

PROJETO METODOLÓGICO PARA *BUSINESS INTELLIGENCE* COM DADOS PARA AVALIAÇÃO DE RISCO À SAÚDE EM CRIANÇAS COM MICROCEFALIA E OUTRAS ANOMALIAS CONGÊNITAS

Data de aceite: 02/10/2023

Jamile Sena Andrade

Faculdade de Ciências e Tecnologia,
Secretaria Municipal de Saúde
Ilhéus – BA

<http://lattes.cnpq.br/6724554926130426>

Sérgio Fred Ribeiro Andrade

Departamento de Engenharias e
Computação - Universidade Estadual de
Santa Cruz (UESC)
Ilhéus – BA

<http://lattes.cnpq.br/2756638501081370>

RESUMO: Este trabalho apresenta um arcabouço metodológico para modelagens conceitual e lógica de um sistema de apoio à decisão com dados de programas de saúde pública para crianças enfermas com microcefalia, sífilis e outras anomalias congênitas. Como também, a estrutura de uma base de dados multidimensional para registros históricos visando descoberta de novos conhecimentos e avaliação de risco do desenvolvimento infantil. Essas informações podem auxiliar a gestão pública na intervenção dos comportamentos atípicos e ajudar na mitigação com ações direcionadas em tempo adequado para promover adaptação e socialização da

criança enferma e o bem-estar materno.

PALAVRAS-CHAVE: Data Warehouse, Microcefalia, Síndrome Congênita

METHODOLOGICAL PROJECT FOR BUSINESS INTELLIGENCE WITH DATA FOR HEALTH RISK ASSESSMENT IN CHILDREN WITH MICROCEPHALY AND OTHER CONGENITAL ANOMALIES

ABSTRACT: This work presents a methodological framework for conceptual and logical modeling of a decision support system with data from public health programs for sick children with microcephaly, syphilis and other congenital anomalies. As well as the structure of a multidimensional database for historical records aimed at discovering new knowledge and risk assessment of child development. This information can help public management in the intervention of atypical behaviors and help in mitigation with targeted actions in a timely manner to promote adaptation and socialization of the sick child and maternal well-being.

KEYWORDS: Data Warehouse, Microcephaly, Congenital Syndrome

1 | INTRODUÇÃO

No segundo semestre do ano de 2015, observou-se no Brasil um aumento expressivo no número de recém-nascidos diagnosticados com microcefalia em locais onde ocorria a circulação do Zika vírus. Essa doença é caracterizada por um conjunto de anomalias congênitas, estruturais e funcionais, com repercussões no crescimento e no desenvolvimento dos embriões ou dos fetos expostos ao vírus durante a gestação. (BRASIL, 2023a).

Entre 2015-2022 o Ministério da Saúde do Brasil notificou 21.196 infectados pelo Zika vírus e 3.732 crianças com problemas no crescimento e desenvolvimento, relacionadas à microcefalia, sífilis, toxoplasmose e outras doenças congênitas. (BRASIL, 2023a).

Dos casos confirmados, 1.716 deveriam receber cuidados em puericultura, estimulação precoce e serviço da atenção especializada. Porém, sabe-se que existe uma precária assistência aos enfermos, mães e famílias de nascidos com essas anomalias, pois de cada 7 crianças apenas 1 recebe atendimento ideal no Brasil. (BRASIL, 2019a).

O desenvolvimento da criança com anomalia congênita depende do grau de comprometimento neurológico, como também, de fatores socioambientais e dos estímulos fornecidos. A falta desse conhecimento pode induzir políticas e programas públicos ineficientes com potencial para gerar saúde materno-infantil debilitada, angústia e estresse aos envolvidos (BRASIL, 2016b).

Embora o Brasil tenha programa e protocolos de acompanhamento e intervenção clínico-terapêutica para crianças com essas anomalias, para a mitigação de sequelas do desenvolvimento neuropsicomotor, efeitos na socialização e acolhimento familiar, ainda não se conhece a real eficácia dessas assistências e o risco do agravamento pelas condições socioambientais das famílias. Especificamente, não se sabe o real desenvolvimento infantil típico, de acordo com as condições socioambientais e não existem indicadores de risco que possam alertar sobre a probabilidade dos grupos ou localidades mais afetadas ou, os transtornos no desenvolvimento da criança com essas patologias em razão da falta ou precariedade de assistência específica.

