

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
DEPARTAMENTO ACADÊMICO DE INFORMÁTICA
BACHARELADO EM SISTEMAS DE INFORMAÇÃO

WAGNER LUÍS PEREIRA

MAPEAMENTO DE DADOS DE RELATÓRIOS QUADRIMESTRAIS DE GESTÃO
DA SECRETARIA DE SAÚDE DE CURITIBA COM O USO DO MODELO DE
DADOS ABERTOS CONECTADOS

CURITIBA
2018

WAGNER LUÍS PEREIRA

MAPEAMENTO DE DADOS DE RELATÓRIOS QUADRIMESTRAIS DE GESTÃO
DA SECRETARIA DE SAÚDE DE CURITIBA COM O USO DO MODELO DE
DADOS ABERTOS CONECTADOS

Monografia apresentada ao
Bacharelado em Sistemas de
Informação da Universidade
Tecnológica Federal do Paraná como
requisito parcial para obtenção do grau
de “Bacharel em Sistemas de
Informação” – Área de Concentração:
Dados Abertos

Orientador: Profª Drª Rita Berardi

CURITIBA
2018

RESUMO

A prefeitura de Curitiba por meio de sua secretaria de saúde gera relatórios quadrimestrais que não seguem um padrão comum entre eles, com isso os relatórios não podem ser facilmente comparados entre si. Além de que os relatórios serem muito específicos para quem os solicita, não sendo facilmente interpretáveis ou utilizáveis por outras pessoas que se interessem pelos seus dados, assim reduzindo o potencial de uso desses dados. Esse fato traz a vista uma demanda de gerar um modelo diferente para o armazenamento destes dados que traga maior visibilidade comparabilidade para os dados. O presente trabalho mapeou as fontes dos dados contidos nestes relatórios, e gerou modelos no padrão de Dados Abertos Conectados, o qual em sua essência pretende que os dados sejam consumidos por qualquer pessoa ou aplicação. Isso é importante para as pessoas e instituições obtenham informações necessárias para o uso e planejamento nos mais diversos fins, assim gerando proporcionando amplo acesso à informação em uma sociedade de informação e se alinhando a conceitos de cidades inteligentes. Para isso foram utilizadas ferramentas de modelagem recomendadas pela W3C como RDF e OWL.

Palavras-chave: Dados abertos, Dados Conectados, Ontologias, Indicadores de saúde, Dados governamentais, RDF, OWL, Cidades Inteligentes

ABSTRACT

The Curitiba city hall through its health department generates four-monthly reports that do not follow a common pattern among them, so the reports can not be easily compared to each other. In addition, the reports are very specific to those requesting them, not being easily interpretable or usable by other people who are interested in your data, thus reducing the potential of using such data. This fact brings the demand to generate a different model for the storage of these data that brings greater visibility comparability to the data. The present work mapped the sources of the data contained in these reports, and generated models in the pattern of Connected Open Data, which in essence intends that the data be consumed by any person or application. This is important for people and institutions to obtain necessary information for use and planning for a wide range of purposes, thus generating broad access to information in an information society and aligning with Smart Cities concepts. W3C-recommended modeling tools such as RDF and OWL were used.

Keywords: Open Data, Linked Data, Ontologies, Health indicators, Government Data, RDF, OWL, Smart Cities

LISTA DE FIGURAS

Figura 1: Diagrama do método	25
Figura 2: Burn Chart Down	28
Figura 3: Gráfico de evolução do número de unidades de saúde	38

LISTA DE TABELAS

Tabela 1: Cronograma de atividades realizadas no TCC	27
Tabela 2: Total de leitos SUS 2017-3	30
Tabela 3: Rede física de serviços 2017-3	31
Tabela 4: Total de leitos SUS 2018-1	31
Tabela 5: Rede física de serviços 2018-1	32
Tabela 6: Tipos de gestão	32
Tabela 7: Tipos de gestão	32
Tabela 8: Tipo de estabelecimentos	33
Tabela 9: Tabela Leitos SUS	35
Tabela 10: Tipo de estabelecimento 2017.3	36
Tabela 11: Tipo de estabelecimento 2018.1	37

SUMÁRIO

1. Introdução	9
1.1. Objeto	10
1.2. Problema computacional	11
1.3. Hipótese	11
1.4. Tese	11
1.5. Objetivo geral	11
1.6. Objetivos específicos	12
1.7. Justificativa	12
1.8. Organização do trabalho	13
2. Referencial Teórico	14
2.1. Dado, informação e conhecimento	14
2.2. Cidades Inteligentes e uso de TIC's	16
2.3. Dados abertos conectados	16
2.4. Gestão da informação na gestão de saúde pública	19
2.5. Estado da arte	20
3. Método aplicado	24
3.1. Análise das informações coletadas	24
3.2. Seleção dos indicadores a serem tratados	24
3.3. Criação de schemas (Modelagem)	24
3.4. Instanciamento dos dados em esquemas gerados	24
3.5. Testes e resultados	25
3.6. Revisão do trabalho e conclusões	25
3.7. Diagrama	25
Modelagem	25
Planejamento	25
Entendimento	25
4. Recursos de Software e Hardware	26
5. Viabilidade e Cronograma	27
5.1. Viabilidade Financeira e técnica	27
5.2. Cronograma de atividades	27
6. Análise da fonte de dados atual	29
7. Construção do schema RDF/OWL	30
8. Resultados obtidos	34
9. Conclusão	39

9.1. Dificuldades encontradas	39
9.2. Trabalhos futuros	40
10.Referências	41

1. Introdução

A prefeitura municipal de Curitiba gera relatórios de gestão na área de saúde em regime quadrimestral e anual, que são disponibilizados no site¹ da secretaria municipal de saúde. Estes relatórios estão disponíveis ao público em geral, o que os torna dados abertos. Porém sua formatação não segue um padrão que torne viável seu uso de forma a auxiliar no processo de gestão. Nos relatórios existem grandes diferenças estruturais, o que os torna de difícil comparação. Essas diferenças ficam evidentes quando nos relatórios existem dados que só estão presentes em algumas versões do relatório. Entre os problemas podem ser citados mudança de nomes e fontes dos indicadores, ordenação e organização dos relatórios. Um caso observado é o do indicador de cobertura vacinal, que se encontra presente no primeiro quadrimestre de 2017, porém não no primeiro quadrimestre de 2016, ainda que no relatório de 2017 constem dados de 2016 para efeito comparativo.

As informações de cada relatório de forma isolada apresentam apenas um retrato do panorama geral da saúde de Curitiba, mas se o objetivo é melhorar a tomada de decisão e torná-la mais transparente, a análise da série histórica deveria compor o processo. Já que essa análise pode auxiliar a tomada de decisão e/ou o planejamento na área da saúde, por trazer o retorno das políticas públicas aplicadas e sua evolução. Ainda sim é importante que esse tipo de dado possa ser acessível desde cidadãos comuns, investidores, governantes ou candidatos postulantes a cargos públicos. Com isso possibilitando a utilização dos dados para os mais diversos fins com intuito de planejar e tomar as melhores decisões, na melhoria ou uso da rede de saúde disponível na cidade.

Um problema relatado na introdução do relatório do terceiro quadrimestre da prefeitura de Curitiba em 2016 (SAÚDE, 2016) é o reúso dos relatórios para modelagem de outro padrão de relatórios, como o existente na base nacional SargSus². Como os dados estão dispostos sem um padrão específico, denotando que eles são gerados a livre critério do solicitante, imaginasse que exista uma

¹ Acessado em 04/09/2018 pelo link: <http://www.saude.curitiba.pr.gov.br/a-secretaria/relatorios-de-gestao.html>

² Acessado em 04/09/2018 pelo link: <https://sargsus.saude.gov.br/sargsus>

preocupação clara com geração de uma estrutura fixa. Com isso a adaptação dos relatórios, a análise de série histórica, ou outras medidas que demandem dados estruturados tem seu custo elevado, podendo inclusive se tornar inviáveis por esta questão. Visto isto o trabalho de geração do relatório perde em eficiência e eficácia no que compete a utilização de informações importantes contidas no relatório para o bem público ou qualquer fim, podendo deixar assim de atender pessoas interessadas em suas informações.

Para o estudo deste problema foi utilizado um conjunto de elementos contextualizados por conceitos e padrões gerados pela W3C. Sendo a W3C uma organização com objetivo de criar padrões para a WEB, fundada pelo professor do MIT da área de Ciência da Computação e criador do conceito de *World Wide Web* (www), Tim Berners-Lee. Tim Berners-Lee já na apresentação da *World Wide Web* (www) (BERNERS-LEE, 1990) onde os Tim Berners-Lee já apresentava a web do futuro com informação conectada entre si e com uma semântica atrelada. Hoje a semântica contida nos dados dos relatórios é em geral implícita e dependente de grande interpretação humana para ser entendida e utilizada.

Para gerar o formato semântico na academia se fala em Dados Abertos Conectados se encaixa muito bem, sendo esse formado pela união conceitos de Dados abertos (W3C, 2009) e Dados Conectados (Bizer; Heath; Berners-Lee, 2006). Para este formato de dados são gerados modelos onde a semântica das conexões possibilita diversas combinações das informação e as consultas a bases de dados se tornam atreladas a lógica contida na informação. Com isso processo uso dos dados mais simples e direto, e não depende de tanto entendimento prévio do modelo, reduzindo o esforço editorial e analítico, além de aumentar a eficácia do uso de mecanismos automatizados para esse fim.

1.1. Objeto

Dados obtidos em consulta às fontes contidas nos relatórios gerados pela secretaria de saúde da prefeitura de Curitiba, na sua distribuição quadrimestral com referências de tempo desde 2012/3 até 2018/1. Sendo os mesmos relatórios apresentados no tema deste trabalho.

