

Uma Arquitetura para Extração, Povoamento e Publicação de Dados Abertos: Um Estudo de Caso com o Anuário Socioeconômico do Estado de Sergipe

Daniel Santos², Wedson Teixeira de Almeida², Methanias Colaço Júnior^{1,2},
André Vinicius R. P. Nascimento²

¹Programa de Pós Graduação em Ciências da Computação (PROCC). Universidade Federal de Sergipe (UFS). São Cristóvão, Sergipe, Brasil.

²Núcleo de Pesquisa e Prática em Inteligência Competitiva – NUPIC, DSI. Universidade Federal de Sergipe (UFS). Itabaiana, Sergipe, Brasil.

{daniell17sants, andreviniciusnascimento}@gmail.com,
wedson.teixeira1998@outlook.com, mjrse@hotmail.com

***Abstract.** Socio-economic atlas are compilations of open data organized into historical series with the aim of presenting indicators on topics of social interest. Collecting data from different sources is the most time-consuming step in creating an atlas. The time spent extracting information from disparate sources is the main cause for some publications to be delayed, not available in a given year, or to present outdated data. This work presents an architecture for an automatic process of extracting open data in the context of the Sergipe state atlas to help the design of data visual representations for socio-economic indicators.*

***Resumo.** Anuários Socioeconômicos são compilações de dados abertos governamentais organizados em séries históricas com o objetivo de apresentar indicadores sobre variados temas de interesse social. A etapa mais onerosa na criação de um anuário é a coleta de dados de diferentes fontes. O tempo despendido para coletar as informações de vários portais é a principal causa para que algumas publicações apresentem atrasos, não sejam disponibilizadas em determinado ano, ou apresentem dados desatualizados. Este trabalho apresenta uma arquitetura para um processo automático de coleta de dados abertos no contexto do anuário do estado de Sergipe para auxiliar a elaboração de visualizações para indicadores socioeconômicos.*

1. Introdução

Nas últimas duas décadas houve um avanço nas discussões sobre Dados Abertos e Transparência nas esferas governamentais [Macedo 2021], como forma de promover a fiscalização de ações do governo, aumentar a confiança da população nos atos públicos

e fornecer informações relevantes que possam ser utilizadas por economistas, especialistas em políticas públicas e o próprio governo para direcionar ações que venham a beneficiar toda a população.

Um problema com os dados disponibilizados pelos vários portais do governo é que eles, em sua maioria, estão em um formato bruto e dependem de um tratamento especializado para que a informação contida neles seja acessível para a população. Até mesmo especialistas precisam efetuar algum tipo de tratamento nos dados para efetuar suas análises.

É nesse contexto que surgem iniciativas para dar maior transparência e significado para os dados abertos. Uma dessas iniciativas é o projeto “Transparência Traduzida — Monitoramento Social da Economia e dos Atos Públicos em Sergipe” [“Transparência Traduzida” 2021], que tem como principal objetivo fornecer informações a respeito de atos públicos, características da população brasileira e da economia, sempre apoiadas por análises de especialistas, dotando o cidadão comum e agentes responsáveis por políticas públicas com conhecimento para poder controlar, fiscalizar, sugerir iniciativas e tomar decisões que venham a beneficiar a população em geral.

Um dos subprojetos do “Transparência Traduzida” é o “Anuário Socioeconômico de Sergipe para Web”. O Anuário Socioeconômico de Sergipe é uma publicação do Grupo de Pesquisa em Análise de Dados Econômicos, vinculado ao Departamento de Economia da Universidade Federal de Sergipe. O Anuário tem sido publicado a cada dois anos, desde 2017 no formato estático, disponível em um arquivo .pdf [“Anuário Socioeconômico de Sergipe” 2022]. O objetivo do subprojeto dentro do “Transparência Traduzida” é promover o Anuário a uma aplicação dinâmica que possa ter seus dados extraídos, transformados e carregados de maneira automática, tornando mais ágil a compilação de informações, a geração de gráficos e tabelas, e, conseqüentemente, a disponibilidade de análises mais rápidas para a população.

