

UNIVERSIDADE FUMEC

Mestrado Profissional em Sistemas de Informação e Gestão do  
Conhecimento

UMA PROPOSTA PARA VISUALIZAÇÃO DE LINKED  
DATA SOBRE ENCHENTES NA BACIA DO RIO DOCE

PATRÍCIA CAROLINA NEVES AZEVEDO

Belo Horizonte  
2014

PATRÍCIA CAROLINA NEVES AZEVEDO

## UMA PROPOSTA PARA VISUALIZAÇÃO DE LINKED DATA SOBRE ENCHENTES NA BACIA DO RIO DOCE

Dissertação apresentada ao curso de Mestrado Profissional em Sistemas de Informação e Gestão do Conhecimento da Faculdade de Ciências Empresariais da Universidade FUMEC, como requisito parcial para obtenção do título de Mestre em Sistemas de Informação e Gestão do Conhecimento.

Linha de Pesquisa: Sistemas de Informação

Orientador: Prof. Dr. Fernando Silva Parreiras

Belo Horizonte  
2014

*Este trabalho é dedicado às minhas avós,  
que são meus exemplos de vida.*

## Agradecimentos

A todos que me incentivaram a realizar este trabalho, em especial meus pais e minha irmã, pelo apoio constante e confiança irrestrita fundamentais à realização dos meus objetivos.

Ao meu namorado, Marcus Souza, que esteve sempre próximo e presente para acompanhar e compartilhar as alegrias e dificuldades diárias desta jornada.

Ao professor Fernando Parreiras, pela orientação e ensinamentos durante a elaboração desta dissertação. Estendo minha gratidão aos professores do Programa de Mestrado em Sistemas de Informação e Gestão do Conhecimento da Universidade FUMEC pelos ensinamentos e ideias que contribuíram para o meu aprimoramento profissional e pessoal.

Aos membros da banca, pela colaboração e análise criteriosa deste trabalho.

A todos os colegas da CPRM, pelo aprendizado sobre a importância da gestão de áreas de risco.

Aos amigos e familiares que, de alguma forma, foram fonte de incentivo e apoio.

*”O único homem que está isento de erros, é aquele que não arrisca acertar.”*

*Albert Einstein*

## RESUMO

Dentre as necessidades da população acerca das enchentes na Bacia do Rio Doce está a transparência das informações de interesse público e a preocupação na capacidade de percepção e compreensão dessas informações. A disponibilização de dados governamentais abertos oferece, para os usuários interessados, a facilidade de combinar e misturar esses dados com o objetivo de agregar valor. Os Sistemas de Informação Geográfica, agregados à tecnologia *Web Semântica*, permitem que a integração de dados e conhecimento, distribuídos por várias fontes heterogêneas, sejam substancialmente mais fáceis e eficientes, potenciando a descoberta e a partilha de novos conhecimentos. Em conjunto com a adoção dessas tecnologias, surge o conceito de *Linked Data*, um conjunto de princípios e técnicas cujo objetivo é interligar dados de fontes distintas e ampliar a rede de dados já existente. O objetivo deste trabalho é elucidar a questão de pesquisa: quais são os conceitos e tecnologias que permitem integrar e disponibilizar os dados relacionados às enchentes na Bacia do Rio Doce? Para isso, utilizando a metodologia experimental criou-se um *dataset* e um protótipo de aplicação capaz de obter dados heterogêneos, relacionados a inundações na Bacia do Rio Doce, de diversas organizações públicas integrá-los e disponibilizá-los para visualização em um Sistema de Informação Geográfica. Outros dados pertinentes foram incluídos com o objetivo de agregar conhecimento ao usuário final. Para tal, os dados foram convertidos para o formato RDF, interligados e visualizados com auxílio de consultas SPARQL. O estudo tem a metodologia experimental, com base em prova de conceito e a Bacia Hidrográfica do Rio Doce é a unidade de análise. Como resultado, obteve-se a implementação das técnicas e princípios da *Web Semântica* em um SIG e a disponibilização, para reuso, de um *dataset* contendo os dados da Bacia do Rio Doce, do ano de 2012, nos padrões RDF. Conclui-se, por meio do experimento, que as tecnologias *Web Semântica* e SIG dão o suporte necessário para a visualização e disponibilização de dados visando à transparência governamental.

**Palavras-chave:** Visualização. Dados Interligados. Sistemas de Informação Geográfica. Enchente. *Web Semântica*. RDF. SPARQL.

## ABSTRACT

Among the needs of the population concerning the floods in Rio Doce basin the transparency of public interest information and the ability of perception and understanding such information is a problem. The availability of open government data provides, to interested users, an easy way to mix and match these data with the purpose of aggregating value. Geographic Information Systems with Semantic Web technology enable an easier and more efficient data integration besides promoting the discovery and sharing of new knowledge. Along with the adoption of these technologies, the study uses the concept of Linked Data, a set of principles and techniques aimed to link data from different sources and expand the data network. The purpose of this study is to elucidate the research question: Which concepts and technologies allow the integration and make available flood related data in Rio Doce Basin? For this, was created a *dataset* and a prototype application which gets heterogeneous data from several public organizations, integrate them and make them available for viewing in a Geographic Information System. Other relevant data were included to add knowledge to users. For such, data were converted to RDF, linked and displayed through SPARQL queries. The methodology of this study is experimental, based on proof of concept, and Rio Doce Basin as the analysis unit. As a result, the implementation of techniques and principles of the Semantic Web in a GIS and availability for reuse of a dataset containing data about Rio Doce basin, of the year 2012, in the RDF standards. We conclude through the experiment that Semantic Web and GIS technologies are applicable to assist visualization and data availability to support government transparency.

**Keywords:** Visualization. Linked Data. Geographic Information System. Flood. Semantic Web. RDF. SPARQL.

## LISTA DE ABREVIATURAS E SIGLAS

AJAX	Asynchronous Javascript and XML
ANA	Agência Nacional de Águas
API	Application Programming Interface
CEMIG	Companhia Energética de Minas Gerais
COPASA	Companhia de Saneamento de Minas Gerais
CPRM	Companhia de Pesquisa de Recursos Minerais
CSV	Comma-Separated Values
DAEE/MG	Departamento de Águas e Energia Elétrica do Estado de Minas Gerais
FTP	File Transfer Protocol
GIS	Geographic Information System
HTML	Hyper Text Markup Language
ICSE	International Conference on Software Engineering
IDH	Índice de Desenvolvimento Humano
IGAM	Instituto Mineiro de Gestão das Águas
INMET	Instituto Nacional de Meteorologia
JDBC	Java Database Connectivity
JSON	JavaScript Object Notation
ODBC	Open Database Connectivity
OWL	Web Ontology Language
PAC	Programa de Aceleração do Crescimento
RDF	Resource Description Framework
RDFS	Resource Description Framework Schema
SI	Sistemas de informação
SIG	Sistemas de Informações Geográficas



SLR	Systematic Literature Review
TIC	Tecnologias de Informação e Comunicação
UNU	Universidade das Nações Unidas
URI	Uniform Resource Description Framework
XML	Extensible Markup Language
W3C	World Wide Web Consortium
WS	Web Semântica

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>13</b>
1.1	Contextualização do Tema	14
1.2	Problema	16
1.3	Justificativa	17
1.4	Objetivos	20
1.4.1	<i>Objetivos Gerais</i>	20
1.4.2	<i>Objetivos Específicos</i>	21
1.5	Estrutura do Trabalho	21
<b>2</b>	<b>REVISÃO SISTEMÁTICA DA LITERATURA</b>	<b>23</b>
2.1	Fundamentação Teórica	23
2.1.1	<i>Sistemas de Informações Geográficas</i>	23
2.1.2	<i>Web Semântica</i>	26
2.1.2.1	<i>Linked Data</i>	29
2.1.2.2	<i>RDF</i>	32
2.1.2.3	<i>SPARQL: Linguagem de Consulta em RDF</i>	35
2.1.3	<i>Visualização de Dados</i>	38
2.1.4	<i>Visualização de Dados Geoespaciais</i>	40
2.1.5	<i>Visualização de Linked Data</i>	42
2.1.6	<i>Autores Citados e Referenciados</i>	45
2.2	Protocolo da Revisão Sistemática da Literatura	46
2.2.1	<i>Planejamento</i>	48
2.2.2	<i>Realização</i>	49
2.2.3	<i>Resultados</i>	51
2.2.3.1	Tipos de Pesquisas	51
2.2.3.2	Análise Temporal	53
2.2.3.3	Trabalhos Relacionados	55
<b>3</b>	<b>METODOLOGIA</b>	<b>57</b>
3.1	Objeto da Pesquisa	59
3.2	Origem dos Dados	60
<b>4</b>	<b>ARCABOUÇO CONCEITUAL</b>	<b>65</b>
4.1	Dados	66
4.2	<i>Dataset</i>	67
4.3	Visualização em um SIG	67

<b>5</b>	<b>IMPLEMENTAÇÃO</b>	<b>68</b>
5.1	Dados	68
5.2	<i>Dataset</i>	74
5.3	Visualização em um SIG	80
<b>6</b>	<b>VALIDAÇÃO</b>	<b>83</b>
6.1	Dados	83
6.2	<i>Dataset</i>	84
6.3	Visualização em um SIG	88
<b>7</b>	<b>CONSIDERAÇÕES FINAIS</b>	<b>92</b>
	Referências	94

## LISTA DE ILUSTRAÇÕES

Figura 1 – Ocorrência dos principais desastres naturais no Brasil entre 1980 e 2010 . . .	18
Figura 2 – Porcentagem de pessoas mortas por tipo de desastre no Brasil entre 1980 e 2010 . . . . .	18
Figura 3 – Visão do SIG: características e relações com a sua visualização . . . . .	24
Figura 4 – Processo de cálculo de mapas de risco utilizando SIG . . . . .	25
Figura 5 – Arquitetura Padrão do W3C para Web Semântica . . . . .	27
Figura 6 – Linked Open Data – Comunidade de dados vinculados . . . . .	31
Figura 7 – Um grafo que descreve o município Governador Valadares . . . . .	34
Figura 8 – RDF/XML descrevendo o município Governador Valadares . . . . .	35
Figura 9 – Representação das instâncias de um domínio . . . . .	37
Figura 10 – O percurso do exército de Napoleão . . . . .	40
Figura 11 – Características dos dados geoespaciais . . . . .	41
Figura 12 – Questões típicas de um SIG sendo respondidas utilizando mapas . . . . .	42
Figura 13 – Visualização de dados interligados . . . . .	44
Figura 14 – Tipos de Pesquisas em Aplicações <i>Linked Data</i> . . . . .	52
Figura 15 – Características das Pesquisas sobre Visualização em SIG e <i>Linked Data</i> . . .	54
Figura 16 – Visão geral da arquitetura da solução . . . . .	66
Figura 17 – Arquitetura da plataforma D2RQ . . . . .	70
Figura 18 – Interface do D2R-Server para consultas no SPARQL Endpoint . . . . .	71
Figura 19 – Interface do D2R-Server - Página Inicial . . . . .	72
Figura 20 – Mapeamento da tabela Cotas . . . . .	76
Figura 21 – Mapeamento das propriedades da tabela Cotas . . . . .	77
Figura 22 – Trecho da representação em RDF/XML da tabela Cotas . . . . .	78
Figura 23 – Representação do <i>dataset</i> em um grafo RDF . . . . .	79
Figura 24 – Visualização do SIG implementado . . . . .	81
Figura 25 – Visualização da consulta Q1 . . . . .	88
Figura 26 – Visualização da consulta Q2 . . . . .	89
Figura 27 – Visualização da consulta Q3 . . . . .	90
Figura 28 – Visualização da consulta Q4 . . . . .	91

## LISTA DE TABELAS

Tabela 1 – Exemplo de Consulta SPARQL . . . . .	38
Tabela 2 – Constructos abordados e autores citados . . . . .	45
Tabela 3 – Strings da pesquisa . . . . .	49
Tabela 4 – Lista das Referências Utilizadas na Revisão Sistemática da Literatura . . . . .	49
Tabela 5 – Fonte, descrição e formato dos dados utilizados no trabalho. . . . .	60
Tabela 6 – Vocabulário Reutilizado . . . . .	74
Tabela 7 – Prefixo dos vocabulários utilizados nas consultas SPARQL. . . . .	75
Tabela 8 – Vocabulário dos Conceitos utilizados no <i>dataset</i> . . . . .	80
Tabela 9 – Classificação dos Dados Disponibilizados . . . . .	83
Tabela 10 – Consulta Q1. . . . .	84
Tabela 11 – Resultado da Consulta Q1. . . . .	85
Tabela 12 – Consulta Q2. . . . .	85
Tabela 13 – Resultado da Consulta Q2. . . . .	85
Tabela 14 – Consulta Q3. . . . .	86
Tabela 15 – Resultado da Consulta Q3. . . . .	86
Tabela 16 – Consulta Q4. . . . .	87
Tabela 17 – Resultado da Consulta Q4. . . . .	87

## 1 INTRODUÇÃO

Os eventos extremos relacionados às Mudanças Climáticas Globais e o crescimento desordenado das cidades, com a ampliação do contingente populacional empobrecido e em precárias condições de territorialização, são um quadro desafiador para a Defesa Civil e aqueles que, sob sua coordenação, lidam com a gestão de desastres no Brasil (VALENCIO; SIENA; MARCHEZINI, 2009).

O governo federal, por meio dos órgãos responsáveis, adota medidas capazes de minimizar os prejuízos causados por cheias nas bacias hidrográficas, como coleta e análise de dados. Porém, apesar da grande quantidade de informações disponíveis, estas estão espalhadas por diversas fontes de dados, em diferentes instituições (agências públicas de variados níveis de governo, empresas privadas, instituições acadêmicas, etc.), bancos de dados, esquemas e formatos heterogêneos.

Alguns dados estão disponibilizados somente em PDF ou em arquivos de imagem digitalizados, indo contra a Lei de Acesso à Informação (Lei nº 12.527) e acarretando retrabalho aos órgãos e entidades que utilizam esses arquivos e são obrigados a recriá-los. A diversidade de formatos e modelos dos dados dificulta a interpretação, integração e reuso. Além disso, não existe uma forma de visualização desses dados para o acompanhamento e compreensão do histórico dos níveis dos rios da bacia do Rio Doce por parte da população. Para isso, o setor da Companhia de Pesquisa de Recursos Minerais (CPRM) responsável pela Bacia do Rio Doce, de maneira desprovida de efeitos práticos, informa as medições por telefone. O interesse social das informações sobre as enchentes não se restringe somente aos usuários e órgãos correlacionados. É recorrente na definição de políticas públicas o acesso às informações, cuja formatação atual não atende aos requisitos para a localização da informação no tempo correto.

Apesar da grande quantidade de informações disponíveis, elas estão espalhadas por diversas fontes de dados, em diferentes instituições (agências públicas de variados níveis de governo, empresas privadas, instituições acadêmicas, etc.), bancos de dados, esquemas e formatos heterogêneos. Essas informações possuem relevância, principalmente em ambientes em que a tomada de decisão se faz necessária, como no planejamento urbano, na gestão de recursos naturais e no gerenciamento de desastres. A diversidade dos formatos, modelos de dados e do seu significado dificulta a interpretação e integração.

Os Sistemas de Informação (SIG) agregados à tecnologia *Web Semântica* permitem a integração de dados distribuídos em diversas fontes heterogêneas e a sua utilização de modo substancialmente mais fácil e eficiente, potencializando a descoberta e a partilha de novos conhecimentos. Além disso, o governo incentiva a publicação de dados ao público, por meio da internet, visando informar a população e apoiar a transparência dos dados governamentais.

Porém, a publicação de dados não estruturados é insuficiente para realizar os objetivos de eficiência, transparência e prestação de contas, já que tais dados não são facilmente encontrados, visualizados e absorvidos como deveriam.

No contexto das enchentes, percebe-se que a visualização, interação e divulgação desses dados são pontos importantes para uma efetiva gestão de desastres. Nesse âmbito, os princípios de *linked data* são um meio para disponibilizar as informações partilhadas na *Web*, ao oferecer vários conjuntos de dados de diversas fontes e temáticas e efetuar ligações. Esses dados utilizam uma família de padrões internacionais e as melhores práticas para a publicação, divulgação e reutilização de dados estruturados.

O presente trabalho tem como objetivo validar conceitos e tecnologias que permitem integrar e disponibilizar dados relacionados às enchentes na Bacia do Rio Doce. Para isso, utilizou-se tecnologias e padrões da *Web* semântica e um SIG para a disponibilização aos usuários interessados.

Assim, ao visualizar os dados integrados da Bacia do Rio Doce, será possível identificar as comunidades que se encontram vulneráveis e desenvolver ações emergenciais e de prevenção, para então, se ter uma gestão de desastre mais eficiente sobre a bacia do Rio Doce.

## 1.1 Contextualização do Tema

O desenvolvimento das Tecnologias de Informação e Comunicação (TICs), nas últimas décadas, proporciona novas possibilidades para a promoção da transparência. Com o auxílio das TICs, como a Internet, a promoção da transparência pode ser potencializada, pois os meios eletrônicos permitem uma maior facilidade de acesso aos dados e informações da Administração Pública. As possibilidades do uso das novas tecnologias pelos governos, para a promoção da transparência, controle social e participação cidadã, são constantemente apresentadas na literatura (PRADO; LOUREIRO, 2006).

Entretanto, mais do que disponibilizar informações, é preciso garantir que elas sejam distribuídas sem restrições de uso e em formato bruto e aberto, sendo passíveis de serem analisadas, processadas e reutilizadas em aplicações digitais desenvolvidas pela sociedade ou pelo próprio governo. Boa parte das informações ainda é disponibilizada em formatos analógicos, sendo de difícil análise e manuseio por *software*, além de, por vezes, ser expostos em padrões proprietários. Um relatório em PDF pode ser elegante quando impresso, mas isso força um analista a aceitar a informação como está, ou a perder muito tempo para realizar referências cruzadas, análises comparativas e estatísticas, desperdiçando e invalidando os esforços de de-

mocratização e cidadania. A iniciativa de dados abertos é capaz de promover a transparência, a colaboração em todo o governo e, ainda, permitir a criação de novos e inovadores serviços com o objetivo de agregar valor e melhorar a qualidade do processo decisório (DING; PERISTERAS; HAUSENBLAS, 2012). A disponibilização de dados governamentais abertos e em formato padronizado promove a transparência das ações do governo e o livre reuso em outras aplicações. E ainda, se esses dados forem "confrontados" com dados de outras fontes, é possível obter novos significados e conhecimento sobre os assuntos envolvidos.

Há um movimento cada vez maior de governos, organizações e pessoas publicando Dados Governamentais Abertos. Ao mesmo tempo, a utilização de dados ligados vem crescendo muito nos últimos anos, sendo fortemente apoiada pelo *World Wide Web Consortium* (W3C). A disponibilização de dados governamentais abertos permite que as informações sejam utilizadas da maneira e conveniência do interessado, de tal forma que elas possam ser misturadas e combinadas para agregar mais valor aos dados (DINIZ, 2010). Para o autor, o objetivo de as informações públicas serem disponibilizadas segundo as regras dos dados abertos é "superar as limitações existentes para que usuários de informações do serviço público possam facilmente encontrar, acessar, entender e utilizar os dados públicos segundo os seus interesses e conveniências"(DINIZ, 2010).

Além disso, a W3C entende que o governo deve incentivar os cidadãos a usar os dados abertos disponíveis pelo governo, ou seja, eles devem ser estimulados a reutilizar os dados conforme as suas necessidades e vontades. Diniz (2010) resume o objetivo desse incentivo: "Não há valor na disponibilização de dados governamentais abertos se a sociedade não tem interesse em reutilizá-los".

Em 2002, Rehbein avalia a importância da eficiência dos órgãos públicos: o setor público, independentemente da esfera, sofre exigência da sociedade para melhorar sua produção de informações, sob o argumento de ampliação da eficiência e da transparência. Também conhecido por sua tradição burocrática e organizado segundo uma estrutura funcional estanque, o setor público encontra-se pressionado a adotar modelos gerenciais de administração pautados pelo subsídio da informação para a eficácia da gestão. A informação é, portanto, pré-condição tanto para o funcionamento quanto para o atingimento dos objetivos de uma organização privada ou pública.

Em se tratando de informação, Rehbein (2002) explica: a informação, como conjunto de dados interpretados, tem como finalidade reproduzir as incertezas sobre algum assunto e permitir ao usuário a busca da solução de suas dificuldades. É desejável que as informações disponibilizadas pelos sistemas permitam a identificação de problemas e forneçam subsídios para avaliação dos impactos das decisões a serem tomadas.

No âmbito governamental, o programa GOVERNO ELETRÔNICO (2013) constitui um conjunto de políticas públicas que priorizam o uso das TICs como forma de democratizar o acesso à informação e incentivam o aperfeiçoamento da qualidade da informação e a partici-



pação cidadã. Este programa (GOVERNO ELETRÔNICO, 2013) brasileiro tem como objetivo a transformação das relações do governo com os cidadãos, empresas e também entre os órgãos do próprio governo, de forma a aprimorar a qualidade dos serviços prestados, promover a interação com empresas e indústrias e fortalecer a participação do cidadão por meio do acesso à informação.

Nesse sentido, o presente trabalho se insere no conjunto de políticas públicas no âmbito do Governo Eletrônico, ao promover o compartilhamento de conhecimentos para a tomada de decisões, o aperfeiçoamento da qualidade da informação e a participação do cidadão na gestão pública.

## 1.2 Problema

Ao analisar o cenário atual dos dados da Bacia do Rio Doce, observa-se que estes não estão abertos nem disponíveis para reuso, não possibilitando ao cidadão a conferência, análise e acompanhamento dos níveis de água da Bacia do Rio Doce.

Atualmente, a Companhia de Pesquisa de Recursos Minerais (CPRM) em parceria com empresas como a Agência Nacional de Águas (ANA) e a Companhia Energética de Minas Gerais (CEMIG), recebe as medições dos sensores instalados ao longo da Bacia do Rio Doce. Esses dados são armazenados, analisados e dispostos em forma de boletins de ocorrência e, dependendo dos resultados, encaminhados aos órgãos competentes, como Defesa Civil, Corpo de Bombeiros, Polícia Militar e Prefeituras. Porém, os órgãos que reutilizam esses dados têm dificuldade na manipulação e muitas vezes refazem grande parte do trabalho em razão da falta de padronização.

Somente esses boletins são disponibilizados na internet, sendo sua linguagem bastante técnica e de difícil entendimento para usuários leigos. No cenário atual, os cidadãos não têm acesso a qualquer tipo de informação sobre o histórico dos níveis de água da bacia do Rio Doce, e o estado atual dos rios desta bacia é informado apenas por telefone.

O trabalho proposto tem o objetivo de atender às atividades de rotina dos atores do setor de atendimento ao usuário, ao sugerir uma solução capaz de exibir informações integradas vindas de vários órgãos e ao disponibilizar novas perspectivas de visão para os gestores e cidadãos interessados, permitindo responder a perguntas como:

(Q1) Das cidades que sofreram com enchentes em janeiro de 2012, quais possuem os menores Índices de Desenvolvimento Humano (IDH)?

(Q2) Quais estações registraram enchente no dia 09/01/2012?

(Q3) Qual a população dos municípios afetados pela enchente no dia 20/01/2012?

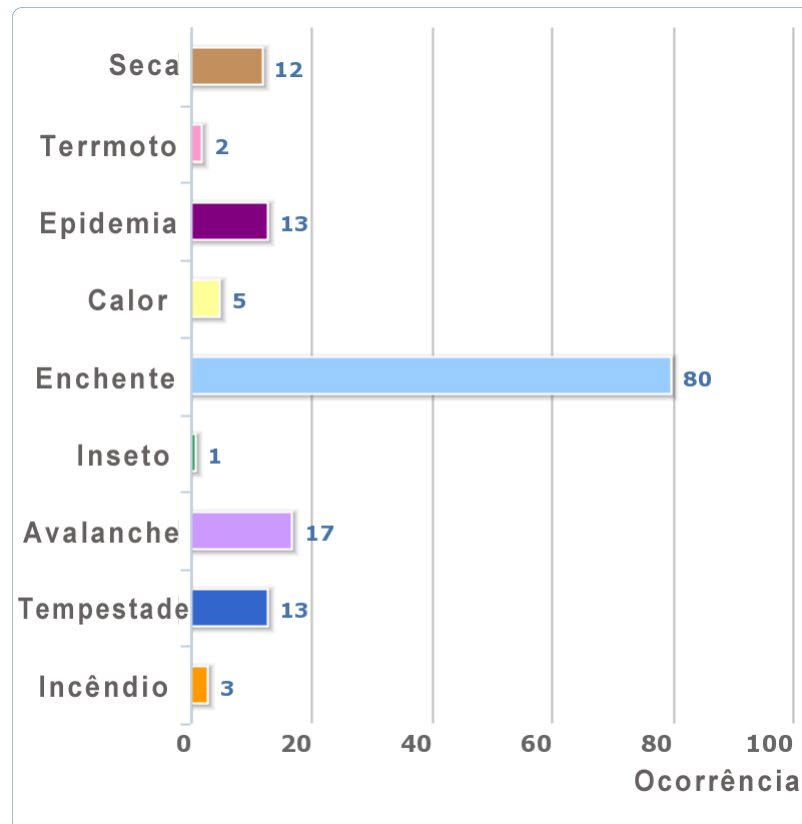
(Q4) Qual a altitude, investimentos do governo e ocorrência de doenças dos municípios afetados por enchente em 2012?

Essas perguntas foram elaboradas juntamente com o setor responsável pelo acesso à informação da CPRM, em consulta aos registros dos atendimentos realizados, sendo as questões mais abordadas pelos agentes do setor. Nesse contexto, pretende-se elucidar a seguinte questão de pesquisa: quais são os conceitos e tecnologias que permitem integrar e disponibilizar os dados relacionados às enchentes na Bacia do Rio Doce?

### **1.3 Justificativa**

Ao analisar os dados sobre desastres naturais no Brasil, no período de 1980 a 2010, disponibilizados pelo principal banco de dados utilizado pela ONU, o International Disaster Database (EM-DAT), observa-se que os principais perigos naturais recorrentes são as enchentes, conforme se observa na Figura 1.

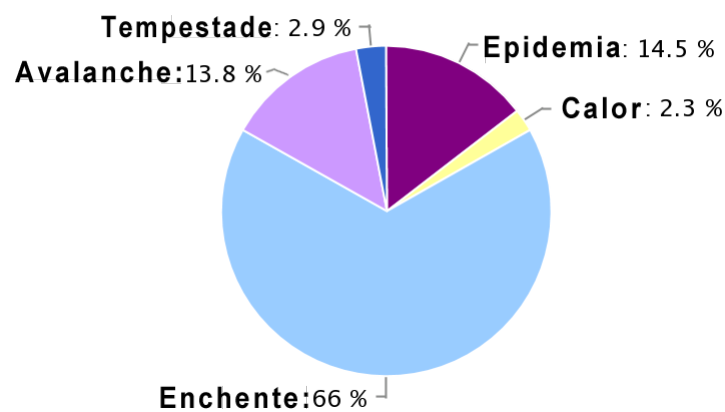
Figura 1 – Ocorrência dos principais desastres naturais no Brasil entre 1980 e 2010



Fonte: Traduzido de EM-DAT (2013).

A porcentagem de pessoas mortas por tipo de desastre, nesse mesmo período, também foi maior na ocorrência de enchentes, conforme a Figura 2. Segundo o EM-DAT, em cada ocorrência de enchente, 170.934 pessoas foram afetadas e 40 morreram.

Figura 2 – Porcentagem de pessoas mortas por tipo de desastre no Brasil entre 1980 e 2010



Fonte: Traduzido de EM-DAT (2013).

Segundo Collins (2004), da Universidade das Nações Unidas (UNU), o Brasil é o sétimo país no ranking mundial sobre número de pessoas atingidas por inundações. O estudo obteve dados de 97 países, entre 1980 e 2000, e relatou que mais de 29 milhões de brasileiros vivem em risco de serem atingidos por inundações. O mesmo estudo mostra ainda que 100 pessoas morrem por serem vítimas de inundações no Brasil.

O trabalho é apoiado pela iniciativa do governo no que se refere à abertura e disseminação dos dados públicos, conforme a Lei de Acesso a Informação Pública (LEI Nº 12.527, 2011): “Todos têm direito a receber dos órgãos públicos informações de seu interesse particular, ou de interesse coletivo ou geral, que serão prestadas no prazo da lei, sob pena de responsabilidade, ressalvadas aquelas cujo sigilo seja imprescindível à segurança da sociedade e do Estado”.

Quanto ao uso das tecnologias, Ruttenberg *et al.* (2007) descrevem a *Web Semântica* como uma das soluções mais propostas atualmente, e mais promissoras, para a integração de dados e conhecimento distribuído por várias fontes heterogêneas. *Web semântica e linked data* oferecem um arcabouço que permite o compartilhamento e integração de dados e sua reutilização. *Resource Description Framework - RDF, Extensible Markup Language - XML, SPARQL* são recursos que permitem que as máquinas entendam os dados com mais facilidade. Segundo Mendes e Cirilo (2001), o SIG proporciona suporte na integração a recursos hídricos ao disponibilizar uma coleção de dados que podem ser utilizados em modelos de simulação e sistemas de suporte à decisão. Por exemplo, vários países já utilizam esses dados em sistemas simulações de enchentes, ao longo de um rio, e seus impactos sobre a região de entorno.

Os Sistemas de Informações Geográficas são bem aceitos como uma tecnologia com a abordagem necessária para a realização de análises com dados espaciais e, assim, para o entendimento e utilização do meio geográfico (SILVA, 1999).

Dados governamentais publicados na *Web*, por si sós, já possuem um grande valor para a população, pois contribuem para uma maior transparência de informações. Mas a disponibilização dessas informações em formatos abertos e acessíveis permite que elas sejam reutilizadas e combinadas com informações de outras fontes para produzir novos significados sobre o desempenho do governo.

Aliar a publicação de dados governamentais abertos às práticas de dados ligados é ainda mais importante, pois proporciona um mecanismo de acesso único e padronizado, permitindo que os dados sejam legíveis por máquinas, facilitando a descoberta e o consumo dos dados, possibilitando que eles sejam ligados a outros conjuntos de dados, aumentando o valor e a utilidade dos dados e abrindo possibilidades de aplicações *Web* mais inteligentes.