Muitas variáveis podem explicar a qualidade da assistência especializada recebida por crianças com essas anomalias. Por exemplo, fatores médico-biológicos como baixo peso ao nascer, prematuridade, doenças crônicas dos pais, uso de artefatos para motricidade fina e ampla e internação hospitalar. E, os fatores socioambientais, como qualidade de moradia, o grau de instrução dos pais, renda familiar, número de adultos e crianças na família e outras (SACCANI; VALENTINI, 2010).

O conhecimento desses parâmetros pode suscitar novas descobertas para auxiliar na intervenção dos comportamentos atípicos e ajudar na mitigação com ações direcionadas em tempo adequado para promover adaptação e socialização da criança enferma e o bem-estar materno.

Nesse contexto, para composição da estrutura multidimensional ora apresentada, são empregadas as bases de dados públicas disponíveis como, a da *Coorte* de 100 milhões de brasileiros do CIDACS - Fiocruz (2019), as originadas dos sistemas SIM (Brasil, 2019c) - como SINASC (Brasil, 2019d), SIH (Brasil, 2019e) e Programa Bolsa Família (Brasil, 2019f), e, as bases do SINAN (Brasil, 2019g) e RESP (Brasil, 2019h). Todas adequadas para aplicação num Sistema de Apoio à Decisão (SAD).

Esse conjunto da base do SAD contempla dados multidimensionais para *Data Mart's* (DM's) e *Data Warehouse* (DW) e serve para posterior aplicação com reconhecimento de padrão, classificação, clusterização, associações, tendências e aprendizado computacional, para saídas com recursos visuais de *-Online Analytical Processing* (OLAP).

Trata-se de um projeto metodológico para *business intelligence* que descreve procedimentos das modelagens conceitual e lógica. Apresenta uma metodologia de apoio a avaliação de risco à saúde de crianças com anomalias congênitas e descreve a estrutura de uma base de dados multidimensional, atualizável e ajustada para métodos de descoberta de novos conhecimentos, com disponibilização de registros históricos ideais para o processo decisório.

Desta forma, o presente trabalho objetiva atender a uma lacuna existente no conhecimento da eficiência dos tratamentos terapêuticos e dos programas assistenciais às crianças enfermas com doenças congênitas, especialmente a microcefalia, e dos riscos envolvidos para seu desenvolvimento e bem-estar materno-infantil.

2 | MÉTODOS

As tarefas metodológicas resumidas na Figura 1 são para consecução do projeto sistêmico, como o conceitual e lógico-multidimensional. Algumas tarefas podem ser simultâneas e não limitadas somente às tecnologias e ferramentas indicadas.

As ferramentas apontadas para implementação do projeto são da marca proprietária em Microsoft (2023), por uma questão de compatibilidade tecnológica. Outras ferramentas proprietárias ou de domínio público podem ser usadas do mesmo modo.

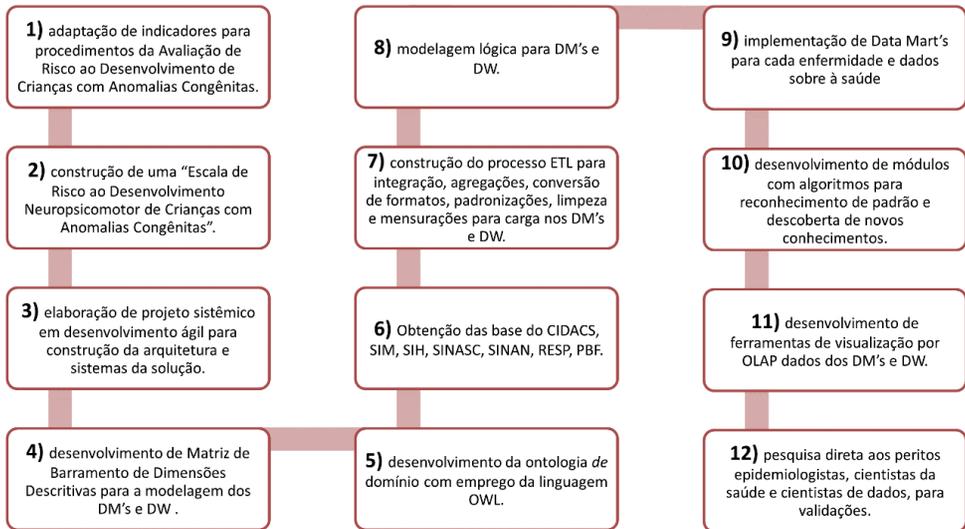


Figura 1. Resumo da proposta da metodologia do sistema.