Dentre os desafios para a transformação do Anuário em uma aplicação, estão a extração e transformação dos dados provenientes dos diversos portais de dados abertos do governo. As informações brutas tratadas pelo Anuário encontram-se em diferentes fontes, em diferentes formatos, e não estão totalmente estruturadas. Logo, faz-se necessária a criação de mecanismos para facilitar a extração automática desses dados, seu armazenamento e sua transformação para que eles possam ser facilmente utilizados para a elaboração de visualizações (gráficos ou tabelas).

Este trabalho apresenta uma arquitetura para a extração, povoamento e transformação de dados abertos necessários à elaboração de visualizações para anuários socioeconômicos. Uma instanciação da arquitetura foi realizada através da implementação de seus componentes e foi iniciado um estudo de caso com o Anuário Socioeconômico do Estado de Sergipe com o objetivo inicial de avaliar a solução proposta.

O artigo está estruturado da seguinte maneira: A seção 2 apresenta a revisão da literatura, apresentando os principais conceitos utilizados no trabalho e os trabalhos relacionados. A seção 3 apresenta a metodologia utilizada. A seção 4 esclarece Ameaças à Validade da solução e, finalmente, na seção 5 são feitas as conclusões e considerações finais.

2. Revisão da Literatura

2.1. Dados Abertos e Transparência

Dados abertos são dados que podem ser livremente utilizados, reutilizados e redistribuídos por qualquer pessoa, sujeitos à exigência de atribuição à fonte original e ao compartilhamento pelas mesmas licenças em que as informações foram apresentadas [“The Open Definition – Defining Open in Open Data, Open Content and Open Knowledge” 2015]. Esses dados são publicados e distribuídos na Internet, compartilhados em formato aberto para que possam ser lidos por pessoas e por máquinas, permitindo o cruzamento com outros dados de diferentes fontes [Dados abertos 2011].

A abertura de dados públicos no formato bruto promove um novo nível de transparência, aumento da fiscalização dos cidadãos sobre seus governos, bem como uma cooperação mais estreita com os mesmos [Reis Júnior 2020]. No entanto, alguns setores do Governo disponibilizam dados na web em formatos não padronizados, não estruturados e sem metadados associados, criando barreiras para o consumo e a análise desses dados por terceiros, reduzindo, dessa forma, a transparência efetiva do governo.

2.1.1 Modelos de abertura de dados no Brasil

O Brasil possui diferentes modelos de abertura de dados sendo utilizados pelo setor público. Alguns portais, a exemplo do “Portal Brasileiro de Dados Abertos”, do “Dados Abertos da Prefeitura de Belo Horizonte” e do “Dados Abertos do DF” utilizam a plataforma CKAN [“CKAN – The open source data management system” 2018], embora aplicada de maneiras diferentes. Dentre os modelos de abertura de dados no Brasil, dois merecem destaque pelo volume, abrangência, importância dos dados publicados, e relação com as informações utilizadas por anuários socioeconômicos: IBGE e o IPEA.

2.1.2 Instituto Brasileiro de Geografia e Estatística (IBGE)

O IBGE (Instituto Brasileiro de Geografia e Estatística) constitui o maior provedor de dados e informações do Brasil, atendendo necessidades dos mais variados segmentos da sociedade civil bem como dos órgãos governamentais. O IBGE possui diversos portais e subdomínios que disponibilizam seus dados. Reformulados com frequência, os portais do IBGE podem apresentar informações redundantes. Dentre os principais portais, podemos citar: a) SIDRA (Sistema IBGE de Recuperação Automática) (<https://sidra.ibge.gov.br/home/pnadct/brasil>); b) Séries Históricas e Estatísticas (<https://seriesestatisticas.ibge.gov.br/>); c) Downloads (<https://downloads.ibge.gov.br/>).

2.1.3 Instituto de Pesquisa Econômica Aplicada (IPEA)

O IPEA (Instituto de Pesquisa Econômica Aplicada) é uma fundação pública federal vinculada ao Ministério do Planejamento. O portal de dados do IPEA, o Ipeadata (<http://www.ipeadata.gov.br>), disponibiliza informações geradas pelo próprio IPEA e outras entidades. As informações são divididas em três seções: macroeconômico, regional e social. O Ipeadata também possui algumas APIs (*Application Program Interface*) para facilitar o acesso aos dados publicados.