Ante a existência de interesse por parte do governo e tendo em vista a demanda na Bacia do Rio Doce, frequentemente atingida por inundações que causam prejuízos econômicos, materiais e perdas humanas, o enfoque será o uso de SIG e *web semântica* como ferramentas

auxiliares na geração de informações sobre a dinâmica do fenômeno na Bacia do Rio Doce.

## 1.4 **Objetivos**

### 1.4.1 *Objetivos Gerais*

O presente trabalho tem como objetivo propor um arcabouço conceitual e validá-lo por meio de tecnologias que permitam a integração e disponibilização de dados relacionados às enchentes na Bacia do Rio Doce. É importante salientar que este trabalho não tem o objetivo de caracterizar todas as variáveis que influenciam na ocorrência e intensidade do fenômeno das enchentes. Essas variáveis já são objeto de estudo em outras áreas. Portanto, o enfoque deste projeto será o uso de SIG e *Web* semântica como ferramentas auxiliares na geração de informações sobre a dinâmica do fenômeno na Bacia do Rio Doce.

### 1.4.2 *Objetivos Específicos*

Os objetivos específicos compreendem:

- (a) Buscar os dados abertos relativos a enchentes na Bacia do Rio Doce;
- (b) Modelar um grafo RDF dos *datasets*;
- (c) Identificar as perguntas de competência;
- (d) Identificar componentes do arcabouço;
- (e) Converter os dados para o formato RDF;
- (f) Delinear a arquitetura conceitual;
- (g) Construir consultas SPARQL;
- (h) Desenvolver o protótipo para a visualização dos dados de forma interativa;
- (i) Validar o arcabouço conceitual.

## 1.5 **Estrutura do Trabalho**

Esta dissertação está estruturada em sete capítulos. No capítulo introdutório, desenvolve-se a contextualização do tema a fim de fornecer informações necessárias como suporte ao conteúdo do trabalho, além de apontar o problema de pesquisa, os objetivos e a justificativa.

O capítulo 2 apresenta a Revisão Sistemática da Literatura utilizada para sustentar a pesquisa do ponto de vista conceitual. A seção 2.1 apresenta conceitos e as linhas de definições sobre Sistemas de Informação Geográficas, seu uso como ferramenta de apoio à decisão e as características e componentes de um SIG, bem como os conceitos e a arquitetura da *Web Semântica*. Esta seção contém as seguintes subseções: *Linked Data*, RDF e SPARQL. Em cada subseção, os respectivos conceitos e definições embasados na literatura.

Ainda na mesma seção, apresenta-se o processo de visualização de dados, suas vantagens e importância para compreensão e transmissão da informação seguida pela visualização de dados geoespaciais e visualização de *linked data*. A primeira discorre sobre as características dos dados geoespaciais e destaca as perguntas que podem ser respondidas com a visualização de dados geoespaciais. A segunda destaca os conceitos e exemplos de visualização de dados ligados. A seção 2.1.6 exibe um quadro com os autores citados e referenciados na revisão da

literatura.

Na seção 2.2, são apresentadas as aplicações existentes de *Linked Data* em SIG com base em uma revisão sistemática da literatura. Destacam-se as etapas dos procedimentos realizados durante a revisão e os resultados obtidos.

O capítulo 3 aborda os procedimentos metodológicos que nortearam o trabalho, seus fundamentos e os argumentos que levaram à sua escolha. Destacam-se as seções 3.1 e 3.2, respectivamente Objeto da Pesquisa e Origem dos Dados. A seção 3.1 caracteriza a unidade de análise e a seção 3.2 discorre sobre a origem e formato dos dados utilizados no trabalho.

O capítulo 4 detalha a arquitetura conceitual por meio de um diagrama e a descrição das camadas que o compõem. Os capítulos 5 e 6 relatam a implementação e a validação da arquitetura proposta, detalhando as camadas em sub-seções. E, finalmente, o capítulo 7 descreve as conclusões obtidas durante a preparação desta dissertação e os trabalhos futuros que podem advir do conteúdo apresentado.

## 2 REVISÃO SISTEMÁTICA DA LITERATURA

### 2.1 Fundamentação Teórica

#### 2.1.1 *Sistemas de Informações Geográficas*

O significado de Sistema de Informação Geográfica (SIG), tradução de GIS (*Geographic Information System*) na literatura, abrange várias formulações, tais como sistemas de informação, sistemas espaciais de informação e sistemas de informação de recursos naturais (ARONOFF, 1989; BERNHARDSEN, 1992; LONGLEY *et al.*, 2005; BONHAM-CARTER, 1994). Estes autores definem SIG como sendo um sistema de *software* computacional com o qual a informação pode ser capturada, armazenada e analisada, combinando dados espaciais de diversas fontes em uma base unificada, empregando estruturas digitais variadas que representam fenômenos espaciais também variados, por meio de uma série de planos de informação que se sobrepõem corretamente em qualquer localização.

Existem três linhas de definições. A primeira, conceituada por Burrough e Mcdonnell (1998) como um conjunto poderoso de ferramentas para coletar, armazenar, recuperar, transformar e exibir dados espaciais partir do mundo real.

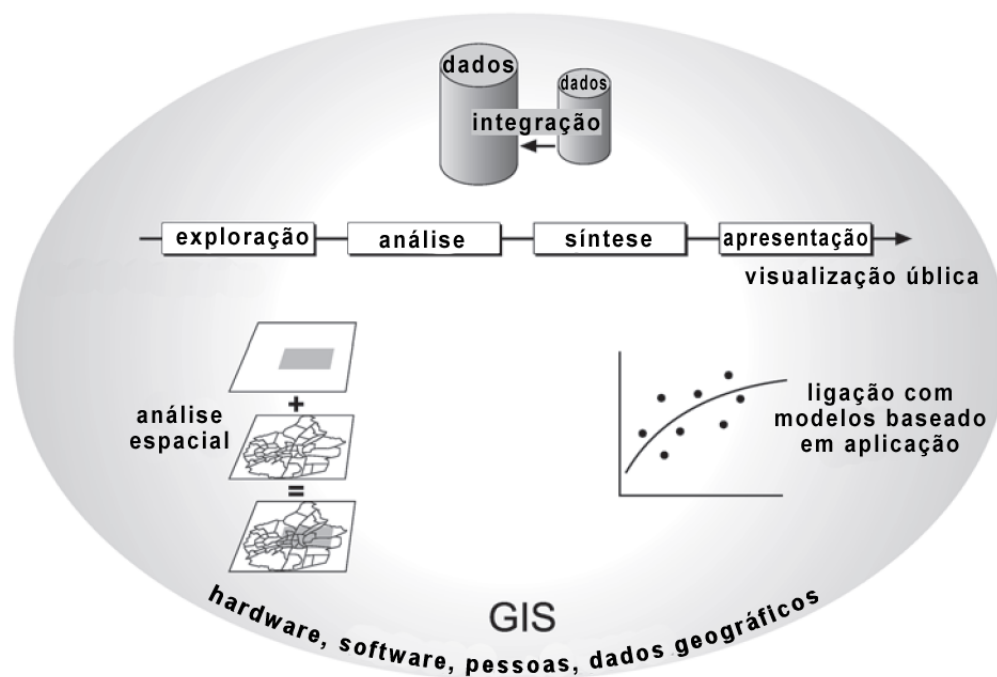
A segunda linha é definida, por Cowen (1988), como sendo um sistema de apoio à decisão que envolve a interação de dados geoespacialmente referenciados em um ambiente de resolução de problemas.

Kraak e Ormeling (2003) lideram a terceira linha e definem GIS como uma derivação e combinação das outras definições: um SIG é um sistema de informação assistido por computador para coletar, armazenar, manipular e exibir dados espaciais no contexto de uma organização, com o objetivo de funcionar como um sistema de apoio à decisão.

Com o objetivo de manipular dados geoespaciais para adquirir valor agregado, um SIG consiste em *software*, *hardware*, dados geográficos e pessoas (ou organização). Esses componentes se comunicam por meio de um conjunto de procedimentos, conforme a Figura 3 (KRAAK; ORMELING, 2003).



Figura 3 – Visão do SIG: características e relações com a sua visualização



Fonte: Traduzido de Kraak e Ormeling (2003).

Na Figura 3, os autores Kraak e Ormeling (2003) resumem um SIG e seus esquemas centrais e ilustram os seguintes componentes: a exploração do problema a se resolver, o potencial da análise geoespacial e a integração de bases de dados geoespaciais. Cada organização vai exigir um SIG com ênfase em um conjunto específico de funções, dependendo da sua área de atuação.

Em geral, as funções necessárias são: entrada de dados e codificação (por exemplo, digitalização e validação de dados), manipulação de dados (por exemplo, conversões geométricas e opções de classificação), recuperação de dados (análise espacial e estatística), apresentação de dados (opções de exibição, principalmente gráficos) e integração de dados.

Ao interpretar os SIGs como ferramenta de apoio à decisão, Eastman *et al.* (1993) exemplificam que a escolha de uma determinada região a ser contemplada com a construção de uma estrada é de caráter político, mas a decisão sobre o melhor traçado e outras características dessa estrada tem caráter técnico. Uma regra de decisão consiste em um procedimento para combinar os critérios selecionados com o objetivo proposto. Com o auxílio de métodos estatísticos é possível atribuir pesos aos critérios envolvidos em uma análise espacial, de modo a ponderar a participação de cada um na análise desejada. Nesse contexto, o SIG constitui-se num instrumento extremamente útil para diminuir a subjetividade no processo de tomada de decisão (HASENACK, 1995).

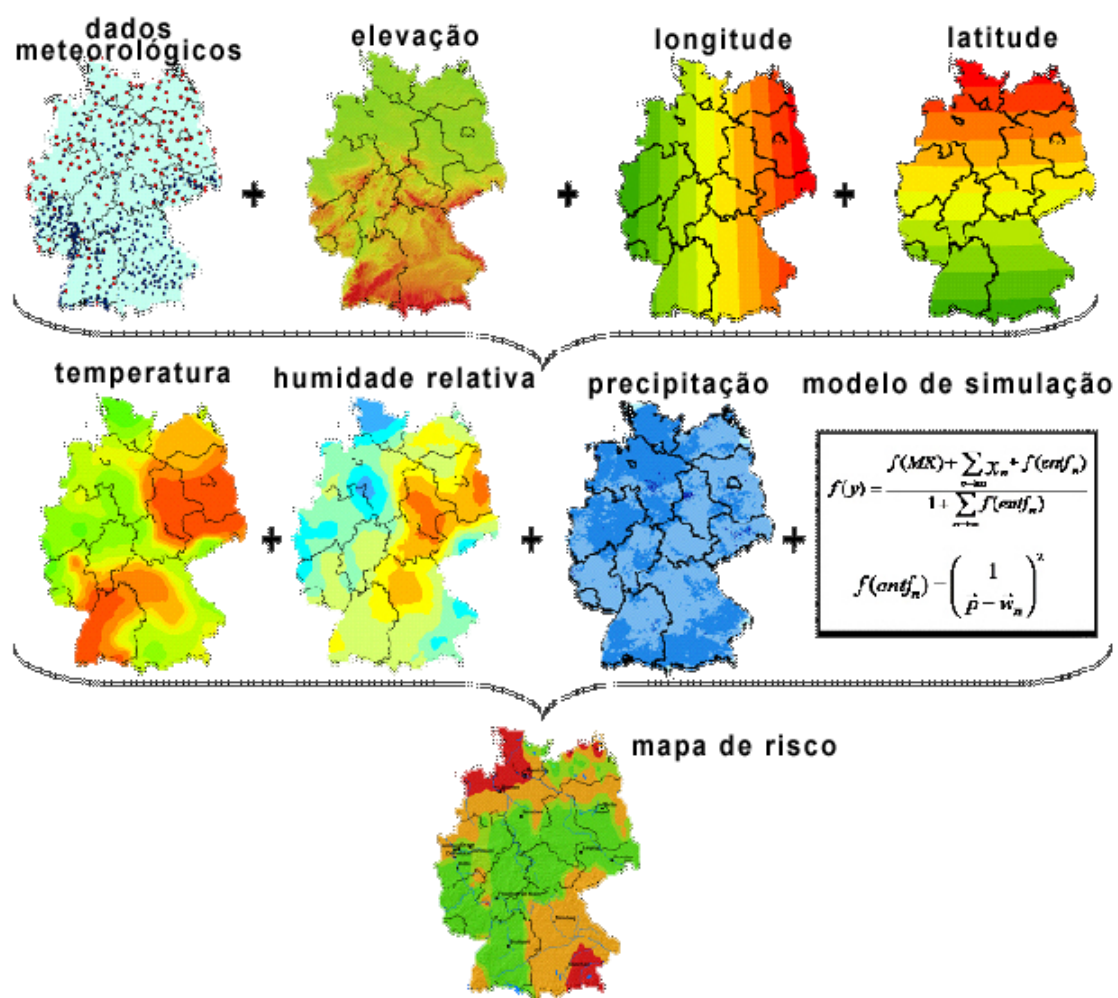
Um sistema de apoio à decisão envolve a integração de dados referenciados espaciais

almente em um ambiente de resolução de problemas (COWEN, 1988). Como observado por Osleeb e Kahn (1999), certas necessidades no apoio à decisão não podem ser efetivamente tratadas sem o uso de GIS.

Cada pessoa, ao se deparar com um problema a ser resolvido, relaciona diferentes valores e seleciona diferentes variáveis, utilizando as informações de maneiras distintas (DENSHAM, 1991). Com a adição de SIG e instrumentos analíticos, tomadores de decisão manipulam os dados em um ambiente de planejamento real (FABER *et al.*, 1997; THOMAS; ROLLER, 1993).

A Figura 4 ilustra um exemplo de uso de vários parâmetros em um SIG, com o objetivo de calcular o mapa de risco de precipitações (ZEUNER, 2008). Na pesquisa dessa imagem, mais de 700 observações foram registradas durante o período de 1994 a 2005 e, para a maioria dos anos, a proporção de previsões corretas atingiu mais de 90% (BENNO *et al.*, 2007). Com o uso de GIS, mapas de riscos espaciais podem ser criados quando os processos espaciais e temporais das aparições regionais do objeto da pesquisa são documentados.

Figura 4 – Processo de cálculo de mapas de risco utilizando SIG



Fonte: Traduzido de Paolo *et al.* (2011).

Na Figura 4, é possível visualizar que a junção de camadas sobrepostas, somada ao uso de fórmulas matemáticas adequadas, resulta no mapa de risco (neste caso, risco de precipitações) de uma região. Um SIG construído para ser utilizado como sistema de apoio à decisão disponibiliza uma interface capaz de explorar várias possibilidades, assim como funções analíticas para gerar soluções viáveis com base em critérios e preferências especificados pelo usuário, que pode repetir o processo de análise quantas vezes desejar.

O principal objetivo dos SIGs é ajudar gestores a tomarem decisões sensatas na administração de recursos naturais ou humanos (MULLER, 1985). O SIG é potencialmente uma ferramenta capaz de ajudar a classe de tomadores de decisão, a qual já está sendo efetivamente utilizada para esses fins, em lugares onde as capacidades avançadas em termos de infraestrutura e pessoal qualificado existem (CARSWELL, 1998).

Conforme Osleeb e Kahn (1999), atualmente o SIG está firmemente estabelecido em empresas comerciais, setores especialistas e no governo de países desenvolvidos. É utilizado em uma grande quantidade de aplicações: de usos tradicionais, como preparação de mapas, a aplicativos que cobrem desde a descoberta de recursos minerais até a avaliação de risco ambiental. O GIS também cresceu na indústria e no governo como um meio de integrar diferentes dados utilizando códigos de endereçamento, mais comumente o código postal. Hart e Dolbear (2013) compreendem que essa integração foi ocasionada principalmente pela detecção de fraudes, verificação de crédito, desenvolvimento do quadro demográfico, conscientização da saúde, entre outros. O crescimento desse mercado é um indicador da importância que a geografia desempenha como um fator comum entre os diferentes conjuntos de dados. Ao visualizar que dados coletados, por uma organização, sobre um lugar especial referem-se ao mesmo lugar que os dados recolhidos por outra organização, estes dados podem ser, então, combinados (HART; DOLBEAR, 2013).

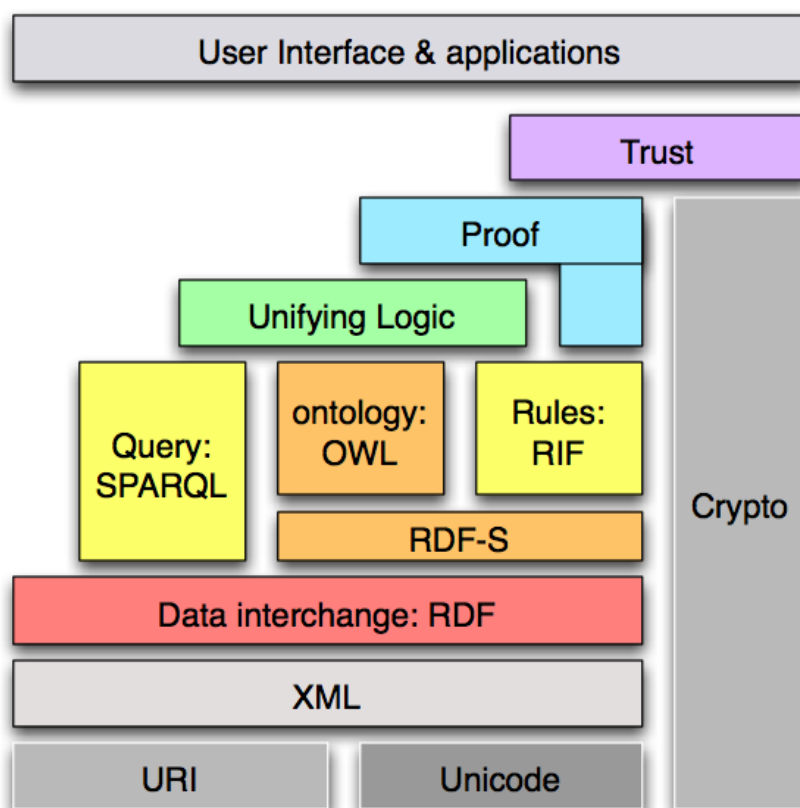
### 2.1.2 *Web Semântica*

Segundo Berners-Lee, Hendler e Lassila (2001), a *Web Semântica* (WS) é uma extensão da *textitWeb* atual, na qual é atribuído à informação um significado bem definido, permitindo uma melhor cooperação entre sistemas computacionais e pessoas. O desenvolvimento de uma *Web* que permita o processamento da informação por humanos e por máquinas possibilitará a resolução de problemas que até agora seriam complexos e muito demorados.

A WS estrutura-se de modo que as páginas *Web* tenham um conteúdo significativo, em que agentes de software ao rastrear página por página podem, facilmente, realizar tarefas sofi-

ticadas para os usuários. (BERNERS-LEE; HENDLER; LASSILA, 2001). Em 2001, Berners-Lee, Hendler e Lassila definiram o conceito de *Web Semântica* e uma possível arquitetura para aplicações sob o mesmo contexto. A arquitetura passou por várias modificações e a sua configuração atual é ilustrada na Figura 5.

Figura 5 – Arquitetura Padrão do W3C para Web Semântica



Fonte: W3C (2011).

A arquitetura é separada em camadas, onde cada uma possui uma determinada função. A alta interoperabilidade entre as camadas é a principal característica desse modelo. Abaixo, a explicação das funcionalidades de cada uma destas camadas:

(a) Camada URI (*Uniform Resource Description Framework*): Permite identificar unicamente cada recurso (ex: figuras e páginas HTML - *Hyper Text Markup Language*) disponível na *Web* por meio de um endereço URI. Um URI possibilita a minimização de problemas relacionados a ambiguidades, principalmente na montagem de triplas RDF (BERNERS-LEE, 1998);

(b) Camada Unicode: Permite a utilização da formatação dos caracteres. Esta camada permite a representação de dados, observando-se os aspectos da internacionalização das aplicações *web*;

(c) Camada XML: Permite a criação de marcações para descrição de informações. A

sintaxe das camadas superiores é baseada em XML. Esta camada garante a interoperabilidade sintática no ambiente da *Web Semântica* (W3C, 2008; CUMMINS, 2002);

(d) Camada RDF: Provê um modelo de descrição lógica de dados, permitindo descrever assertivas e informações sobre um determinado recurso. Esta camada garante a interoperabilidade estrutural, proporcionando o compartilhamento de recursos de informação “compreendidos” por máquina (LASSILA *et al.*, 1998);

(e) Camada RDFS (*Resource Description Framework Schema*): Permite a criação de um vocabulário para a camada RDF. Com o RDFS, é possível criar hierarquia de classes e propriedades;

(f) Camada da Ontologia: Estende a camada RDFS, provendo um maior nível de expressividade para a definição da semântica das informações;

(g) Camada de Consulta: Responsável por prover meios para a realização de consultas sobre o modelo de dados RDFS/Ontologia. Uma ontologia, segundo Gruber, é uma especificação explícita e formal de uma conceitualização compartilhada (GRUBER, 1995). Essa semântica formal pode ser materializada no ambiente da *Web Semântica* através da linguagem padrão para construção de ontologias OWL (*Web Ontology Language*) (BERNERS-LEE; HENDLER; LASSILA, 2001);

(h) Camada de Regras: Provê um mecanismo de criação de relações entre recursos, as quais não podem ser descritas diretamente na ontologia;

(i) Camada Lógica: É responsável pelo raciocínio e execução de inferências lógicas a partir da semântica previamente descrita;

(j) Camadas Prova e Confiança: Camadas que proveem um mecanismo para avaliar o nível de confiabilidade das fontes de recursos e informações. São responsáveis pelas verificações necessárias para comprovar e exibir uma inferência;

(k) Camada de Criptografia: Camada responsável por dar suporte à assinatura, identificação e criptografia das descrições e definições que implicam diretamente na realização de inferências.

(l) Camada de Aplicação: Camada que permite a interação entre o usuário e a aplicação *Web Semântica*.

A estrutura e sintaxe da WS têm a preocupação justamente em estabelecer como o conteúdo da *Web* deve ser escrito para permitir seu acesso e interpretação consensual. A partir do uso dessas tecnologias, pode-se garantir a interoperabilidade estrutural e sintática da WS. Com isso, diferentes aplicações computacionais podem acessar a informação e identificar o seu significado, assim como executado neste trabalho.

Um aspecto importante sobre o uso de ontologias para uma descrição explícita e formal dos dados é a independência entre estes e o código da aplicação. Isso torna a ontologia mais

visível e provê uma maior flexibilidade, já que ontologias são fáceis de compartilhar e reutilizar ou adaptar para situações específicas (HART; DOLBEAR, 2013).

Dessa forma, deverá existir uma preocupação crescente em encontrar, aceder e processar a informação disponibilizada na *Web*. Dentre os resultados dos esforços para a criação da *Web Semântica*, destacam-se o padrão RDF, a linguagem de consultas SPARQL e as práticas *Linked Data*, que serão descritos nas seções adiantes.

### 2.1.2.1 *Linked Data*

O movimento *Linked Data* começou com um retorno à prancheta de desenho a fim de expor, para a *Web*, dados que estavam escondidos em bases de dados proprietárias e estruturados de diversas formas. Para isso, recomendou-se a estruturação de dados em um formato padrão: RDF, que também poderia ser usado para especificar as ligações de entrada e saída de cada conjunto de dados (HART; DOLBEAR, 2013). Conforme descrito por Heath e Bizer (2011), *Linked Data* fornece um paradigma de publicação em que documentos e dados são os sujeitos da *Web* e utilizam uma extensão baseada em padrões abertos. Por exemplo, se um especialista deseja compartilhar dados na *Web*, poderia publicá-los como RDF e incluir *links* para outros conjuntos de dados, permitindo que estes sejam descobertos por motores de busca.

Recentemente, com a maturidade da WS, tem-se assistido a um movimento de criação de uma verdadeira plataforma de conhecimento por meio da interligação de repositórios, o que conduziu ao paradigma *Linked Data* (HEATH; BIZER, 2011), um conjunto de princípios e tecnologias que visam à partilha e reutilização de informação de modo massivo, num espaço de dados global, a que as aplicações podem aceder, permitindo também a descoberta de novos dados.

Conforme Bizer, Heath e Berners-Lee (2009), *Linked Data* refere-se aos dados disponibilizados na *Web* de tal forma que são facilmente processados por máquinas. Se significado é definido explicitamente, e esses dados são ligados bidirecionalmente a outros conjuntos de dados externos. O autor complementa que as aplicações tenderão a operar sobre esse vasto conjunto de dados distintos, por meio de mecanismos de acesso padronizados.

Assim, *Linked Data* é o meio para se alcançar o objetivo da WS, da construção de uma *Web* global de dados, em que esses dados possam ser automaticamente processados e integrados por sistemas computacionais.

Foram várias as organizações que adotaram a *Linked Data* como um meio de disponibilizar a sua informação na *Web*. Esse espaço global, designado por *Web* de dados (*Web of*

*Data*), forma um grafo global, constituído por bilhões de declarações RDF de inúmeras fontes, cobrindo tópicos como localizações geográficas, pessoas, companhias, livros, genes, proteínas, fármacos e testes clínicos, entre outros (HEATH; BIZER, 2011).

Portanto, a ideia básica sobre *Linked Data* é aplicar a arquitetura geral da *World Wide Web* para distribuir melhor os dados, de forma estruturada, em uma escala global. Em 2001, Berners-Lee, Hendler e Lassila introduziram princípios de *Linked Data*, que descrevem este conjunto de melhores práticas:

- (a) Usar URI como nome para recursos;
- (b) Usar URI's HTTP para que as pessoas possam encontrar esses nomes;
- (c) Quando alguém procura por uma URI, garantir que informações úteis possam ser obtidas por meio dessas URI, as quais deve estar representadas no formato RDF;
- (d) Incluir *links* para outros URIs, de forma que outros recursos possam ser descobertos;

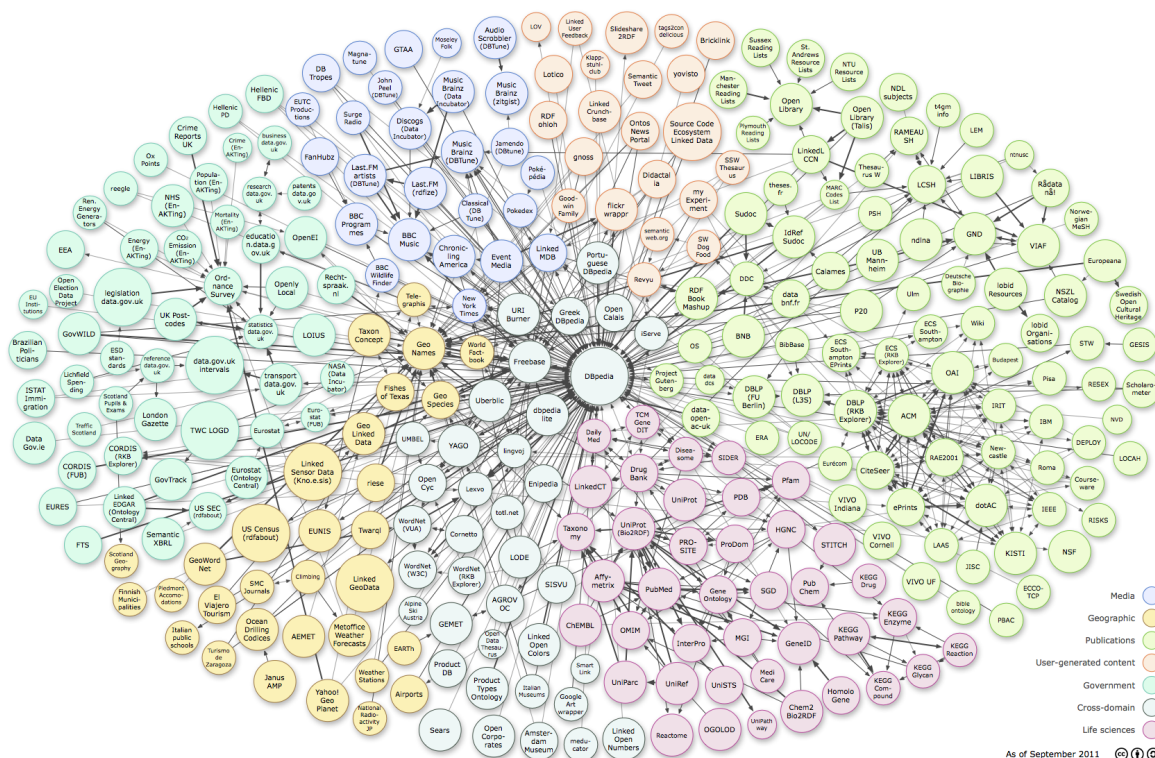
Além disso, Berners-Lee, Hendler e Lassila (2001) criaram a escala das 5 estrelas, a qual classifica os dados abertos publicados. Assim, ganha uma estrela a iniciativa de tornar a informação disponível em formato aberto, e, a partir daí, as iniciativas recebem progressivamente mais estrelas de acordo com o grau de abertura e acessibilidade dos dados, conforme abaixo:

- (a) 1 estrela: disponível na internet (em qualquer formato, por exemplo, PDF) desde que com licença aberta;
- (b) 2 estrelas: disponível na internet de maneira estruturada (por exemplo, em um arquivo com extensão XLS);
- (c) 3 estrelas: disponível na internet, de maneira estruturada e em formato não proprietário (CSV no lugar de XLS);
- (d) 4 estrelas: seguindo todas as regras acima, mas dentro dos padrões estabelecidos pelo W3C (RDF e SPARQL) e usando URL para identificar coisas e propriedades, de forma que as pessoas possam direcionar para suas publicações;
- (e) 5 estrelas: todas as regras acima, mais vínculo dos dados aos de outras pessoas, de forma a fornecer um contexto;

O exemplo mais visível da adoção e aplicação dos princípios *Linked Data* tem sido o projeto *Linking Open Data*, fundado em janeiro de 2007 e apoiado pela *W3C Semantic Web Education and Outreach Group*. O objetivo principal desse projeto é identificar conjuntos de dados disponíveis sob licenças abertas e convertê-los para RDF, de acordo com os princípios *Linked Data* (HEATH; BIZER, 2011).