Nos itens enumerados a seguir estão as especificações dos procedimentos para desenvolvimento da modelagem conceitual e lógica na construção do SAD.

1) uso dos indicadores de Denver II em Frankenburg (1992) e das Avaliações à Exposição de Risco à Saúde Humana em USEPA (1989), WHO (2007), WHO (2010) e ATSDR (2005), através da técnica de clusterização, como parâmetros para a composição da Escala de Risco ao Desenvolvimento Neuropsicomotor de Crianças com Anomalias Congênitas (ERDN);

2) aplicação da ERDN para atribuição de classes ordinais de risco, baseada no item 1, para aplicação com algoritmos de regras de decisão, redes neurais e comitê de máquina de suporte vetorial, visando o processo de “Avaliação de Risco ao Desenvolvimento de Crianças com Anomalias Congênitas e ao Bem-estar Materno”;

3) elaboração de projeto sistêmico com desenvolvimento ágil em engenharia de software baseada em componentes, para construção da arquitetura e sistemas da solução, com etapas para elicitação de requisitos, implementação, documentação, testes, manutenção e implantação;

4) desenvolvimento de Matriz de Barramento de Dimensões Descritivas, com “dimensões descritivas *versus* processos do domínio”, como representação do modelo conceitual para a construção dos DM's e DW;

5) desenvolvimento da ontologia de domínio sobre doenças congênitas com emprego da linguagem OWL (*Ontology Web Language*), de McGuinness et al. (2004), para caracterização da dimensão identificação do perigo e possibilitar reconhecimento por algoritmos de buscas semânticas, com aplicação do software Protégé (2019);

6) obtenção e estruturação das bases públicas do SIM (Brasil, 2019c), SINASC (Brasil 2019d), SIH (Brasil, 2019e), PBF (Brasil, 2019f), CIDACS (Fiocruz, 2019),

SINAN (Brasil, 2019g) e RESP (Brasil, 2019h), com aplicação do MS SQL Server, MS SQL Server Management Studio (SSMS) ou outros congêneres;

7) construção do processo ETL (*Extract, Transform and Load*) para integração, agregações, conversão de formatos, padronizações, limpeza e mensurações para carga de dados nos DM's e DW, com uso do MS SQL Server Integration Services (SSIS) ou outra ferramenta congênera;

8) modelagem lógica para DM's e DW, na abordagem *bottom-up*, fundamentada no modelo descrito em Datta e Thomas (1999), com aplicação do MS SQL Server, MS SQL Server Management Studio (SSMS) ou outros congêneres;

9) implementação física para bases dos DM's, conforme estruturas das bases citadas no item 6, dos domínios sobre microcefalia, sífilis, toxoplasmose e outras doenças congênitas, com dados de internação, dos agravos, dados socioeconômicos e socioambientais, de intervenção clínico-terapêutica, do desenvolvimento da criança, puericultura e serviços da atenção especializada, com aplicação do MS SQL Server, MS SQL Server Reporting Services (SSRS), MS SQL Service Data Tools, MS SQL Server Analysis Services (SSAS) ou outras congêneres;

10) desenvolvimento de módulos com algoritmos em Python ou Java, para reconhecimento de padrão e descoberta de novos conhecimentos com aplicação de MS AI Platform, MS R Tools for Visual Studio, MS Python Tools for Visual Studio ou outras ferramentas afins;

11) desenvolvimento de ferramentas de visualização por OLAP, sobre a base de dados dos DM's e DW, aplicação com MS Power BI, MS PowerApps, MS Visual Studio App Center, ou outros congêneres;

12) pesquisa aos peritos epidemiologistas, cientistas da saúde e cientistas de dados, para validar a confiabilidade dos resultados do método de avaliação de risco à saúde, arquitetura e sistema propostos, com aplicação do método estatístico Delphi (Astigarraga, 2003), coeficientes de concordância de Kendall (Abdi, 2007) e Kappa (Wood, 2007), com análise de *qui-quadrado*.