1.2. Problema computacional

Como modelar os dados que são objetos deste trabalho, com o intuito de que possam ser gerados tanto relatórios semelhantes e aos apresentados pela prefeitura de Curitiba, quanto outros diversos, relatórios e análises. Para que estes dados possam vir a ter um armazenamento que os torne de alguma forma serem mais livres, reutilizáveis e comparáveis entre si.

1.3. Hipótese

A utilização do conceito de dados abertos conectados apresentado por Isotani e Bittencourt (2015) na modelagem de uma base de dados, é capaz de gerar dados que possam ser utilizados de forma livre por qualquer pessoa ou entidade sobre os dados contidos naquela base.

1.4. Tese

A base gerada pela hipótese apresentada, com os dados que são objeto deste trabalho, poderá trazer alguma contribuição para a liberdade de uso, reusabilidade e análises históricas dos dados contidos nos relatórios da prefeitura. Com isso será possível gerar novamente estes mesmos relatórios ou outras diversas análises, assim sendo uma ponte de acesso aos dados de forma a contribuir de alguma forma na utilização desses dados para os mais diversos fins.

1.5. Objetivo geral

- Gerar um modelo para um conjunto de indicadores contidos em relatórios trimestrais da saúde com base no formato de Dados Abertos Conectados. Com intuito de possibilitar o reúso, integração e comparação dos dados contidos nos relatórios da prefeitura de Curitiba, possibilitando assim formas de uso que possam agregar ao processo de obtenção de informações sobre a situação da saúde na cidade de Curitiba a qualquer um que se interessar pelo tema.

1.6. Objetivos específicos

- Mapear os indicadores e suas respectivas fontes de dados baseada no estudo de relatórios referentes à gestão da Secretaria Municipal de Saúde de Curitiba disponíveis no site¹ da secretaria municipal de saúde;
- Selecionar indicadores a serem tratados;
- Padronizar os dados conforme modelo da W3C, utilizando os formatos, *Resource Description Framework* (RDF), W3C (2014) e *Web Ontology Language* (OWL), W3C (2009);
- Realizar testes dos esquemas gerados, com objetivo de verificar o efeitos e fidelidade do modelo gerado com instanciamento de dados reais, tanto para fins de geração de relatórios quanto para fins de análises de séries históricas.

1.7. Justificativa

O governo brasileiro tem feito um esforço para aumentar a transparência dos dados de interesse público, fato que é citado por trabalhos acadêmicos que serão apresentados neste trabalho. Para que isso ocorra de forma eficaz e eficiente existe a necessidade é a de tratar dados referentes à administração pública a utilizar as melhores abordagens para este fim na área da Ciência da Computação. Uma abordagem que tem destaque nesta área é a dos Dados Abertos Conectados, por isso esta irá ser empregada por esta pesquisa para estruturar dados pertencentes aos relatórios da Prefeitura Municipal de Curitiba.

A existência de dificuldade de comparação entre os relatórios, pela sua mudança estrutural, no que tange tanto quais dados são abordados, quanto os nomes dos indicadores e sua localização. Causa uma grande redução no potencial de aproveitamento dos dados constantes nestes relatórios teriam se fosse possível a sua fácil comparação. Anteriormente citamos que no terceiro quadrimestre da prefeitura de Curitiba em 2016 (SAÚDE, 2016) é descrita uma dificuldade na integração da prefeitura com os relatórios da base nacional SargSus². Estes relatórios do SargSus são descritos pela secretaria de saúde de Curitiba como muitos diferentes, assim não atendendo as demandas e nem podendo ser integrados processo de geração de relatórios local. As dificuldades de integração acabam se tornando recorrentes ao pensar em qualquer processo integrações ou em comparações que queiram ser feitas com os dados presentes no relatório para

quaisquer fins. Pensando nisto o presente trabalho buscou estudar uma alternativa que possa ser útil na mitigação destes problemas e analisar sua viabilidade.

Além disso o mapeamento e comparação dos relatórios como estão hoje se torna um processo extremamente custoso e demorado, já que os indicadores são alterados mudando sua nomenclatura, posição e até disponibilidade nos relatórios. Com isso o potencial de auxílio nas decisões dos relatórios se torna muito menor, em Lacerda et al. (2011) é citada a importância de instrumentos que auxiliem a gestão da saúde. A agilidade e o baixo custo, na geração de informação são importantes para a viabilidade de grande parte das atividades organizacionais no mundo, na saúde pública não é diferente.

Para exemplificar, Tanaka e Tamaki (2012) falam sobre o papel da avaliação na gestão de saúde pública. No texto há explicações de como o processo de avaliação necessita de informações e conhecimentos disponíveis em tempo real e que não tenha um alto custo, para que haja viabilidade no processo de avaliação. As características citadas para bom uso do processo que são utilidade, oportunidade, factibilidade, confiabilidade e objetividade. Ela tem seus efeitos diretamente alterados pela informação, sua viabilidade e disponibilidade na hora correta.

1.8. Organização do trabalho

Este trabalho é distribuído em nove capítulos, sendo o primeiro e presente capítulo o de introdução. Em seguida há um capítulo dedicado ao referencial teórico, o qual busca trazer conceitos importantes para o entendimento do trabalho e trabalhos correlatos no estado da arte sobre o tema. Em seguida se encontram capítulos dedicados ao planejamento do experimento, sendo eles método aplicado, recursos de hardware e software e viabilidade e cronograma. Logo após o desenvolvimento do experimento é apresentado em capítulos dedicados à análise da base de dados, construção de *schemas* RDF e OWL e resultados obtidos. Por fim se encontra a conclusão, que inclui ainda dificuldades e trabalhos futuros a serem realizados.

2. Referencial Teórico

2.1. Dado, informação e conhecimento

Não é possível falar em geração de dados com semântica falar nos conceitos de dado, informação e conhecimento. Já que esta forma de geração de dados tem por objetivo produzir informação de acordo com sentidos e enriquecer tornar a informação mais relevante para seu uso na disseminação do conhecimento.

O conceito de dado, definido por Davenport e Pruzak (2003), descreve os dados como sendo um conjunto de fatos distintos e objetivos, relativos a eventos. Para os autores os dados importantes, e devem ser cuidadosamente armazenados. Apesar disso, é afirmado no texto que os dados apenas descrevem parcialmente os fatos ocorridos não possuem por si só grande relevância ou propósito. Para Ackoff (1989), os dados são símbolos que representam as propriedades de objetos e eventos. Ainda sobre o tema Setzer (1999) afirma que dados podem ser algo como um texto, um áudio, uma imagem, entre outros, porém devem ser algo quantificável. Assim é a compreensão de que o dado em si não carrega um sentido próprio e com isso não pode ser interpretado individualmente. Apesar de ser possível sua quantificação ela não pode ser interpretada sem conhecer a semântica envolvida no dado.

Já a informação carrega além do dado quantificável a semântica atrelada, para Ackoff (1989) a informação se diferencia do dado por conter algo que provenha sentido ao dado. Por exemplo uma unidade de medida, que traz a uma informação numérica, de apenas um símbolo para uma quantificação de uma determinada métrica específica, transformando o dado em informação. Davenport e Pruzak (2003) tratam a informação como uma mensagem com emissor e receptor e ainda afirmam que a informação é capaz de impactar e mudar a forma com que o receptor vê algo. Portanto a informação para os autores se diferencia do dado por ter o significado e a relevância, não encontradas nos dados. Com isso se torna compreensível que a definição informação tem relação com o sentido empregado ao dado e o que ele representa para algo ou alguém.

Sobre o conhecimento Davenport e Pruzak (2003) citam que para a maioria das pessoas o conhecimento é algo mais profundo que a informação. Também é citado pelos autores que o poucos acreditam que este pode ser encontrado em

memorandos ou em bancos dados densos. O que os autores afirmam ser claro é que o conhecimento não é simples, já que muitos estudiosos passam anos para entender o que é conhecer. Com isto os autores afirmam que o conhecimento é intrínseco às pessoas e o mesmo deriva das informações, assim como elas derivam de dados. Na visão abordada por Ackoff (1989) o conceito de conhecimento é transmitido por respostas para as perguntas sobre determinado dado, ou seja, por informações como descritas acima. Por exemplo não adianta se ler todo um livro sobre determinada tecnologia, se não se entender os conceitos nele contidos. Assim o livro é mero auxiliar na aprendizagem da tecnologia, pois nele está contida informação, e quem o lê e pratica os conceitos nele contidos apreende sobre a tecnologia, sendo inclusive capaz de ensinar a mesma. Com base nestes dados é denotado que o conhecimento é algo relativo ao ser humano e não algo armazenável em uma base de dados. O que é armazenável são as informações que podem ser usadas para se obter o conhecimento e que precisa de sentido dado às informações para ser eficientemente transmitido.

O conhecimento gerado pode ser entendido como o objetivo de informações transparentes e claras, o processo de obtenção deste é descrito por Nonaka e Takeuchi (1997) no ambiente organizacional. Ainda em Nonaka e Takeuchi (1997), são explicadas diferenças entre a informação e o conhecimento, são citadas características do conhecimento como ser carregado de crenças e compromissos, ter uma ação e ter um significado. Já a informação para os autores é tratada como um meio para obtenção de conhecimento, pois traz uma nova visão sobre a interpretação de eventos ou objetos. Apesar de tratar do ambiente organizacional Nonaka e Takeuchi (1997) trazem teorias de obtenção do conhecimento que podem ser usadas no contexto geral, (NONAKA e TAKEUCHI, pg 57, 1997) é citado que a criação de conhecimento, em sentido rígido é um processo do individual e não organizacional. Com isso os processos nas empresas listados são estratégias para criação do conhecimento no indivíduo e perpetuação dessa criação no ambiente organizacional. Dentre as estratégias listadas (NONAKA e TAKEUCHI, pg 65-66, 1997) é lembrado que a combinação e troca de conhecimento ocorre conhecimento é dada por trocas e combinação entre as pessoas, mas isso só é possível por meio de reconfiguração da informação existente.