Figura 6 – Linked Open Data – Comunidade de dados vinculados



Fonte: W3C (2011)

A Figura 6 demonstra os *datasets* disponíveis em *linked open data*, bem como suas ligações. Cada nó nesse diagrama representa um distinto conjunto de dados publicados como *linked data*. O tamanho dos círculos corresponde ao número de triplas de cada *dataset*. As setas indicam a existência de pelo menos 50 links entre dois *dataset*, e cada *link* é uma tripla RDF em que sujeito e objeto são URIs em *namespaces* de *datasets* diferentes. A direção das setas indica o conjunto de dados que contém os links, por exemplo, uma seta de A para B significa que um conjunto de dados contém triplas RDF que usam identificadores de setas B. Setas bidirecionais indicam que os links são espelhados em ambos os conjuntos de dados e a espessura dessas setas corresponde ao número de ligações.

Os 295 conjuntos de dados estão ilustrados na Figura 6 e consistem em mais de 31 bilhões de triplas RDF, interligadas por cerca de 504 milhões de ligações RDF (W3C, 2011).

A legenda desta figura exhibe, em azul, os *datasets* referentes à mídia; em amarelo, os *datasets* geográficos; em verde, as publicações; em vermelho, os conteúdos gerados por usuários; em verde esmeralda, os *datasets* governamentais; na cor cinza, estão os de domínios cruzados e os que estão na cor rosa referem-se à área da ciência da vida. É frequente a utilização de SIGs como plano de fundo para outras fontes de informação. SIGs estão sendo amplamente aproveitados na *Web* de dados para interligar outros conjuntos de dados RDF que contêm um



elemento geográfico. Na Figura 6, os conjuntos de dados geográficos, como GeoNames<sup>1</sup> e LinkedGeoData<sup>2</sup>, são fortemente interligados e agem como intermediários entre outros *datasets*.

Há um movimento global de governos e autoridades locais disponibilizando seus dados na *Web*. Projetos de dados governamentais abertos surgiram em vários países do mundo, como Estados Unidos, Reino Unido, Austrália, Nova Zelândia, Noruega, Holanda, Suécia, Espanha, Estônia, Áustria, Grécia, Canadá e Dinamarca, existindo também um número crescente de iniciativas locais de estados e cidades (SHERIDAN; TENNISON, 2010). Alguns governos criaram catálogos ou portais para tornar a localização e a utilização desses dados mais fácil para o público (BENNETT; HARVEY, 2009), como o portal data.gov e data.gov.uk. Além disso, pessoas e organizações vêm publicando dados governamentais por conta própria em vários formatos (BIZER; HEATH; BERNERS-LEE, 2009). O Brasil tem uma boa oferta de dados, oferecidos pública e gratuitamente, em todas as esferas e poderes, mas existem poucas iniciativas do governo que se propõem a dar acesso à base integral estruturada e em linguagem aberta. Os exemplos mais recentes da iniciativa brasileira, nesse sentido, são o projeto Governo Aberto SP, em fase de implantação (GOVERNO ABERTO, 2010), e o Governo Eletrônico (2013), cujo objetivo é estimular o desenvolvimento, a padronização, a integração, a interoperabilidade, a normalização dos serviços de produção e a disseminação de informações de forma descentralizada. Enquanto o governo não disponibiliza mais dados em formato aberto, estão surgindo no Brasil iniciativas no sentido de extrair os dados de sites e portais governamentais, reorganizá-los, torná-los abertos e, ou, conferir novo valor a eles, por meio de diferentes aplicações, como o Congresso Aberto, o Parlamento Aberto, o Legisdados, entre outros (THACKER, 2011).

#### 2.1.2.2 RDF

Em 2004, Manola e Miller conceituam e explicam RDF: o *Resource Description Framework* (RDF) é uma linguagem para a representação de informações na *Web*. O RDF é particularmente projetado para representar metadados sobre recursos, como o título, autor e data de alteração de uma página *Web*, direitos autorais e licenciamento sobre um documento na *Web*, o cronograma de disponibilidade de algum recurso compartilhado, ou a descrição das preferências de um usuário da *Web* para entrega de informação. Assim, o RDF pode ser usado para representar informações sobre qualquer coisa que possa ser identificada na *Web*, mesmo que não seja diretamente recuperada pela *Web*. Exemplos incluem a informação sobre itens disponíveis em uma página de comércio eletrônico ou a descrição das preferências de um usuário da *Web*

---

<sup>1</sup> <http://www.geonames.org/>

<sup>2</sup> <http://linkedgeodata.org/>

para entrega de informações.

O RDF foi projetado para situações em que a informação precisa ser processada por aplicações, em vez de simplesmente ser mostrada para pessoas. Seu objetivo é ser uma maneira de ver as coisas e uma linguagem real para descrever recursos da *Web*, independentemente de qualquer domínio particular ou área temática. O propósito é permitir que aplicações troquem informações compreensíveis por máquinas, para uma melhor interação na *Web*. Para isso, essa estrutura precisa ser escalável, de modo que funcione para qualquer quantidade de dados, seja flexível, expressiva o suficiente para codificar qualquer tipo de informação, e ainda assim simples, de forma que seja fácil para qualquer pessoa ler, escrever e consultar. A capacidade de troca de informações entre aplicações diferentes significa que estas podem ser disponibilizadas para outras aplicações que não aquelas para as quais foram originalmente criadas (HART; DOLBEAR, 2013).

A cola que une os documentos tradicionais da *Web* são os *links* de hipertexto entre as páginas HTML. A cola da *web* de dados são as ligações RDF (BIZER; CYGANIAK; HEATH, 2007). O RDF pode ser interpretado em três níveis distintos de abstração (KLYNE; CARROLL, 2004):

(a) Nível sintático – trata-se essencialmente de documento XML;

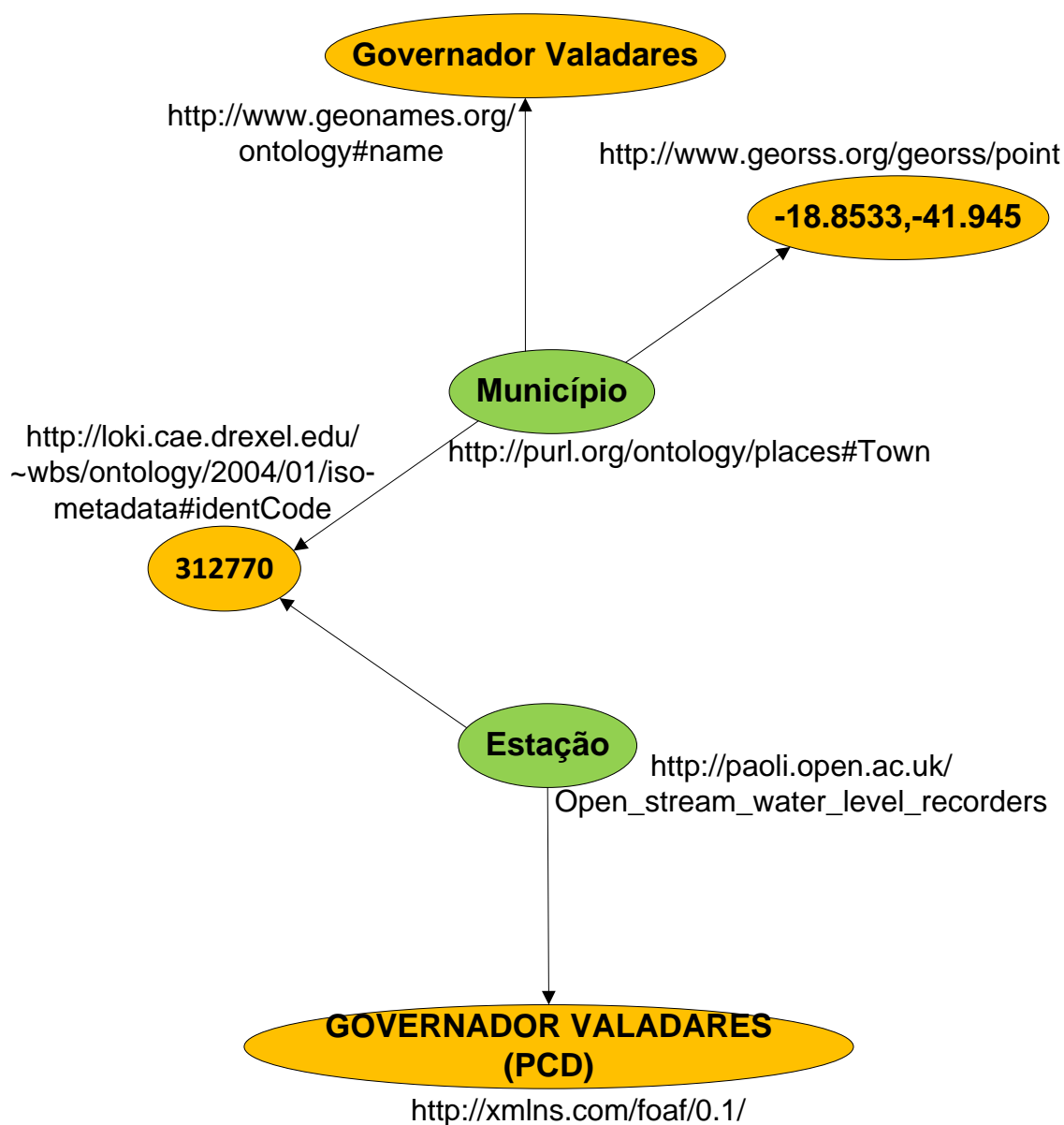
(b) Nível estrutural – conjunto de triplas na forma (sujeito-predicado-objeto) que codificam fatos conhecidos. Os predicados codificam relacionamentos binários entre um sujeito e um objeto e estão rotulados com *Uniform Resource Identifier* (URI). Um sujeito é um recurso identificado por um espaço de nomes globais fornecido pelo uso de URI. Um objeto pode ser outro recurso relacionado, ou o valor da propriedade do sujeito;

(c) Nível semântico – grafos dirigidos com semântica predefinida associada aos nós e arcos.

Os links RDF são a base dos dados ligados. Eles permitem que as aplicações cliente naveguem entre as fontes de dados e descubram dados adicionais. Para fazer parte da *Web* de Dados, fontes de dados devem definir links RDF para relacionar as entidades em outras fontes de dados (BIZER; HEATH; BERNERS-LEE, 2009).

A Figura 7 exibe a representação gráfica de parte do trabalho, como exemplo, onde existe um Município identificado por `<http:purl.orgontologyplaces#Town>`, cujo nome é Governador Valadares, o código é 312770 e os valores -18.8533, -41.945 são as coordenadas geográficas.

Figura 7 – Um grafo que descreve o município Governador Valadares



Fonte: Elaborado pela autora, 2013.

A Figura 7 ilustra os URIs utilizados para identificar:

- (a) Indivíduos: Governador Valadares, identificado por `<http://www.geonames.org/ontology#name>`;
- (b) Tipos de coisas: Município, identificado por `<http://purl.org/ontology/places#Town>`;
- (c) Propriedades dessas coisas: código do município, identificado por

<http://loki.cae.drexel.edu/~wbs/ontology/2004/01/iso-metadata#identCode>;

(d) Valores dessas propriedades: 312770 como valor da propriedade código do município.

O RDF fornece uma sintaxe baseada em XML para a representação desses grafos. A Figura 8 é um trecho na notação RDF/XML, que corresponde ao grafo apresentado na Figura 7.

Figura 8 – RDF/XML descrevendo o município Governador Valadares

```
<?xml version="1.0"?>
<rdf:RDF
  xmlns:paoli="http://paoli.open.ac.uk/"
  xmlns:loki="http://loki.cae.drexel.edu/~wbs/ontology/2004/01/iso-metadata"
  xmlns:geonames="http://www.geonames.org/ontology#"
  xmlns:grs="http://www.georss.org/georss/point"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
  xmlns:idc="http://loki.cae.drexel.edu/~wbs/ontology/2004/01/"
  <info:Municipio rdf:about="http://purl.org/ontology/placesTown">
    <grs:pointpoint>-18.8533, -41.945</grs:pointpoint>
    <geonames:name>Governador Valadares</geonames:name>
    <idc:identCode>312770</idc:identCode>
  </info:Municipio>
</rdf:RDF>
```

Fonte: Elaborado pela autora, 2013.

A maneira mais simples de publicar dados do governo na internet é disponibilizá-los de maneira bem estruturada em sua forma bruta. As estruturas e formatos XML, RDF, Comma-Separated Values (CSV) permitem o sucesso do uso automatizado dos dados. Quando possível, utilizar padrões abertos e ferramentas que possibilitam a fácil e eficiente produção e publicação de dados (BENNETT; HARVEY, 2009).

### 2.1.2.3 SPARQL: Linguagem de Consulta em RDF

SPARQL é um protocolo (KENDALL; FEIGENBAUM; TORRES, 2008) e uma linguagem de consultas para RDF (PRUD'HOMMEAUX; SEABORNE, 2008). Assim como os sistemas de bancos de dados relacionais fazem uso do SQL para consultar registros nas suas bases de dados, SPARQL é a linguagem de consulta padrão recomendada pelo W3C para recuperação de informações contidas em grafos RDF.

Semelhante ao SQL, o SPARQL possui uma estrutura *Select-From-Where* onde:

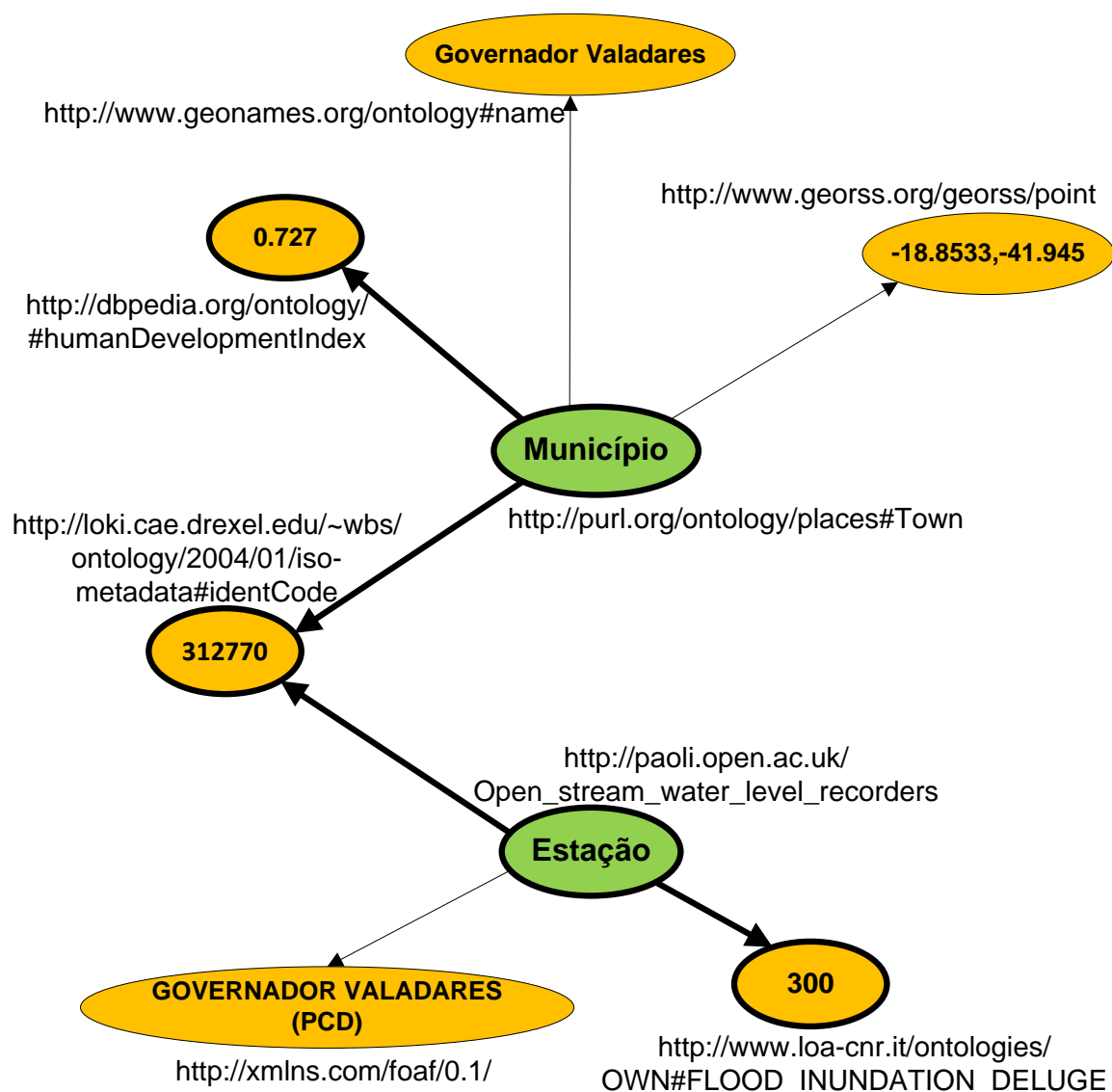
(a) *Select*: Especifica uma projeção sobre os dados como a ordem e a quantidade de atributos e, ou, instâncias que serão retornados;

(b) *From*: Esta cláusula é opcional e declara as fontes que serão consultadas. Quando não especificada, a busca será feita em um documento RDF/RDFS particular;

(c) *Where*: Impõe regras e restrições na consulta. Os registros retornados pela consulta deverão satisfazer as restrições impostas por esta cláusula.

O resultado de uma consulta SPARQL pode ser visto como um subgrafo resultante da execução da consulta sobre o grafo que representa o modelo. Considere-se, por exemplo, o grafo apresentado na Figura 9.

Figura 9 – Representação das instâncias de um domínio



Fonte: Elaborado pela autora, 2013.

Este grafo representa a relação entre as instâncias de uma ontologia cujo domínio é focado na descrição e formalização de escritores. O subgrafo destacado em negrito é o resultado da consulta exibida na Tabela 1, cujo retorno é a informação de que a estação, localizada no município de código 312770, tem nível de enchente de 300cm e o município onde se localiza a estação tem o IDH de 0,727.

Tabela 1 – Exemplo de Consulta SPARQL

<b>Exemplo de Consulta SPARQL</b>
<pre> SELECT ?cod_municipio ?cod_municipio2 ?idh ?nivel_alerta WHERE { ?cota geonames:featureCode ?cod_estacao . ?cota loa:#FLOOD__INUNDATION__DELUGE ?medicao . ?cota loki:identCode ?cod_municipio . ?mun loki:identCode ?cod_municipio2 . ?mun dbpedia-owl:humanDevelopmentIndex ?idh FILTER (?cod_municipio = ?cod_municipio2 &amp;&amp; ?cod_municipio = 312770)} </pre>

Fonte: Elaborado pela autora, 2014.

No exemplo descrito na Tabela 1, a linha 1 é conhecida como a cláusula do resultado e especifica que, nesse caso, deseja-se retornar somente a seleção das variáveis incluídas após o comando SELECT. A linha 2 impõe as condições para os resultados, descritas nas linhas 3-7, como, por exemplo, a classe *dbpedia-owl:humanDevelopmentIndex* do recurso ?mun será contida na variável ?idh e, na última linha, um filtro indica que os códigos do município dos recursos ?cota e ?mun devem ter o mesmo valor e este deve ser 312770.

Assim como outras linguagens de consulta de dados, como SQL, a palavra-chave DISTINCT pode ser adicionada após o SELECT, para retornar somente resultados únicos, evitando duplicatas. Modificadores do resultado podem ser adicionados após a cláusula condicional WHERE. Por exemplo, ORDER BY classifica os resultados em ordem crescente ou decrescente de uma das variáveis; LIMIT limita o número de resultados retornados, e OFFSET determina os primeiros resultados ignorados. O comando FILTER permite excluir resultados cujos valores não atendem às restrições especificadas e a palavra-chave UNION retorna resultados que correspondem a uma das exigências apresentadas na cláusula WHERE (HART; DOLBEAR, 2013).

### 2.1.3 Visualização de Dados

Steele e Iliinsky (2011), dissertam sobre a visualização de dados como um eficiente e eficaz meio de comunicação para um grande volume de informações, e conceituam que os

termos de visualização de dados e de visualização de informações são úteis para se referir a qualquer representação visual dos dados. Esses termos são:

(a) algoritmicamente desenhados (podem ter toques personalizados, mas são amplamente renderizado com a ajuda de métodos computadorizados);

(b) fáceis de se regenerar com dados diferentes (o mesmo formulário pode ser reaproveitado para representar conjuntos de dados diferentes, com dimensões ou características semelhantes);

(c) muitas vezes esteticamente áridos (dados não decorados);

(d) relativamente ricos em dados (grandes volumes de dados são bem vindos e viáveis).

Visualizações de dados são inicialmente projetadas por um humano, mas são desenhadas graficamente por algoritmos ou *software* de diagramação. A vantagem dessa abordagem é o fato de ser relativamente simples para atualizar ou gerar novamente a visualização, incluindo novos dados.

A visualização dinâmica de dados é uma das formas culturais genuinamente novas, que se tornou possível graças à computação. Com os computadores, pode-se visualizar conjuntos de dados muito mais amplos, criar visualizações dinâmicas, alimentar dados em tempo real, basear as representações gráficas de dados em sua análise matemática usando vários métodos, da estatística clássica à prospecção de dados, mapear um tipo de representação em outro (imagens em sons, sons em espaços tridimensionais, etc.) (MANOVICH, 2009).

Os autores são explícitos sobre o motivo de a visualização ser um meio útil para examinar, compreender e transmitir a informação:

(a) Visualização aproveita as capacidades incríveis do sistema visual para mover uma enorme quantidade de informações para o cérebro muito rapidamente;

(b) Visualização permite identificar padrões, relacionamentos e seus significados;

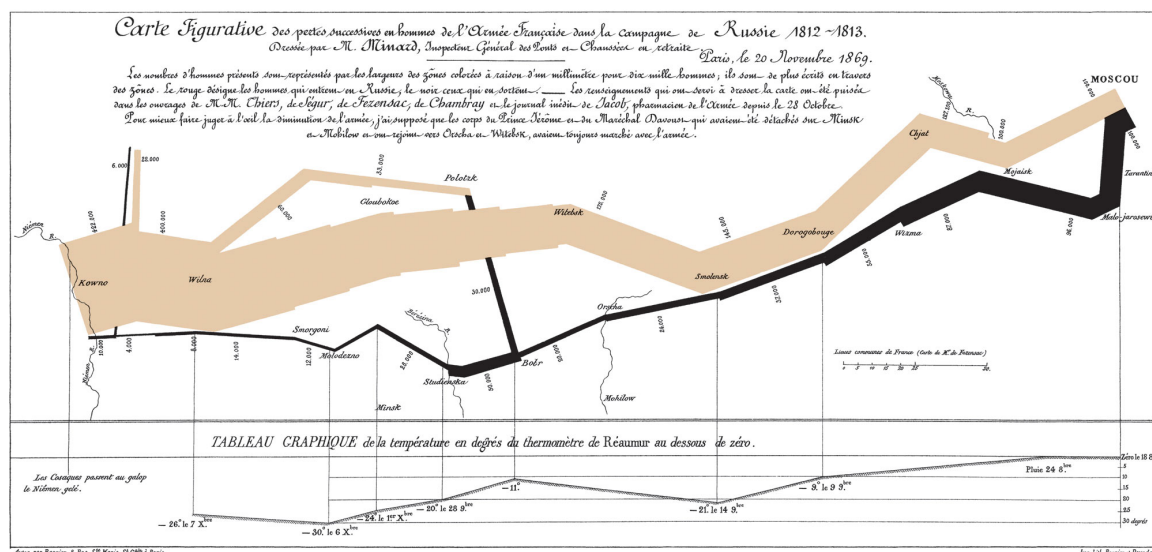
(c) Visualização ajuda a identificar subproblemas;

(d) Visualização é algo realmente bom para identificação de tendências ou produtos fora de série, descobrindo pontos específicos ou interessantes em um campo maior, etc.

Um exemplo de visualização de dados é o clássico gráfico de Charles Joseph Minard (1781-1870), que revela o avanço das tropas do exército de Napoleão, pela Rússia e as perdas desse exército, em 1812 (Figura 10). No gráfico, são representadas seis variáveis: o tamanho do exército, a sua localização em uma superfície bidimensional, a direção do movimento do exército e a temperatura em diversas datas durante a retirada de Moscou.



Figura 10 – O percurso do exército de Napoleão



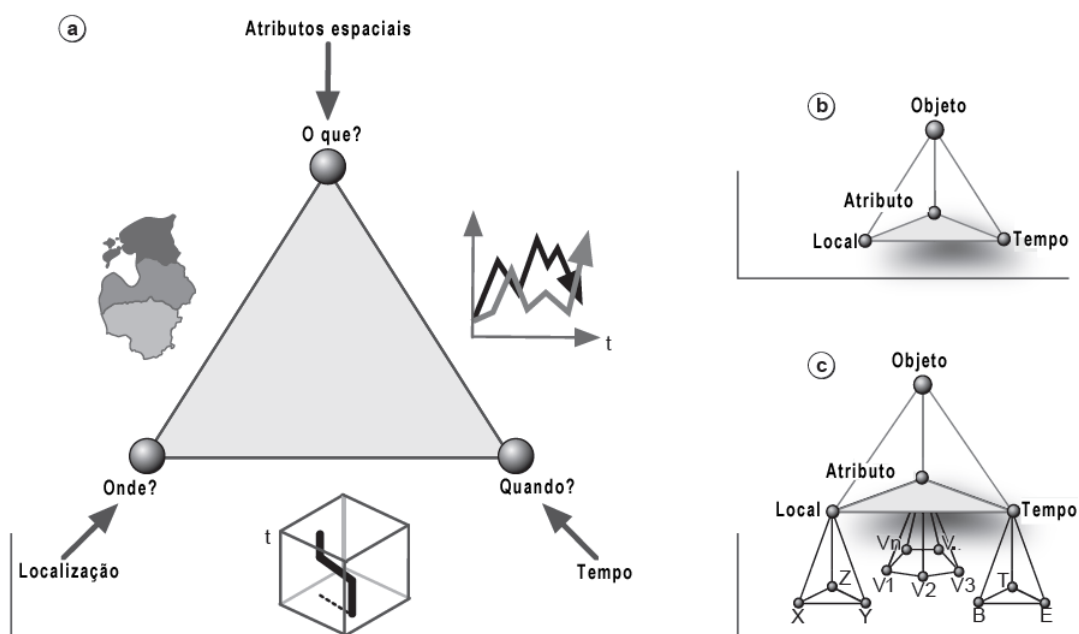
Fonte: Gráfico de Minard (1869).

Esse gráfico é famoso pela sua visualização concisa e elegante de diversas variáveis. O tamanho do exército de Napoleão é mostrado com a largura da banda do mapa, a partir da fronteira russo-polonesa com 422.000 homens. Até o momento em que chegaram a Moscou, o tamanho do exército havia reduzido para 100.000 homens. Eventualmente, apenas uma fração do exército original de Napoleão sobreviveu.

#### 2.1.4 Visualização de Dados Geoespaciais

A informação geográfica se distingue de outras informações por referir-se a objetos ou fenômenos com uma localização específica no espaço e, portanto, tem um endereço espacial (KRAAK; ORMELING, 2003). Os mesmos autores explicam que, devido a essa característica, os locais dos objetos ou fenômenos podem ser visualizados, e essas visualizações, chamadas de mapas, mostram como os objetos do mundo real (como casas, estradas, campos ou montanhas) podem ser abstraídos como um modelo de paisagem digital, de acordo com alguns critérios pré-determinados, e armazenados em SIGs (como pontos, linhas, áreas ou volumes). Quando armazenados em um banco de dados, esses dados geoespaciais se dividem em dados de localização, dados de atributos e dados temporais. A Figura 11 exhibe essas características.

Figura 11 – Características dos dados geoespaciais



Fonte: Traduzido de Kraak e Ormeling (2003).

(a) Componentes localização, atributo e tempo e suas perguntas relacionadas: onde, o quê e quando;

(b) A visualização do objeto;

(c) Características detalhadas dos componentes dos dados.

Em 2003, Kraak e Ormeling justificam a unicidade de um SIG pela capacidade de combinar dados geoespaciais e não geoespaciais de diferentes fontes de dados em uma operação de análise geoespacial, a fim de responder a vários tipos de perguntas. A Figura 12 demonstra os tipos de perguntas que podem ser respondidas por um SIG.

Figura 12 – Questões típicas de um SIG sendo respondidas utilizando mapas



Fonte: KRAAK; ORMELING, 2003.

O desenvolvimento de SIGs foi estimulado por áreas individuais, tais como a defesa civil, cadastros, serviços públicos e planejamento regional. Já que todas as áreas têm origens e necessidades diferentes, a funcionalidade do *software* SIG se torna diferente a cada tipo de necessidade (KRAAK; ORMELING, 2003).

A qualidade dos dados é outro aspecto importante, para o apoio à decisão, na informação processada e apresentada por um SIG. Os SIGs são eficientes na combinação de conjuntos de dados. Não obstante o fato de esses dados serem de épocas e resoluções diferentes, ou até não passíveis de combinação, o *software* combina esses dados e apresenta os resultados (KRAAK; ORMELING, 2003).

### 2.1.5 Visualização de Linked Data

A visualização e a interação de *linked data* é uma questão que tem sido reconhecida desde o início da *Web* semântica (GEROIMENKO; CHEN, 2003). Ao aplicar técnicas de visualização de informação, a *Web* semântica auxilia os usuários na exploração e interação dos dados. A transformação e apresentação visual desses dados são os principais objetivos da visualização de informação, de tal modo que os usuários possam obter uma melhor compreensão dos dados (CARD; MACKINLAY; SCHNEIDERMAN, 1999). Visualizações são úteis para a obtenção de uma visão geral dos *datasets*, seus tipos principais e as relações entre eles.