3 | MODELAGEM CONCEITUAL E MULTIDIMENSIONAL

A modelagem descrita abrange o aspecto conceitual e lógico-dimensional para desenvolvimento de um sistema de *business intelligence*, que aplica metodologia para avaliação de risco à saúde materno-infantil de portadores de doenças congênitas, e especifica a arquitetura sistêmica para dados multidimensionais na composição de DM's e integração num DW.

Para tanto, contempla a escala ERDN que consiste num indicador com classes ordinais (de 1 a 5) conforme a probabilidade de ocorrência, indicadores de saúde, parâmetros terapêuticos e socioambientais, com base no Teste de Triagem de Desenvolvimento de Denver Revisado – TTDD-R ou Denver II. (FRANKENBURG, 1992).

A ERDN permite a atribuição de classes ordinais sobre os efeitos adversos na psicomotricidade em crianças enfermas e por precárias condições socioambientais e econômicas dos familiares, de acordo com uma composição de indicadores extraídos por clusterização das metodologias de avaliação citadas. A justificativa recai pela necessidade de categorização ordinal de risco para a fase de treinamento de dados no ambiente supervisionado por máquina de vetor suporte e outros classificadores, que serão utilizados para a avaliação do risco à saúde.

A Figura 2 mostra o esquema geral das etapas para construção do modelo sistêmico que começa pelo modelo conceitual, o qual é baseado numa ontologia de domínio com taxonomias retiradas das caracterizações das doenças microcefalia, sífilis, toxoplasmose e outras congênicas, que afetam à saúde materno-infantil. Essa ontologia é representada numa linguagem formal *OWL*, que permite compartilhamento semântico via *web* e *mobile*, de fácil operação e recuperação de informações.

A base multidimensional do DW disponibiliza dados históricos e mensuráveis para o processo de avaliação de risco ao desenvolvimento neuropsicomotor da criança com anomalias congênicas e do bem-estar materno. A arquitetura e os dados carregados são ajustados para as operações com algoritmos de classificação, associação, clusterização, reconhecimento de padrão e predições e, possibilita visualização dos dados por OLAP através de *dashboards*, gráficos e tabelas.

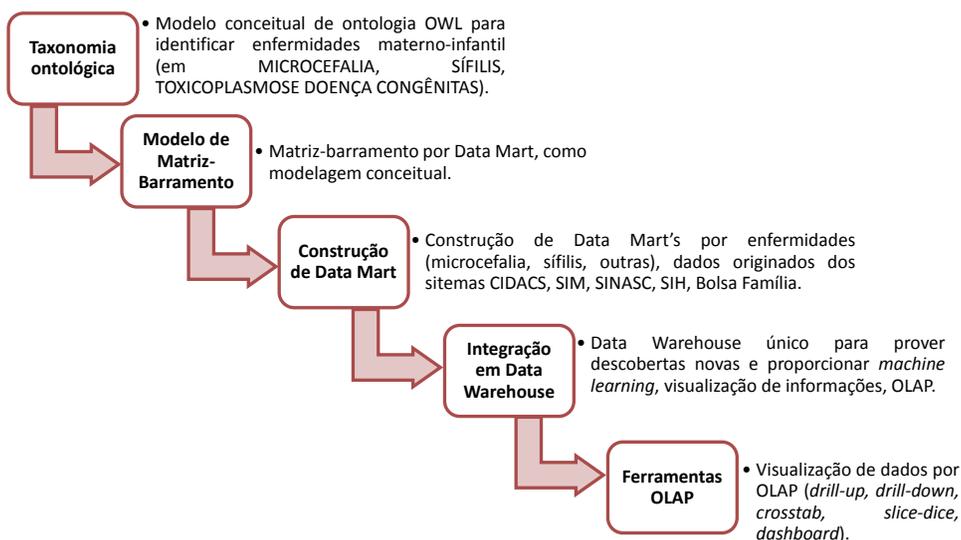


Figura 2. Esquema da modelagem sistêmica.