2.2 Data Warehouse e Data Mart

Um *Data Warehouse* (DW) é um repositório de dados histórico, integrado, não volátil, baseado em assuntos e variante no tempo [Inmon 2005]. A principal característica de um *Data Warehouse* é ser um Banco de Dados separado dos ambientes operacionais, voltado exclusivamente para consultas.

A ideia básica por trás de um ambiente de *Data Warehouse* é extrair dados provenientes de bases operacionais ou fontes externas, limpá-los, integrá-los e consolidá-los em um repositório especialmente projetado para consultas.

A arquitetura mais utilizada para a criação de *Data Warehouses* é definida por Kimball et al. (1998). Nela, um *Data Warehouse* é formado por um conjunto de esquemas (divisão do *Data Warehouse*) conhecidos como *Data Marts*. Cada *Data Mart* representa um esquema lógico que armazena informações associadas a um tema específico.

2.2.1 Modelagem Dimensional

A Modelagem dimensional é o nome de uma técnica de projeto lógico frequentemente usada para construção de um *Data Warehouse*. Para Kimball et al. (1998), a abordagem de modelagem dimensional é a melhor maneira de modelar dados de apoio à decisão, pois ela oferece facilidade no uso e consegue balancear desempenho e volume de informações.

Cada modelo dimensional é composto por uma tabela central, chamada de tabela de fatos, e um conjunto de tabelas, chamadas de dimensões, que possuem um relacionamento de 1:n com a tabela de fatos. A ideia central do modelo é armazenar fatos (medidas ou indicadores), caracterizados por vários contextos (dimensões). O esquema resultante da modelagem dimensional é o esquema estrela [Kimball et al. 1998]. Podemos observar um exemplo de esquema estrela na Figura 1.

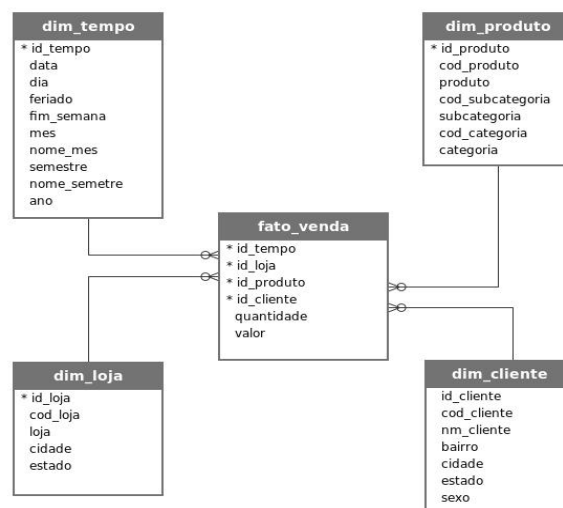


Figura 1. Exemplo de esquema estrela.

2.3 ETL (*Extract, Transform and Load*)

O processo ETL (*Extract, Transform and Load*) é a base para a construção de um *Data Warehouse*. Esse processo tem como função extrair dados de fontes diversas, transformar, integrar e armazenar esses dados em um formato apropriado para apresentação [Kimball 2004]. O processo ETL compreende três principais etapas ou atividades: Extração, Transformação e Armazenamento.

2.3.1 Extração

A Extração, primeiro passo do processo ETL, representa a obtenção de dados brutos nos sistemas de origem. Esses sistemas podem ser sistemas operacionais, sistemas legados ou fontes externas, e os dados podem estar armazenados em diferentes formatos. Devido a essa diversidade de fontes e formatos, essa é a etapa mais crítica de todo o processo.

2.3.2 Transformação

A etapa de Transformação realiza uma série de atividades sobre os dados a fim de convertê-los e adequá-los a um formato para a carga no repositório de destino. Nessa etapa são realizados um ou mais processos a depender da necessidade: Limpeza, Remoção, Combinação, Normalização, Cálculos e Derivação.