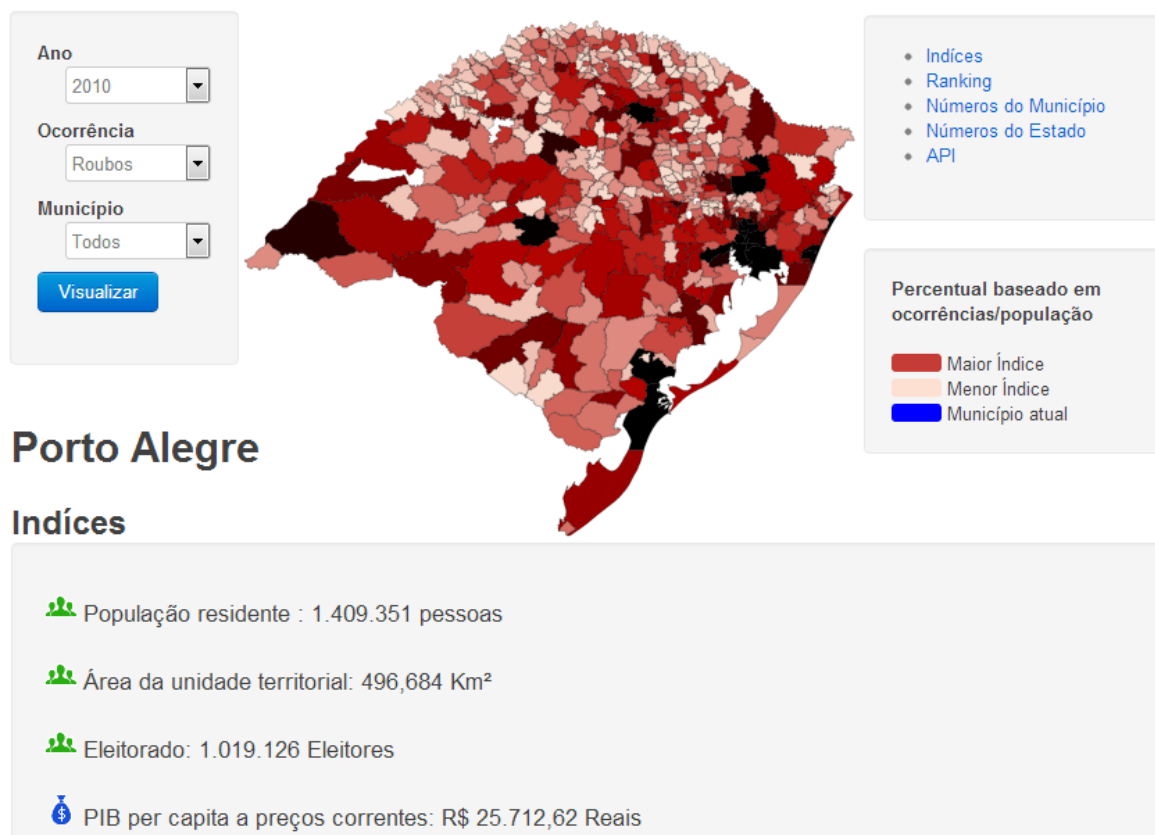
A visualização de dados pode ser definida como algo que dá ao usuário uma maneira de analisar os dados, de modo a obter conhecimento e entendimento. Já a visualização de dados vinculados é uma exibição de dados que se comunica com outra visão. Se uma modificação é feita para uma das visões, o outro ponto de vista vai mudar sua aparência em reação àquela modificação (CHEN; HÄRDLE; UNWIN, 2007).

A visualização de dados ligados se enquadra na categoria de navegação baseada em ontologia em busca de informações, em que a anotação semântica de dados é utilizada para apoiar a exploração desses dados (PAULHEIM; PROBST, 2010).

Para uma utilização eficaz, é essencial fornecer mecanismos simples para consultar os *dataset*. Ahlberg, Williamson e Shneiderman (1992), conceituam consultas dinâmicas como sendo a interface gráfica com manipulação direta, como, por exemplo, listas ou *slide-bars* que, quando alterados, consultam automaticamente o banco de dados e os dados do filtro são exibidos. Em 1996, Shneiderman explica que, primeiramente, todos os dados são exibidos; então, o usuário utiliza os filtros para selecionar o subconjunto de interesse, sendo então visualizados os detalhes destes novos dados.

A Figura 13 ilustra um exemplo de uma aplicação de visualização de *linked data*, em que as informações são baseadas em *open data* disponibilizados pelo governo e institutos nacionais, com o objetivo de promover e incentivar que diversos governos disponibilizem seus dados em uma plataforma que possa ser utilizada pela sociedade. Todo o conteúdo está disponível para compartilhamento, distribuição e reuso, com o propósito de promover uma plataforma aberta, em que é possível criar visualizações de dados como ferramenta de investigação.

Figura 13 – Visualização de dados interligados



Fonte: COMUNIDADE DE DADOS ABERTOS (2013).

A Figura 13 foi produto de um projeto desenvolvido pela comunidade “São Paulo Perl Mongers” e “Opendata-BR”, com o objetivo de divulgar dados sobre segurança pública do país, dos estados e municípios, para permitir a comparação entre regiões através da análise de índices de ocorrência de crime. Foram interligados dados do IBGE e da Secretaria de Segurança Pública do Rio Grande do Sul, a partir dos seguintes *datasets*:

- <http://www.ibge.gov.br/cidadesat/link.php?uf=rs;>
- [http://www.ssp.rs.gov.br/;](http://www.ssp.rs.gov.br/)
- [ftp://geofpt.ibge.gov.br/mapas/malhas\\_digitais/;](ftp://geofpt.ibge.gov.br/mapas/malhas_digitais/)
- [http://mapicons.nicolasmollet.com/.](http://mapicons.nicolasmollet.com/)

De acordo com Hart e Dolbear (2013), um mapa, representação panorâmica bidimensional, é a forma mais objetiva de representar informações geográficas. É comumente utilizado como mapas generalizados que não mostram com precisão uma área, mas permitem clareza na visualização de informações sobre uma posição exata. Esse tipo de mapa minimiza objetos menos relevantes e enfatiza construções importantes, como, por exemplo, o uso de símbolos para destacar as representações. Os mapas são usados para vários propósitos. Os mais comuns

são: navegação, cenário para sobrepor outras informações ou forma de visualizar os aspectos de uma paisagem (HART; DOLBEAR, 2013). Os mapas são, portanto, um meio de visualização de informações relevantes quando combinados com as tecnologias *Linked Data* e *WS*, ao disponibilizar o plano de fundo e as funções dinâmicas e ao agregar conteúdo às informações que serão sobrepostas.

### 2.1.6 Autores Citados e Referenciados

Na Tabela 2, são apresentados os constructos e os autores citados e referenciados em cada uma das seções e subseções do Referencial Teórico.

Tabela 2 – Constructos abordados e autores citados

<b>Constructo</b>	<b>Autores</b>
Sistema de Informação Geográfica	Bernhardsen (1992) Burrough e McDonnell (1998) Cowen (1988) Kraak e Ormeling (2003) Eastman <i>et al.</i> (1993) Hasenack (1995) Osleeb e Kahn (1999) Densham (1991) Faber <i>et al.</i> (1997) Zeuner (2008) Benno <i>et al.</i> (2007) Paolo <i>et al.</i> (2011) Muller (1985) Carswell (1998) Bonham-Carter (1994) Longley <i>et al.</i> (2005) Aronoff (1989) Hart e Dolbear (2013) Thomas e Roller (1993)
Web Semântica - <i>Linked Data</i>	Berners-Lee, Hendler e Lassila (2001) Heath e Bizer (2011) Bizer, Heath e Berners-Lee (2009)

<p>- RDF</p> <p>- SPARQL</p>	<p>Sheridan e Tennison (2010)</p> <p>Bennett e Harvey (2009)</p> <p>Hart e Dolbear (2013)</p> <p>Manola e Miller (2004)</p> <p>Bizer, Cyganiak e Heath (2007)</p> <p>Klyne e Carroll (2004)</p> <p>Bizer, Heath e Berners-Lee (2009)</p> <p>Hart e Dolbear (2013)</p> <p>Kendall, Feigenbaum e Torres (2008)</p> <p>Prud’Hommeaux e Seaborne (2008)</p> <p>Hart e Dolbear (2013)</p> <p>Allemang e Hendler (2008)</p>
<p>Visualização de Dados</p> <p>- Visualização de Dados Geoespaciais</p> <p>- Visualização de <i>Linked Data</i></p>	<p>Steele e Iliinsky (2011)</p> <p>Kraak e Ormeling (2003)</p> <p>Manovich (2009)</p> <p>Geroimenko e Chen (2003)</p> <p>Card, MacKinlay e Schneiderman (1999)</p> <p>Chen, Härdle e Unwin (2007)</p> <p>Paulheim e Probst (2010)</p> <p>Shneiderman (1996)</p> <p>Ahlberg, Williamson e Shneiderman (1992)</p> <p>Hart e Dolbear (2013)</p>

Fonte: Elaborado pela autora, 2013.

## 2.2 Protocolo da Revisão Sistemática da Literatura

A disponibilização de dados governamentais, de forma acessível, pode representar um avanço no processo democrático, uma vez que possibilita o aumento da transparência na gestão pública e permite que a população participe da interação governo-sociedade, por meio de documentos em formatos de fácil manipulação para humanos e máquinas. A publicação dos dados, de forma que as pessoas possam reutilizá-los, é o passo operacional mais relevante para caracterizá-los como dados governamentais abertos. Os dados na *Web* são facilmente publicados, porém, ao estarem disponíveis conforme os padrões de dados abertos, garante-se que

possam ser acessados e reutilizados por agentes de *software*. Além disso, a apresentação visual dos dados, de forma clara, permite que os usuários possam obter uma melhor compreensão dos mesmos. A literatura aponta os Sistemas de Informação Geográfica (SIG) como importante ferramenta para visualização de *linked data*, por permitirem a integração de dados vindos de várias fontes heterogêneas, com o propósito de potencializar a descoberta e a divulgação de novos conhecimentos.

Ao publicar dados em ambientes geoespaciais, é necessário que esse ambiente contenha especificações semânticas para alcançar a interoperabilidade (KUHN, 2005). O potencial de definir conceitos para dados geográficos explícitos, com a semântica, leva ao desenvolvimento de conjunto de dados abertos ligados e fontes semânticas para GIS.

Existe uma grande lista de aplicações de *Web* semântica, que são implantadas para consumirem *Linked Data*, fundamentadas direta ou indiretamente em informações geográficas. Sejam governamentais ou aplicações de domínio específico, os temas são sempre interdisciplinares. No entanto, como normalmente ocorre na engenharia de *software*, o desenvolvimento aparenta ser guiado mais por opinião de especialistas que baseado em evidências empíricas ou modelos cognitivos, levando a dúvidas sobre o que realmente se sabe, os benefícios e as limitações do objeto de pesquisa.

Assim, foi feita uma revisão sistemática da literatura (tradução de *Systematic Literature Review* - SLR) com o propósito de sintetizar e apresentar as buscas empíricas sobre visualização de *linked data* em SIGs, assim como uma visão geral sobre o estado da arte, que se acredita ser importante para a comunidade científica construir um entendimento comum dos desafios que devem ser enfrentados sobre os tópicos abordados.

O paradigma baseado em evidências é amplamente utilizado na medicina clínica e na educação, como uma ferramenta para apoiar a prática e formulação de políticas. O conceito básico que sustenta esta técnica é a realização de um estudo secundário que sistematicamente localiza, avalia e agrega os resultados de um conjunto de estudos empíricos, a fim de reunir as melhores evidências disponíveis para responder a uma pergunta de pesquisa de forma imparcial. A ideia de adaptá-lo para uso em engenharia de *software* foi proposta pela primeira vez em 2004, por Kitchenham *et al.* (2004) e, desde então, o paradigma tornou-se cada vez mais aceito como um complemento útil para o conjunto de ferramentas metodológicas utilizadas na engenharia de *software*.

A revisão sistemática da literatura é um método importante para resumir e fornecer uma visão geral da maturidade da disciplina (KITCHENHAM; DYBA; JORGENSEN, 2004), que busca um sentido em grandes volumes de informação, e um meio de contribuição para as respostas às questões sobre o que funciona e o que não funciona - entre vários outros tipos de perguntas. É um método de mapeamento e identificação de áreas de incerteza e em que ainda são necessários estudos, por não haver nenhuma ou pouca pesquisa relevante sobre o assunto. A revisão sistemática também sinaliza áreas onde existem falsas certezas, áreas em que se pensa



saber mais do que realmente se sabe e em que, na verdade, há poucas evidências para apoiar essas crenças (PETTICREW; ROBERTS, 2006).

O objetivo desse método é fornecer uma oportunidade alternativa de melhor visualização do contexto da pesquisa em questão, combinando e analisando resultados quantitativos de estudos empíricos, a fim de dar sentido à literatura em constante evolução (GLASS, 1976).

Diante da literatura em crescimento, cujo conhecimento encontra-se inexplorado, a SLR merece maior prioridade que a adição de um novo experimento ou *survey* (GLASS, 1976). O acúmulo de conhecimento depende cada vez mais da integração entre estudos anteriores e descobertas empíricas (KING; HE, 2005).

A SLR é definida por um protocolo que estabelece as etapas dos procedimentos a serem realizados durante a revisão. Os procedimentos metodológicos, presentes no protocolo, representam as “forças” da SLR, permitindo tanto avaliar o estado atual dos conhecimentos da área, como manter a atualização de pesquisas em base avançada (COOK; MULROW; HAYNES, 1997). Uma SLR difere de uma revisão da literatura simples ou de um *survey* por ser um estudo replicável, científico e transparente, evitando assim os vieses.

### 2.2.1 *Planejamento*

Este estudo foi realizado como uma revisão sistemática da literatura, com base nas diretrizes originais propostas por Kitchenham (KITCHENHAM *et al.*, 2009), e com o propósito de responder às seguintes perguntas:

P1. Quais são os tipos de pesquisa mais utilizados ao se tratar de visualização em SIG e *linked data*?

P2. A partir de 2010, qual a frequência do uso de dados governamentais em pesquisas que relacionam visualização em SIG e *linked data*?

P3. Quais foram os tipos de resultado obtidos com o uso de dados governamentais?

### 2.2.2 Realização

Na pesquisa, utilizou-se o *Google Scholar*, por ser um motor de busca em bases de dados confiáveis, de documentos, artigos científicos, revisões, *papers* de conferências, repositórios de documentos digitais, institucionais e multidisciplinares, reconhecidos pela comunidade acadêmica internacional.

A definição dos termos para as buscas procedeu-se por meio da combinação das seguintes palavras-chave: *linked data*, *visualization*, *geovisualization*, *maps*, *geographic information system*, *gis*, *semantic web*. Utilizando os operadores booleanos OR e AND, foram feitas combinações de termos para formação da *string* de pesquisa, segundo a Tabela 3.

Tabela 3 – Strings da pesquisa

<b>Strings da Pesquisa</b>
(“geographic information system” OR gis OR geographic) AND (visualization OR geovisualization OR “data visualization”) AND (“web semantic” OR semantic) AND (“linked data”)

Fonte: Elaborado pela autora, 2013.

A pesquisa realizada nas bases de dados permitiu a seleção de 55 publicações, após a eliminação de 58 publicações, seguindo os seguintes critérios de exclusão:

- (a) Monografias, editoriais, prefácios, sumários, entrevistas, notícias, revisões, tutoriais, *workshops*, painéis e *pôsteres*;
- (b) Publicações que não estivessem em inglês ou português;
- (c) Publicações pagas.

A Tabela 4 exhibe as referências utilizadas na Revisão Sistemática da Literatura.

Tabela 4 – Lista das Referências Utilizadas na Revisão Sistemática da Literatura

#	Referência
1	Koubarakis <i>et al.</i> (2011)
2	Janowicz <i>et al.</i> (2013)
3	Abargues <i>et al.</i> (2010)
4	Reitz e Templer (2012)
5	Garbis <i>et al.</i> (2012)
6	Galiotou e Fragkou (2013)

7	Florczyk <i>et al.</i> (2010)
8	Gliozzo (2010)
9	Koubarakis <i>et al.</i> (2012a)
10	Mentes (2012)
11	Latif <i>et al.</i> (2010)
12	Szekely <i>et al.</i> (2011)
13	Simon <i>et al.</i> (2010)
14	Keßler, Janowicz e Kauppinen (2012)
15	Vilches-Blázquez <i>et al.</i> (2010)
16	Marino, Nascimento e Borges (2012)
17	Paulheim (2011)
18	Bandholtz e Fock (2011)
19	Knap, Michelfeit e Necasky (2012)
20	Phuoc e Hauswirth (2009)
21	Hage, Erp e Malaisé (2012)
22	Kauppinen e Espindola (2011)
23	Mendes <i>et al.</i> (2010)
24	Poblet, Casanovas e López-Cobo (2010)
25	Pauliina (2011)
26	Krabina <i>et al.</i> (2012)
27	Lebo <i>et al.</i> (2011)
28	García <i>et al.</i> (2011)
29	Vos <i>et al.</i> (2012)
30	Hitzler <i>et al.</i> (2012)
31	Ding <i>et al.</i> (2010b)
32	Hidayat <i>et al.</i> (2012)
33	Isaac <i>et al.</i> (2012)
34	Pschorr <i>et al.</i> (2010)
35	Yue <i>et al.</i> (2011)
36	Stasch <i>et al.</i> (2012)
37	Logean <i>et al.</i> (2012)
38	Koubarakis <i>et al.</i> (2012b)
39	Janowicz e Hitzler (2012)
40	Haase <i>et al.</i> (2009)
41	Waitelonis <i>et al.</i> (2010)
42	Hahmann, Burghardt e Weber (2011)
43	Shvaiko <i>et al.</i> (2012)

44	Fensel <i>et al.</i> (2013)
45	Le-Phuoc <i>et al.</i> (2010)
46	Ding <i>et al.</i> (2010a)
47	Dell'Aglio, Celino e Valle (2013)
48	Steinberg e Brehm (2009)
49	Harth (2009)
50	Soares (2011)
51	Gür, Díaz e Kauppinen (2012)
52	Huertas (2012)
53	Grossner (2010)
54	Roth (2011)
55	Rojas (2009)

Fonte: Elaborado pela autora, 2014.

Na etapa seguinte da revisão sistemática da literatura, foi feita a leitura e análise dos textos completos das publicações selecionadas para classificá-las de acordo com o tipo de publicação, resultado da pesquisa e ano de publicação.

### 2.2.3 Resultados

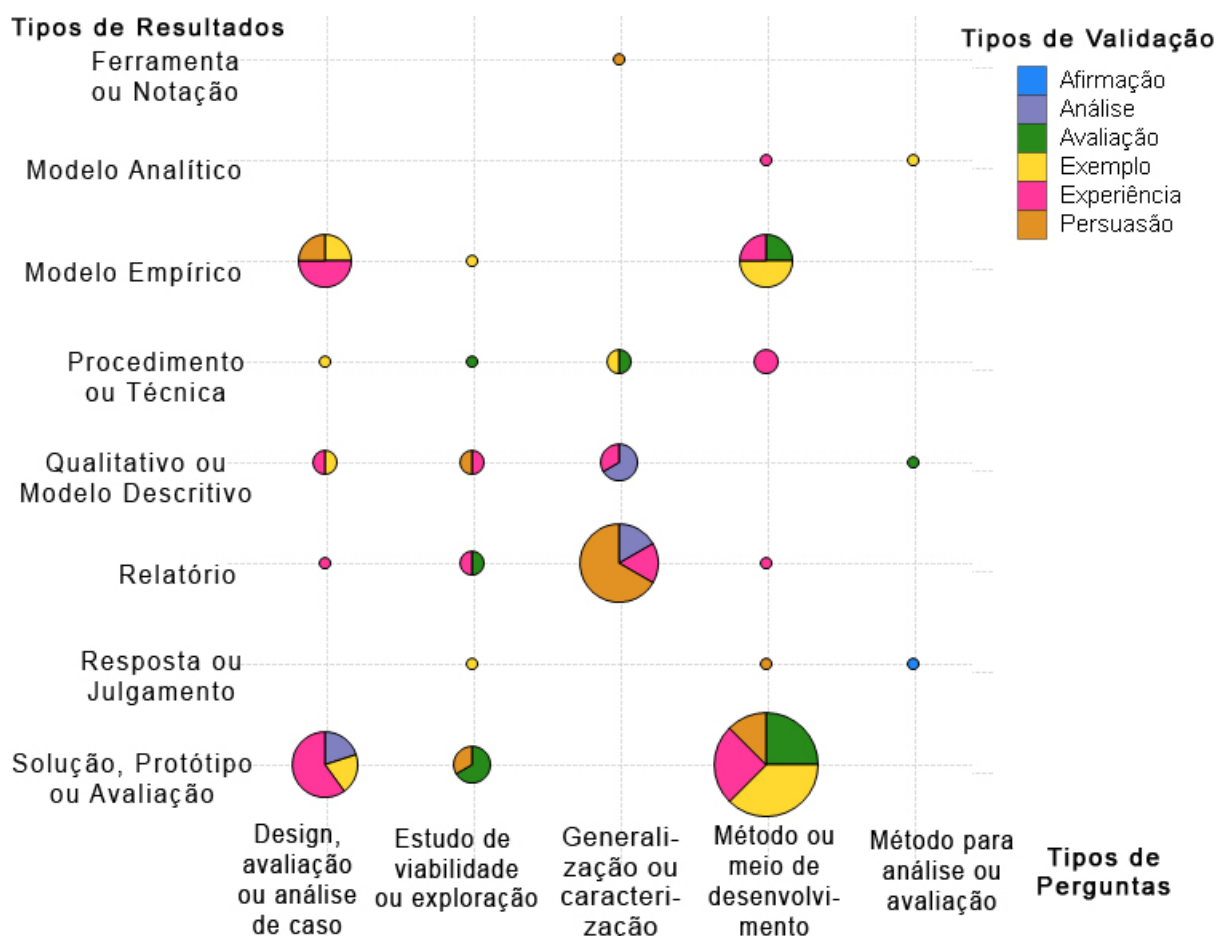
#### 2.2.3.1 Tipos de Pesquisas

A Figura 14 apresenta um modelo que explica trabalhos de pesquisa em aplicações que envolvem *linked data*, classificando-os em três níveis: os tipos de questões de investigação que solicitam, os tipos de resultados que produzem e o caráter da validação que fornecem. Este modelo pertence à engenharia de *software* e vem evoluindo ao longo de vários anos, desde a versão apresentada inicialmente por Mary Shaw, na *International Conference on Software Engineering (ICSE)*, em 2001.

As pesquisas em engenharia de *software* respondem a perguntas sobre métodos de desenvolvimento ou análise, sobre detalhes do projeto ou avaliação de um caso particular, sobre generalizações, classes de sistemas ou técnicas, ou sobre questões exploratórias visando à

existência ou à viabilidade de uma tarefa (SHAW, 2002).

Figura 14 – Tipos de Pesquisas em Aplicações *Linked Data*



As contribuições tangíveis nas pesquisas em engenharia de *software* podem ser procedimentos ou técnicas para o desenvolvimento ou análise, podem ser modelos que generalizam a partir de exemplos ou podem ser ferramentas específicas, soluções ou resultados sobre sistemas particulares (SHAW, 2002).

O último nível do modelo descreve os tipos de validação para suportar os resultados da pesquisa. É essencial selecionar a forma de validação apropriada para o tipo de resultado e o método utilizado para obter o resultado (SHAW, 2002).

A Figura 14 exhibe a resposta da questão P1, onde se visualiza que a combinação mais utilizada nos tipos de pesquisa que tratam de visualização em SIG e *linked data* foram perguntas sobre o método ou meio de desenvolvimento; soluções, protótipos ou avaliações como resultado e exemplos como forma de validação. Nesse caso, há a tendência em saber como criar ou automatizar e qual o melhor jeito de fazê-lo, sendo testado por meio de um sistema que, em execução, incorpore ou seja portador do resultado, ou, ainda, que a sua implementação ilustre um princípio que pode ser aplicado em outros lugares. O uso de exemplo é adequado a esta combinação, sendo uma evidência convincente da validação do resultado obtido, como um

sistema desenvolvido.

Pode-se observar também que a maioria dos relatórios, que respondem a perguntas sobre generalização ou caracterização, utiliza a persuasão como forma de validação. Nesse cenário, a validação apenas pela persuasão raramente é suficiente para um trabalho de pesquisa. Porém, se a pergunta original for sobre viabilidade, um sistema em funcionamento, mesmo sem análise, pode ser suficiente (SHAW, 2002). Ao verificar esse novo cenário na Figura 14, conclui-se que menos da metade das publicações que estudam a viabilidade, e resultam em um sistema em funcionamento, utilizam a persuasão como forma de validação.

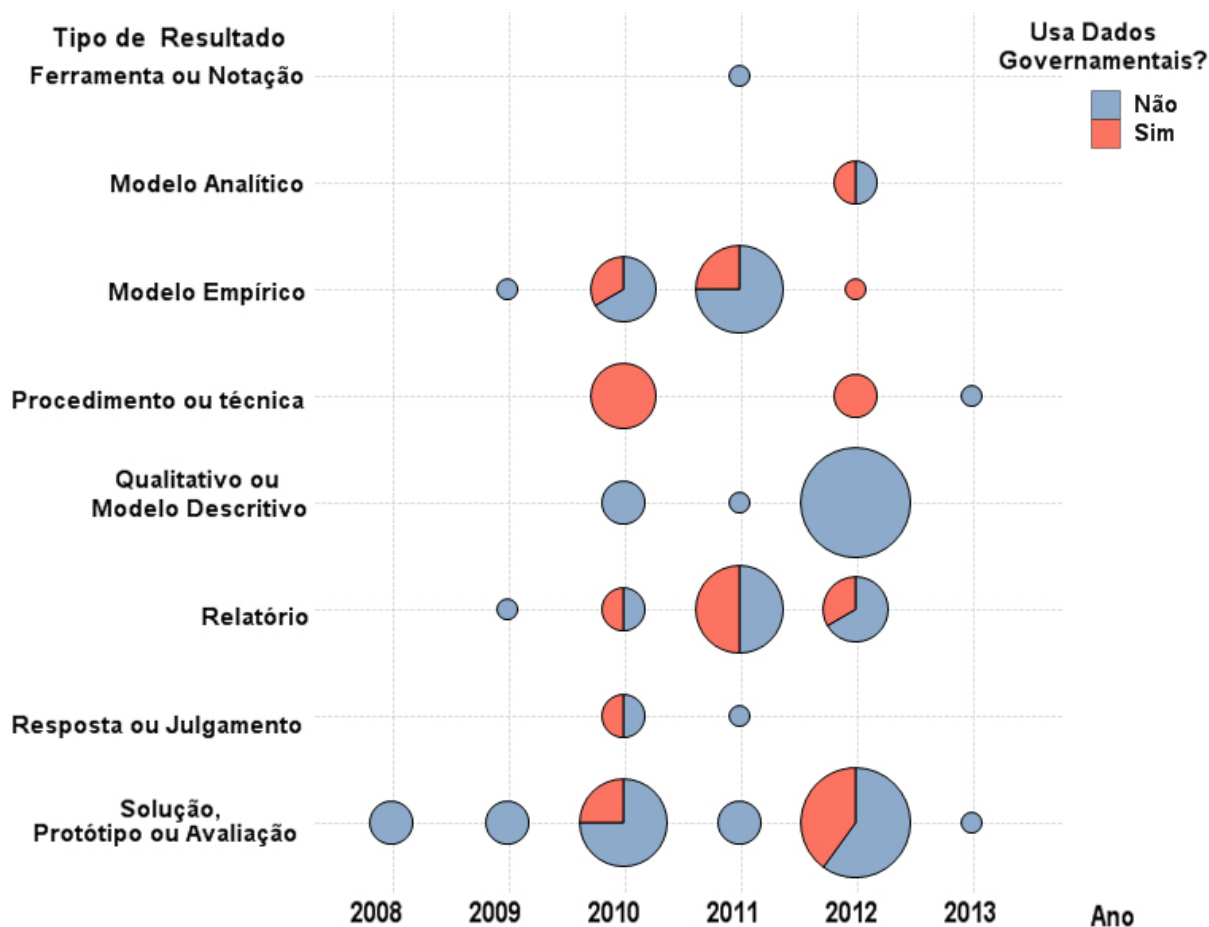
Quanto aos tipos de resultados, prevaleceram aqueles que abordavam uma solução de aplicação para um problema que demonstra o uso dos princípios da engenharia de *software*. Os tipos de perguntas que mais têm sido explorados são sobre método ou meio de desenvolvimento. Os métodos para análise ou avaliação são os tipos de perguntas menos explorados. Em relação ao tipo de validação, as soluções apresentadas têm explorado mais a experiência, seguida de exemplo e persuasão. Somente um artigo abordou a afirmação como forma de validação e deve ser visto como um ponto positivo, já que, nesse caso, nenhuma tentativa séria foi utilizada para avaliar o resultado.

Percebe-se uma lacuna nas pesquisas que geram uma ferramenta ou notação. Somente um artigo foi caracterizado como uma nova ou melhor maneira de fazer alguma tarefa, medição técnica ou avaliação, incluindo técnicas operacionais para execução, representação, gestão e análise, mas excluindo os que recomendam diretrizes.

### 2.2.3.2 Análise Temporal

A Figura 15 ilustra a combinação entre tipos de resultados, ano das pesquisas e se estas fizeram uso de dados governamentais. Ao observar esta figura, é possível notar que, a partir de 2010, houve um crescimento no interesse em uso de dados governamentais nas pesquisas aplicadas à *linked data*. Esse ponto tende a aumentar com a influência da Lei nº 12.527, a Lei de Acesso à Informação, sancionada em 18 de novembro de 2011 e em vigor em 16 de maio de 2012. De acordo com o seu regulamento, "é dever dos órgãos e entidades públicas promover, independentemente de requerimentos, a divulgação em local de fácil acesso, no âmbito de suas competências, de informações de interesse coletivo ou geral por eles produzidas ou custodiadas".

A Figura 15 responde a questão P2, ao apontar que grande parte das publicações sobre a temática enfocada na pesquisa foi em 2012, 2010 e 2011 com 19, 16 e 13 artigos, respectivamente, sendo possível analisar o desenvolvimento do enfoque da pesquisa no decorrer do tempo, as características, resultados e utilização de conhecimentos acadêmicos e científicos produzidos por diversos pesquisadores. Como é possível perceber, o estudo gera várias possibilidades de futuras pesquisas e contribui para uma visão mais ampla sobre o assunto *linked data*. Além disso, fornece vários insumos para enriquecer a discussão sobre o rumo das pesquisas e

Figura 15 – Características das Pesquisas sobre Visualização em SIG e *Linked Data*

as prováveis tendências nesse campo de pesquisa.

Como resposta à P3, observa-se na Figura 15, que o tipo de resultado mais obtido, utilizando-se dados governamentais, foi procedimento ou técnica, seguido por relatório. Ferramenta ou notação e modelo descritivo ou qualitativo não obtiveram nenhuma pesquisa com uso de dados governamentais, o que indica lacunas a serem exploradas por pesquisadores ao relacionar o uso de dados governamentais com visualização de *linked data* e SIG.

