \quad (\text{A.1})$$

Onde V são os valores, e t é o tempo em anos.