2.3.3 Carga

A última etapa do processo ETL é a Carga. Ela é responsável pela inclusão e atualização de dados no repositório de destino. A depender da arquitetura e dos esquemas finais, o processo de carga precisa lidar com tratamento de histórico, tratamento de erros, agregações e povoamento de tabelas derivadas.

2.4 Trabalhos Relacionados

Para a pesquisa foi utilizado o acesso café capes junto ao Google Scholar para encontrar documentos científicos nas principais bases de pesquisas científicas. Não foram encontrados trabalhos fortemente relacionados à proposta do trabalho. No entanto, foram encontrados trabalhos moderadamente relacionados, vinculados aos seguintes temas: Dados Abertos, Visualização de Informações e Processos ETL.

Em Ribeiro et al. (2021) é desenvolvida uma aplicação no contexto do projeto VisPublica com o objetivo de analisar diferentes técnicas de visualização de dados que melhoram o entendimento dos dados públicos, promovendo mais transparência para os dados disponibilizados. A aplicação é descrita considerando seus principais componentes funcionais e apresenta um exemplo da utilização do VisPublica na visualização de um conjunto de dados do projeto DadosGov (Portal Brasileiro de Dados Abertos).

O trabalho de Quingerski, Marques e Freire (2016) apresenta uma proposta de processo ETL (*Extract, Transform, Load*) utilizando dados abertos de fonte governamental para descoberta e representação de conhecimento. Esses dados foram padronizados e processados utilizando um conjunto de tarefas e ferramentas, resultando em um processo capaz de integrar dados e informações, gerando novas informações e

conhecimentos, além de permitir diferentes representações, não possíveis através dos dados e informações da fonte original.

Lima e Souza (2019) desenvolveram um *Data Mart* para automatizar o processo ETL no contexto da produção acadêmica do Centro Universitário Luterano de Palmas (CEULP/ULBRA). O procedimento realizado tem como intuito centralizar os dados referentes à produção acadêmica que estão armazenados em diversas fontes de dados da instituição.

Em outro trabalho com menor relacionamento, Oliva (2017) propôs desenvolver ferramentas de apoio à decisão baseadas nos dados de indicadores da saúde, definidos pelo Mapa de Saúde e disponibilizados pelo Observatório Regional de Assistência Hospitalar (ORAH).

Embora fuja aos vieses de trabalhos científicos é possível considerar outros anuários socioeconômicos como trabalhos relacionados com o estudo de caso da proposta. Ao realizar buscas por anuários socioeconômicos percebe-se que há publicações por todo o território nacional, divididos por cada estado, sendo que muitos publicados no formato digital PDF, desatualizados e sem possibilidade de extrair ou compartilhar os dados exibidos.

O CEPERJ (Centro Estadual de Estatísticas, Pesquisas e Formação de Servidores Públicos do Rio de Janeiro) publicou a última versão do Anuário Estatístico do Rio de Janeiro em 2013. Na página que o cidadão possui acesso é possível escolher um ano, dos disponíveis para o anuário, e navegar entre opções que dão acesso às análises socioeconômicas, categorias e downloads de arquivos que, em sua maioria, apresentam formato proprietário (xls) (“Anuários Estatísticos do estado do Rio de Janeiro”, 2013).

A publicação do Anuário estatístico de Mato Grosso mais recente é do ano de 2016. O visitante da página web pode realizar a visualização em PDF do Anuário, efetuar o download do próprio PDF e/ou realizar o download dos dados acerca das categorias socioeconômicas. O site dessas informações é da Secretaria de Estado de Planejamento e Coordenação Geral de Mato Grosso (“Anuário estatístico de Mato Grosso”, 2016).

Outro trabalho relacionado é da Agência Estadual de Planejamento e Pesquisas de Pernambuco — CONDEPE/FIDEM. A CONDEPE/FIDEM coloca à disposição da sociedade o Anuário Estatístico de Pernambuco, cuja série de publicação iniciou no ano de 2011. Sua última publicação foi em 2016. O site mantém uma estrutura de diretórios para navegação e permite o download dos artefatos do anuário em formato proprietário (xls) [CONDEPE FIDEM 2016].