No entanto, é importante reconhecer a limitação da pesquisa, no que diz respeito às palavras-chave, que, na área de engenharia de *software*, não são padronizadas, podendo ser específicas de um segmento de conhecimento ou idioma. Portanto, devido à escolha em utilizar palavras-chave e *strings* de pesquisa, há um risco de que alguns estudos relevantes tenham sido omitidos.

O resultado aponta que grande parte das publicações sobre a temática enfocada na pesquisa foram em 2012, 2010 e 2011 com 19, 16 e 13 artigos, respectivamente, sendo possível analisar o desenvolvimento do enfoque da pesquisa no decorrer do tempo, as características, resultados e utilização de conhecimentos acadêmicos e científicos produzidos por diversos pesquisadores.

Esta pesquisa apresentou os resultados de um estudo bibliométrico sobre os termos “*web* semântica”, “visualização de *linked data*” e “Sistema de Informação Geográfica”. Foi desenvolvido com o propósito de disponibilizar aos pesquisadores e interessados nos temas um mapeamento sobre as características das pesquisas que tratam dos assuntos abordados.

O estudo utilizou técnicas de revisão sistematizada da literatura para a captura de dados que, contextualizados, possibilitaram a identificação de padrões e tendências da literatura científica. O estudo mostrou, assim, que pesquisas desse tipo podem ser promissoras por auxiliarem os pesquisadores a identificar embasamento teórico nessa área de estudo. Essa técnica permite lidar com o desafio de agrupar informações e traçar perfis representativos no campo de estudo de *linked data* e GIS, além do enfoque da pesquisa no decorrer do tempo. Apesar de emergente, conclui-se que o número de pesquisas nesta área é crescente e se concentra principalmente no desenvolvimento de soluções como exemplos, mas ainda com pouco uso de dados governamentais. Esses pontos servem de insumo para os pesquisadores desta área nos próximos anos e contribui para uma visão ampla sobre o assunto *linked data*. Para novas pesquisas nesse campo, esta revisão sistemática da bibliografia é relevante ao permitir a visualização do enquadramento de futuras pesquisas em relação aos trabalhos já realizados.

Para o proposto neste trabalho, esta revisão sistemática da bibliografia é relevante, por permitir a visualização do enquadramento do trabalho em relação aos trabalhos já realizados. Esta pesquisa, segundo o modelo de Shaw, situa-se no quadrante método ou meio de desenvolvimento, no que corresponde ao tipo de pergunta, solução, protótipo ou avaliação, no que tange ao tipo de resultado e avaliação como forma de validação. Este é o quadrante onde há a maioria dos trabalhos sobre *linked data*, porém, a forma de validação mais utilizada é o exemplo. Quanto ao uso de dados governamentais, o trabalho proposto encontra-se no quadrante solução, protótipo ou avaliação, no que se refere ao tipo de resultado e no ano de 2014, utilizando dados governamentais na pesquisa.

### 2.2.3.3 Trabalhos Relacionados

Um dos primeiros exemplos do uso de sistema de informação geográfica foi feito por John Snow, mostrando a relação entre a distribuição de água e os surtos de cólera em Londres, em 1854, realizado por meio da ligação de dados públicos sobre água contaminada e doenças (JOHNSON, 2006).

Na pesquisa de Nurefşan Gür, Laura Diaz e Tomi Kauppinen, utilizou-se *linked open data* para a publicação de dados relacionados à saúde, como doenças, transtornos, genes, drogas e a visualização em uma tecnologia denominada *geo web*. Para isso, o caso de uso estudado foi o CSISP - *Research Center of Public Health* (Centro de Pesquisa de Saúde Pública), baseado em três domínios conceituais: saúde, espacial e estatístico e seguindo os princípios *linked data*. Por fim, utilizou-se uma infraestrutura integrando as tecnologias geoespaciais e *web* semântica para exibir os índices de mortalidade de doenças específicas em um formato espaço-temporal



(GüR; DÍAZ; KAUPPINEN, 2012).

Na publicação de Vilches-Blázquez *et al.* (2010), foi apresentada uma sequência de procedimentos utilizados para o desenvolvimento de um aplicativo que faz uso de vários *datasets* públicos heterogêneos, sobre a Espanha, relacionados especificamente a unidades administrativas, hidrografia e unidades estatísticas. O aplicativo tem como objetivo analisar as relações existentes entre a zona costeira espanhola e diferentes variáveis estatísticas, tais como população, desemprego, habitação, indústria, comércio e construção. Além de fornecer orientações metodológicas para a geração, publicação e exploração de *linked data* a partir desses *datasets*, utilizaram-se recursos para lidar com a informação geométrica dos dados.

Observa-se que todos os trabalhos relacionados geram um arquivo RDF e a visualização em um SIG. Porém, nenhum combinou dados de um nicho específico com dados estatísticos do local em questão, como visto nesta pesquisa. É importante notar o uso de dados governamentais em todos os trabalhos relacionados.

### 3 METODOLOGIA

O presente estudo realizou uma pesquisa de caráter experimental, com base em prova de conceito. A pesquisa é do tipo experimental, caracterizada por verificar a relação entre causa e efeito (KIDDER, 2007). Conforme Gil (1996)s, a pesquisa experimental consiste em determinar um objeto de estudo, selecionar as variáveis que seriam capazes de influenciá-lo, definir as formas de controle e observação dos efeitos que a variável produz no objeto. A pesquisa genuinamente experimental pressupõe algum tipo de intervenção sobre o grupo estudado e a verificação dos efeitos dessa intervenção.

Realizar estudos experimentais provê um meio de testar, refinar e refutar a compreensão sobre um determinado fenômeno. Essa compreensão permite determinar a influência dos vários fatores sobre os efeitos de tecnologias e métodos (JURISTO; MORENO, 2001). O desenvolvimento de *software* é dependente de diversas tecnologias, muitas das quais não possuem evidências suficientes sobre potenciais benefícios, limitações, custo de implantação e riscos associados. Caso a engenharia de *software* fizesse um uso intenso e sistemático de uma abordagem baseada em evidência, seria possível caracterizar uma determinada tecnologia em uso. Através dessa caracterização, seria possível indicar, com níveis razoáveis de segurança, o que funciona, o que não funciona e sob quais circunstâncias (TRAVASSOS *et al.*, 2008).

Para atender a essa finalidade, a engenharia de *software* deve prover meios pelos quais melhores evidências, provenientes da pesquisa, possam ser integradas com experiência prática e valores humanos no processo de tomada de decisão, considerando o desenvolvimento e a manutenção do *software*. A Engenharia de *Software* baseada em evidência faz uso, essencialmente, de estudos experimentais (KITCHENHAM; DYBA; JORGENSEN, 2004).

Segundo Wohlin *et al.* (2000), a experimentação provê, de forma sistemática, disciplinada e controlada, a avaliação de processos e atividades humanas. Estudos experimentais são úteis para entender as influências nos vários fatores que determinam os efeitos das tecnologias e métodos. Os resultados de pesquisas experimentais são informações chave para muitos tipos de atividades de melhoria de *software*. O conhecimento sobre o comportamento de técnicas, métodos e processos auxilia a antecipar seus impactos no processo de *software* e na qualidade do produto. Estudos experimentais são importantes nas pesquisas em engenharia de *software*, pois permitem avaliar e comparar novas teorias, métodos, técnicas e linguagens (JURISTO; MORENO, 2001).

O método experimental é visto como um movimento para aumentar a quantidade de experimentos em ciência da computação (TEDRE, 2011). Nessa área, a experimentação visa a desenvolver uma base de evidência para compreensão e intervenção científica nos processos de desenvolvimento de tecnologias de *software*.

A pesquisa experimental caracteriza-se por manipular diretamente as variáveis relacionadas com o objeto de estudo. Neste tipo de pesquisa, a manipulação das variáveis proporciona o estudo da relação entre as causas e os efeitos de determinado fenômeno. Esta pesquisa pretende dizer de que modo ou por que o fenômeno é produzido (CERVO; BERVIAN; SILVA, 2007).

O núcleo da noção de experimentos, em ciência da computação, é a construção de sistemas, seja de *hardware*, seja de *software*. Isso é feito não só para estudar esses sistemas, como também para demonstrar sua viabilidade (HARTMANIS, 1994). Constitui também como objetivo deste trabalho desenvolver um protótipo de ferramenta que atue como uma prova de conceito para ilustrar como o conhecimento será utilizado. O cenário de utilização está na etapa de planejamento, na qual o conhecimento dos subdomínios de estratégia e ambiente de estudo e método de pesquisa são relevantes (TRAVASSOS *et al.*, 2008).

Quanto à natureza, a pesquisa foi aplicada por ser uma investigação original, concebida pelo interesse em adquirir novos conhecimentos, porém, orientada para uma aplicação prática (CASARIN; CASARIN, 2011). Segundo esses autores, a pesquisa aplicada é realizada ou para determinar os possíveis usos para as descobertas da pesquisa básica ou para definir novos métodos ou maneiras de alcançar a solução de problemas específicos.

Utiliza-se a pesquisa aplicada para estudar o problema em um contexto e buscar soluções para os problemas no ambiente específico. Este tipo de pesquisa é relacionado à prática, mas deve estar também atrelado a uma reflexão teórica (MASCARENHAS, 2012).

Em relação à abordagem do problema a ser investigado, a pesquisa foi considerada qualitativa, por ser predominantemente descritiva. Conforme Casarin e Casarin (2011), os objetivos de uma pesquisa qualitativa envolvem a descrição de certo fenômeno, caracterizando sua ocorrência e relacionando-o com outros fatores. Há também a preocupação de explicar a sua ocorrência baseando-se em modelos contextuais variados. Assim, o objetivo da pesquisa está relacionado ao contexto no qual o objeto pesquisado está inserido. Além disso, existe uma grande preocupação em fazer associações entre as variáveis que possam contribuir para explicar o que está sendo pesquisado (CASARIN; CASARIN, 2011).

Entende-se por objetivo a busca de solução para um problema, a explicação para um determinado fenômeno ou, simplesmente, novos conhecimentos que venham a enriquecer os já existentes sobre um determinado tema (CASARIN; CASARIN, 2011). Nesse aspecto, a pesquisa será explicativa, pois procura identificar fatores que determinam ou contribuem para a ocorrência dos fenômenos (GIL, 1996).

Ao analisar as definições de tipologias de pesquisas apresentadas, observa-se que a pesquisa do tipo prova de conceito e experimento apresenta-se como uma boa opção para o desenvolvimento deste trabalho. Nas pesquisas explicativas, o principal método utilizado é o experimental, sendo mais comum nas áreas de saúde e ciências exatas (CASARIN; CASARIN,

2011).

### 3.1 Objeto da Pesquisa

A unidade de análise desta pesquisa é a bacia hidrográfica do rio Doce que, segundo o próprio Comitê (CBH-DOCE), apresenta uma significativa extensão territorial, cerca de 83.400 km<sup>2</sup>, dos quais 86% pertencem ao estado de Minas Gerais e, o restante, ao estado do Espírito Santo. Abrange, total ou parcialmente, áreas de 228 municípios, sendo 202 em Minas Gerais e 26 no Espírito Santo e possui uma população total da ordem de 3,1 milhões de habitantes.

O rio Doce, com uma extensão de 853 km, tem como formadores os rios Piranga e Carmo, cujas nascentes estão situadas nas encostas das serras da Mantiqueira e Espinhaço, onde as altitudes atingem cerca de 1.200 m. Seus principais afluentes são: pela margem esquerda os rios Piracicaba, Santo Antônio e Suaçuí Grande, em Minas Gerais, Pancas e São José, no Espírito Santo; pela margem direita, os rios Casca, Matipó, Caratinga-Cuieté e Manhuaçu, em Minas Gerais, e Guandu, no Espírito Santo (CBH-DOCE, 2013).

A partir de dados da FJP, em 2011, o PIB da bacia do rio Doce representa em torno de 15% do PIB do Estado de Minas Gerais (estimado em 122 bilhões em 2001), sendo que somente o município de Ipatinga contribui com 5,4% daquele valor.

Segundo o Anuário Estatístico do Brasil (IBGE, 2013), residem na bacia cerca de 3.100.000 habitantes, com a população urbana representando 68,7% da população total. Observa-se que a taxa de crescimento urbano é inferior às verificadas nos dois Estados. O êxodo rural é generalizado na área da bacia. Entre os anos de 1970 e 1991, a região perdeu 615.000 habitantes (IBGE). Em Minas Gerais, a bacia do rio Doce é caracterizada como a região que mais perdeu população: 615.259 habitantes entre 1970 e 1991 (UFMG, IBGE). Na região do médio rio Doce, entre Tumiritinga e Aimorés, houve uma redução demográfica da ordem de 40% no mesmo período (IBGE).

### 3.2 Origem dos Dados

Como o objetivo desta dissertação é criar um novo *dataset* com dados de enchentes da Bacia do Rio Doce, torna-se necessária a coleta de dados de diferentes fontes, inclusive dados governamentais. Neste caso, coletou-se dados da ANA, ANEEL, CEMIG, Instituto Mineiro de Gestão das Águas (IGAM) e CPRM. Em cada órgão, os dados foram coletados via *File Transfer Protocol* (FTP) ou diretamente, pelo site da organização.

Conforme visualizado na Tabela 5, outras fontes de dados foram utilizadas objetivando agregar informações, como, por exemplo, o *Geonames*, que é um tipo de banco de dados que disponibiliza metadados geográficos, e contém um vocabulário para adicionar a semântica necessária para a distribuição de dados. A *DBpedia* é outro repositório que extrai informações estruturadas da *Wikipedia*, permitindo efetuar consultas e associar outros conjuntos de dados e disponibilizá-los na *web*.

Tabela 5 – Fonte, descrição e formato dos dados utilizados no trabalho.

Fonte	Descrição	Formato
ANA	Precipitação e Nível dos Rios	DAT
ANEEL	Precipitação e Nível dos Rios	CSV
CEMIG	Precipitação	CSV
IGAM	Precipitação	TXT
CPRM	Nível dos Rios	Banco de Dados
Portal Transparência MG	Repasso de Investimentos	CSV
IBGE	População e Altitude	CSV
Portal da Saúde	Doenças	CSV
PNUD	IDH da Cidade	CSV
Portal Dados Abertos	Obras do PAC	XML
Geonames	Nomes geográficos, latitude e longitude	RDF
DBPEDIA	Dados gerais de cidades	RDF

Fonte: Elaborado pela autora, 2013.

Segundo Nathwani, Lind e Pandey (2009), evidências empíricas sugerem que os ativos econômicos das cidades, incluindo edifícios, sistemas de transporte, serviços públicos, entre outros, são tipicamente cinco vezes o PIB produzido pela população. Essa experiência também mostra que, em geral, as perdas em inundações são cinco vezes maiores do que o PIB da população afetada (NICHOLLS *et al.*, 2007).

Nathwani, Lind e Pandey (2009) afirmam que, em geral, cidades de países mais ricos têm níveis bem melhores de proteção do que aqueles no mundo ainda em desenvolvimento, e essa proteção é também fortemente influenciada por questões culturais, políticas e históricas.

As cidades situadas nos deltas dos rios tendem a ter maior risco de inundação, como resultado de sua tendência a se localizarem em altitudes mais baixas. Esta tendência poderá ser confirmada na aplicação desenvolvida, cruzando os dados de altitude dos municípios e dos pontos de enchentes na Bacia do Rio Doce.

A inclusão dos dados sobre o IDH parte do cenário descrito por Chaudhry e Ruyschaert (2007), em que inundações e tempestades mais intensas poderiam retardar o progresso do desenvolvimento humano nas áreas populacionais afetadas.

As enchentes, secas e outros choques climáticos podem forçar as famílias pobres a retirar as crianças da escola para aumentar a oferta de trabalho, ou para reduzir os gastos em saúde e nutrição. Muitas pessoas não têm a proteção básica de saúde e, na escola, a taxa de evasão de filhos é elevada. Para esse grupo, até mesmo uma pequena queda na renda ou perda de oportunidades de emprego relacionados aos efeitos das inundações teria consequências adversas para a nutrição, saúde e educação. Os pobres enfrentam um risco duplo. Além de serem muito mais propensos a viver em áreas vulneráveis a inundações, eles são menos propensos a viver em casas permanentes e mais robustas (CHAUDHRY; RUYSSCHAERT, 2007).

O programa Cidade Melhor, do Programa de Aceleração do Crescimento (PAC), destina recursos à prevenção de áreas de risco, como prevenção de novos deslizamentos, contenção de encostas em áreas de risco, controle de enchentes e inundações com obras de drenagem, além da redução de áreas vulneráveis a deslizamentos. A justificativa para a inclusão desses dados na aplicação segue a linha de pensamento de Ellis (1995), em que o planejamento de atividades urbanas relacionadas à água deve estar integrado ao próprio planejamento urbano, incluindo o desenho da malha urbana e sua expansão, o zoneamento de atividades, a rede viária e de transportes, aspectos paisagísticos, entre outros.

Quanto à gestão, Pompeo (2000) ressalta a importância da definição de atribuições e formas de financiamento que permitam garantir a ação permanente e direta, em programas de longo prazo. Deve-se reconhecer que a drenagem urbana é parte de um sistema de gestão e não apenas algumas obras de controle de enchentes que são definidas após a constatação de algum problema. Além disso, deve-se lembrar também que é necessária a existência de uma organização institucional apta a oferecer respostas durante e após uma situação de enchente. Um evento hidrológico extremo é capaz de criar um colapso generalizado no sistema de drenagem, sendo necessária uma estrutura de defesa civil para organizar ações nesse contexto (POMPEO, 2000).

Pompeo (2000) esclarece a importância de ações diretas para tratamento de cheias em áreas urbanas, como administração e gestão, planejamento, projeto e execução de obras

de engenharia, operação de comportas e bombas, e recuperação e limpeza dos sistemas de drenagem.

O indicador de doenças foi adicionado ao *dataset* pela relação com as enchentes e o aumento do surto de doenças após o desastre. Apesar da ausência de evidências de que restos mortais, após desastres naturais, representam um risco para epidemias (MORGAN, 2004), a súbita presença de grande número de corpos mortos na área afetada pelo desastre pode aumentar as preocupações de surtos da doença (GOYET, 2004).

As inundações podem causar traumatismos, afogamentos e outros agravos à saúde que excedem a capacidade de resposta dos serviços locais de saúde. No Brasil, as principais ocorrências em relação à saúde humana, após as inundações, são os surtos de leptospirose, transmitida pelo contato com água ou lama contaminada pela urina de roedores contaminados. Frequentemente, as inundações levam à contaminação da rede pública de abastecimento de água e, diversas vezes, a população utiliza essa água expondo-se ao risco de ingerir bactérias, vírus e parasitas. Com isto, a população poderá desenvolver doenças como a cólera, febre tifoide, hepatite A e infecções por parasitas intestinais. Quanto à dengue, posteriormente à inundação ocorre a formação de muitos criadouros em recipientes naturais e artificiais, favorecendo, dessa forma, o desenvolvimento do vetor e aumentando o índice de infestação por *Aedes aegypti* (BRASIL, 2011).

O risco da propagação de doenças transmissíveis após os desastres está associado principalmente ao tamanho e às características da população, especificamente à proximidade de água potável e rede de esgoto, ao estado nutricional da população deslocada, ao nível de imunidade a doenças, e o acesso aos serviços de saúde (NOJI, 1996).

Surtos de doenças diarreicas podem ocorrer depois que a água potável foi contaminada e têm sido relatados após enchentes. Em 2004, um surto de doença diarreica após uma inundação, em Bangladesh, envolveu mais de 17.000 casos (QADRI *et al.*, 2004). Em 1998, uma epidemia de cólera (mais de 16.000 casos) em Bengala Ocidental foi atribuída a inundações anteriores (SUR *et al.*, 2000), e em janeiro-março de 2000, uma inundação em Moçambique levou ao aumento na incidência de diarreia (KONDO *et al.*, 2002). Um estudo realizado na Indonésia, em 1992-1993, identificou as inundações como um fator de risco para doenças diarreicas causadas por *Salmonella enterica* tipo A (febre paratifoide) (VOLLAARD *et al.*, 2004).

Em regiões de ocorrência de hepatite E, os surtos são frequentemente precedidos de chuvas fortes e inundações (AGGARWAL; KRAWCZYNSKI, 2000). A leptospirose é uma doença bacteriana com potencial epidêmico, que pode ser transmitida por contato direto com a água contaminada. Inundações facilitam a propagação do organismo pela proliferação de roedores e a proximidade destes com os seres humanos, em terrenos mais altos, fora da área inundada. Surtos de leptospirose ocorreram no Taiwan, China, em 2001 (YANG *et al.*, 2005), em Mumbai, na Índia, após alagamento em 2000 (KARANDE *et al.*, 2003), na Argentina, após a inundação em 1998 (VANASCO *et al.*, 2002), e na região de Krasnodar na Rússia, em 1997

(KALASHNIKOV *et al.*, 2003). Depois de um surto de leptospirose relacionada às inundações no Brasil, em 1996, a análise espacial indicou que as taxas de incidência de leptospirose dobraram dentro das áreas sujeitas a inundações no Rio de Janeiro (BARCELLOS; SABROZA, 2001).

Inicialmente, as inundações podem lavar os criadouros de mosquitos existentes, porém, a água parada, causada por chuvas fortes ou transbordamento de rios, pode criar novos locais de reprodução (principalmente abrigos artificiais). Essa situação pode resultar (com atraso de algumas semanas) em um aumento da população de vetores e do potencial para a transmissão de doenças. A aglomeração de hospedeiros infectados, somados à infraestrutura de saúde pública debilitada, e às interrupções dos programas de controle em curso são todos fatores de risco para a transmissão da dengue e febre amarela (LIFSON, 1996).

Os surtos de malária, após as inundações, são um fenômeno bastante conhecido. As inundações periódicas relacionadas ao *El Niño* têm sido associadas a epidemias de malária na região costeira seca do norte do Peru (GAGNON; SMOYER-TOMIC; BUSH, 2002). Já o tétano não é transmitido de pessoa para pessoa, mas é causado por uma toxina liberada pelo bacilo anaeróbico *Clostridium tetani*. Feridas contaminadas são encontradas particularmente em populações onde os níveis de cobertura vacinal são baixos e estão associados com a doença e morte por tétano. Um conjunto de 106 casos de tétano, incluindo 20 mortes, ocorreu em Aceh e chegou duas semanas e meia depois do tsunami (WORLD HEALTH ORGANIZATION, 2005).

A população dos municípios foi incluída no *dataset* por ser a parte mais afetada pelas enchentes. Bouwer e Vellinga (2007) afirmam que as perdas econômicas causadas por desastres naturais, em geral, aumentaram mais rapidamente do que o crescimento econômico médio nacional, o que indica a possibilidade de a exposição a desastres naturais ter aumentado desproporcionalmente em áreas costeiras densamente urbanizadas. Como o Brasil é densamente povoado e situado em uma área propensa a inundações, o estudo da alta exposição causada pelo crescimento habitacional e acumulação de bens em áreas de baixa altitude será particularmente relevante para o país (MAASKANT; JONKMAN; BOUWER, 2009).

Huq *et al.* (2007) descreve que as áreas urbanas apresentam sempre algum risco de inundação quando as chuvas ocorrem. Edifícios, estradas, infraestrutura e áreas pavimentadas evitam que a chuva se infiltre no solo. Nas cidades bem governadas, isso raramente é um problema, pois a boa disposição do solo é preparada para receber tempestades e a drenagem superficial pode facilmente ser incorporada ao tecido urbano, somado a medidas complementares para proteção contra inundações - por exemplo, o uso de parques e outras áreas abertas como lugares para acomodar, com segurança, a água vinda de tempestades inesperadas. Mas, em cidades mal governadas, isso não acontece - e é comum que os edifícios ou infraestrutura a serem construídos obstruam os canais de drenagem naturais. Em cidades com rápido crescimento, é comum a reconstrução dos canais naturais de drenagem. Chuvas fortes e, ou, prolongadas produzem grandes volumes de água na superfície de toda a cidade, o que pode facilmente sobrecarregar os



sistemas de drenagem - especialmente se houver disposição inadequada ou falta de manutenção dos esgotos (por exemplo, muitos estão cheios de lodo ou obstruídos pelo lixo) (HUQ *et al.*, 2007).

Muitas cidades estão localizadas em locais perigosos, com relação aos riscos de tempestades e inundações. Uma estimativa indica que cerca de 10 milhões de pessoas são atualmente afetadas a cada ano por inundações e que esse número irá aumentar em todos os cenários de mudanças climáticas (NICHOLLS, 2004).

Após a estruturação de todos esses dados, foi usado o modelo RDF para representar a informação, visto que esse é o padrão para representação de informações na *Web*, conforme descrito no Capítulo Revisão Sistemática da Literatura.

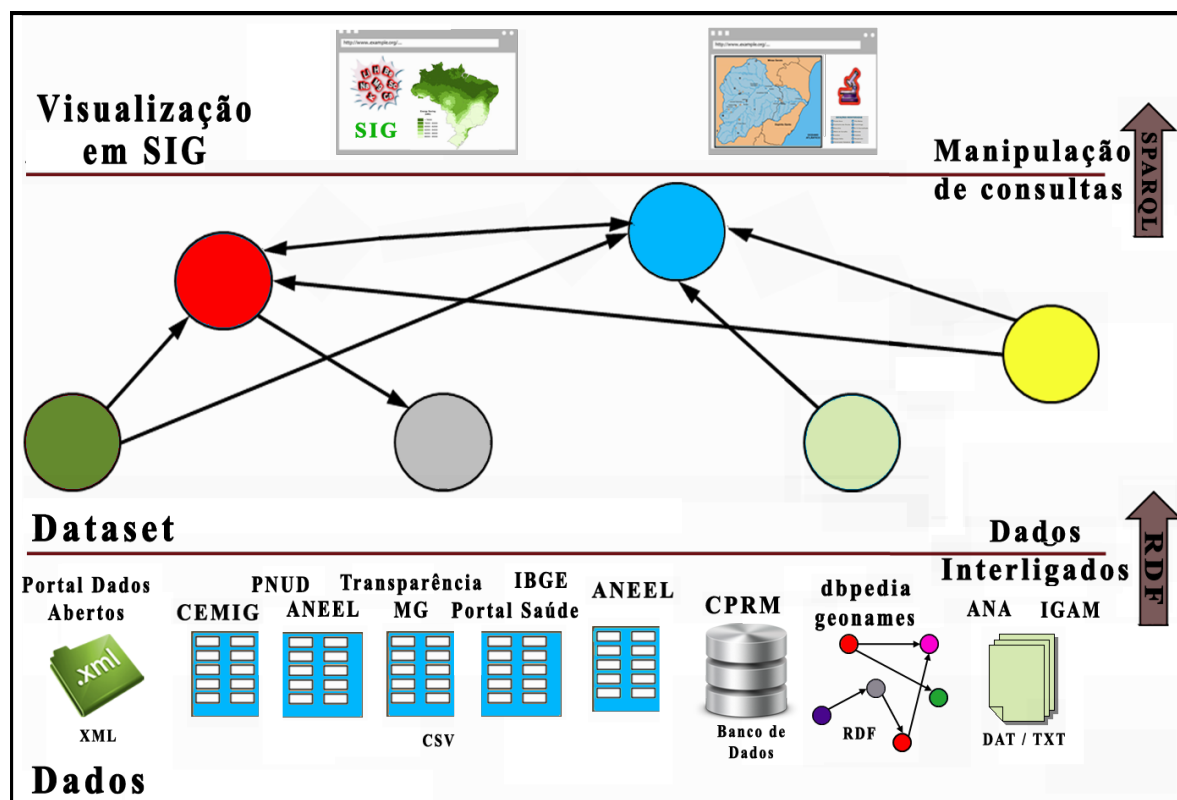
## 4 ARCABOUÇO CONCEITUAL

O propósito da arquitetura conceitual é direcionar o foco em uma decomposição do sistema, destacando os principais elementos e sem se aprofundar nos detalhes da especificação da interface. Os principais componentes são identificados, bem como a relação entre eles e os mecanismos da arquitetura. Ao focalizar os elementos e abstrações mais importantes, em lugar de detalhes técnicos, a arquitetura conceitual fornece um veículo útil para comunicar com o público não técnico, como gestão, marketing e usuários em geral (MALAN; BREDEMEYER, 2006).

O diagrama de arquitetura conceitual identifica os componentes do sistema, suas interconexões e responsabilidades, sendo análogo à visualização de plantas baixas que os arquitetos de construção utilizam para seus clientes (MALAN; BREDEMEYER, 2006).

A arquitetura proposta contempla os objetivos do trabalho, as respectivas metodologias e os conceitos técnicos apresentados no referencial teórico. Com essa arquitetura é possível, por meio das tecnologias e princípios de *linked data*, desenvolver uma solução capaz de receber dados de várias organizações, integrá-los e disponibilizá-los visualmente. A Figura 16 exhibe, em alto nível, a decomposição dos componentes que fazem parte da solução, assim como a interconexão entre eles.

Figura 16 – Visão geral da arquitetura da solução



Fonte: Elaborado pela autora, 2013.

Na Figura 16, estão as três camadas da arquitetura conceitual da solução proposta: na primeira camada estão os dados que serão utilizados, conforme descrito no tópico 3.2 deste trabalho. Esses dados, relativos a enchentes na Bacia do Rio Doce, estão em diferentes formatos e serão convertidos para o padrão RDF com o propósito de serem interligados e, assim, gerar o grafo RDF, que está ilustrado na segunda camada da arquitetura. Na última camada, será utilizado o SPARQL para efetuar as consultas nesses dados. O resultado será a combinação de todos os dados e a visualização geográfica destes em um SIG.

#### 4.1 Dados

Os dados, provenientes de diversos órgãos públicos e da nuvem de *Linked Open Data*, estavam em formatos variados (txt, dat, csv, xml, rdf). Esses dados foram armazenados em um

banco de dados e convertidos para o padrão RDF.

## 4.2 **Dataset**

O arquivo gerado a partir da conversão já é um dos resultados da pesquisa, o *dataset* contendo todas as informações dos níveis dos rios que compõem a Bacia do Rio Doce, assim como os níveis de atenção e alerta e as informações dos municípios, de forma conectada. Para responder às perguntas da pesquisa, as consultas SPARQL foram manipuladas e o resultado enviado ao SIG.

## 4.3 **Visualização em um SIG**

No topo da arquitetura, está a camada de aplicação, onde as informações são exibidas, através do SIG, em uma interface amigável e capaz de responder às perguntas propostas inicialmente.