2.2. Cidades Inteligentes e uso de TIC's

É importante entender um pouco sobre o uso de tecnologias de informação e comunicação (TIC's) em cidades inteligentes e como a informação é importante nesse novo meio que é vivido por muitos.

Segundo Lemos (2013) desde a década de 90 é fortemente debatida a criação de cidades digitais e o uso das novas TIC's. No livro de Warschauer (2006) em sua introdução são citados exemplos em contexto global da adoção das TIC's e como elas devem ser implementadas com cuidado para ter um efeito concreto e positivo. Nestes exemplos há casos onde a implementação de TIC's sem o devido planejamento e discussão geraram conflitos, sociais e organizacionais, ainda que mesmo com grande quantidade de recursos as mesmas não foram absorvidas de forma a trazer benefícios.

Nos anos seguintes essas questões foram abordadas de forma a entender e pensar no problema do acesso ao acesso das TIC's pela população. Neste contexto o debate de cidades digitais evoluiu para o debate de cidades inteligentes (Lemos, 2013). No texto é abordado quer um acesso a informação sensível ao contexto e a grande produção de informação atrelada ao grande acesso às TIC's pela população ao inerente a popularização das mesmas. Por fim é enfatizada a importância da participação de todos na geração e uso de informação. Com essa informação em vista fica visível que a qualidade da informação a ser acessada pela população e seu acesso são extremamente importantes ao se pensar em cidades inteligentes.

2.3. Dados abertos conectados

O conceito de Dados Abertos (W3C, 2009) que será utilizado para o desenvolvimento deste trabalho, é apresentado pelo texto de Isotani e Bittencourt (2015) com três normas fundamentais que são disponibilidade e acesso, reuso, redistribuição e participação universal. Essas características visam permitir o acesso e uso universal, inclusive comercial, destes dados. Essas características visam a livre utilização para desenvolvimento de quaisquer aplicações, independente de seu propósito final.

Pensando na importância da disponibilidade de dados e na transparência da gestão pública que é cobrada com maior intensidade a cada dia pela sociedade, o governo tem por meio de leis e decretos regulamentando a disponibilização de

dados abertos. Um destes marcos regulatórios foi o decreto nº 8.777³, de 11 de maio de 2016, assinado pela então presidente Dilma Rousseff. Neste decreto as únicas restrições para que estes dados sejam disponibilizados são que não haja nenhum impeditivo ou sigilo atrelado aos dados, sendo assim de responsabilidade do poder público fazer essa disponibilização. Segundo publicação⁴ da prefeitura de Curitiba o marco regulatório que é seguido é a lei Federal 12.527/2011⁵ a qual é regulamentada no âmbito municipal por pelo decreto 1.135/2012⁶.

O conceito de Dados Abertos (W3C, 2009) é apresentado por diversos autores e trabalhos da área. Lopes, Vidal e Oliveira (2016), afirmam que esse tipo de disponibilização de dados tem sido feita em países como Brasil, Estados Unidos e Reino Unido. Além disso os autores trazem a informação de que dados de diversas áreas, como clima, educação, finanças e outras áreas, além da saúde, tem sido alvo dessa disponibilização de dados.

Já em relação aos dados conectados Bizer, Heath, Berners-Lee (2006) que é um conceito que surge para integrar a quantidade de dados existente na internet. Segundo Isotani e Bittencourt (2015) já em 2002 existiam dados na internet no nível de exabytes (10^{18} bytes). Número que cresceu para a casa dos zettabytes (10^{21} bytes) e que até 2020 o crescimento do volume de dados é estimado em cerca a 40 zettabytes, pelo estudo. Por fim o estudo afirma que o uso dessas informações é relevante em cerca de 70% do produto interno bruto (PIB) dos países do G7, essa característica os torna sociedades de informação.

Quando foi anunciada a criação a *World Wide Web* (WWW) em 1989 pelo físico inglês Tim Berners-Lee, ele já citava a necessidade de semântica entre os dados para viabilizar sua conexão. Tim Berners-Lee ainda criou um consórcio visando o desenvolvimento de padrões e tecnologias relacionadas à Web, este denominado de *World Wide Web Consortium* (W3C). Esta é a grande responsável por padrões de dados conectados, como a Web Semântica, conceito introduzido

³Acessado em 09/10/2018 pelo link: http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2016/Decreto/D8777.htm

⁴ Acessado em 14/04/2019 pelo link: <http://www.curitiba.pr.gov.br/leiacessoinformacao>

⁵ Acessado em 14/04/2019 pelo link: http://www.planalto.gov.br/ccivil_03/_Ato2011-2014/2011/Lei/L12527.htm

⁶ Acessado em 14/04/2019 pelo link: <http://multimidia.curitiba.pr.gov.br/2018/00228941.pdf>

por Tim Berners-Lee em 1994, que procura estender a Web clássica criando uma estrutura semântica que pessoas e máquinas possam navegar com eficiência.

Isso se deve a percepção da possibilidade de usar essa grande massa de dados de forma a gerar informações com semântica embutida, pode ser vista como uma oportunidade de gerar informação útil em várias áreas.

Tim Berners-Lee (2010) ainda propôs um “sistema de 5 estrelas” para os dados conectados em que:

- 1 estrela: Formato disponível na internet em qualquer formato imagem ou texto, desde com licença aberta;
- 2 estrelas: Disponível na internet em modo estruturado, como em uma planilhas, podendo ser em formato proprietário;
- 3 estrelas: Cumprir o formato de 2 estrelas, só que em formato não proprietário;
- 4 estrelas: Cumprir as 3 estrelas e usar padrões da W3C, através da utilização do *Resource Description Framework* (RDF), ontologias com o padrão *Ontology Web Language* (OWL) e com isso possibilitando consultas via SPARQL, linguagem semelhante ao SQL tradicional para obter dados em bases Web Semânticas;
- 5 estrelas: Todas as regras anteriores e conexão com outras bases de dados, fornecendo contexto aos dados.

Sobre o framework RDF (W3C, 2014) que foi criado pela W3C para criação de relacionamentos, através de expressões triplas, descrevendo a semântica contida. Segundo o levantamento feito por Miller (1998), o framework oferece acesso precedentes a informações fornecidas globalmente. Miller (1998) ainda diz em sua descrição do RDF que o uso de metadados, sendo essas estruturas sobre os dados que o framework tem por objetivo produzir, podem melhorar a descoberta e o acesso à informação.

Já sobre o padrão *Ontology Web Language* (OWL) sua primeira versão foi formalizada em 2004, pela W3C, a mesma em 2009 através de um documento não normativo, com um alto nível de abstração descreve a utilização da OWL 2.0. Neste documento são apresentados os aspectos básicos do uso da OWL 2.0 e feedback sobre a OWL 1.1, versão anterior da linguagem. O documento explica como a OWL 2.0 define uma estrutura para a construção de semântica com uso do framework

RDF. A grande vantagem apresentada pela W3C para o uso desta linguagem é que ela cria uma estrutura lógica forte para as semânticas produzidas pelo RDF.

O novo conceito de Dados Abertos Conectados (*Open Linked Data*) (Isotani; Bittencourt, 2015) proporciona aos dados uma forma de consulta mais eficiente pelas pessoas ou aplicações esse aspecto é muito interessante para dados governamentais abertos. Já que quanto mais consultados maior o potencial de que esses dados sejam úteis e maior, podendo inclusive aumentar a percepção social de transparência governamental. A abertura de dados é um grande passo, porém para seu uso ser feito de forma a realmente contribuir é necessário alguma forma de consultar e comparar esses dados. Sendo que há um esforço governamental para atender uma demanda da população por maior transparência, tendo então esse tipo de tecnologia uma aliada neste esforço para que a transparência ocorra na prática.

2.4. Gestão da informação na gestão de saúde pública

Para realizar o estudo ainda é necessário entender como o tema de gestão de saúde pública e como a gestão da informação é importante o tema, para isso existem vários artigos. Um deste artigo é o de Carreno et al. (2015) que descreve o Sistema de Informação de Atenção Básica (SIAB) como um dos principais recursos computacionais utilizados para a gestão da saúde pública no Brasil, tendo este sistema por objetivo ser uma base de dados de indicadores de saúde da saúde pública do país. No texto é feita uma explanação sobre a necessidade de melhora e as potencialidades existentes. São abordados problemas no cadastro das informações, em quais informações são coletadas e no retorno gerado por estas informações. Algo que chama a atenção é a potencialidade que esses recursos podem ter segundo o artigo, o sistema traz indicadores e informações relevantes e é a principal fonte de informação para entender a realidade da saúde pública. A partir disso é importante salientar que dados mais acessíveis a usuários e outras aplicações podem maximizar essa potencialidade citada.

Na pesquisa de Lacerda et al. (2011) é citada a importância do planejamento para área de saúde, sendo que o texto explica que para realizar o planejamento há dependência de dados atualizados e confiáveis. O método empregado neste artigo teve 100% de respostas, porém é citado pelos autores, que esta média apenas 11% de respostas e que eles só conseguiram todas as respostas por contatos e

facilidades específicas que eles tinham. Isso leva a crer que se os dados estiverem disponíveis sem necessidade de consultas e questionários esse tipo de pesquisa pode ser otimizado.

Lopes, Vidal e Oliveira (2016) realizam um trabalho visando integração de bases heterogêneas para auxílio na localização de correlações entre óbitos e anomalias congênitas em recém nascidos. Assim podendo obter informações relevantes que possam ser utilizadas em campanhas de conscientização sobre maus hábitos na gravidez. E essas campanhas tornem-se mais eficientes na redução do número de óbitos na maternidade da cidade de Tauá/CE por decorrência das anomalias estudadas.