O Anuário Estatístico do Estado de Alagoas é mantido pela Secretaria de Estado do Planejamento, Gestão e Patrimônio — SEPLAG, criada em 2015. A SEPLAG publicou o anuário pela última vez em 2017. Só é possível realizar o download de um único arquivo com as informações do anuário [SEPLAG 2015].

3. Metodologia

O método de pesquisa escolhido para o trabalho foi o *Design Science Research* (DSR), uma vez que a proposta do trabalho é a criação de um artefato (arquitetura), que precisa ser instanciado e avaliado, para a solução de um problema real (extração de dados abertos para anuários socioeconômicos).

Uma vez escolhido o DSR como método de pesquisa, o presente trabalho adotou o ciclo apresentado em Hevner (2010), que aponta as seguintes etapas: a) Identificação do Problema; b) Sugestão de Solução; c) Desenvolvimento do Artefato; d) Avaliação; e) Conclusão.

O trabalho atualmente encontra-se no início da etapa de Avaliação. Nessa etapa, a estratégia de pesquisa escolhida foi o Estudo de Caso, por se tratar de uma avaliação da arquitetura em um contexto real do anuário socioeconômico do estado de Sergipe.

A seguir serão apresentados os elementos das três primeiras etapas do ciclo DSR.

3.1 Identificação do Problema

A primeira fase do ciclo DSR iniciou com reuniões com o grupo de pesquisa do Departamento de Economia para que fosse conhecido e detalhado o processo atual (manual) de elaboração do anuário. Após revisão da literatura e análise de anuários de outros estados ou cidades disponíveis na Internet, foi possível identificar que os anuários, independente do formato, compartilhavam o mesmo problema: a atualização das informações. Conhecendo o processo manual, foi possível perceber que esse problema é causado principalmente pela etapa onerosa de coleta de dados das fontes de dados abertos do governo. Some-se a isso as etapas de integração e transformação, uma vez que alguns indicadores não podem ser obtidos diretamente dos dados originais. Como exemplo, podemos citar o Produto Interno Bruto, que precisa levar em conta variações econômicas como os índices de inflação. Logo, valores de uma série histórica, para serem corretamente comparados, precisam ser deflacionados. Para efetuar esse cálculo, são coletados dados sobre os índices de preços a serem considerados, e são implementados algoritmos para efetuar a deflação.

3.2 Sugestão de Solução

Uma vez que a revisão da literatura não identificou trabalhos que apresentassem uma arquitetura ou soluções para o problema identificado, foi iniciado a segunda etapa do ciclo DSR. Nessa etapa, após serem identificados os requisitos que deveriam ser atendidos, foi elaborada uma proposta de arquitetura para a extração automática de dados abertos.

3.3 Desenvolvimento do Artefato

Após o levantamento e elicitação de requisitos, e também através de alguns protótipos exploratórios, chegou-se a uma arquitetura que pudesse ser adequada para o problema identificado. A arquitetura procurou ser genérica o suficiente para que diferentes tecnologias pudessem ser utilizados ou trocadas durante uma instanciação. A Figura 2 apresenta a arquitetura proposta.

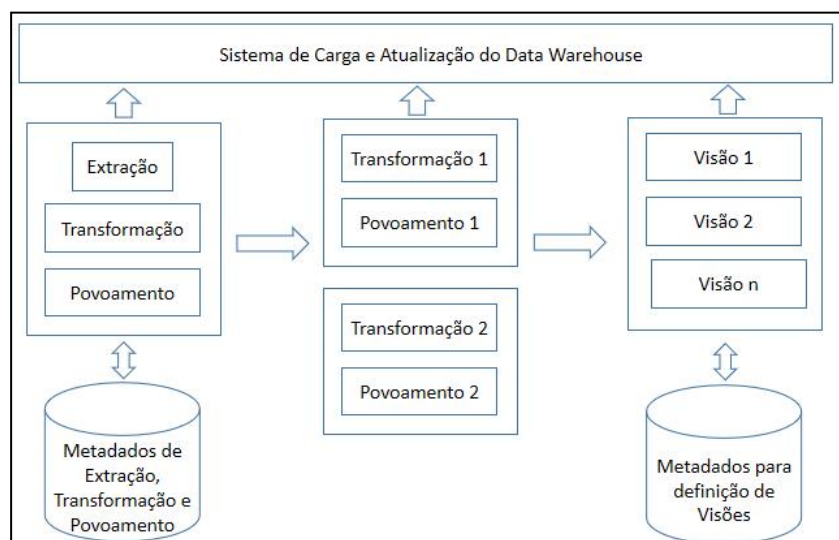


Figura 2. Arquitetura para Processo Automático de extração de dados abertos para um anuário socioeconômico.