## 5 IMPLEMENTAÇÃO

### 5.1 Dados

Os dados utilizados neste trabalho vieram de diversas fontes. Entre elas, agências governamentais. Todos esses dados estavam em formatos não estruturados e passaram por um tratamento antes da sua utilização no protótipo. Visto o esforço em promover a *Web* semântica, procurou-se seguir os padrões abertos conforme as recomendações da W3C, representando os *datasets* na forma de *linked data*. Com o uso de RDF, é possível expressar informações específicas sobre os recursos e relacionamentos entre eles. O processo de preparação dos dados consistiu nos seguintes passos:

- (a) Identificação dos dados necessários;
- (b) Busca e *download*;
- (c) Extração das informações desejadas;
- (d) Inserção dos dados em um banco de dados relacional.

No *website* <<http://hidroweb.ana.gov.br>>, coletaram-se dados gerados pelos seguintes órgãos governamentais:

- (a) ANA (Agência Nacional de Águas)
- (b) CEMIG (Companhia Energética de Minas Gerais)
- (c) COPASA (Companhia de Saneamento de Minas Gerais)
- (d) CPRM (Companhia de Pesquisa de Recursos Minerais)
- (e) DAEE/MG (Departamento de Águas e Energia Elétrica do Estado de Minas Gerais)
- (f) IGAM (Instituto Mineiro de Gestão das Águas)
- (g) INMET (Instituto Nacional de Meteorologia)

Cada fonte opera as medições em diferentes municípios e trechos dos rios. Os dados foram extraídos em formato TXT e foram convertidos em CSV, utilizando o *software* MS Excel. Outros dados, também referentes aos níveis dos rios, foram coletados diretamente em um servidor da CPRM, com a ajuda de um funcionário da empresa. Estes estavam em formato DAT e também foram convertidos em CSV, utilizando o mesmo *software*. Do *website* <http://dados.gov.br/>, coletaram-se dados referentes a obras do Programa de Aceleração do Crescimento (PAC) em Minas Gerais, no formato XML. Os dados sobre o IDH dos municípios fo-

ram obtidos através do *website* do PNUD (<http://www.pnud.org.br/>), e também já estavam em formato CSV. No Portal da Saúde, coletaram-se dados sobre ocorrência das seguintes doenças relacionadas a enchentes: Tétano, dengue, leptospirose, malária, hepatite A e C, febre tifóide e cólera. Esses dados foram encontrados em formato CSV. Os dados de população e altitude de cada município foram coletados diretamente no *website* do IBGE (<http://cidades.ibge.gov.br/>), em formato CSV.

Para converter as planilhas, arquivos CSV, arquivos XML, dados relacionais e outros documentos para o formato RDF, utilizou-se a plataforma D2RQ. O D2RQ foi escolhido para ser utilizado neste trabalho por uma série de fatores, dentre os quais destacam-se: a flexibilidade da linguagem de mapeamentos, a simplicidade dos comandos, o suporte da comunidade ativo e a geração de *dumps* RDF, tornando possível o reuso do *dataset* criado.

O D2RQ é uma plataforma para acessar bancos de dados relacionais na forma de grafos RDF (BIZER; SEABORNE, 2004). A plataforma é formada pelos seguintes componentes:

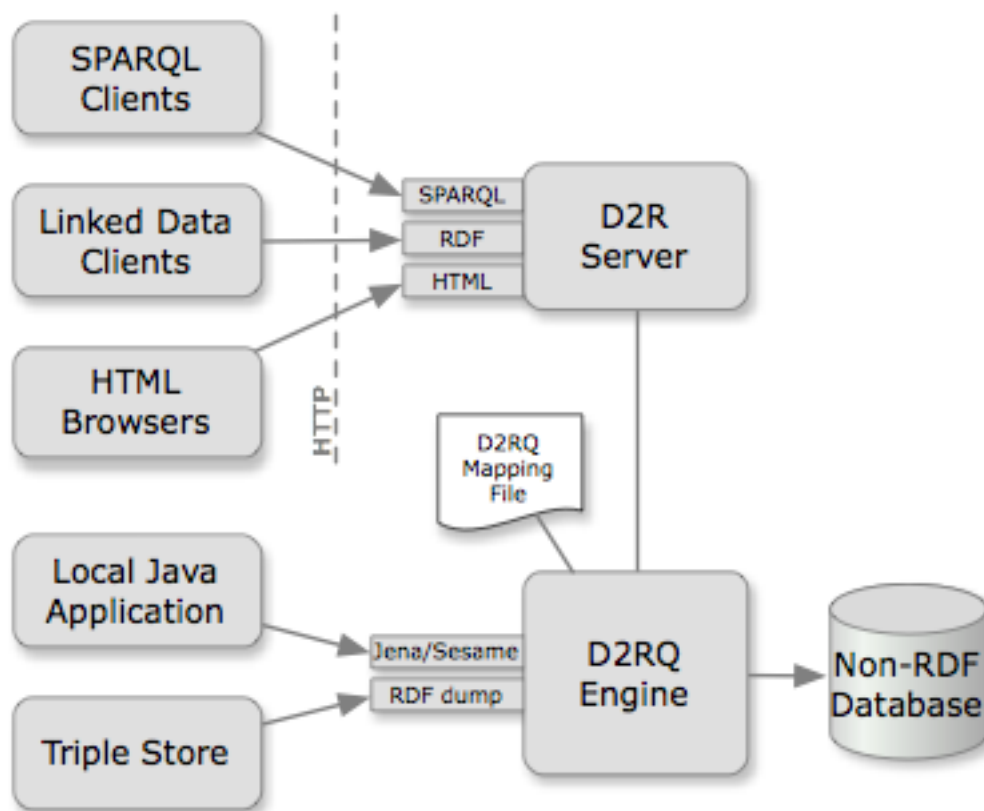
(a) Linguagem D2RQ: uma linguagem de mapeamentos declarativa para criação de correspondências entre conceitos do modelo relacional em termos do modelo RDF. Os mapeamentos nessa linguagem são documentos RDF, escritos usando a sintaxe N3 (BIZER; SEABORNE, 2004).

(b) Servidor D2R: um servidor HTTP que disponibiliza uma interface *Linked Data* e um *Spargl Endpoint* sobre o banco relacional (BIZER; CYGANIAK; HEATH, 2007).

(c) Motor de Regras D2RQ: é responsável por interpretar os mapeamentos D2RM. Utilizado pelo Servidor D2R para transformar consultas SPARQL em consultas SQL.

Esses componentes e a integração entre eles podem ser observados na Figura 17 que apresenta a arquitetura geral da plataforma D2RQ.

Figura 17 – Arquitetura da plataforma D2RQ



Fonte: Arquitetura criada por Cyganiak *et al.* (2012).

Conforme ilustrado na Figura 17, o banco de dados mapeado pelo D2RQ pode ser acessado de diversas maneiras, como por exemplo:

(a) *Dumps* RDF em RDF/XML ou em N-triplas.

(b) *Application Programming Interface* (API) RDF: D2RQ pode ser incorporado em aplicações Java para fornecer acesso a dados relacionais através do Jena e APIs do Sesame. As requisições de APIs são reescritas em instruções SQL e executadas no banco de dados. Esta tradução permite a publicação de RDF a partir de grandes bases de dados, em tempo real, e elimina a necessidade de replicar os dados em um armazenamento dedicado de triplas.

(c) *SPARQL Endpoint*: D2R Server fornece acesso remoto a um banco de dados D2RQ mapeado através do protocolo SPARQL.

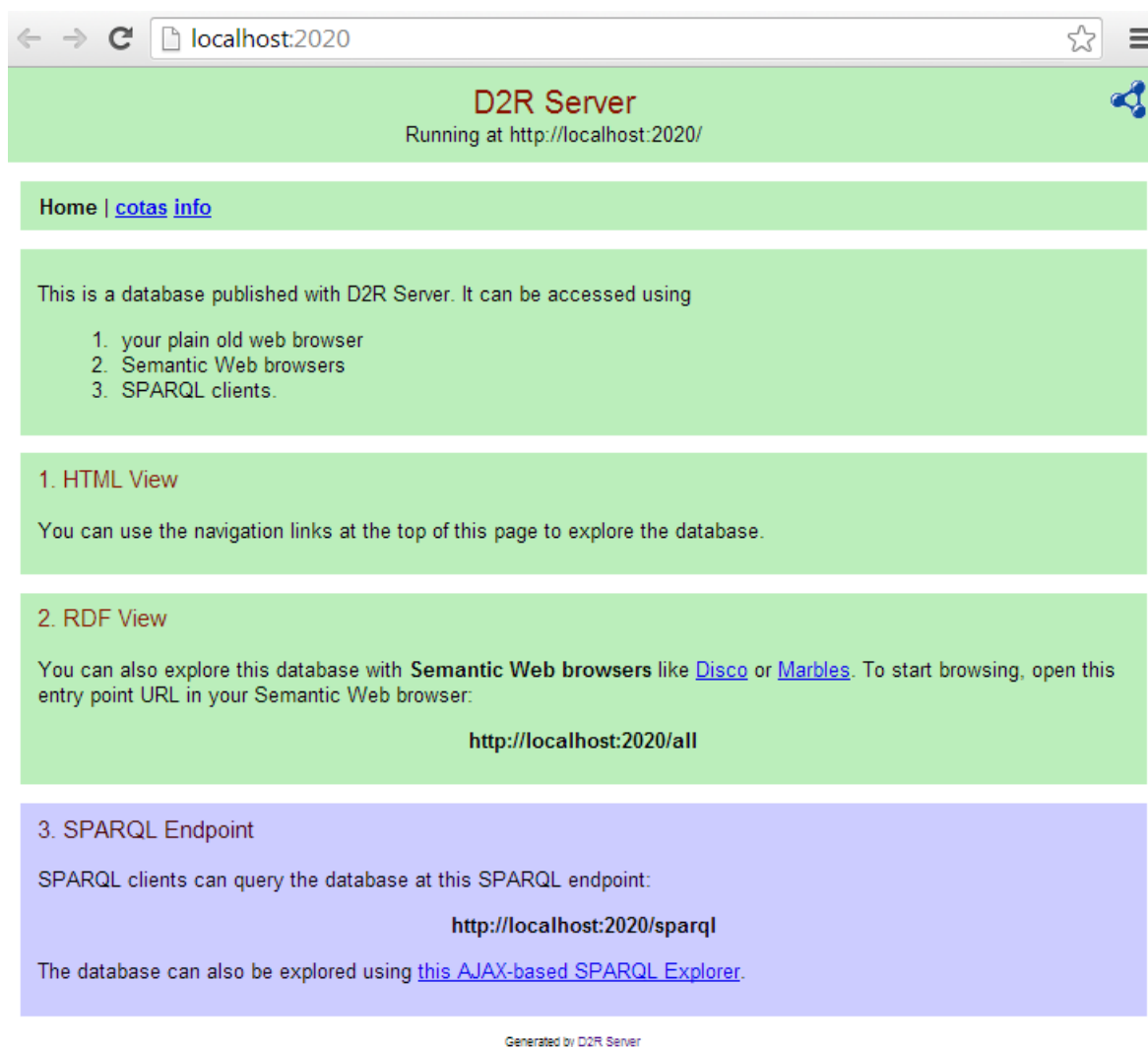
(d) *Linked Data*: Descrições RDF de entidades individuais no banco de dados podem ser acessadas por referência de URI.

(e) *Visão HTML*: D2R Server fornece uma visão simples em HTML para ajudar a escrever e depurar o mapeamento.

O *SPARQL Endpoint* criado durante o trabalho está ilustrado na Figura 18. Esta plata-

forma D2RQ permite a publicação do conteúdo em bases relacionais na *Web Semântica* através da ferramenta D2R-Server, que provê uma interface para navegar nos dados RDF.

Figura 18 – Interface do D2R-Server para consultas no SPARQL Endpoint



Fonte: Elaborado pela autora.

Além disso, essa ferramenta permite a submissão de consultas SPARQL. As requisições *Web* recebidas neste *endpoint* são reescritas em consultas SQL e submetidas ao banco de dados. A Figura 19 exibe a interface para a realização dessas consultas. Os resultados podem ser exibidos nos formatos HTML, XML ou JavaScript Object Notation (JSON).



Figura 19 – Interface do D2R-Server - Página Inicial

The screenshot shows the D2R-Server interface in a web browser. The address bar contains the URL: localhost:2020/snorql/?query=SELECT+DISTINCT+%3Fnor. The page title is "Snorql: Exploring http://localhost:2020/sparql".

**SPARQL:**

```

PREFIX db: <http://localhost:2020/resource/>
PREFIX dbpedia-owl: <http://dbpedia.org/ontology/>
PREFIX geo: <http://www.w3.org/2003/01/geo/wgs84_pos#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX map: <http://localhost:2020/resource/#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX dbpprop: <http://dbpedia.org/property/>
PREFIX vocab: <http://localhost:2020/resource/vocab/>
PREFIX grs: <http://www.georss.org/georss/point>

SELECT DISTINCT ?nomes
WHERE
{
  ?resource <http://xmlns.com/foaf/0.1/name> ?nomes
}
order by ?nomes

```

**Browse:**

- [Classes](#)
- [Properties](#)

Results:

**SPARQL results:**

nomes
"ABRE CAMPO"
"ACAIACA-JUSANTE"
"AFONSO CLAUDIO-MONTANTE"
"ASSARAI-MONTANTE"
"Abadia dos Dourados"
"Abaeté"
"Abre Campo"
"Acaiaca"
"Aguanil"
"Aimorés"
"Aiuruoca"
"Alagoa"
"Albertina"
"Alfenas"
"Alfredo Vasconcelos"
"Almenara"

Fonte: Elaborado pela autora.

O mapeamento define um grafo RDF virtual que contém informações do banco de dados de forma semelhante ao conceito de visualizações no SQL, exceto pela estrutura de dados

ser um grafo em vez de uma tabela relacional. O grafo RDF pode ser acessado de várias maneiras, dependendo do que for oferecido pela implementação. A plataforma D2RQ fornece acesso ao SPARQL, um servidor *Linked Data*, uma interface HTML simples e acesso, através da API Jena, aos bancos de dados D2RQ mapeados (BIZER; CYGANIAK, 2007).

O banco de dados é mapeado para os termos RDF usando *ClassMaps* e *PropertyBridges*. Um mapa de classe representa uma classe ou um grupo de classes similares da ontologia. Um mapa de classes especifica como os URIs são gerados para as instâncias da classe, através dos conjuntos de pontes de propriedades, que especificam como as propriedades de uma instância são criados.

A plataforma D2RQ pode se conectar a fontes de dados *Open Database Connectivity* (ODBC) através de um *driver Java Database Connectivity* (JDBC) específico para o banco de dados em questão. Os bancos de dados suportados são: Oracle, MySQL, PostgreSQL, SQL Server, HSQLDB e Interbase/Firebird (BIZER; CYGANIAK, 2007). Para este trabalho, o banco de dados utilizado foi o MySQL, pela simplicidade na instalação e na utilização e pelo prévio conhecimento da autora.

Conforme detalhado, essa ferramenta teve um impacto positivo e relevante no povoamento do *dataset* criado e na sua disponibilização como RDF e *linked data* através da *Web*.

Os dados foram inseridos no banco de dados MySQL através da ferramenta *freeware* Excel2MySQL, versão 1.7 e algumas modificações na estrutura da tabela gerada foram feitas após a conversão, como a alteração dos tipos de campos.

Consultando vocabulários de caráter geral e específicos da área de enchentes, todas as classes foram reutilizadas, não sendo necessária a criação de vocabulário. A Tabela 6 exibe os termos reusados.

Tabela 6 – Vocabulário Reutilizado

<b>Conceito</b>	<b>Termo Reutilizado</b>	<b>Tipo</b>
Município	purl:Town	<i>String</i>
Código do Município	loki:identCode	<i>Integer</i>
Nome do Município	geonames:name	<i>String</i>
Localização	grs:point	<i>String</i>
População	dbpedia-owl:populationTotal	<i>Integer</i>
Altitude	dbpedia-owl:elevation	<i>Double</i>
IDH	dbpedia-owl:humanDevelopmentIndex	<i>Double</i>
Unidade do SUS	dbpedia-owl:Hospital	<i>Integer</i>
Investimento do PAC	paoli2:Government_aid	<i>Double</i>
Doença	dbpedia-owl:Disease	<i>Integer</i>
Rio	purl:River	<i>String</i>
Estação	paoli:Open_stream_water_level_recorders	<i>String</i>
Código da Estacao	geonames:featureCode	<i>Integer</i>
Nome da Estação	foaf:name	<i>String</i>
Latitude	geo:lat	<i>Double</i>
Longitude	geo:long	<i>Double</i>
Data da Medição	dbpprop:date	<i>Date</i>
Nível atencao	loa:FLOOD__INUNDATION__DELUGE	<i>Integer</i>
Nível Alerta	kmi:ALERT_TIME_DUR	<i>Integer</i>
Nível Enchente	ontosem:flood	<i>Integer</i>
Cota	loa:WATER_LEVEL_2	<i>Double</i>

Fonte: Elaborado pela autora.

## 5.2 **Dataset**

A criação do *dataset* envolveu duas linhas de ações: a extração manual de dados coletados via FTP ou dos *websites* das organizações e a conversão de dados em bases relacionais para o modelo RDF.

O *SPARQL Endpoint* possibilita, ao usuário, escrever suas próprias consultas SPARQL a partir do *dataset* criado. Foi possível explorar o conjunto de dados desenvolvido utilizando os

recursos e a flexibilidade do SPARQL. A partir dessa interface, foram realizadas consultas sobre as classes e propriedades definidas no mapeamento. A Tabela 7 exhibe os prefixos utilizados para a construção das consultas.

Tabela 7 – Prefixo dos vocabulários utilizados nas consultas SPARQL.

Prefixo	Vocabulário
db	<http://localhost:2020/resource/>
dbpedia-owl	<http://dbpedia.org/ontology/>
ontosem	<http://ontosem.org/>
geo	<http://www.w3.org/2003/01/geo/wgs84_pos#>
paoli2	<http://paoli.open.ac.uk/watson-cache/9/feb/1411/72815/d669a668d7/393dee7719f320bbe>
geonames	<http://www.geonames.org/ontology#>
loa	<http://www.loa-cnr.it/ontologies/OWN#>
rdfs	<http://www.w3.org/2000/01/rdf-schema#>
loki	<http://loki.cae.drexel.edu/wbs/ontology/2004/01/iso-metadata>
kmi	<http://kmi-web05.open.ac.uk:81/cache/f/d22/e37d/afaf7/7c67bcb3dd/7037c791f22062a84>
purl	<http://purl.org/ontology/places>
owl	<http://www.w3.org/2002/07/owl#>
xsd	<http://www.w3.org/2001/XMLSchema#>
map	<http://localhost:2020/resource/#>
rdf	<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
dbpprop	<http://dbpedia.org/property/>
vocab	<http://localhost:2020/resource/vocab/>
paoli	<http://paoli.open.ac.uk/watson-cache/8/3e0/8cea/dfd33/9a9a4cbcd/ffeb4a390e9303b28>
grs	<http://www.georss.org/georss/point>

Fonte: Elaborado pela autora.

A ferramenta *generate-mapping* da plataforma D2RQ foi utilizada para a criação de um arquivo de mapeamento a partir da análise do esquema do banco de dados. Para isso, executou-se o comando, passando como parâmetro as credenciais do banco, para que a aplicação tenha acesso ao esquema, e o nome do arquivo de saída. A partir desse arquivo, foi possível configurar os mapeamentos das tabelas e colunas do banco de dados para as classes e propriedades de ontologias. Por padrão, o *generate-mapping* cria termos locais para cada elemento do banco, mas é possível editar para reaproveitar termos já existentes.

Para ilustrar o funcionamento do mapeamento, a Figura 20 exibe uma parte desse arquivo gerado pelo componente *generate-mapping* da ferramenta D2RQ, a partir do banco de dados criado anteriormente. Esse trecho do arquivo refere-se à tradução da tabela Cotas, que armazena os dados dos níveis dos rios da Bacia do Rio Doce.

Figura 20 – Mapeamento da tabela Cotas

```
# Table cotas
map:cotas a d2rq:ClassMap;
  d2rq:dataStorage map:database;
  d2rq:uriPattern "cotas/@@cotas.pk@";
  d2rq:class vocab:cotas;
  d2rq:classDefinitionLabel "cotas";
```

Fonte: Elaborado pela autora, 2013.

A primeira linha é um comentário, a segunda é o mapeamento de *map:cotas* para a classe *d2rq:ClassMap* e a terceira linha define que as consultas são feitas no banco de dados configurado em *map:database*. As próximas linhas definem o padrão da URI (*d2rq:uriPattern*), a classe que está sendo representada (*d2rq:Class*) e uma identificação opcional (*d2rq:classDefinitionLabel*). O endereço do servidor, onde o *D2R Server* estará em execução, define o padrão URI base. A referência ao campo *pk* da tabela *cotas* é definido na expressão *"cotas/@@cotas.pk@"* e a classe padrão é a *vocab:cotas*, criada a partir do vocabulário padrão local(*vocab*) em referência ao nome da tabela.

Já a Figura 21 ilustra o mapeamento das propriedades (*d2rq:PropertyBridge*) oriundas das colunas da tabela *Cotas*. O nome das propriedades (*d2rq:property*) também foi pré-definido com o vocabulário padrão. Os atributos *d2rq:column* e *d2rq:datatype* são, respectivamente, a coluna e o tipo definidos para cada propriedade.

Figura 21 – Mapeamento das propriedades da tabela Cotas

```

map:cotas_cod_estacao a d2rq:PropertyBridge;
  d2rq:belongsToClassMap map:cotas;
  d2rq:property vocab:cotas_cod_estacao;
  d2rq:propertyDefinitionLabel "cotas cod_estacao";
  d2rq:column "cotas.cod_estacao";
  d2rq:datatype xsd:integer;
-
map:cotas_nome_estacao a d2rq:PropertyBridge;
  d2rq:belongsToClassMap map:cotas;
  d2rq:property foaf:name;
  d2rq:column "cotas.nome_estacao";
-
map:cotas_latidade a d2rq:PropertyBridge;
  d2rq:belongsToClassMap map:cotas;
  d2rq:property geo:lat;
  d2rq:column "cotas.latidade";
  d2rq:datatype xsd:double;
-
map:cotas_longitude a d2rq:PropertyBridge;
  d2rq:belongsToClassMap map:cotas;
  d2rq:property geo:long;
  d2rq:column "cotas.longitude";
  d2rq:datatype xsd:double;
-
map:cotas_maxima_mes a d2rq:PropertyBridge;
  d2rq:belongsToClassMap map:cotas;
  d2rq:property vocab:cotas_maxima_mes;
  d2rq:propertyDefinitionLabel "cotas maxima_mes";
  d2rq:column "cotas.maxima_mes";
  d2rq:datatype xsd:integer;

```

Fonte: Elaborado pela autora, 2013.

O próximo passo foi a geração do arquivo *dump RDF*, a partir do arquivo de mapeamento e através da ferramenta *dump-rdf* da plataforma D2RQ. O comando disponibiliza os seguintes tipos de formato de saída: *Turtle*, *RDF/XML*, *RDF/XML-Abbrev*, *N3* ou *N-Triple*. Neste trabalho, foi utilizado o *RDF/XML*, conforme ilustrado na Figura 22.

Figura 22 – Trecho da representação em RDF/XML da tabela Cotas

```

<rdf:RDF
  xmlns:dbpprop="http://dbpedia.org/property/"
  xmlns:vocab="http://localhost:2020/vocab/"
  xmlns:map="http://localhost:2020/#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:foaf="http://xmlns.com/foaf/0.1/"
  xmlns:db="http://localhost:2020/"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:grs="http://www.georss.org/georss/point"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
  xmlns:dbpedia-owl="http://dbpedia.org/ontology/"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:geo="http://www.w3.org/2003/01/geo/wgs84_pos#"
  xmlns:j.0="http://www.georss.org/georss/"
  xml:base="http://localhost:2020/" >
  <rdf:Description rdf:about="cotas/648">
    <geo:long rdf:datatype="http://www.w3.org/2001/XMLSchema#double">
      -42.359167E0</geo:long>
    <vocab:cotas_nivel_alerta rdf:datatype="http://www.w3.org/2001/XMLSchema#integer">
      400</vocab:cotas_nivel_alerta>
    <vocab:cotas_nivel_atencao rdf:datatype="http://www.w3.org/2001/XMLSchema#integer">
      0</vocab:cotas_nivel_atencao>
    <vocab:cotas_cota22 rdf:datatype="http://www.w3.org/2001/XMLSchema#double">
      320.0E0</vocab:cotas_cota22>
    <vocab:cotas_cota21 rdf:datatype="http://www.w3.org/2001/XMLSchema#double">
      321.5E0</vocab:cotas_cota21>
    <vocab:cotas_cota28 rdf:datatype="http://www.w3.org/2001/XMLSchema#double">
      321.0E0</vocab:cotas_cota28>
    <vocab:cotas_maxima_mes rdf:datatype="http://www.w3.org/2001/XMLSchema#integer">
      343</vocab:cotas_maxima_mes>
    <vocab:cotas_cota04 rdf:datatype="http://www.w3.org/2001/XMLSchema#double">
      336.5E0</vocab:cotas_cota04>
  </rdf:Description>

```

Fonte: Elaborado pela autora, 2013.

O arquivo RDF é iniciado pela tag `<rdf:RDF>` seguido pelos prefixos e o elemento `<rdf:Description>`, que descreve o recurso cuja URI é `<http://localhost:2020/cotas/648>`. Assim, o arquivo está pronto para ser manipulado como um *dataset*.

A Figura 23 ilustra o grafo RDF, gerado a partir do arquivo RDF, onde a classe *Município* e *Rio* herdam da classe superior *Thing*.

Figura 23 – Representação do *dataset* em um grafo RDF

Fonte: Elaborado pela autora, 2014.

A Tabela 8 corresponde à Figura 23 e exibe o vocabulário utilizado para cada nó do grafo.



Tabela 8 – Vocabulário dos Conceitos utilizados no *dataset*

Conceito	URI
Município	<a href="http://purl.org/ontology/places#Town">http://purl.org/ontology/places#Town</a>
nome_município	<a href="http://www.geonames.org/ontology#name">http://www.geonames.org/ontology#name</a>
população	<a href="http://dbpedia.org/ontology/populationTotal">http://dbpedia.org/ontology/populationTotal</a>
altitude	<a href="http://dbpedia.org/ontology/elevation">http://dbpedia.org/ontology/elevation</a>
lat_long	<a href="http://www.georss.org/georss/point">http://www.georss.org/georss/point</a>
idh	<a href="http://dbpedia.org/ontology/humanDevelopmentIndex">http://dbpedia.org/ontology/humanDevelopmentIndex</a>
doença	<a href="http://dbpedia.org/ontology/Disease">http://dbpedia.org/ontology/Disease</a>
investimento_pac	<a href="http://paoli.open.ac.uk/watson-cache#Government_aid">http://paoli.open.ac.uk/watson-cache#Government_aid</a>
Rio	<a href="http://dbpedia.org/ontology/River">http://dbpedia.org/ontology/River</a>
Estação	<a href="http://paoli.open.ac.uk/Open_stream_water_level_recorders">http://paoli.open.ac.uk/Open_stream_water_level_recorders</a>
cod_município	<a href="http://loki.cae.drexel.edu/wbs/ontology/2004/01/iso-metadata#identCode">http://loki.cae.drexel.edu/wbs/ontology/2004/01/iso-metadata#identCode</a>
cota	<a href="http://www.loa-cnr.it/ontologies/OWN/OWN.owl#WATER_LEVEL_2">http://www.loa-cnr.it/ontologies/OWN/OWN.owl#WATER_LEVEL_2</a>
nivel_alerta	<a href="http://kmi-web05.open.ac.uk:81/cache#ALERT_TIME_DUR">http://kmi-web05.open.ac.uk:81/cache#ALERT_TIME_DUR</a>
nivel_atenção	<a href="http://www.loa-cnr.it/ontologies/OWN/OWN.owl#FLOOD__INUNDATION__DELUGE">http://www.loa-cnr.it/ontologies/OWN/OWN.owl#FLOOD__INUNDATION__DELUGE</a>
nivel_enchente	<a href="http://ontosem.org/#flood">http://ontosem.org/#flood</a>
longitude	<a href="http://dbpedia.org/resource/Longitude">http://dbpedia.org/resource/Longitude</a>
latitude	<a href="http://dbpedia.org/resource/Latitude">http://dbpedia.org/resource/Latitude</a>
nome_estação	<a href="http://xmlns.com/foaf/0.1/">http://xmlns.com/foaf/0.1/</a>
cod_estação	<a href="http://www.geonames.org/ontology#featureCode">http://www.geonames.org/ontology#featureCode</a>

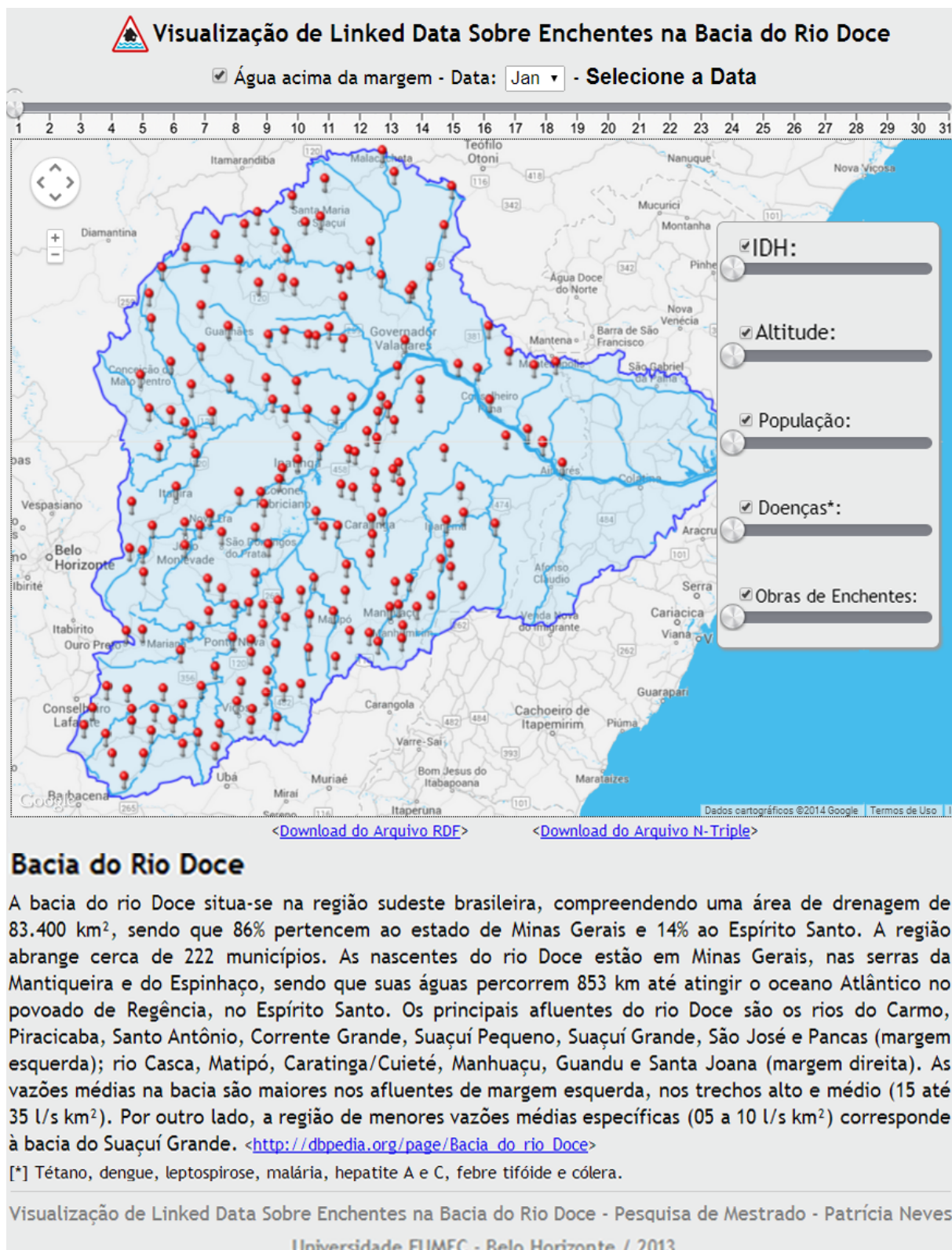
Fonte: Elaborado pela autora.