Para Zaidan e Bax (2013) o uso de sistemas de informação pode enriquecer as análises clínicas por isso propõe o uso de Dados Abertos Conectados podem contribuir para essa área. São apresentados alguns frameworks e ferramentas a serem utilizadas para o uso de dados abertos na área. A conclusão do artigo cita questões como a interoperabilidade dos dados que as ferramentas semânticas, conseguindo geração de uma maior massa de dados, tratando com maiores massas de dados porém ainda mantendo a qualidade dos mesmo. Dessa forma esse tipo de tecnologia pode trazer e que seu uso pode agregar valor a informação dos Sistemas de Informação Clínico.

Com isso podemos entender que tanto no planejamento público da saúde, quanto no desenvolvimento da pesquisa a disponibilidade de dados é extremamente útil. Esse tipo de disponibilidade torna viável diversas ações e com isso gerando valor agregado ao processo de planejamento e pesquisa, com intuito de facilitar e melhorar as decisões tomadas neste processo.

2.5. Estado da arte

Visto todos estes conceitos é agora necessário entender como está se fazendo sua aplicação no cenário acadêmico atual, com isso contextualizando o presente trabalho neste cenário. Algo importante é entender como os dados abertos, no contexto das cidades inteligentes, podem auxiliar a tornar possível a difusão de informações de interesse público à quem possa interessar. Um exemplo é o estudo de Duarte (2005) em que é pesquisado o papel da inovação tecnológica nas cidades em uma sociedade de informação, sendo abordada importância da cidade nesse tipo de sociedade. No texto é ressaltado que tanto para empresas

como para pessoas que buscam inovar são dependentes de um ambiente informacionalmente rico e analisa os benefícios que um fluxo eficiente de informações pode vir a trazer para este processo.

Outra publicação interessante de Gaspar, Azevedo e Teixeira (2016) traz uma análise dos rankings das cidades inteligentes e abordando quais os critérios relevantes para que uma cidade seja considerada inteligente. Neste artigo é citado que o processamento das informações nas cidades inteligentes serve como referência nas tomadas de decisão de empresas, governos e cidadãos, visando tornar seu dia a dia melhor e mais eficiente. Ainda é explicado que a troca de informações entre os órgãos públicos e privados é importante e que visualizar as necessidades do cidadão é um objetivo que é levado em conta. Nesse ponto, se torna visível a necessidade de um meio mais eficiente para transmitir essas informações.

Já sobre dados abertos conectados e suas ferramentas se levantam pontos como o fato de que na web tradicional, os dados não observam que as relações semânticas e isto torna muito difícil reúso e desambiguação, tornando inviável o uso de dados em vários tipos de aplicações. Lopes, Vidal e Oliveira (2016), afirmam que dados abertos em geral não possuem um formato fixo. Com isso as bases acabam ficando isoladas, e com isso há dificuldade na integração de dados diferentes bases e que essa integração é um processo difícil e custoso. Levantada esta questão pode ser pensado sobre a importância de existir um mecanismo, para os dados abertos, que possibilite seu uso integrado. Neste sentido a tecnologia de dados abertos conectados depois de empregada aos dados pode auxiliar para que o processo de geração de informação. Além disto pode aumentar escalabilidade e reduzir custo deste processo, e com isso auxiliando na viabilidade da geração de informações e ainda sendo de alguma forma mais acessível a todos.

Sobre a geração de dados neste modelo, utilizando as ferramentas que são indicadas pela W3C, que foram acima apresentadas, podemos citar algumas publicações. Uma destas é o estudo de Victorino et al (2016) em que é apresentada uma abordagem do uso de dados abertos conectados para criação de modelos que possibilitem o reúso de dados governamentais em grande quantidade. O estudo é uma proposta de criação de bases de 5 estrelas, através um levantamento bibliográfico buscando a integração de propostas. Com isso foi gerada uma visão holística do domínio apresentado. Por fim o estudo conclui que a criação de dados

abertos conectados é uma tarefa árdua, mas que enriquece a transparência da informação pública podendo proporcionar inúmeros benefícios aos cidadãos.

Já no artigo de Alencar et al. (2018) propôs uma aplicação com características semântica com objetivo de criar uma vitrine de currículos lattes, através da base de dados existente no Instituto Federal da Paraíba (IFPB). Neste trabalho foi desenvolvida uma ontologia denominada OUAL, já que não foi encontrada uma ontologia completa que atendesse o problema abordado pelo trabalho. Nessa ontologia foram utilizadas partes de ontologias consolidadas e criadas novas características próprias para outros dados necessários para o instanciamento dos dados. Criada a ontologia os dados precisavam ser passados do formato original, csv e xml para triplas RDF, o que foi feito através de uma aplicação codificada pelos pesquisadores. Por fim foi criada uma aplicação para consumir os dados conforme especificações e demandas criadas pelos usuários da futura plataforma. As conclusões do trabalho citam que a modelagem e implantação de bases de dados semânticas requerem grandes esforços, mas geram um artefato que pode ser reutilizado por diversas aplicações.

O artigo de Pierin e Sichman (2018) traz uma proposição de uma arquitetura para a Web que seja integrada apresentando suas vantagens e ilustrando o processo com casos reais. Segundo este trabalho a informação constante na internet é concentrada em silos, como Facebook, Wikipedia e Google Maps. E como a Web é carregada de dados sem uma semântica clara, há dificuldade de correlacionar os dados para produzir novas informações relevantes, carecendo a internet de um mecanismo que crie correlações automáticas entre os dados em diversas fontes. Neste contexto é citada a Web semântica como uma alternativa para o problema, já que tem a capacidade de gerar consultas em um contexto mais próximo ao do usuário. O trabalho se preocupou em ontologias ligadas a eventos delimitados em determinada geolocalização. Os dados são recuperados por um sistema inteligente de recuperação de informação o que os interpreta e gera triplas RDF. Para essa recuperação de informação são utilizadas expressões regulares contidas nos documentos para encaixar os dados nas triplas. Os dados são armazenados em uma base centralizada e disponibilizados em um endpoint SPARQL para serem acessíveis. Sendo que a ontologia OWL proporciona a verificação automática pelo conjunto de regras de domínio a fim de, por exemplo, não ser ter um mesmo evento ocorrendo em lugares diferentes, minimizando as

inconsistências. Como resultado o trabalho obteve um massa de dados de eventos na USP, e citou que a resolução de conflitos é essencial para não gerar inconsistência. Como conclusão o trabalho cita que a Web Semântica tem um enorme potencial de gerar correlações entre dados na Web, porém ainda tem grandes desafios na extração dos dados da Web Clássica para tornar-lhes semânticos.

3. Método aplicado

3.1. Análise das informações coletadas

Para o desenvolvimento deste trabalho foi realizada análise dos indicadores e fontes através de um mapeamento sistemático dos indicadores. As informações que foram coletadas de cada indicador são descrição, fonte, relatório (descrito por ano e quadrimestre), página(s) no relatório indicado, assim gerando uma fonte concisa para análise dos indicadores para padronização.

3.2. Seleção dos indicadores a serem tratados

A partir deste mapeamento foram selecionados os indicadores a serem estudados. Para este fim foi adotados como primeiro critério de seleção a análise de quais indicadores permanecem presentes nos relatórios do durante o todo ou a maior parte do período de tempo estudado. Posteriormente a viabilidade do estudo desses dados foi estudada através de visitas às fontes para entender a disponibilidade e conformidade dos dados com o relatório. Neste processo foram acrescentadas no mapeamento comentários sobre o uso ou não dos dados além de terem sido inseridos os links das bases de dados das fontes que foram localizados no processo.

3.3. Criação de schemas (Modelagem)

Selecionado os indicadores foi possível realizar a padronização deles com base no padrão apresentado pela W3C citado anteriormente. Para isso foi criada uma modelagem no modelo RDF dos dados como apresentada no referencial teórico, gerando triplas que tem um formato definido em sujeito, predicado e objeto, dando um sentido semântico às relações dos dados. Além disso foi criada a ontologia no padrão OWL, que tem por objetivo gerar um conjunto de restrições que fazem com que os dados sejam validados e tenham representação com sentido similar ao mundo real, dando coesão aos dados.

3.4. Instanciamento dos dados em esquemas gerados

Após a criação do schema é necessário inserir os dados constantes nas fontes para que o seu uso seja realizado na fase de testes. Com isso nessa etapa

da pesquisa foi realizado o processo de instanciamento os dados no modelo gerado, com isso os dados puderam ser acessados em *endpoint* SPARQL.

3.5. Testes e resultados

Com os dados instanciado foi feita a análise dos resultados obtidos, foram realizadas consultas no *endpoint* e com a massa de dados obtidas foram gerados tabelas e um gráfico. Para mostrar como os dados obtidos podem gerar estruturas semelhantes às dos relatórios e também novas visualizações.

3.6. Revisão do trabalho e conclusões

Por fim foi realizada a revisão do trabalho, ajustando de acordo com os resultados e geradas as conclusões do mesmo, com informações como viabilidade e dificuldades no desenvolvimento.

Ainda foram relatados trabalhos futuros, com bases nas proposições apresentadas neste trabalho, com isso foi obtido conhecimento necessário para realizar ajustes nos modelos gerados e construção de novos em trabalhos futuros. Com essa base poderão ser novos experimentos, ficando assim mais claro quais pontos ainda precisam ser melhorados no desenvolvimento de ontologias e das triplas para esta base. E sendo possível o melhor aproveitamento dos dados foram estudados no processo de desenvolvimento dessa monografia.