Um dos elementos cruciais dessa arquitetura é o primeiro conjunto de processos de Extração, Transformação e Povoamento (localizados na parte esquerda da Figura 2). Esse primeiro conjunto ETL é auxiliado por um esquema de dados lógico que armazena informações necessárias para identificar todas as fontes de dados necessárias para os indicadores do anuário. Essas fontes podem apresentar os dados em diferentes formatos como .zip, .csv, .xlsx ou .json. Cada fonte de dado vai ter um conjunto de metadados que descrevem sua origem, seu formato, seu *layout*, quais dados, linhas ou células devem ser recuperados, seus arquivos intermediários de transformação e os esquemas dimensionais finais para os quais os dados devem ser carregados.

O segundo conjunto de processos de transformação e povoamento são elementos essenciais para a criação de estruturas derivadas dos dados brutos. Também apoiados por metadados, esses processos estão diretamente relacionados com os indicadores socioeconômicos a serem trabalhados. São esses processos, por exemplo, responsáveis pela transformação necessária dos dados brutos para atender cenários em que índices de preço precisam ser levados em consideração.

Finalmente, o último elemento da arquitetura são as visões que podem ser utilizadas diretamente para a geração de gráficos ou tabelas.

Todos os elementos mencionados são coordenados por um subsistema ETL (elemento superior da Figura 2), responsável por coordenar a execução de todos os processos.

A arquitetura apresentada na Figura 2 foi instanciada e foram criados artefatos de código com diferentes tecnologias. Os processos de extração de dados de fontes externas foram implementados na linguagem Python. Os processos de transformação e povoamento de estruturas derivadas foram implementados utilizando a linguagem pgsq do sistema gerenciador de banco de dados PostgreSQL. As visões foram utilizadas pela ferramenta PowerBI da Microsoft para elaboração de gráficos, tabelas e publicação dos visuais na web.

4. Ameaças à Validade

Embora o trabalho ainda esteja no início da fase de Avaliação, a natureza do trabalho permite que sejam apontadas ameaças à validade da solução proposta. Seriam necessários outros Estudos de Caso, replicando a utilização da arquitetura, para confirmar os benefícios esperados. O Anuário Socioeconômico de Sergipe também possui peculiaridades que podem ter contribuído para que a arquitetura proposta apresentasse bons resultados nas avaliações iniciais. Uma dessas particularidades é a reutilização de fontes de dados para grupos de indicadores socioeconômicos. Essa característica facilita a instanciação da arquitetura para um número razoável de visuais (gráficos e tabelas).

5. Conclusões

O presente trabalho, ainda em andamento, apresentou uma proposta de arquitetura para a criação de processos automáticos de extração de dados abertos para anuários socioeconômicos. A arquitetura proposta teve seus componentes instanciados e utilizados em um contexto real do Anuário Socioeconômico do Estado de Sergipe, embora restrito a alguns indicadores. O próximo passo, através da conclusão do Estudo de Caso de maneira mais rigorosa, é avaliar a arquitetura e concluir o ciclo DSR, identificando ou não o artefato como um elemento útil e inovador. Com a utilização da arquitetura, o Anuário, antes estático, poderá apresentar as seguintes características: a) Possuir uma base histórica que poderá ser consultada através de APIs ou diretamente através de acesso ao seu banco de dados; b) Dados extraídos e transformados de maneira mais rápida, facilitando a elaboração de novas visualizações; c) Disponibilidade dos dados do anuário em formato aberto, acessível a qualquer pessoa que queira utilizá-los para efetuar novas análises ou cruzamento de informações.