### 5.3 Visualização em um SIG

O resultado das consultas SPARQL foi exibido no SIG, uma aplicação *Web* implementada utilizando a linguagem Javascript e a interface de programação de aplicativos (API, em inglês *Application Programming Interface*), disponibilizada pelo *Google Maps*, em que o usuário seleciona os dados referentes à Bacia do Rio Doce, a serem visualizados no mapa, conforme

ilustrado na Figura 24.

Figura 24 – Visualização do SIG implementado



Fonte: Elaborado pela autora, 2014.

O protótipo foi implementado utilizando *Asynchronous Javascript and XML* (AJAX), em que o usuário seleciona os dados, referentes à Bacia do Rio Doce, a serem visualizados no mapa. Combinações podem ser feitas com o objetivo de vincular dados de diversas fontes, simultaneamente. Por exemplo, é possível visualizar se os trechos com mais ocorrência de enchente são os mesmos com maiores ocorrências de doenças relativas às enchentes, ou com baixo IDH.

Conforme ilustrado na Figura 24, ao selecionar o mês e o dia na linha do tempo, são exibidos no SIG os locais que sofreram com enchente, por meio de um mapa de calor. Se somente o mês for selecionado, será exibido o mapa de calor de todo o mês. Também é possível fazer combinações com as opções à direita, como IDH, altitude e população. Assim, o mapa de calor combinará os locais com ocorrência de enchente e o atributo selecionado, podendo ser mais de um. Abaixo, os arquivos RDF disponíveis para *download* nos formatos RDF/XML e N-Triple. Após, um texto informativo sobre a Bacia do Rio Doce, retirado da *Wikipédia*.

## 6 VALIDAÇÃO

Com o intuito de validar a abordagem proposta, foi realizada uma prova de conceito com as perguntas de competência apresentadas na Seção 1.2. A seguir, a demonstração da utilização das consultas na aplicação e seus resultados.

### 6.1 Dados

Acredita-se que a disponibilização de dados abertos e padronizados possibilita a descoberta de novos conhecimentos, por meio da reutilização desses dados em novas aplicações. A publicação de dados sobre as enchentes na Bacia do Rio Doce segue os princípios *linked data* e possibilita a descoberta, integração e buscas por outras fontes de dados.

Os dados disponibilizados seguiram as boas práticas de publicação e consumo de *linked data* e foram validados segundo o princípio das 5 estrelas (descrito na seção 2.1.2.1) criado por Berners-Lee, Hendler e Lassila (2001), conforme comparação na Tabela 9.

Tabela 9 – Classificação dos Dados Disponibilizados

Classificação	Dados Disponibilizados
1 estrela	Podem ser acessados e compartilhados por qualquer pessoa
2 estrelas	Podem ser agregados e exportados em formato estruturado
3 estrelas	Podem ser manipulados sem precisar de um <i>software</i> específico
4 estrelas	Podem ser reutilizados e combinados com outros
5 estrelas	Podem ser descobertos através de outros dados vinculados

Fonte: Elaborado pela autora.

O arquivo criado durante o trabalho foi validado pelas 5 estrelas e está pronto para ser consumido e vinculado com outras fontes de dados na *web*. Além dessa validação, verificou-se o arquivo RDF através da ferramenta *online* de validação *W3C RDF Validation Service*, que foi executada com sucesso. Por questões de legibilidade, o grafo RDF gerado pelo serviço de validação W3C não foi inserido no trabalho e se encontra nesse link: <<http://goo.gl/fQ4ZAX>>. Os arquivos RDF estão disponíveis nos formatos RDF/XML e N/Triple nos seguintes links:

RDF/XML: <https://db.tt/pJ0r78qw>; N/Triple: <https://db.tt/DKx7dkK4>. Assim, os dados estão preparados para serem consumidos como dados ligados, através de *browsers*, motores de buscas ou aplicações para domínios específicos.

## 6.2 *Dataset*

As tabelas a seguir demonstram as consultas e resultados, limitados a 10 linhas e sem ordenação, das perguntas de competência Q1, Q2, Q3 e Q4, respectivamente, como forma de validação dos conceitos citados anteriormente.

Tabela 10 – Consulta Q1.

<b>Das cidades que sofreram com enchentes em janeiro de 2012, quais possuem os menores IDHs ?</b>
<pre> SELECT ?cod_municipio ?cod_municipio2 ?idh ?data ?medicao ?nive l_alerta WHERE { ?cota dbpprop:date ?data . ?cota loa:WATER_LEVEL_2 ?medicao . ?cota ontosem:flood ?nivel_alerta . ?cota loki:identCode ?cod_municipio . ?mun loki:identCode ?cod_municipio2 . ?mun dbpedia-owl:humanDevelopmentIndex ?idh FILTER (?data &lt;= "2012-31-01"^^xsd:date &amp;&amp; ?medicao &gt; ?nivel_alerta &amp;&amp; ?cod_municipio = ?cod_municipio2)} </pre>

Fonte: Elaborado pela autora, 2014.

Tabela 11 – Resultado da Consulta Q1.

<b>município</b>	<b>idh</b>
"Mathias Lobato"	0.612
"Frei Inocênciao"	0.648
"Imbé de Minas"	0.553
"Araponga"	0.536

Fonte: Elaborado pela autora, 2014.

Tabela 12 – Consulta Q2.

<b>Quais estações registraram enchente no dia 09/01/2012 ?</b>
<pre> SELECT ?resource ?cod_estacao ?estacao ?medicao ?nivel_alerta WHERE { ?resource geonames:featureCode ?cod_estacao . ?resource paoli:Open_stream_water_level_recorders ?estacao . ?resource dbpprop:date ?data . ?resource loa:WATER_LEVEL_2 ?medicao . ?resource ontosem:flood ?nivel_alerta . FILTER (?data = "2012-09-01"^^xsd:date)} </pre>

Fonte: Elaborado pela autora, 2014.

Tabela 13 – Resultado da Consulta Q2.

<b>cod_estacao</b>	<b>estacao</b>	<b>medicao</b>	<b>nivel alerta</b>
56500000	"FAZENDA CACHOEIRA D'ANTAS"	640	450
56335001	"SANTO ANTONIO DO MANHUACU"	572.5	400
56989400	"MARIO DE CARVALHO"	571	520
56940002	"TUMIRITINGA"	120	550
56055000	"PORTO SANTA RITA"	484.5	450
56539000	"GOVERNADOR VALADARES"	448	360
56900000	"PORTO FIRME"	463	350
56640000	"DOM JOAQUIM"	426	400
56719998	"IPANEMA"	406.	300
56750000	"RIO PIRACICABA"	414	350

Fonte: Elaborado pela autora, 2014.

Tabela 14 – Consulta Q3.

<b>Qual a população dos municípios afetados pela enchente no dia 20/01/2012 ?</b>
<pre> SELECT ?cod_municipio ?cod_municipio2 ?populacao ?cod_estacao ?estacao ?data ?medicao ?nivel_alerta WHERE { ?cota geonames:featureCode ?cod_estacao . ?cota paoli:Open_stream_water_level_recorders ?estacao . ?cota dbpprop:date ?data . ?cota loa:WATER_LEVEL_2 ?medicao . ?cota ontosem:flood ?nivel_alerta . ?cota loki:identCode ?cod_municipio . ?mun loki:identCode ?cod_municipio2 . ?mun dbpedia-owl:populationTotal ?populacao FILTER (?data = "2012-01-20"^^xsd:date &amp;&amp; ?medicao &gt; ?nivel_alerta &amp;&amp; ?cod_municipio = ?cod_municipio2)} ORDER BY DESC (?populacao) </pre>

Fonte: Elaborado pela autora, 2014.

Tabela 15 – Resultado da Consulta Q3.

<b>cod_mun</b>	<b>estações</b>	<b>população</b>
315800	FAZENDA BARRACA	10552
312770	GOVERNADOR VALADARES (PCD)	263689
314090	MATIPO	17639
311340	SANTO ANTONIO DO MANHUACU	85239
313120	IPANEMA	18170
316340	FAZENDA CACHOEIRA D'ANTAS	5636
317130	FAZENDA VARGINHA	72220
310050	PORTO SANTA RITA	10276
315230	PORTO FIRME	10417
316950	TUMIRITINGA	6293

Fonte: Elaborado pela autora, 2014.

Tabela 16 – Consulta Q4.

<b>Qual a altitude, investimentos do governo e ocorrência de doenças dos municípios afetados por enchente em 2012?</b>
<pre> SELECT ?cod_municipio ?cod_municipio2 ?municipio ?doencas ?altitude ?investimento ?cod_estacao ?estacao ?medicao ?nivel_alerta WHERE { ?cota geonames:featureCode ?cod_estacao . ?cota paoli:Open_stream_water_level_recorders ?estacao . ?cota loa:WATER_LEVEL_2 ?medicao . ?cota kmi:ALERT_TIME_DUR ?nivel_alerta . ?cota loki:identCode ?cod_municipio . ?mun geonames:name ?municipio . ?mun loki:identCode ?cod_municipio2 . ?mun paoli2:Government_aid ?investimento . ?mun dbpedia-owl:elevation ?altitude. ?mun dbpedia-owl:Disease ?doencas FILTER (?nivel_alerta != 0 &amp;&amp; ?medicao &gt;= ?nivel_alerta &amp;&amp; ?cod_municipio = ?cod_municipio2)} ORDER BY DESC (?investimento) </pre>

Fonte: Elaborado pela autora, 2014.

Tabela 17 – Resultado da Consulta Q4.

<b>cod</b>	<b>municipio</b>	<b>doencas</b>	<b>investimento</b>	<b>altitude</b>
315210	Ponte Nova	258	9546757.00	431
315230	Porto Firme	0	803819.62	595
310050	Açucena	9	52036.25	472
317130	Viçosa	6	19055909.94	648
316340	São José do Goiabal	1	752351.56	287
313120	Ipanema	365	108894.89	241
311340	Caratinga	53	130219616.70	578
314090	Matipó	0	6347298.30	615
312770	Governador Valadares	1162	257931755.90	170
315800	Santa Maria de Itabira	0	794564.46	506

Fonte: Elaborado pela autora, 2014.

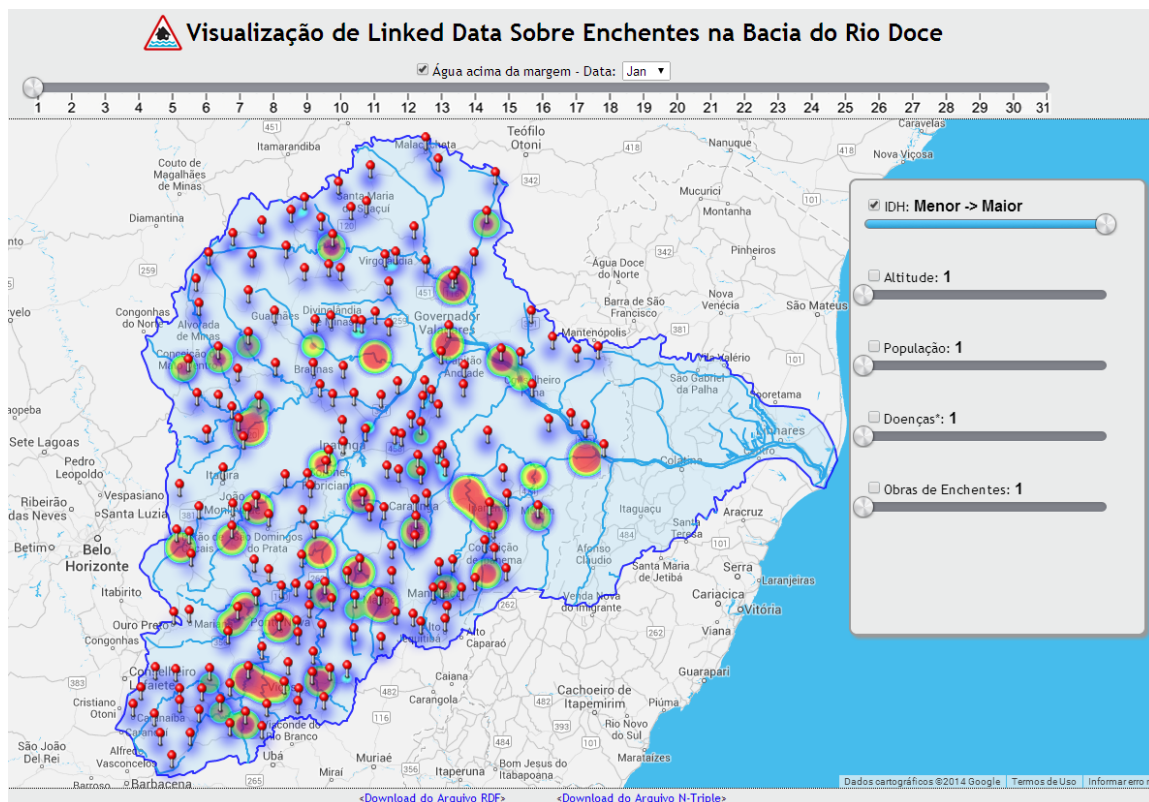


Com este *dataset* RDF criado sobre as enchentes na Bacia do Rio Doce e os dados agregados, essas informações passam a fazer parte da *Web* de dados, onde máquinas e humanos podem fazer buscas e utilizar esse conjunto de dados como uma de suas fontes de dados.

### 6.3 Visualização em um SIG

Esta aplicação de visualização de dados traz duas contribuições principais: a visualização de informações em um formato de mapas e a comprovação de que é possível construir aplicações consistentes a partir do *dataset* criado neste trabalho. As figuras 25, 26, 27 e 28 apresentam o protótipo resultante das consultas SPARQL descritas na seção anterior, visualizadas em um SIG.

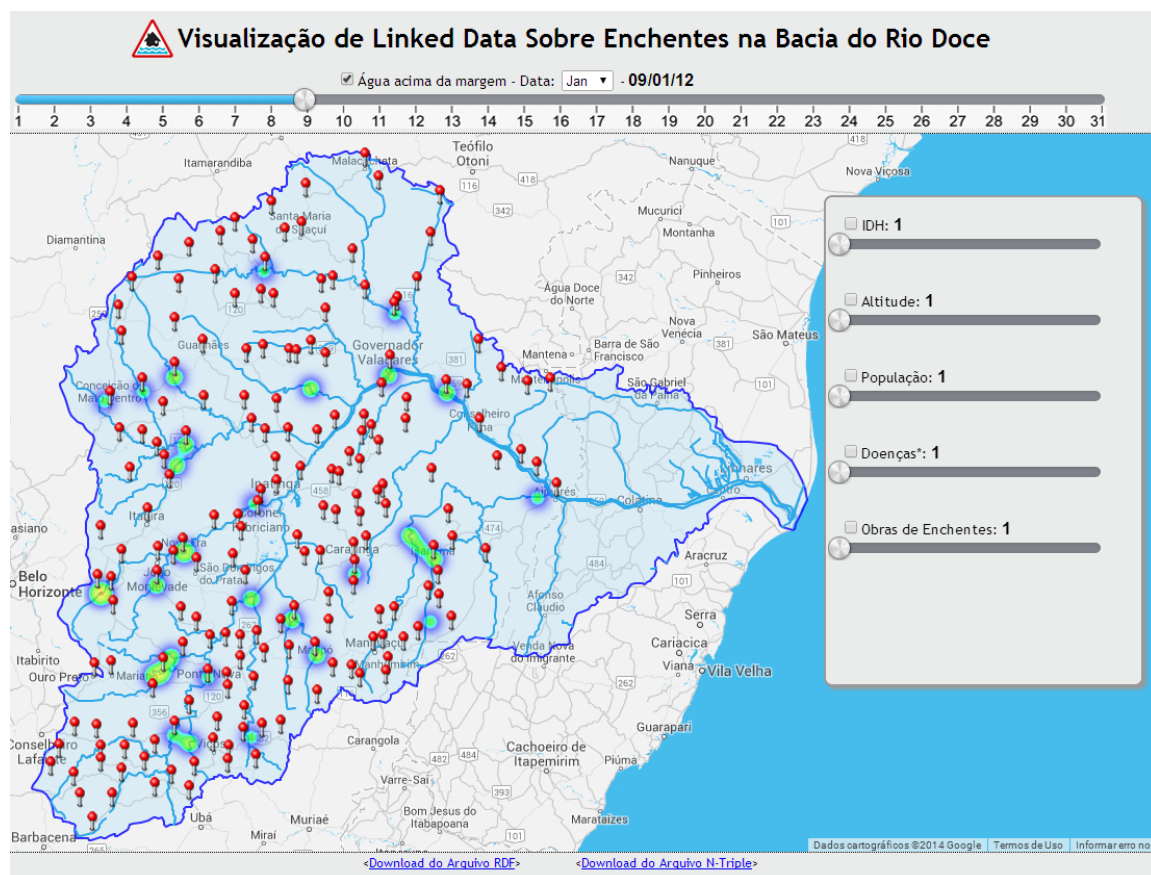
Figura 25 – Visualização da consulta Q1



Fonte: Elaborado pela autora, 2014.

Com as exceções apresentadas na Tabela 11, é possível visualizar, na Figura 25, que as cidades com menores IDHs não são as cidades afetadas por enchente no mês de janeiro.

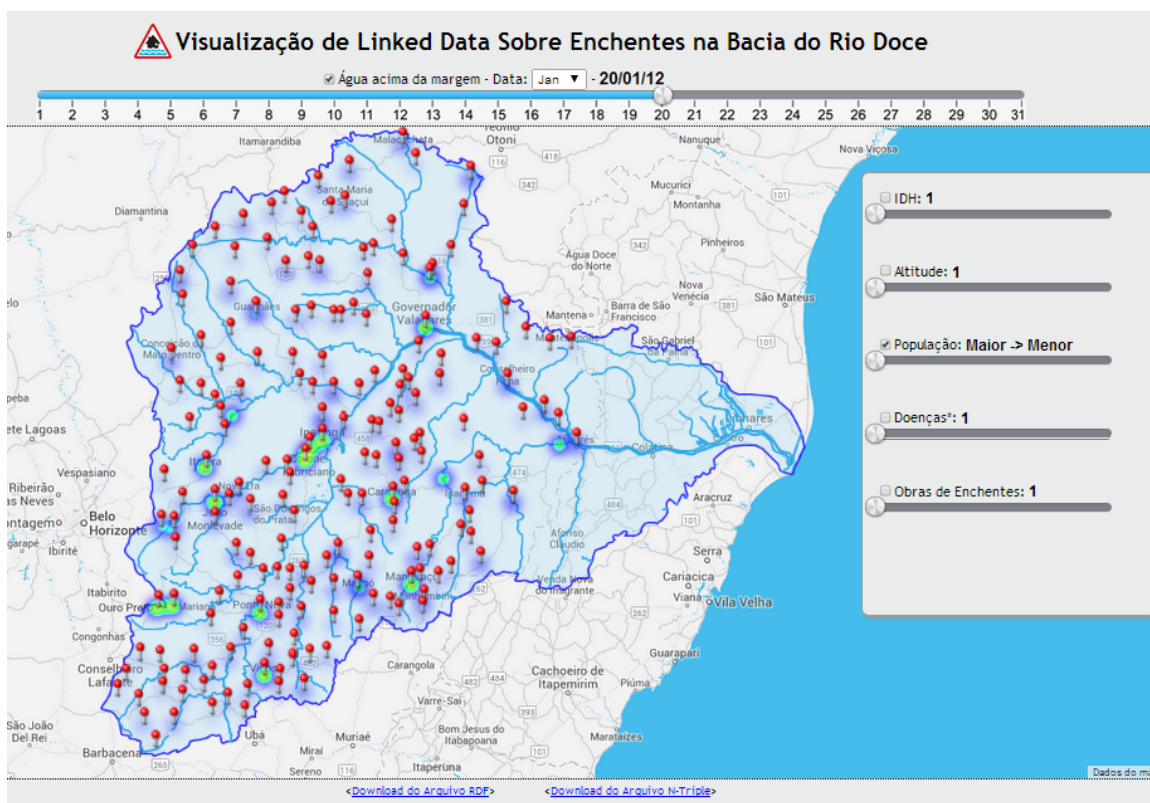
Figura 26 – Visualização da consulta Q2



Fonte: Elaborado pela autora, 2014.

De acordo com a Figura 26, as estações que mais se destacaram com medições acima do nível de alerta, no dia nove de janeiro de 2012, estão localizadas nos municípios Ipanema, Governador Valadares, Acaiaca, Ponte Nova, Porto Firme, Ferros e Aimorés.

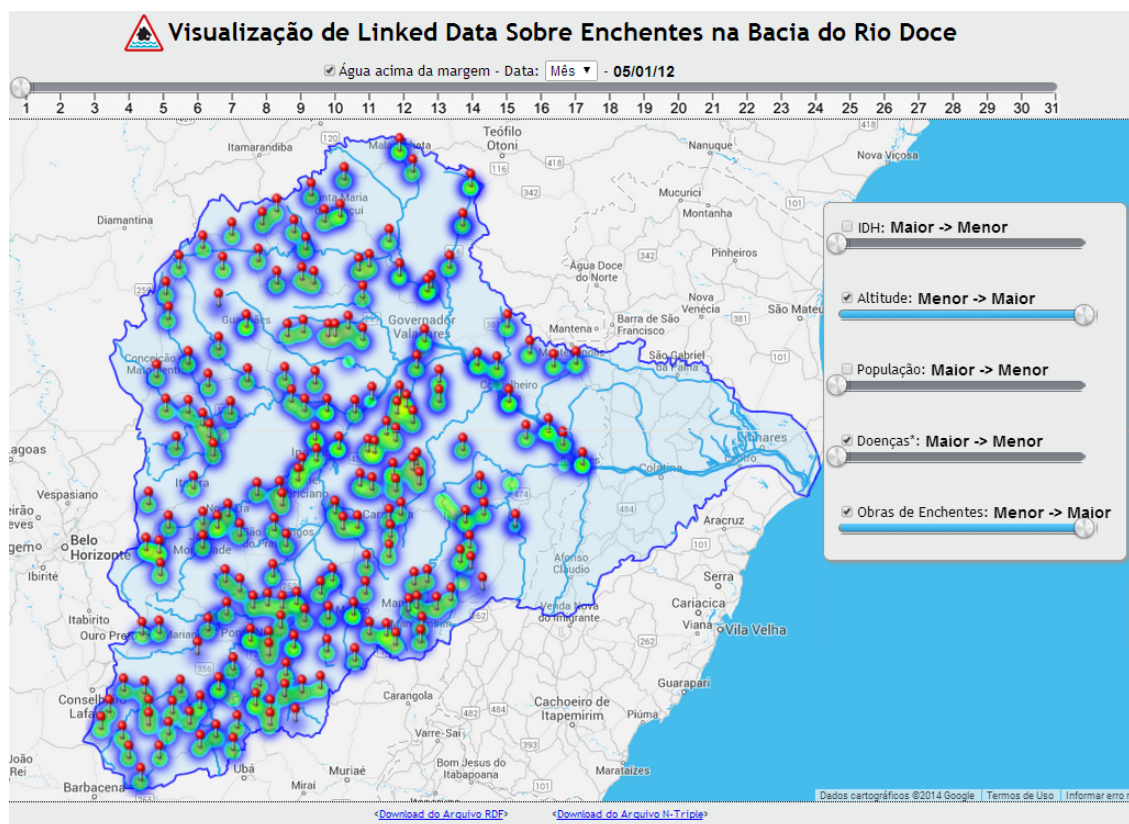
Figura 27 – Visualização da consulta Q3



Fonte: Elaborado pela autora, 2014.

A Figura 27 exibe os municípios mais populosos, afetados por enchente no dia 20 de janeiro de 2012, destacando Mathias Lobato, Frei Inocência, Aimorés, Governador Valadares e Ipanema.

Figura 28 – Visualização da consulta Q4



Fonte: Elaborado pela autora, 2014.

A Figura 28 ilustra os municípios que mais foram afetados por enchente, com menores altitudes, maiores ocorrências de doenças e menores investimentos do governo. Como resultado, destacam-se os municípios de Governador Valadares, Ponte Nova, Aimorés, Nova Era, Sobrália, Iapu, Ipanema, São João do Manhuaçu e Resplendor. Pode-se inferir que esses municípios são pontos críticos na gestão de desastres na Bacia do Rio Doce.

## 7 CONSIDERAÇÕES FINAIS

Atualmente, quanto à publicação de dados abertos, observa-se que, de um modo geral, o governo vem disponibilizando uma quantidade significativa de dados. Entretanto, esses dados ainda são apresentados de maneira incompleta, sem estrutura e padronização, o que dificulta a sua manipulação e o entendimento dos usuários de forma adequada.

De forma a tratar esses problemas no contexto de dados das enchentes na Bacia do Rio Doce, este trabalho dividiu-se em três partes: a criação de um conjunto de dados em RDF; a publicação desse conjunto de dados em um *SPARQL Endpoint*, disponível para a execução de consultas; e o desenvolvimento de uma aplicação *Web* para a visualização, em um SIG, do *dataset* criado. Portanto, conclui-se que os resultados do experimento proposto são válidos não somente para o contexto utilizado como para outras abordagens e até para uma população mais ampla. As tecnologias da *web* semântica juntamente com um SIG fornecem o suporte necessário para a integração, disponibilização e visualização do conjunto de dados sobre a Bacia do Rio Doce.

Com o *dataset* RDF criado com os dados sobre as enchentes na Bacia do Rio Doce e os dados agregados, essas informações passam a fazer parte da *Web* de dados, onde máquinas e humanos podem fazer buscas e utilizar esse conjunto de dados padronizado como uma de suas fontes de dados.

Desse modo, a contribuição deste experimento abrange a utilização de métodos e ferramentas para a publicação de dados conforme os princípios e padrões *linked data*. Acredita-se que a disponibilização de dados abertos e padronizados possibilita a descoberta de novos conhecimentos, pela reutilização desses dados em novas aplicações.

Este trabalho empenha-se em publicar dados sobre as enchentes na Bacia do Rio Doce, seguindo os princípios *linked data*, a fim de possibilitar a descoberta, integração e buscas por outras fontes de dados. Para o cidadão, a aplicação desenvolvida permite uma visualização amigável dos dados envolvidos na pesquisa e a construção de conhecimento a partir deles. Essa aplicação já está sendo usada internamente na CPRM, com previsão para ser disponibilizada aos cidadãos, em agosto de 2014,

Futuramente, sugere-se adicionar os dados do ano de 2013, para comparar com os dados de 2012, identificar avanços nas medidas governamentais contra enchentes, o controle das doenças e os níveis dos rios nas mesmas épocas do ano. Além disso, outras futuras linhas de ação são destacadas:

(a) Ampliação do *dataset*: A inclusão de dados pertinentes torna a aplicação ainda mais

relevante, principalmente ao cruzar com os dados existentes.

(b) Modificação da ferramenta para disponibilização do *SPARQL Endpoint*: A ferramenta oferecida pelo *D2R-Server* tem a desvantagem de utilizar uma base relacional como pré-requisito para que os mapeamentos sejam feitos em tempo real. Seria interessante que não constasse essa dependência para executar as consultas, mas somente um conjunto de dados RDF. Assim, uma melhoria para este trabalho seria a adoção de uma ferramenta independente de um banco de dados relacional para o acesso ao *SPARQL Endpoint*.

(c) Melhorias na visualização de dados da aplicação: A ampliação do *dataset* possibilita novas formas de representar os dados de forma mais amigável. Portanto, a visualização das informações pode ser melhorada com uma quantidade maior de dados, tornando a aplicação mais dinâmica e interativa para o usuário final.