3.7. Diagrama

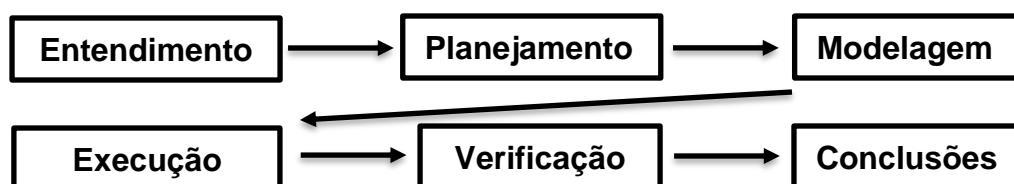


Figura 1: Diagrama do método

A figura 1 apresenta um diagrama com as fases do método. A fase de entendimento representada pela análise das informações coletadas, o planejamento corresponde a seleção dos indicadores, em relação a modelagem está a criação dos esquemas, na execução é realizado o instanciamento dos dados, para a fase de verificação serão realizados os testes e resultados e por fim a fase de conclusões é representada pela revisão do trabalho e conclusões.

4. Recursos de Software e Hardware

Para o desenvolvimento do trabalho foi utilizado um notebook com as seguintes especificações 8GB de memória ram, processador Intel(R) Core(TM) i5-3317U@1,7GHz e sistema operacional Windows 10, que é de posse do pesquisador.

Já para o desenvolvimento do Schema foi utilizado o software Protege, sendo este é um software gratuito e de código aberto amplamente utilizado para desenvolvimento de RDF e OWL, o qual tem integrado o *endpoint* SPARQL.

5. Viabilidade e Cronograma

5.1. Viabilidade Financeira e técnica

Tendo em vista a disponibilidade livre dos software e a que o hardware necessário é de posse do pesquisador ou de baixo custo, no aspecto econômico o projeto é considerado viável. Já no que tange a disponibilidade de dados e bases no estudo prévio se foi verificado que é sim viável realizar o estudo, portanto no que tange às informações o estudo é viável. Com essas informações o trabalho é considerado viável para continuidade do seu desenvolvimento.

5.2. Cronograma de atividades

Id	Nome da tarefa	Dias	Início	Término	Horas	Pred.
	Total Trabalho	290	18/08/2018	03/06/2019	350	
	Definições	247	18/08/2018	21/04/2019	290	
1	Estudo inicial dos relatórios de dados	18	18/08/2018	04/09/2018	40	
2	Referencial teórico / Levantamento Bibliográfico	68	05/09/2018	11/11/2018	120	
3	Definição de indicadores a serem tratados na primeira fase	39	05/09/2018	13/10/2018	40	1
4	Definição da método a ser aplicado	62	05/09/2018	05/11/2018	20	1
5	Planejamento e viabilidade	14	23/10/2018	05/11/2018	10	1
6	Revisão das definições / Enriquecimento levantamento bibliográfico	161	12/11/2018	21/04/2019	60	2
	Desenvolvimento	42	22/04/2019	02/06/2019	45	1,6
7	Relato do desenvolvimento	42	22/04/2019	02/06/2019	30	3,4,5
8	Criação de schemas para indicadores selecionados	14	22/04/2019	05/05/2019	10	3,4,5
9	Instanciamento dos dados em esquemas gerados na primeira fase	10	06/05/2019	15/05/2019	20	6
10	Testes e resultados	17	16/05/2019	01/06/2019	15	7
	Conclusões	9	26/05/2019	03/06/2019	15	10
11	Viabilidade do uso de dados abertos para o caso apresentado	9	26/05/2019	03/06/2019	5	10
12	Dificuldades no desenvolvimento do trabalho	9	26/05/2019	03/06/2019	5	11

13	Trabalhos Futuros	9	26/05/2019	03/06/2019	5	12
----	-------------------	---	------------	------------	---	----

Tabela 1 - Cronograma de atividades realizadas no TCC Fonte: Autoria própria

A tabela 1, detalha o cronograma de atividades a serem realizadas neste projeto. A fase de definição engloba o entendimento do problema, planejamento de estudo e referencial bibliográfico necessário para desenvolvimento da pesquisa. A fase de desenvolvimento engloba a criação e utilização de modelos para os dados analisados. Por fim a conclusão engloba as considerações sobre a pesquisa, tratando de viabilidade do uso do modelo, dificuldades encontradas na utilização do mesmo e trabalhos futuros a serem realizados sobre o tema.

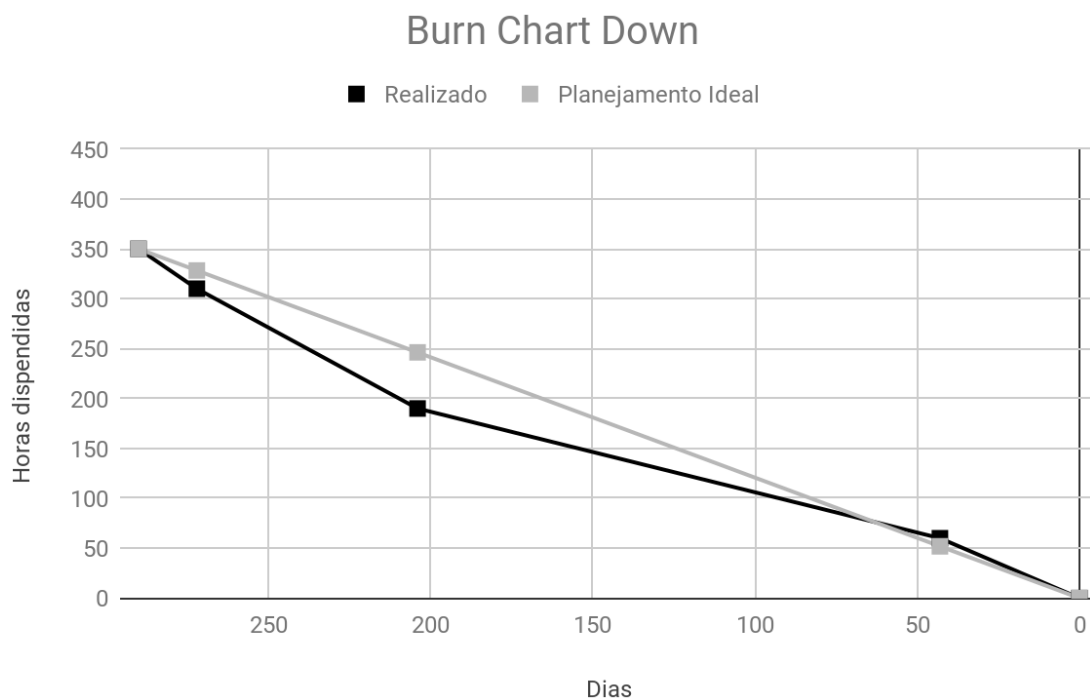


Figura 2: Burn Chart Down Fonte: Autoria Própria

6. Análise da fonte de dados atual

Para construção do schema relatado neste trabalho foram selecionados os indicadores “Total de Leitos SUS Curitiba” e “Rede Física de Serviços no SUS Curitiba”. Eles estão disponíveis na base de dados do datasus, na área que compete ao cadastros nacional de estabelecimentos de saúde (CNES), na seção de rede assistencial. Os indicadores se encontram em sub-seções diferentes, que estão disponíveis respectivamente para os indicadores na seção de recursos físicos⁷ e na seção de estabelecimentos⁸. Foi observado que os registros estão divididos em sub-seções e contam com características específicas. Em ambos os casos existem nas sub-seções opções de tipos de dados que podem ser obtidos conforme o indicador que queira ser pesquisado e essas opções possuem linhas e colunas que podem ser obtidas. No caso do primeiro indicador pesquisado foi necessário obter informações utilizando as opções “Hospitalar - Leitos Internação”, selecionando linha “Município” e coluna “Ano/mês compet.” gestão e “Hospitalar - Leitos Complementares” foi selecionado como linha “Ano/mês compet.” e coluna “Leito Complementar”. Já para o segundo todas as informações foram encontradas na opção “Tipos de Estabelecimentos”, utilizando linha “Tipos de Estabelecimento” e coluna “Tipo de Gestão” sendo necessário fazer uma filtragem por mês requerido. Para obtenção dos dados é necessário escolher o estado do paraná e em seguida realizar filtragem por período mês a mês e escolher o município “410690 Curitiba” para se obter os dados necessários na consulta.

Após as consultas à base do DataSUS foi possível entender que os registros são exibidos de forma agrupada, com características quantitativas o que facilita o uso em relatórios, porém com isso não são informadas identificação características específicas de cada estabelecimento na base. Com isso a base modelada teve o mesmo enfoque de macro indicadores, assim como nos relatórios de saúde. Além disso como a base se encontra distribuída mês a mês e deve ser assim mantida essa característica, com dados instanciados para cada mês.

⁷ Acessado em 10/12/2018

<http://www2.datasus.gov.br/DATASUS/index.php?area=0204&id=11665&VObj=http://tabnet.datasus.gov.br/cgi/defthtm.exe?cnes/cnv/leiint>

⁸ Acessado em 01/06/2019

<http://www2.datasus.gov.br/DATASUS/index.php?area=0204&id=6906&VObj=http://tabnet.datasus.gov.br/cgi/defthtm.exe?cnes/cnv/estab>

7. Construção do schema RDF/OWL

Sendo assim da forma que se encontrava distribuída a base de dados e que os dados se encontravam, para que seja feita a construção foram utilizados registros para cada mês relatado no DataSus. Tendo este registro características semelhantes aos dados utilizados, como o registro é mensal e os relatórios são quadrimestrais foram instanciados na base de triplas para esse experimento registros sempre de meses de abril, agosto e dezembro. Nas tabelas 2 a 5 temos exemplos dos dados apresentados nos relatórios da secretaria de saúde de Curitiba.