6. Referências

- Adamson, C. Star Schema: The Complete Reference. United States: McGraw-Hill, 2010.
- Anuário estatístico de Mato Grosso, 2016. Disponível em: <<http://www.dados.mt.gov.br/publicacoes/anuarios/>>. Acesso em: 21 set. 2022.
- Anuário Socioeconômico de Sergipe. Café com Dados, 2022. Disponível em: <<http://cafecomdados.com/anuario/>>. Acesso em: 14 jun. 2021.
- CEPERJ. Anuários Estatísticos do estado do Rio de Janeiro. 2013. Disponível em: <https://www.ceperj.rj.gov.br/?page_id=243>. Acesso em: 21 set. 2022.
- CKAN — The open source data management system, 2018. Disponível em: <<http://ckan.org/>>. Acesso em: 5 mar. 2022.
- CONDEPE FIDEM. Anuários Estatísticos, 2016. Disponível em: <<http://www.condepefidem.pe.gov.br/web/condepe-fidem/anuarios>>. Acesso em: 21 set. 2022.
- Dados abertos. Manual dos dados abertos: governo (2011) Disponível em: <https://www.w3c.br/pub/Materiais/PublicacoesW3C/Manual_Dados_Abertos_WEB.pdf>. Acesso em: 17 mar. 2022.
- Hevner, A.; Chatterjee S. Design Research in Information Systems: Theory and Practice. London: Springer, 2010.

- Inmon, W. H.; Building the Data Warehouse. 4ª edição. Indianapolis, Indiana: Wiley Publishing, Inc., 2005.
- Kimball, R. Ross, M.; The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling. 3ª edição. Indianapolis, Indiana: John Wiley & Sons, Inc., 2013.
- Kimball, R. Ross, M. Thornthwaite, W.; The Data Warehouse Lifecycle Toolkit. 2ª edição. Indianapolis, Indiana: Wiley Publishing Inc., 2008.
- Kimball, R.; The Data Warehouse ETL Toolkit. 1ª edição. Indianapolis, Indiana: Wiley Publishing, Inc., 2004.
- Lima, M. R. F.; Souza, J.. Desenvolvimento de um Data Mart e Automatização do Processo ETL no contexto da Produção Acadêmica do CEULP/ULBRA. In: ENCOINFO - Congresso de Computação e Tecnologias da Informação, 21., 2019, Palmas - TO.
- Macedo, D. F.; Lemos, D. L. S. Dados abertos governamentais: iniciativas e desafios na abertura de dados no Brasil e outras esferas internacionais. AtoZ: novas práticas em informação e conhecimento, 10(2),14 – 26. 2021.
- Oliva, S. Z. Ambiente de Data Warehousing para integração de dados de saúde pública em âmbito de gestão regional. [s.l.] Universidade de São Paulo, Agência USP de Gestão da Informação Acadêmica (AGUIA), 2017.
- Quingerski, L.; Marques, D.; Freire, P. DE S. ETL process using open data from e-gov for discovery and knowledge representation. Proceedings of the 13th CONTECSI International Conference on Information Systems and Technology Management. Anais...TECSI, 2016.
- Reis Júnior, Cleyton Peixoto dos; Holanda, M. T. d. UnBGOLDProv: Arquitetura De Proveniência De Dados Para Um Workflow De Publicação De Dados Abertos Governamentais. 2020.
- Ribeiro, F. C. et al. VisPublica: Uma proposta para aprimorar a transparência de dados públicos. Anais do VIII Simpósio Brasileiro de Sistemas de Informação (SBSI 2012). Anais...Sociedade Brasileira de Computação, 2012.
- SEPLAG. Anuário Estatístico do Estado de Alagoas — Alagoas em Dados e Informações, 2015. Disponível em: <<https://dados.al.gov.br/catalogo/dataset/anuario-estatistico-do-estado-de-alagoas>>. Acesso em: 21 set. 2022.
- The Open Definition — Open Definition - Defining Open in Open Data, Open Content and Open Knowledge, 2015. Disponível em: <<http://opendefinition.org/>>. Acesso em: 20 fev. 2022.
- Transparência Traduzida, 2021. Disponível em: <<http://www.transparenciatraduzida.com.br>>. Acesso em: 17 out. 2022.