Especificamente, a matriz de barramento na Figura 3 é um modelo conceitual que relaciona os indicadores e dados da exposição de risco sobre desenvolvimento motor e

aspectos socioambientais, por enfermidade – que compõem as linhas da matriz, e, as dimensões indicadas nas colunas da mesma, que retratam o domínio de cada enfermidade (microcefalia, sífilis, toxoplasmose e outras congênicas), e formam os DM's.

Essas dimensões são baseadas nas etapas de identificação do perigo, avaliação do desenvolvimento, caracterização neuropsicomotor e comunicação de risco, e adaptadas das metodologias de avaliação de risco de *Human Health Risk Assessment* (USEPA, 1989), *Communicable Disease Risk Assessment* (WHO, 2010), *Health Assessments and Consultations* (ATSDR, 2005) e *Rapid Risk Assessment* (WHO, 2019), e da escala ERDN.

As intersecções entre linhas e colunas estão marcadas com um “X” quando as dimensões, possivelmente, forem utilizadas para formação dos cubos de dados multidimensionais para aplicação nas ferramentas visuais de OLAP. Na linha OLAP estão citadas algumas técnicas para visualização da informação (*drill-up, drill-down, crosstab, rollup, grouping, pivot, slice-dice*), que permitem a interpretação e análise para o processo decisório.

Dimensões	Identificação do perigo	Avaliação do desenvolvimento infantil	Caracterização neuropsicomotor	Comunicação de risco	Caracterização de risco	Caracterização socioambiental	Escala de risco do desenvolvimento neuropsicomotor	Data
OLAP Processos	Drill-up, drill-down, crosstab, pivot, grouping	Drill-up, drill-down crosstab, pivot, grouping	Drill-up, drill-down crosstab, pivot, grouping	Drill-up, drill-down, crosstab, pivot, grouping	Drill-up, drill-down, rollup, grouping	Drill-up, drill-down, rollup, grouping, slice-dice	Drill-up, drill-down, rollup, grouping	Drill-up, drill-down, grouping, crosstab, pivot, slice-dice
Motricidade infantil	X	X	X		X		X	X
Indicador neuropsicomotor	X	X	X		X		X	X
Influência pessoal-social	X	X	X	X	X	X	X	X
Indicador adaptativo	X	X	X		X	X	X	X
Indicador uso da linguagem	X	X	X		X	X	X	X
Indicador influência socioambiental	X	X		X	X	X	X	X
Indicador escolaridade da mãe	X			X	X	X	X	X
Indicador econômico familiar	X			X	X	X	X	X

Figura 3. Matriz de barramento do modelo multidimensional.

Na Figura 4 é mostrado um modelo lógico-dimensional de entidades-relacionamentos para DM's – em cada enfermidade que são periféricos, e para DW central – que tem o mesmo modelo esquemático. É representado por um esquema “estrela” com oito entidades de dimensões descritivas e uma entidade de fato quantitativo, relacionados pela cardinalidade um-para-muitos.

As dimensões têm atributos descritivos e fatos têm atributos quantitativos, Esses atributos foram originados dos termos da taxonomia, ou seja, um glossário que visa a padronização e o reconhecimento por algoritmos.

Para representar essa definição, foi considerado o modelo de algebra relacional

adaptado de Datta e Thomas (1999), que propõe a notação quintupla $DW = \{D, M, A, f, R\}$, onde os elementos integrados indicam as características que lembram um “cubo multidimensional”. Essas características são:

- a) $D = \{d_1, d_2, \dots, d_n\}$, um conjunto de n entidades dimensões que representam os atributos descritivos do domínio, onde cada d_i é uma dimensão extraída do domínio;
- b) $M = \{m_1, m_2, \dots, m_k\}$ um conjunto de k entidades fatos que representam os dados numéricos e mensurações do domínio, onde cada m_i é um fato extraído do domínio ou DM, ou, $M = \{m_1\}$, onde é uma entidade fato para DW;
- c) $A = \{a_1, a_2, \dots, a_t\}$, é um conjunto de t atributos onde cada a_i é um atributo extraído de uma dimensão ou um fato do domínio;
- d) A função f é mapeada como $f: D \rightarrow A$ e $f: M \rightarrow A$, i.e., para cada entidade dimensão e fato existe um conjunto de atributos. Cada f é disjuto para atributos, i.e., $\forall a_p, a_i, a_i \neq a_p, f(d_i, a) \cap f(d_p, a) = \emptyset$ e $f(m_r, a) \cap f(m_s, a) = \emptyset$; e,
- e) $R = \{r_1, r_2, \dots, r_m\}$, é um conjunto de relacionamentos um-para-muitos entre dimensões e fatos, através de chaves substitutas, *primary key (pk)* e *foreign key (fk)*, na seguinte relação algébrica: $d_i \cdot pk = m_i \cdot fk$. O conjunto de todas as fk 's concatenadas em fato torna-se pk composta.