Rede Física de Serviços no SUS Curitiba - 2017				
Tipo de Estabelecimento	Total	Tipo de Gestão		
		Municipal	Gestão Mista (Dupla)	Estabelecimento de Gestão Estadual
Centro de Regulação de Serviços de Saúde	01	01		
Central de Regulação Médica das Urgências	01	01		
Centro de Atenção Hemoterapia e ou Hematologia	01			01
Centro de Atenção Psicossocial	13	12	01	
Centro de Saúde/ Unidade de Saúde	110	110 ³		
Clínica Especializada/ Ambulatório de Especialidades	44	43	01	
Consultório isolado	01	01		
Hospital Especializado	08 *	06	02	
Hospital Geral	16 **	10	06	
Laboratório Central de Saúde Pública – LACEN	01			01
Policlínica	18	16	02	
Pronto Atendimento (UPA)	09	09		
Secretaria de Saúde (DS + Central de Vacinas + CSA + SMS)	13	13		
Serviço de Atenção Domiciliar (08 EMAD + 2EMAP)	13	13		
Unidade de Apoio Diagnose e Terapia (SADT isolado)	43	27	16	
Unidade Móvel de Nível Pré-Hospitalar na área de Urgência/ SAMU	28	28		
Telessaúde – NUTES	01	01		
Oficina Ortopédica	01	01		
Central de transplantes	01			01

Fonte: CNES -dados atualizados em 22/01/2018

*Centro Hospitalar de reabilitação – só tem fatura ambulatorial

**hospital Santa Madalena Sofia – não está mais faturando pelo SUS

Tabela 2: Rede física de serviços 2017-3 Fonte: Relatório de saúde de secretária municipal de curitiba 2017-3

Total de Leitos SUS Curitiba cadastrados no CNES		
	3º quadrimestre de 2016	3º quadrimestre de 2017
Leitos Gerais	2.671	2.659
Leitos UTI	321	321
Leitos UCI + isolamento	135	128
Total	3.127	3.108

Fonte: CNES Base local – CCAA e RRQA referente ao 3º quadrimestre 2017
 Dados atualizados 22/01/2018

Tabela 3: Total de leitos SUS 2017-3 Fonte: Relatório de saúde de secretária municipal de Curitiba 2017-3

Rede Física de Serviços no SUS Curitiba - 2018				
Tipo de Estabelecimento	Total	Tipo de Gestão		
		Municipal	Gestão Mista (Dupla)	Estabelecimento de Gestão Estadual
Centro de Regulação de Serviços de Saúde	01	01		
Central de Regulação Médica das Urgências	01	01		
Centro de Atenção Hemoterapia e ou Hematologia	01			01
Centro de Atenção Psicossocial	13	12	01	
Centro de Saúde/ Unidade de Saúde	111	111		
Clínica Especializada/ Ambulatório de Especialidades	47	43	02	02
Consultório isolado	01	01		
Hospital Especializado	08 *	06	02	
Hospital Geral	18 **	11	06	01
Laboratório Central de Saúde Pública - LACEN	01			01
Policlínica	16	14	02	
Pronto Atendimento (UPA)	09	09		
Secretaria de Saúde (DS + Central de Vacinas + CSA + SMS)	13	13		
Serviço de Atenção Domiciliar (10 EMAD + 03 EMAP)	13	13		
Unidade de Apoio Diagnóstico e Terapia (SADT isolado)	50	30	16	04
Unidade Móvel de Nível Pré-Hospitalar na área de Urgência/ SAMU	28	28		
Telessaúde - NUTES	01	01		
Oficina Ortopédica	02	02		
Central de transplantes	01			01

Fonte: CNES - dados atualizados em 09/05/2018

*Centro Hospitalar de reabilitação – só tem fatura ambulatorial **hospital Santa Madalena Sofia – não está mais faturando pelo SUS

Tabela 4: Rede física de serviços 2018-1 Fonte: Relatório de saúde de secretária municipal de Curitiba 2018-1

6.2 Total de Leitos SUS Curitiba cadastrados no CNES

Total de Leitos SUS Curitiba cadastrados no CNES		
Tipo de Leito	1º quadrimestre 2017	1º quadrimestre 2018
Leitos Gerais	2.656	2.664
Leitos UTI	321	320
Leitos UCI + isolamento	135	128
Total	3.112	3.112

Fonte: CNES - dados atualizados em 09/05/2018 e RRQA referente ao 1º quadrimestre 2017

Tabela 5: Total de leitos SUS 2018-1 Fonte: Relatório de saúde de secretária municipal de Curitiba 2018-1

Pode ser então observado nas tabelas 3 e 5 que para modelar os dados contidos dos totais de leitos SUS são necessários dados dos anos anteriores além do quadrimestre do relatório. Esses dados podem ser distribuídos em registros com as seguintes informações:

Forma de obtenção pelo datasus	Linha relatório	Nome propriedade RDF
Hospitalar - Leitos Internação	Leitos Gerais	Gerais
Hospitalar - Leitos Complementares - somar valores ref UTI	Leitos UTI	Uti
Hospitalar - Leitos Complementares - somar valores ref UCI e isolamento	Leitos UCI + isolamento	Ucisol

Tabela 6: Tipos de gestão Fonte: Própria

Já no caso do indicador de “Rede Física de Serviços”, podemos observar nas tabelas 2 e 4, que para reconstruir a tabela será necessário em registro mensal inserir um ou mais tipos de gestão. Os tipos de gestão serão caracterizados pelas propriedades RDF que constam na tabela 7:

Tipos de gestão (Objetos)	Nome propriedade RDF
Municipal	Municipal
Gestão Mista (Dupla)	Mista
Gestão estadual	Estadual

Tabela 7: Tipos de gestão Fonte: Própria

Cada tipo de gestão contém as informações que estão descritas na tabela 8 para poder comportar a massa de dados que é necessária para gerar as tabelas do relatório.

Linha DataSUS	Linha relatório	Nome propriedade RDF
CENTRAL DE REGULAÇÃO	Centro de Regulação de Serviços de Saúde	CntrRegSrvSd
CENTRAL DE REGULAÇÃO MÉDICA DAS URGÊNCIAS	Central de Regulação Médica das Urgências	CntrRegMedUr
CENTRO DE ATENÇÃO HEMOTERÁPICA E/OU HEMATOLÓGICA	Centro de Atenção Hemoterapia e ou Hematologia	CntrAtcHem
CENTRO DE ATENÇÃO PSICOSSOCIAL CAPS	Centro de Atenção Psicossocial	CntrAtcPsic
CENTRO DE SAÚDE/UNIDADE BÁSICA DE SAÚDE	Centro de Saúde/ Unidade de Saúde	CntrSaude
CLÍNICA ESPECIALIZADA/AMBULATÓRIO ESPECIALIZADO	Clínica Especializada/ Ambulatório de Especialidades	ClincEspc
UNIDADE MISTA	Consultório isolado	CnsltIso
HOSPITAL ESPECIALIZADO	Hospital Especializado	HospEsp
HOSPITAL GERAL	Hospital Geral	HospGeral
LABORATÓRIO DE SAÚDE PÚBLICA	Laboratório Central de Saúde Pública – LACEN	Lacen
POLICLÍNICA	Policlínica	Policli
PRONTO ATENDIMENTO	Pronto Atendimento (UPA)	Upa
SECRETARIA DE SAÚDE	Secretaria de Saúde (DS + Central de Vacinas + CSA + SMS)	Sms
SERVIÇO DE ATENÇÃO DOMICILIAR ISOLADO (HOME CARE)	Serviço de Atenção Domiciliar (10 EMAD + 03 EMAP)	ServAtDom
UNIDADE DE SERVIÇO DE APOIO DE DIAGNOSE E TERAPIA	Unidade de Apoio Diagnose e Terapia (SADT isolado)	SadtIso
UNIDADE MÓVEL DE NÍVEL PRÉ-HOSPITALAR na área de Urgência/EMERGÊNCIA	Unidade Móvel de Nível Pré-Hospitalar na área de Urgência (SAMU)	Samu
TELESAÚDE	Telessaúde – NUTES	Nutes
OFICINA ORTOPÉDICA	Oficina Ortopédica	OfOrto
CENTRAL DE NOTIF. CAPTAÇÃO E DISTR. ÓRGÃOS ESTADUA	Central de transplantes	CnrtTransp

Tabela 8: Tipo de estabelecimentos Fonte: Autoria própria

8. Resultados obtidos

Com base na análise e na geração do uso *schema* foi realizado um experimento focado em testar a eficiência e eficácia no uso desse tipo de modelo da base de dados de forma preliminar. Para este fim foram gerados os registros suficientes recriar as tabelas constantes relatórios de um período 2017/2 até 2018/1, sendo o instanciamento de dos dados feito de forma manual. Já que os dados de leitos sus, necessitam do ano anterior, foram inseridos só com esses dados os registro de:

- Agosto/2016;
- Dezembro/2016.
- Abril/2017;

E com os dois indicadores foram criados registros para:

- Agosto/2017;
- Dezembro/2017;
- Abril/2018.

Para comparar com os dados do levantamento foram geradas pelo modelo através de consultas no modelo SPARQL, a ontologia criada foi nomeada como “cnes” para representar os dados captados do CNES DataSUS. A primeira consulta que gerou massa de dados para o indicador de leitos SUS foi:

Select

?mes ?leitosGerais ?Uti ?Ucilsol

where {

?mes a cnes:MesRegistro . ?mes cnes:Gerais ?leitosGerais . ?mes cnes:Ucilsol ?Ucilsol . ?mes cnes:Uti ?Uti }

A primeira vista se a consulta se assemelha com as consultas SQL tradicionais, porém em uma visão mais profunda se percebe que não são usadas tabelas limitadas por condições dadas e sim clausulas que tem sua massa de dados definidos por condições definidas com as triplas RDF. Com o uso da massa de dados produzida pela consulta e após algum tratamento, foi gerada a tabela 9 com estrutura similar às geradas para nos relatórios.

Competência	Gerais	UTI	UCI e Isolamento
2016.2	2745	316	136
2016.3	2745	321	136
2017.2	2734	321	129
2017.3	2734	321	136
2017.3	2731	321	129
2018.1	2736	320	129

Tabela 9: Tabela Leitos SUS

O tratamento que foi realizado nos dados manualmente através de uma planilha poderia ser facilmente resultado de uma rotina em uma aplicação com acesso ao endpoint SPARQL, para gerar automaticamente esses dados.

Para gerar o segundo indicador foi utilizada a seguinte consulta:

Select

```
?lbmes ?lbtpgest ?CntrRegSrvSd ?CntrRegMedUr ?CntrAtcHem
?CntrAtcPsic ?CntrSaude ?ClnclEspc ?CnsltIso ?HospEsp ?HospGeral ?Lacen
?Policli ?Upa ?Sms ?ServAtDom ?Sadtlso ?Samu ?Nutes ?OfOrto ?CnrtTransp
Where {
    ?mes a cnes:MesRegistro . ?mes rdfs:label ?lbmes . ?tpgest
cnes:pertenceMes ?mes . ?tpgest rdfs:label ?lbtpgest . ?tpgest
cnes:CntrRegSrvSd ?CntrRegSrvSd . ?tpgest cnes:CntrRegMedUr
?CntrRegMedUr . ?tpgest cnes:CntrAtcHem ?CntrAtcHem . ?tpgest
cnes:CntrAtcPsic ?CntrAtcPsic . ?tpgest cnes:CntrSaude ?CntrSaude . ?tpgest
cnes:ClnclEspc ?ClnclEspc . ?tpgest cnes:CnsltIso ?CnsltIso . ?tpgest
cnes:HospEsp ?HospEsp . ?tpgest cnes:HospGeral ?HospGeral . ?tpgest
cnes:Lacen ?Lacen . ?tpgest cnes:Policli ?Policli . ?tpgest cnes:Upa ?Upa .
?tpgest cnes:Sms ?Sms . ?tpgest cnes:ServAtDom ?ServAtDom . ?tpgest
cnes:Sadtlso ?Sadtlso . ?tpgest cnes:Samu ?Samu . ?tpgest cnes:Nutes ?Nutes .
?tpgest cnes:OfOrto ?OfOrto . ?tpgest cnes:CnrtTransp ?CnrtTransp}
```

Como pode ser observado essa consulta por conter muitos dados ficou relativamente extensa, isso se deve ao fato de que cada dado precisou ser explicitamente explicado pela tripla rdf. Possivelmente a linguagem possua recursos que facilite a consulta, mas não está tão madura e bem documentada como o SQL tradicional.

Com a massa de dados da consulta descrita foram construídas tabelas 10 e 11 e a figura 3, com intuito de ilustrar alguns dos artefatos que podem ser gerados utilizando a base de dados. A tabela 10 têm a mesma estrutura das tabelas 2 dados similares às mesmas.

Tipo de estabelecimento 2017.3	Estadual	Mista	Municipal	Total
Centro de Regulação de Serviços de Saúde	1	0	1	2
Central de Regulação Médica das Urgências	0	0	1	1
Centro de Atenção Hemoterapia e ou Hematologia	1	0	0	1
Centro de Atenção Psicossocial	0	1	12	13
Centro de Saúde/ Unidade de Saúde	0	0	112	112
Clínica Especializada/ Ambulatório de Especialidades	2	2	369	373
Consultório isolado	0	0	1	1
Hospital Especializado	0	2	16	18
Hospital Geral	1	6	35	42
Laboratório Central de Saúde Pública – LACEN	1	0	0	1
Policlínica	0	3	420	423
Pronto Atendimento (UPA)	0	0	9	9
Secretaria de Saúde (DS + Central de Vacinas + CSA + SMS)	2	0	12	14
Serviço de Atenção Domiciliar (10 EMAD + 03 EMAP)	0	0	19	19
Unidade de Apoio Diagnose e Terapia (SADT isolado)	4	16	469	489
Unidade Móvel de Nível Pré-Hospitalar na área de Urgência (SAMU)	0	0	29	29
Telessaúde – NUTES	0	1	0	1
Oficina Ortopédica	0	0	2	2
Central de transplantes	1	0	0	1

Tabela 10: Tipo de estabelecimento 2017.3 Fonte: CNES DataSUS

Já a tabela 11 tem a mesma similaridade em relação a tabela 4.

Tipo de estabelecimento 2018.1	Estadual	Mista	Municipal	Total
Centro de Regulação de Serviços de Saúde	1	0	1	2
Central de Regulação Médica das Urgências	0	0	1	1
Centro de Atenção Hemoterapia e ou Hematologia	1	0	0	1
Centro de Atenção Psicossocial	0	1	12	13
Centro de Saúde/ Unidade de Saúde	0	0	112	112
Clínica Especializada/ Ambulatório de Especialidades	2	2	369	373
Consultório isolado	0	0	1	1
Hospital Especializado	0	2	15	17
Hospital Geral	1	6	35	42
Laboratório Central de Saúde Pública – LACEN	1	0	0	1
Policlínica	0	3	423	426
Pronto Atendimento (UPA)	0	0	9	9
Secretaria de Saúde (DS + Central de Vacinas + CSA + SMS)	2	0	12	14
Serviço de Atenção Domiciliar (10 EMAD + 03 EMAP)	0	0	20	20
Unidade de Apoio Diagnose e Terapia (SADT isolado)	4	16	474	494
Unidade Móvel de Nível Pré-Hospitalar na área de Urgência (SAMU)	0	0	29	29
Telessaúde – NUTES	0	1	0	1
Oficina Ortopédica	0	0	2	2
Central de transplantes	1	0	0	1

Tabela 11: Tipo de estabelecimento 2018.1 Fonte: CNES DataSUS

Por fim a figura 2 é um o gráfico mostrando uma análise da evolução do número de unidades de saúde levando em conta os três tipos de estabelecimento com maior número absoluto de unidades na cidade de Curitiba, sendo um exemplo de série histórica.

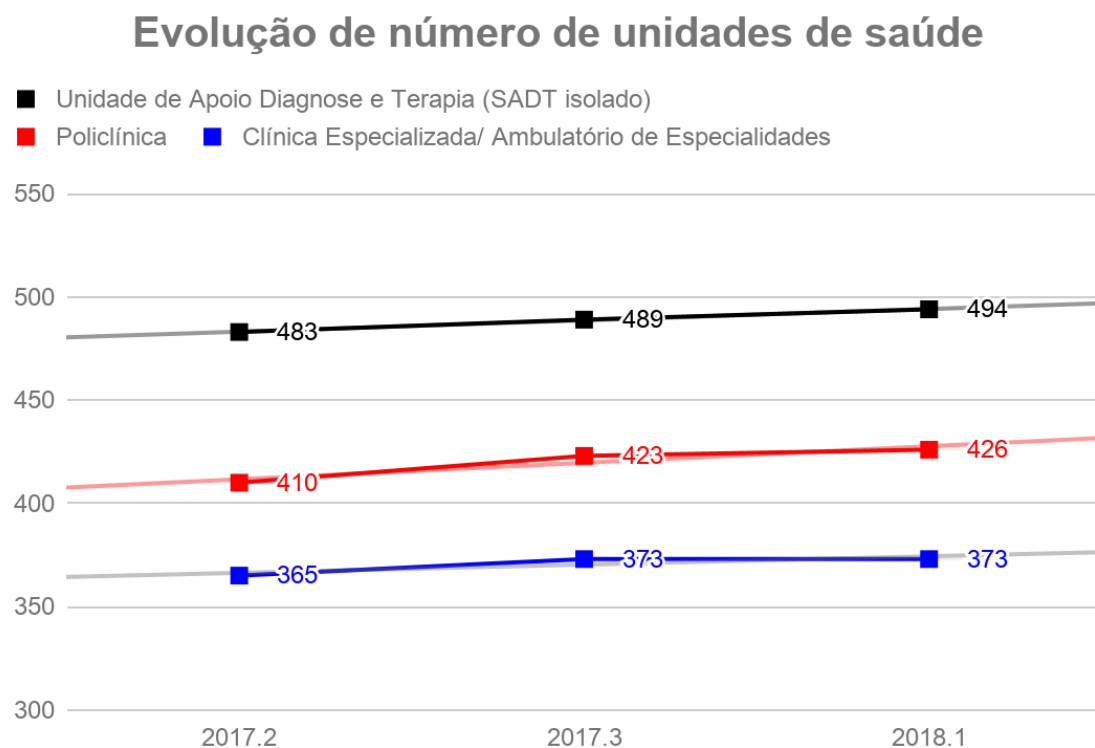


Figura 3: Gráfico de evolução do número de unidades de saúde Fonte: CNES DataSUS

9. Conclusão

Apesar de haver de terem sido verificadas algumas diferenças nos dados dos relatórios para os dados localizados nas bases do DataSUS, possivelmente por mudanças no momento de geração dos relatórios para o momento atual da base do DataSUS. O modelo RDF, OWL, se mostra promissor ao instanciamento dos dados, porém requer um grande esforço e teste para ser corretamente utilizado. A construção da ontologia demanda grande trabalho e teste para expressar de forma concreta o cenário real a ser apresentado e o instanciamento dos dados requer uma grande quantidade de estudo da base de dados apresentada para ter uma consistência significativa entre o modelo e o mundo real. O benefício de as relações serem geradas por propriedades que se assemelham a linguagem humana ao invés não apenas por chaves de relação contidas em tabelas, se mostra muito interessante para o uso dos dados. Com isso se pode observar que gerar dados com estas propriedades acaba de certa forma se assemelhando a classes de uma linguagem de programação.

Com esta análise foi possível entender que os dados estão disponíveis, apesar de não haver facilidade para realizar consultas, para geração de grandes massas de dados, por isso não são acessíveis. Com isso uma forma para auxiliar na acessibilidade dos dados é a construção de um recurso computacional para realizar de forma automática a migração dos dados. Este modelo exige grande complexidade da geração do novo modelo dos dados, sendo assim necessário maior tempo de estudo. Atendendo a essas questões é possível gerar um recurso computacional condizente com as demandas levantadas neste trabalho, seguindo assim em aberto a questão da viabilidade do emprego da técnica dessa implementação, a qual será motivo de novos trabalhos.