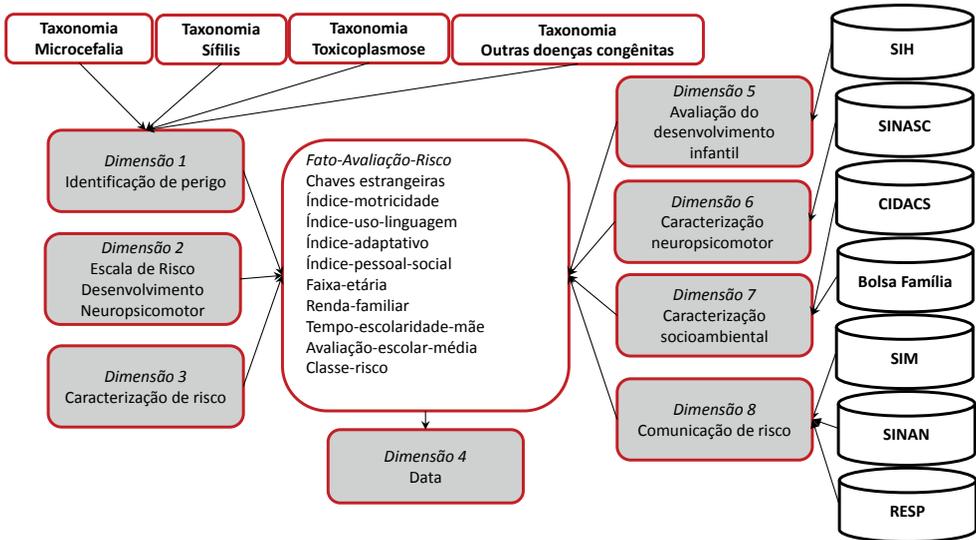


Figura 4. Esquema estrela multidimensional.

CONSIDERAÇÕES FINAIS

Com esse arcabouço espera-se auxiliar para a formação do conhecimento dos tratamentos terapêuticos e da gestão dos programas assistenciais às crianças enfermas com doenças congênicas, especialmente a microcefalia, e, dos riscos envolvidos para seu

desenvolvimento e bem-estar materno-infantil.

Através dos registros históricos pode-se fazer planejamento e acompanhamento com mais acurácia do desenvolvimento infantil, da prevenção de agravos consequentes e a identificação de atrasos no desenvolvimento neuropsicomotor. O que pode permitir o mais breve possível, o diagnóstico, o tratamento, a estimulação precoce e a reabilitação das crianças que necessitam de cuidados especializados.

Na prática, trata-se da possibilidade de oferta de uma aplicação tecnológica para a gestão da saúde pública que pode servir para programas sociais que enfrentam esse problema.

Para o futuro, espera-se aprimorar o ambiente para reconhecimento de padrão e predição de alto risco para auxiliar no diagnóstico dessas anomalias congênitas no recém-nascido e afastar falsos-positivos, como também, desenvolver artefatos para auxiliar no acompanhamento e intervenções desses agravos pelos profissionais à distância.

REFERÊNCIAS

Abdi, Hervé. “**The Kendall rank correlation coefficient.**” *Encyclopedia of Measurement and Statistics*. Sage, Thousand Oaks, CA (2007): 508-510.

Agency for Toxic Substances and Disease Registry - ATSDR. (2019). **Public Health Assessment Guidance Manual, 2005**. Disponível em: <http://www.atsdr.cdc.gov/HAC/PHAMannual/ch2.html>.

Astigarraga, Eneko. “**El método delphi.**” *San Sebastián, Spain: Universidad de Deusto* (2003).