9.1. Dificuldades encontradas

A análise dos relatórios se mostra bem trabalhosa, além de não existir um canal que proporcione de forma simples a comunicação com a equipe que gera os relatórios, acabando deixando o trabalho mais raso e menos condizente com as reais demandas da prefeitura. Em próximas etapas apesar disso, pretende-se

estabelecer este tipo de contato e com isso verificar as demandas, contribuições e viabilidade de implementação deste tipo de modelo.

Outra dificuldade é que por não se tratar de um modelo maduro de dados, e ser relativamente pouco utilizado pela indústria, estando no âmbito acadêmico como estado da arte, há alguma dificuldade de localizar documentações e tutoriais sobre as implementações e linguagens utilizadas. Com isso o trabalho se torna mais empírico em algumas situações encontradas.

9.2. Trabalhos futuros

Visto este cenário em trabalhos futuros serão abordados os desafios apresentados neste contexto para realizar a automatização e modelagem necessárias para que as ontologias sejam corretamente estabelecidas e o instanciamento ocorra de forma correta. Para os próximos trabalhos serão desenvolvidas um algoritmo que busque de forma automática os dados e instância ele nas bases de dados abertos conectados, e criados mais modelos para instanciar mais dados, além de contemplar a maior cobertura possível de indicadores nos relatórios e de dados abertos disponíveis para carga, no que tange ao domínio de fontes do relatório.

Além disso pretende-se realizar questionários e testes a serem distribuídos a funcionários da prefeitura que já utilizam as bases para relatórios e outras pessoas que possam ser público alvo dos dados gerados pelas bases. Com isso visa-se uma visão consistente da utilidade desse tipo de abordagem e sua contribuição para o melhor utilização dos dados abertos.

Outra frente a ser atacada para que o potencial do modelo RDF e OWL seja utilizado de forma plena é a integração dessa base com outras similares no mundo, com isso possibilitando comparações em escala muito maior.

Este projeto buscou trazer uma experiência de uso dos conceitos de dados abertos conectados e o problema encontrado nos relatórios da secretaria de saúde pertencente à prefeitura municipal de Curitiba. Nele foram abordadas questões estruturais dos relatórios gerados e proposto um modelo para mitigação do problema. Com isso visando uma contribuição acadêmica com o acesso à informação, tanto pela gestão pública como pela população em geral. Alinhando-se assim com a missão da UTFPR, que visa uma interação ética e produtiva com a sociedade a fim do desenvolvimento tecnológico.

10. Referências

SAÚDE, Secretaria Municipal. **Relatórios de Gestão**. Disponível em: <<http://www.saude.curitiba.pr.gov.br/a-secretaria/relatorios-de-gestao.html>>. Acesso em: 04 set. 2018.

LACERDA, Josimari Telino et al. Avaliação da gestão para o planejamento em saúde em municípios catarinenses. **Ciência & Saúde Coletiva**, Rio de Janeiro - RJ, v. 17, n. 4, p. 851-859, abr. 2012. Disponível em: <<https://www.scielo.org/pdf/csc/2012.v17n4/851-859>>. Acesso em: 18 set. 2018.

Ackoff, R. L. *From Data to Wisdom*, **Journal of Applied Systems Analysis**, Volume 16, 1989 p 3-9.

SETZER, Valdemar W. Dado, Informação, Conhecimento e Competência. **DataGramZero**, São Paulo, p. 1-2, dez. 1999. Disponível em: <http://www.academia.edu/download/44270487/ART_2_GEST.pdf>. Acesso em: 10 out. 2018.

ISOTANI, Seiji; BITTENCOURT, Ig Ibert. **Dados Abertos Conectados**. São Paulo: Novatec, 2015. 176 p. Disponível em: <<http://ceweb.br/livros/dados-abertos-conectados/>>. Acesso em: 11 out. 2018.

VICTORINO, Marcio de Carvalho et al. Uma proposta de ecossistema de big data para a análise de dados abertos governamentais conectados. **Informação & Sociedade: Estudos**, João Pessoa - PB, v. 27, n. 1, p. 225-242, abr. 2017. Disponível em: <<http://www.periodicos.ufpb.br/ojs2/index.php/ies/article/view/29299/17505>>. Acesso em: 13 out. 2018.

MILLER, E. An Introduction to the Resource Description Framework. **D-Lib Magazine**, v. 4, n. 5, May, 1998. Disponível em: <<http://www.dlib.org/dlib/may98/miller/05miller.html>>. Acesso em: 23 out. 2018.

W3C. OWL 2 web ontology language: document overview. 27 Oct. 2009. Disponível em: <<https://www.w3.org/TR/owl2-overview/>>. Acesso em: 23 out. 2018.

ALENCAR, André et al. Publicação e consumo de dados abertos conectados acadêmicos. **Revista Principia** - Divulgação Científica e Tecnológica do IFPB,

[S.I.], n. 42, p. 136-145, ago. 2018. ISSN 2447-9187. Disponível em: <<http://periodicos.ifpb.edu.br/index.php/principia/article/view/1988>>. Acesso em: 04 nov. 2018.

PIERIN, Felipe L.; SICHMAN, Jaime S. IntegraWeb: uma arquitetura baseada em mapeamentos semânticos. In: ONTOBRAS, 2018, São Paulo - SP. **XI Seminar on Ontology Research in Brazil ...** São Paulo - SP: CEUR-WS, 2018. p. 116-127. v. 2228.

Disponível em: <<http://ceur-ws.org/Vol-2228/paper7.pdf>>.

Acesso em: 04 nov. 2018.

NONAKA, Ikujiro; TAKEUCHI, Hirotaka – **Criação do conhecimento na empresa**. Rio de Janeiro, Elsevier, 1997

LOPES, Gabriel; VIDAL, Vania; OLIVEIRA, Mauro. Construção de Linked Data Mashup para Integração de Dados da Saúde Pública. 2016, Salvador - BA. **31º Simpósio Brasileiro de Banco de Dados**. Salvador - BA: [s.n.], 2016. p. 145-150.

Disponível em: <http://sbbd2016.fpc.ufba.br/sbbd2016/sbbd/sbbd_s008.pdf>.

Acesso em: 18 nov. 2018.

ZAIDAN, F. H.; BAX, M. P. Linked Open Data como forma de agregar valor às informações clínicas. **AtoZ: novas práticas em informação e conhecimento**, Curitiba, v. 2, n. 1, p. 44-59, jan./jun. 2013. Disponível em:

<<https://revistas.ufpr.br/atoz/article/view/41319/25258>>. Acesso em: 19 nov. 2018.

W3C. Resource Description Framework (RDF). 2014. Disponível em:

<<https://www.w3.org/2001/sw/wiki/RDF>>. Acesso em: 20 nov. 2018.

DAVENPORT, Thomas H.; PRUSAK, Laurence. O que queremos dizer com conhecimento? In: **Conhecimento empresarial: como as organizações gerenciam o seu capital intelectual**. Rio de Janeiro: Elsevier, 2003. Cap. 1, p. 1-28.

TANAKA, Oswaldo Yoshimi; TAMAKI, Edson Mamoru. O papel da avaliação para a tomada de decisão na gestão de serviços de saúde. **Ciência & Saúde Coletiva**, Rio de Janeiro, v. 17, n. 4, p. 821-828, abr. 2012.

Disponível em: <<https://www.scielo.org/pdf/csc/2012.v17n4/821-828/pt>>.

Acesso em: 22 nov. 2018.

CARRENO, Ioná et al. Análise da utilização das informações do Sistema de Informação de Atenção Básica (SIAB): uma revisão integrativa. **Ciência & Saúde Coletiva**, Lajeado - RS, v. 20, n. 3, p. 947-956, mar. 2015. Disponível em:

<<https://www.scielo.org/pdf/csc/2015.v20n3/947-956/pt>>. Acesso em: 23 nov. 2018.

W3C. **Publishing Open Government Data**. [S. l.], 2009. Disponível em: <https://www.w3.org/TR/gov-data/>. Acesso em: 14 abr. 2019.

BIZER, C.; HEATH, T.; BERNERS-LEE, T. **Linked Data - The story so far**. [ed.] Tim Heath, M. Hepp and Christian Bizer. International Journal on Semantic Web and Information System, Special Issue on Linked Data, 2006.

BERNERS-LEE, Tim. **Information Management: A Proposal**. [S. l.], 1990. Disponível em: <https://www.w3.org/History/1989/proposal.html>. Acesso em: 20 abr. 2019.

Warschauer, M. Carlos Szlak (tradução) **Tecnologia e Inclusão Social**. 2006 Editora SENAC, SP, p. 15-27

Lemos, A. (2013). **Cidades inteligentes**. GVexecutivo, v. 12, n. 2, p. 46-49. Disponível em: <http://bibliotecadigital.fgv.br/ojs/index.php/gvexecutivo/article/view/20720>. Acesso em: 20 abr. 2019.

Gaspar, J. V.; Azevedo, I. S. C.; Teixeira, C. S. **Análise do Ranking Connected Smart Cities**. In: Congresso Internacional de Conhecimento e Inovação, 2016. Anais... Ciki, 2016. Disponível em: <http://via.ufsc.br/wp-content/uploads/2016/12/AN%C3%81LISE-DO-RANKING-CONNECTED-SMART-CITIES.pdf>. Acesso em: 21 abr. 2019.

DUARTE, F. **Cidades Inteligentes: inovação tecnológica no meio urbano**. São Paulo, 2005. Disponível em: < <http://www.scielo.br/pdf/spp/v19n1/v19n1a11.pdf> >. Acesso em: 21 abr. 2019.

BERNERS-LEE, Tim. **Linked Data**. [S. l.], 2010. Disponível em: <https://www.w3.org/DesignIssues/LinkedData.html>. Acesso em: 28 maio 2019.