Brasil. Ministério da Saúde(a), Secretaria de Vigilância em Saúde em ambiente. **Boletim Epidemiológico**, n. 5, Volume 54, 3 de abril de 2023.

Brasil. Ministério da Saúde(b). Secretaria de Atenção à Saúde. **Diretrizes de estimulação precoce: crianças de zero a 3 anos com atraso no desenvolvimento neuropsicomotor/** Ministério da Saúde, Secretaria de Atenção à Saúde. – Brasília: Ministério da Saúde, 2016. 184 p. I.

Brasil. Ministério da Saúde(c). DATASUS – Departamento de Informática do SUS. SIM – Sistema de Informações de Mortalidade. Disponível em: <http://www2.datasus.gov.br/DATASUS/index.php?area=060701>. Março/2019.

Brasil. Ministério da Saúde(d). DATASUS – Departamento de Informática do SUS. SINASC – Sistema de Informações de Nascidos Vivos. Disponível em: <http://datasus.saude.gov.br/sistemas-e-aplicativos/eventos-v/sinasc-sistema-de-informacoes-de-nascidos-vivos>. Março/2019.

Brasil. Ministério da Saúde(e). DATASUS – Departamento de Informática do SUS. SIH – Sistema de Informações Hospitalares do SJUS. Disponível em: <http://datasus.saude.gov.br/sistemas-e-aplicativos/hospitalares/sihsus>. Março/2019.

Brasil. Ministério da Saúde(f). PBF - Programa do Bolsa Família. Disponível em: <https://bfa.saude.gov.br/>. Março/2019.

Brasil. Ministério da Saúde(g). DATASUS – Departamento de Informática do SUS. SINAN – Sistema de Informação de Agravos de Notificações. Disponível em: <http://portalsinan.saude.gov.br/>. Março/2019.

Brasil. Ministério da Saúde(h). RESP – Registro de Eventos em Saúde Pública. Disponível em: <http://www.resp.saude.gov.br/microcefalia#/painel>. Março/2019.

Datta, Anindya, and Helen Thomas. “The cube data model: a conceptual model and algebra for on-line analytical processing in data warehouses.” *Decision Support Systems* 27.3 (1999): 289-301.

FIOCRUZ–Fundação Oswaldo Cruz. (2019). CIDACS – Centro de Integração de Dados e Conhecimentos para Saúde. Acessado em: <https://cidacs.bahia.fiocruz.br/plataforma-de-dados/>. Março/2019.

Frankenburg, William K., et al. “**The Denver II: a major revision and restandardization of the Denver Developmental Screening Test.**” *Pediatrics* 89.1 (1992): 91-97.

McGuinness, Deborah L., and Frank Van Harmelen. “**OWL web ontology language overview.**” *W3C recommendation*10.10 (2004): 2004.

Microsoft. Documentações e especificações de ferramentas de desenvolvimento. Disponível em: <https://developer.microsoft.com/pt-br/>. Maio/2023.

Protégé Software. (2019). *Protégé is a core component of The National Center for Biomedical Ontology.* (Version 5.5.0) [Software]. Stanford Center for Biomedical Informatics Research. Available from <http://protege.stanford.edu/>.

Saccani, Raquel e. Valentini, Nadia C. “**Análise do desenvolvimento motor de crianças de zero a 18 meses de idade: representatividade dos itens da Alberta Infant Motor Scale por faixa etária e postura.**” *Journal of Human Growth and Development* 20.3 (2010): 711-722.

United States Environmental Protection Agency - USEPA. (1989). **Risk Assessment Guidance for Superfund. Vol I: Human Health Evaluation Manual (Part A).** Washington - DC.

Wood, J. M. “**Understanding and computing Cohen’s Kappa: a tutorial.** *WebPsychEmpiricist.* Retrieved September 14, 2009 from http://wpe.info/papers_table.html (2007).

World Health Organization - WHO. (2010). **Risk Assessment Toolkit: chemical hazards.** USA, WHO/ IPCS, 2010.

World Health Organization - WHO. RAPID RISK ASSESSMENT. **Communicable disease risk assessment: protocol for humanitarian emergencies.** 2007. https://www.who.int/diseasecontrol_emergencies/guidelines/Com_dis_risk_ass_oct07.pdf. Acessado em Março/2019.