## Referências

- ABARGUES, C. *et al.* Aggregating geoprocessing services using the OAI-ORE data model. international journal on advances in intelligent services,. *International Journal On Advances in Intelligent Services*, [S. l.], n . 3-4, p. 1–11, 2010.
- AGGARWAL, R.; KRAWCZYNSKI, K. Hepatitis E: an overview and recent advances in clinical and laboratory research. *J Gastroenterol Hepatol*, v. 15, n. 1, 2000.
- AHLBERG, C.; WILLIAMSON, C.; SHNEIDERMAN, B. Dynamic queries for information exploration: an implementation and evaluation. New York, NY; ACM Press, p. 619–626, 1992. Disponível em: <<http://doi.acm.org/10.1145/142750.143054>>.
- ALLEMANG, D.; HENDLER, J. *Semantic web for the working ontologist : effective modeling in RDF, RDFS and OWL*. Amsterdam: Morgan Kaufmann Publishers/Elsevier, 2008.
- ARONOFF, S. *Geographical Information Systems: A management perspective*. Ottawa, Canada: W.D.L. Publications, 1989.
- BANDHOLTZ, T.; FOCK, J. Linked environment data. Citeseer, 2011. Disponível em: <<http://citeseerx.ist.psu.edu/index;jsessionid=2510FE6A81B6F93E065B9B944A8F18EE>>.
- BARCELLOS, C.; SABROZA, P. C. The place behind the case: leptospirosis risks and associated environmental conditions in a flood-related outbreak in rio de janeiro. *Cadernos de Saúde Pública*, v. 17, p. S59–S67, 2001.
- BENNETT, D.; HARVEY, A. Publishing Open Government Data. 2009. Disponível em: <<http://www.w3.org/TR/gov-data/>>.
- BENNO, K. *et al.* Simblight1 a new model to predict first occurrence of potato late blight. *EPPO/OEPP Bulletin*, Blackwell Publishing, [S. l.], v. 37, n. 2, p. 339–343, 2007. Disponível em: <<http://dx.doi.org/10.1111/j.1365-2338.2007.01135.x>>
- BERNERS-LEE, T. *Uniform Resource Identifiers (URI): Generic Syntax*. 1998.
- BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The semantic web. *Scientific American*, [S. l.], v. 284, n. 5, p. 34–43, 2001. Disponível em: <<http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21>>
- BERNHARDBSEN, T. *Geographic information systems*. [S. l.]: Arendal/Cambridge: Viak IT and Norwegian Mapping Authority, 1992.
- BIZER, C.; CYGANIAK, R. *D2RQ . Lessons Learned. W3C Workshop on RDF Access to Relational Databases*. 2007.
- BIZER, C.; CYGANIAK, R.; HEATH, T. *How to publish Linked Data on the Web*. 2007. Disponível em: <<http://www4.wiwiw.fu-berlin.de/bizer/pub/LinkedDataTutorial/>>.
- BIZER, C.; HEATH, T.; BERNERS-LEE, T. Linked data - the story so far. *Int. J. Semantic Web Inf. Syst.*, v. 5, n. 3, p. 1–22, 2009.



- BIZER, C.; SEABORNE, A. D2rq - treating non-rdf databases as virtual rdf graphs. In: *ISWC2004 (posters)*. 2004. Disponível em: <<http://sites.wiwiss.fu-berlin.de/suhl/bizer/pub/Bizer-D2RQ-ISWC2004-Poster.pdf>>.
- BONHAM-CARTER, G. F. *Geographic information systems for geoscientists: modelling with GIS*. [S. l.]: Pergamon Press, 1994. (Computer methods in the geosciences).
- BOUWER, L. M.; VELLINGA, P. On the flood risk in the netherlands. In: BEGUM, S. et al. (Ed.). *Flood Risk Management in Europe* [S. l.], Springer, v. 25, p. 469-484, 2007.
- BRASIL. *Guia de preparação e resposta aos desastres associados às inundações para a gestão municipal do sistema único de saúde*. 2011. Disponível em: <[http://portal.saude.gov.br/portal/arquivos/pdf/guia\\_sms\\_desastres\\_jan2011\\_2.pdf](http://portal.saude.gov.br/portal/arquivos/pdf/guia_sms_desastres_jan2011_2.pdf)> .
- BRASIL. GOVERNO ELETRÔNICO. *Programa de Governo Eletrônico Brasileiro*. 2013. Disponível em: <<http://www.governoeletronico.gov.br>>.
- BURROUGH, P. A.; MCDONNELL, R. A. *Principles Of Geographical Information Systems*. [S. l.]: Oxford University Press, 1998. (Spatial Information Systems and Geostatistics).
- CARD, S. K.; MACKINLAY, J. D.; SCHNEIDERMAN, B. *Readings in information visualization: using vision to think*. San Francisco, CA: Morgan Kaufmann Publishers Inc., 1999. (Interactive Technologies Series).
- CARSWELL, B. *Bcas: an information system for aquaculture and marine resource planning. Unpublished report*:Minister of Agriculture Food and Fisheries (MAFF), British Columbia, 1998.
- CASARIN, H. C. S.; CASARIN, S. J. C. *Pesquisa Científica: da teoria à prática*. Curitiba: Ibpx, 2011.
- CBH-DOCE. *Comitê da Bacia Hidrográfica do Rio Doce*. 2013. Disponível em: <<http://www.riodoce.cbh.gov.br>>.
- CERVO, A. L.; BERVIAN; SILVA, R. *Metodologia Científica*. São Paulo: Pearson Prentice Hall, 2007.
- CHAUDHRY, P.; RUYSSCHAERT, G. *Climate Change and Human Development in Viet Nam*. [S. l.], 2007. Disponível em: <<http://ideas.repec.org/p/hdr/hdocpa/hdocpa-2007-46.html>>.
- CHEN, C.; HÄRDLE, W.; UNWIN, A. *Handbook of data visualization*. [S. l.]: Springer London, Limited, 2007. (Springer Handbooks of Computational Statistics).
- COLLINS, T. *Disaster risk for floods*. 2004. Disponível em: <<http://www.unu.edu/news/ehs/floods.doc>>.
- COMUNIDADE de dados abertos. *Segurança Pública em Porto Alegre*. 2013. Disponível em: <<http://ondeacontece.com.br/seguranca/RS>>.
- COOK, D.; MULROW, C.; HAYNES, B. Systematic Reviews: Synthesis of Best Evidence for Clinical Decisions. *Annals of International Medicine*, [S. l.], v. 126, n. 5, 1 Mar.1997.
- COWEN, D. J. GIS versus CAD versus DBMS: What are the differences. *Photogrammetric Engineering and Remote Sensing*, [S. l.], v. 54, p. 1551–1554, 1988.



- CUMMINS, F. *Enterprise Integration: An Architecture for Enterprise Application and Systems Integration*. New York: Wiley, 2002. (OMG press).
- CYGANIAK, R. *et al. The D2RQ Mapping Language*. 2012. Disponível em: <<http://d2rq.org/d2rq-language>>.
- DELL'AGLIO, D.; CELINO, I.; VALLE, E. D. Urban mashups. In: EDRES-NIGGERMAYER, Brigitte (Ed.). *Semantic Mashups: reuse of resources*. Berlin: Springer-Verlag, 2013. p. 287-319.
- DENSHAM, P. *Spatial Decision Support Systems*. In: MAGUIRE, D. J.; GOODCHILD, M. F.; RHINMD, D.W.(Ed.). *Geographical Information Systems*: Cambridge: Longman, p.1991. 403-412
- DING, L.; PERISTERAS, V.; HAUSENBLAS, M. Linked open government data. *IEEE Intelligent Systems*, [S. l.], v. 27, n. 3, p. 11–15, May 2012. Disponível em: <<http://doi.ieeecomputersociety.org/10.1109/MIS.2012.56>>.
- DING, Y. *et al.* Upper tag ontology for integrating social tagging data. *Journal of the American Society for Information Science and Technology*, Wiley Online Library, [S. l.], v. 61, n. 3, p.505–521, 2010.
- DING, Y. *et al.* Semantic web portal: a platform for better browsing and visualizing semantic data. In: *Active Media Technology*. Toronto: Springer, 2010. p. 448–460.
- DINIZ, V. Como conseguir dados governamentais abertos. In: *Congresso CONSAD de Gestão Pública*, 3., 2010, Brasília: Consad, 2010.
- EASTMAN, J. *et al. Explorations in geographic systems technology v. 4: GIS and Decision Making*. Explorations in geographic systems technology. Geneva: UNITAR, 1993. v. 4.
- ELLIS, J. B. T Sustainable integrated development of storm drainage in urban landscapes. In: INTERNATIONAL CONFERENCE ON INNOVATIVE TECHNOLOGIES, 2., p. 19-25, 1995.
- EM-DAT. *The OFDA/CRED International Disaster Database*. 2013. Disponível em: <<http://www.emdat.be>>.
- FABER, B. *et al.* Enhancing stakeholder involvement in environmental decision-making: active response geographic information system. In: *Proceedings of the 22nd Annual Conference of National Association of Environmental Professionals*. Washington, D.C., [s. n.], 1997. p. 174-18.
- FENSEL, D. *et al.* Ubiquitous interaction and collaboration with touristic services. Innsbruck, 2013.
- FJP. *Fundação João Pinheiro; Informativo PIB nas regiões de planejamento em MG*. 2011. Disponível em: <<http://fjp.mg.gov.br/>>.
- FLORCZYK, A. J. *et al.* Applying semantic linkage in the geospatial web. In: *Geospatial Thinking: Lecture Notes in Geoinformation and Cartography*. [S. l.]: Springer, 2010. p. 201-219.
- GAGNON, A. S.; SMOYER-TOMIC, K. E.; BUSH, A. B. The el niño southern oscillation and malaria epidemics in South America. *Int J Biometeorol*, [S. l.], v. 46, n. 2, p. 81-89, May 2002.
- GALIOTOU, E.; FRAGKOU, P. Applying linked data technologies to greek open government data: A case study. *Procedia-Social Behavioral Sciences*, [S. l.], Elsevier, v. 73, p. 479-486, 2013.
- GARBIS, G. *et al. An Implementation of a Temporal and Spatial Extension of RDF and SPARQL on top of MonetDB-Phase I*. Teleios, [S. l.], Feb. 2012.

- GARCÍA, R. *et al.* Publishing and interacting with linked data. In: ACM. *Proceedings of the International Conference Web Intelligence, Mining and Semantics*. Norway: [s. n.], 2011. p. 18.
- GEROIMENKO, V.; CHEN, C. *Visualizing the Semantic Web: Xml-Based Internet and Information Visualization*. [S.l.]: Springer-Verlag GmbH, 2003.
- GIL, A. C. *Como elaborar projetos de pesquisa*. São Paulo: Atlas, 1996.
- GLASS, G. V. Primary, Secondary, and Meta-Analysis of Research. *Educational Researcher*, v. 5, n. 10, p. 3-8, 1976. Disponível em: <<http://dx.doi.org/10.2307/1174772>>.
- GLIOZZO, G. *Bridging the gap between user generated spatial content and the semantic web*. Tese (Doutorado) — Delft University of Technology, 2010.
- GOYET, C. de Ville de. Epidemics caused by dead bodies: A disaster myth that does not want to die. May, 2004. Disponível em: <<http://books.google.com.br/books?id=LbfQSAACAAM>>.
- GROSSNER, K. E. *Representing Historical Knowledge in Geographic Information Systems*. 2010. 283f. Dissertação (Doctorate in Geography) University of California, Santa Barbara, 2010.
- GRUBER, T. R. Toward principles for the design of ontologies used for knowledge sharing. *Int. J. Hum.-Comput. Stud.*, Academic Press, Inc., Duluth, MN, USA, v. 43, n. 5-6, p. 907-928, Dec. 1995.
- GÜR, N.; DÍAZ, L.; KAUPPINEN, T. Gi systems for public health with an ontology based approach. In: *AGILE International Conference on Geographic Information Science (AGILE2012), 15., 2012, Avignon. Proceedings of the 15th AGILE International Conference on Geographic Information Science*. Avignon, France: [s. n.], 2012.
- HAASE, P. *et al.* The information workbench interacting with the web of data. Citeseer, 2009.
- HAGE, W. R. V.; ERP, M. V.; MALAISÉ, V. Linked open piracy: A story about e-science, linked data, and statistics. *Journal on Data Semantics*, Springer, v. 1, n. 3, p. 187–201, 2012.
- HAHMANN, S.; BURGHARDT, D.; WEBER, B. *Towards a Research Framework: Using the Semantic Web for (In) Validating this Famous Geo Assertion*. [S. l.]: Universiteit Utrecht, 2011.
- HART, G.; DOLBEAR, C. *Linked Data: A Geographic Perspective*. [S. l.]: Taylor & Francis, 2013.
- HARTH, A. Visinav: Visual web data search and navigation. In: SPRINGER. *Database and Expert Systems Applications*. [S. l.], 2009. p. 214–228.
- HARTMANIS, J. Turing award lecture on computational complexity and the nature of computer science. *Commun. ACM*, ACM, New York, NY, USA, v. 37, n. 10, p. 37-43, out. 1994. Disponível em: <<http://doi.acm.org/10.1145/194313.214781>>.
- HASENACK, H. O geoprocessamento no processo de tomada de decisão. *Boletim Gaúcho de Geografia*, Porto Alegre, v. 20, p. 185-188, 1995.
- HEATH, T.; BIZER, C. *Linked Data: Evolving the Web into a Global Data Space*. Morgan & Claypool, 2011. (Synthesis Lectures on Web Engineering Series). Disponível em: <<http://linkeddatabook.com/>>.

- HIDAYAT, R. *et al.* Semantic web portal in university research community framework. *International Journal on Advanced Science, Engineering and Information Technology*, v. 2, n. 6, p. 39–43, 2012.
- HITZLER, P. *et al.* Semantic aspects of earthcube. 2012.
- HUERTAS, A. S. Providing energy efficiency location-based strategies for buildings using linked open data. (Master in Geospatial Technologies) - Universidade Nova, Lisboa, 2012.
- HUQ, S. *et al.* Editorial reducing risks to cities from disasters and climate change. *Environment and Urbanization*, n. 1, p. 3–15, 2007.
- IBGE. *Instituto Brasileiro de Geografia e Estatística*. 2013. Disponível em: <<http://www.ibge.gov.br/>>.
- ISAAC, Y. *et al.* Semunit-french unit and linked data. In: *Proceedings of the 2nd International Workshop on Learning and Education with the Web of Data*. [S. l.: s. n.], 2012. p. 840.
- JANOWICZ, K. *et al.* A restful proxy and data model for linked sensor data. *Int. J. Digital Earth*, v. 6, n. 3, p. 233–254, 2013.
- JANOWICZ, K.; HITZLER, P. The digital earth as knowledge engine. *Semantic Web*, IOS Press, v. 3, n. 3, p. 213–221, 2012.
- JOHNSON, S. *The Ghost Map: The Story of London's Most Terrifying Epidemic-and how it Changed Science, Cities, and the Modern World*. Riverhead Books, 2006. Disponível em: <<http://books.google.com.br/books?id=cWtgIGzhPPEC>>.
- JURISTO, N.; MORENO, A. *Basics of Software Engineering Experimentation*. Dordrecht: Kluwer Academic Publishers, 2001.
- KALASHNIKOV, I. A. *et al.* Features of leptospirosis in the Krasnodar Territory. n. 6, 2003.
- KARANDE, S. *et al.* An observational study to detect leptospirosis in Mumbai, India. 2003.
- KAUPPINEN, T.; ESPINDOLA, G. M. d. Linked open science-communicating, sharing and evaluating data, methods and results for executable papers. *Procedia Computer Science*, [S. l.], v. 4, p. 726-731, 2011.
- KENDALL, G. C.; FEIGENBAUM, L.; TORRES, E. *SPARQL Protocol for RDF*. 2008. World Wide Web Consortium Recommendation REC-rdf-sparql-protocol-20080115. Disponível em: <<http://www.w3.org/TR/2008/REC-rdf-sparql-protocol-20080115>>.
- KESSLER, C.; JANOWICZ, K.; KAUPPINEN, T. spatial@ linkedscience-exploring the research field of giscience with linked data. In: *Geographic Information Science*. [S.l.]: Springer, 2012. p. 102–115.
- KIDDER, L. H. *Métodos de pesquisa nas relações sociais*. São Paulo: Editora Pedagógica e Universitária, 2007. v. 1 (delineamentos de pesquisa).
- KING, W.; HE, J. Understanding the role and methods of meta-analysis in is research. *Communications of the Association for Information Systems*, v. 16, n. 1, p. 665–686, 2005.

- KITCHENHAM, B. *et al.* Systematic literature reviews in software engineering - a systematic literature review. *Inf. Softw. Technol.*, Butterworth-Heinemann, Newton, MA, v. 51, n. 1, p. 7–15, jan. 2009. Disponível em: <<http://dx.doi.org/10.1016/j.infsof.2008.09.009>>.
- KITCHENHAM, B.; DYBA, T.; JORGENSEN, M. Evidence-based software engineering. In: *Proceedings of INTERNATIONAL CONFERENCE ON SOFTWARE ENGINEERING*, 26., 2004. Edinburgh, Scotland, UK, IEEE Computer Society Press, 2004. p. 273 - 281.
- KLYNE, G.; CARROLL, J. J. *Resource Description Framework (RDF): Concepts and Abstract Syntax*. 2004. World Wide Web Consortium, Recommendation REC-rdf-concepts-20040210. Disponível em: <<http://www.w3.org/TR/2004/REC-rdf-concepts-20040210>>.
- KNAP, T.; MICHELFEIT, J.; NECASKY, M. Linked open data aggregation: Conflict resolution and aggregate quality. In: IEEE. Computer Software and Applications Conference Workshops (*COMPSACW*), 2012 *IEEE 36th Annual*. [S.l.], 2012. p. 106–111.
- KONDO, H. *et al.* Post-flood infectious diseases in Mozambique. [S.l.] v. 17, n. 3, Sept. 2002.
- KOUBARAKIS, M. *et al.* Teleios: A database-powered virtual earth observatory. *Proceedings of the VLDB Endowment*, [S. l.]: VLDB Endowment, v. 5, n. 12, p. 2010-2013, 2012.
- KOUBARAKIS, M. *et al.* Data models and query languages for linked geospatial data. In: *Reasoning Web. Semantic Technologies for Advanced Query Answering*. [S.l.]: Springer, 2012. v. 7487, p. 290-328.
- KOUBARAKIS, M. *et al.* A data model and query language for an extension of rdf with time and space. *Deliverable D2*, v. 1, 2011.
- KRAAK, J.; ORMELING, F. J. *Cartography: visualization of geospatial data*. [S. l.]: Prentice Hall, 2003.
- KRABINA, B. *et al.* *Open government implementation model*. Vienna: KDZ , 2012.
- KUHN, W. Geospatial Semantics: Why, of What, and How? *Journal on Data Semantics III*, Berlin, v. 3534, p. 1–24, 2005. Disponível em: <[http://dx.doi.org/10.1007/11496168\\_1](http://dx.doi.org/10.1007/11496168_1)>.
- LASSILA, O. *et al.* *Resource Description Framework (RDF) Model and Syntax Specification*. W3C Recommendation, World Wide Web Consortium, Cambridge (MA), Feb. 1998.
- LATIF, A. *et al.* Discovery and construction of authors' profile from linked data (a case study for open digital journal). In: *LDOW* [S. l.: s. n.], 2010.
- LE-PHUOC, D. *et al.* Unifying stream and linked open data.[S. l.]: Deril Technical Report, 2010.
- LEBO, T. *et al.* Producing and using linked open government data in the twc logd portal. In: *Linking Government Data*. [S. l.]: Springer, 2011. p. 51-72.
- LEI Nº 12.527/2011. *Lei de Acesso a Informação Pública*. [S.l.], 2011. Disponível em: <[http://www.planalto.gov.br/ccivil\\_03/\\_ato2011-2014/2011/lei/l12527.htm](http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm)>.
- LIFSON, A. R. Mosquitoes, models, and dengue. University of Minnesota, 1996. p. 1201-2.
- LOGEAN, A. *et al.* Sustainability through Open Data: Examples from Switzerland. 2012.

- LONGLEY, P. A. *et al. Geographic Information Systems and Science*. [S.l.]: Wiley, 2005.
- MAASKANT, B.; JONKMAN, S. N.; BOUWER, L. M. Future risk of flooding: an analysis of changes in potential loss of life in South Holland. *Environmental Science And Policy*, [S.l.], v.12, n. 2, p. 157–169, 2009.
- MALAN, R.; BREDEMEYER, D. *Software Architecture Action Guide*. [S.l.]: Bredemeyer Consulting, 2006.
- MANOVICH, L. Information as an Aesthetic Event. p. 8, jan. 2009. Disponível em: <<http://www.manovich.net/>>.
- MARINO, T. B.; NASCIMENTO, B. S. do; BORGES, M. R. Gis supporting data gathering and fast decision making in emergencies situations. In: *Proceedings of the 9th International Conference on Information Systems for Crisis Response and Management*. [S. l.: s. n.], 2012.
- MASCARENHAS, S. A. *Metodologia científica*. São Paulo: Pearson Brasil, 2012.
- MENDES, C. A. B.; CIRILO, J. A. *Geoprocessamento em Recursos Hídricos: Princípios, Integração e Aplicação*. [S.l.: s.n.], 2001.
- MENDES, P. N. *et al.* Linked open social signals. In: IEEE. *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on*. [S. l.], 2010. v. 1, p. 224–231.
- MENTES, H. S. Design and development of a mineral exploration ontology. 2012.
- MINARD, C. J. *Napoleon's Invasion of Russia Map*. 1869. Disponível em: <<http://www.mappery.com/map-of/Napoleon's-Invasion-of-Russia-Map>>.
- MORGAN, O. Infectious disease risks from dead bodies following natural disasters. *Rev Panam Salud Publica*, [S. l.], v. 15, n. 5, p. 307-312, 2004.
- MULLER, J. C. *Geographic information systems: a unifying force for geography*. Johannes Gutenberg Universität, Mainz: The Operational Geographer, 1985. 41 p.
- NATHWANI, J.; LIND, N.; PANDEY, M. *Engineering Decisions for Life Quality: How Safe is Safe Enough?* [S.l.]: Springer, 2009. (Springer series in reliability engineering).
- NICHOLLS, R. *et al. Ranking port cities with high exposure and vulnerability to climate extremes: exposure estimates*. OECD Working Papers, n. 1, [S. l.]: OECD Publishing, 2007.
- NICHOLLS, R. J. Coastal flooding and wetland loss in the 21st century: changes under the sres climate and socio-economic scenarios. *Global Environmental Change*, v. 14, n. 1, p. 69–86, 2004.
- NOJI, E. K. *Public Health Consequences of Disasters*. New York: Oxford University Press, 1996.
- OSLEEB, J. P.; KAHN, S. *Integration of geographic information*. In *Tools to Aid Environmental Decision Making*. New York: Springer-Verlag, 1999. 161-189 p.
- PAOLO, R. *et al.* Decision support systems in agriculture: Administration of meteorological data, use of geographic information systems(gis) and validation methods in crop protection warning service, efficient decision support systems - practice and challenges from current to future. *Prof. Chiang Jao (Ed.)*, [S. l.]: Intech Open Science, 2011.

- PAULHEIM, H. Improving the usability of integrated applications by using interactive visualizations of linked data. In: ACM. *Proceedings of the International Conference on Web Intelligence, Mining and Semantics*. [S. l.], 2011. p. 19.
- PAULHEIM, H.; PROBST, F. Ontology-enhanced user interfaces: A survey. *Int. J. Semantic Web Inf. Syst.*, [S.l.], v. 6, n. 2, p. 36-59, 2010.
- PAULIINA, L. Open data in finland - public sector perspectives on open data. In: HACKER, Kenneth L.; DIJK, Jan. *Digital democracy: issues of theory and practice*. [S. l.]: Sage, 2011.
- PETTICREW, M.; ROBERTS, H. *Systematic Reviews In The Social Sciences: A Practical Guide*. [S. l.]: Blackwell Publishing, 2006.
- PHUOC, D. L.; HAUSWIRTH, M. Linked open data in sensor data mashups. CEUR, 2009.
- POBLET, M.; CASANOVAS, P.; LÓPEZ-COBO, J.-M. Online dispute resolution for the next web decade: the ontomedia approach. In: *Journal of universal computer science, Proceedings of the 10th international conference on knowledge management and knowledge technologies, Graz, Austria*. [S. l.: s. n.], 2010. p. 117-125.
- POMPEO, C. A. Drenagem urbana sustentável. RBRH, v. 5, n. 1, p. 15–23, jan./mar. 2000.
- PRADO, O.; LOUREIRO, M. R. Governo eletrônico e transparência: avaliação da publicização das contas públicas das capitais brasileiras. *Revista Alcance*, [S. l.], v.13, n. 3, p. 355-372. 2006.
- PROGRAMA de Aceleração do Crescimento (PAC). 2007. Disponível em: <<http://www.pac.gov.br>>.
- PRUD'HOMMEAUX, E.; SEABORNE, A. Sparql query language for rdf. Jan. 2008. Disponível em: <<http://www.w3.org/TR/rdf-sparql-query/>>.
- PSCHORR, J. *et al.* Sensor discovery on linked data. In: *Proceedings of the 7th Extended Semantic Web Conference, ESWC2010, Heraklion, Greece*: [s.n.], 2010. v. 30.
- QADRI, F. *et al.* Enterotoxigenic escherichia coli and vibrio cholerae diarrhea, Bangladesh. *Emerg Infect Dis.*, [S. l.], v. 11, n. 7, p. 1104- 1107, July 2004.
- REHBEIN, A. R. Avaliação de sistemas de informação. UFRGS, Porto Alegre, RS, 2002.
- REITZ, T.; TEMPLER, S. *An Environment for the Conceptual Harmonisation of Geospatial Schemas and Data*. In: AGILE, 2012, Avignon: [s. n.], 2012. p. 24-27.
- ROJAS, R. A. Modelos de base de datos de grafo y rdf. 1 4 0 f . Universidad de Chile, 2009.
- ROTH, M. Geographic feature pipes. 2011.
- RUTTENBERG, A. *et al.* Advancing translational research with the semantic web. *BMC Bioinformatics*, p. -1-1, 2007. Disponível em: <<http://www.biomedcentral.com/1471>>.
- SHAW, M. What makes good research in software engineering. *for Technology Transfer (STTT)*. Springer Berlin / Heidelberg, v. 4, n. 1, p. 1-7, 2002.
- SHERIDAN, J.; TENNISON, J. Linking uk government data. In: WWW2010, 2010.
- SHNEIDERMAN, B. The eyes have it: A task by data type taxonomy for information visualizations. IEEE Computer Society, Washington, DC, USA, p. 336-, 1996.

- SHVAIKO, P. *et al.* Trentino government linked open geo-data: a case study. In: *The Semantic Web-ISWC 2012*. [S.l.]: Springer, 2012. p. 196–211.
- SILVA, A. B. *Sistemas de Informações Geo-referenciadas: conceitos e fundamentos*. Campinas: Unicamp, 1999.
- SIMON, R. *et al.* Explorative user interfaces for browsing historical maps on the web. *e-Perimtron*, [S.l.] v. 5, n. 3, p. 132-143, 2010.
- SÃO PAULO. GOVERNO ABERTO - DECRETO n 55.559. *Portal Governo Aberto de São Paulo*. [S. l.], 2010. Disponível em: <<http://www.governoaberto.sp.gov.br/view/legis.php>>.
- SOARES, A. Design and implementation of a gazetteer. Universidade Técnica, Lisboa, 2011.
- STASCH, C. *et al.* Spatio-temporal aggregation of european air quality observations in the sensor web. *Computers & Geosciences*, [S. l.], v. 47, p. 111–118, 2012.
- STEELE, J.; ILIINSKY, N. *Designing Data Visualizations: Representing Informational Relationships*. O'Reilly Media, 2011. (Oreilly and Associate Series). Disponível em: <<http://books.google.com.br/books?id=Mp\R-vs00EoC>>.
- STEINBERG, M.; BREHM, J. Utilizing open content for higher-layered rich client applications. *International Journal On Advances in Intelligent Systems*, [S.l.], v. 2, n. 2 and 3, p. 303-316, 2009.
- SUR, D. *et al.* Severe cholera outbreak following floods in a northern district of west bengal. *Indian J Med Res.*, [S. l.], v. 112, p. 178-182, Nov. 2000.
- SZEKELY, P. *et al.* Exploiting semantics of web services for geospatial data fusion. In: ACM. *Proceedings of the 1st ACM SIGSPATIAL International Workshop on Spatial Semantics and Ontologies*. [S.l.], 2011. p. 32–39.
- TEDRE, M. Computing as a science: A survey of competing viewpoints. *Minds Mach.*, Kluwer Academic Publishers, Hingham, MA, v. 21, n. 3, p. 361–387, ago. 2011. Disponível em: <<http://dx.doi.org/10.1007/s11023-011-9240-4>>.
- THACKER. *Comunidade Transparência Hacker*. 2011. Disponível em: <<http://thacker.com.br/node/>>.
- THOMAS, M.; ROLLER, N. Information systems for integrated global change research. *25th International Symposium for Remote Sensing and Global Environmental Change*, Austria, p. 294-305, 1993.
- TRAVASSOS, G. H. *et al.* A environment to support large scale experimentation in software engineering. In: IEEE ICECCS. Belfast, Northern Ireland: IEEE, 2008. p. 193-202.
- VALENCIO, N.; SIENA, M.; MARCHEZINI, V. *Sociologia dos desastres: construção, interfaces e perspectivas no brasil*. São Carlos: Rima Editora, 2009. 280p.
- VANASCO, N. B. *et al.* Outbreak of human leptospirosis after a flood in reconquista, Santa Fe. *Rev Argent Microbiol*, [S. l.], v. 34, n. 3, p. 124–131, 2002.
- VILCHES-BLÁZQUEZ, L. M. *et al.* Geolinked data and inspire through an application case. In: ACM. *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*. [S.l.], 2010. p. 446–449.

- VOLLAARD, A. *et al.* Risk factors for typhoid and paratyphoid fever in Jakarta, Indonesia, JAMA - Journal of American Medical Association, [S. l.], v. 2, n. 29, June 2004.
- VOS, M. D. *et al.* Reconstructing semantics of scientific models: a case study. In: *Proceedings of the OEDW workshop on Ontology engineering in a data driven world, EKAW*. [S. l.: s. n.], 2012.
- MANOLA, F.; MILLER, E. (Ed.). *RDF Primer, W3C Recommendation*. 2004. Disponível em: <<http://www.w3.org/TR/2004/REC-rdf-primer-20040210/>>.
- W3C. *Extensible Markup Language (XML) 1.0 (Fifth Edition)*. 2008. World Wide Web Consortium, Recommendation REC-xml-20081126.
- W3C. *W3C SWEOW Linking Open Data; Comunidade de dados vinculados*. 2011. Disponível em: <<http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>>.
- WAITELONIS, J. *et al.* The path is the destination-enabling a new search paradigm with linked data. *Linked Data in the Future Internet at the Future Internet Assembly, Ghent*, p. 8, 2010.
- WOHLIN, C. *et al.* *Experimentation in software engineering: An introduction*. Boston, 2000.
- WORLD HEALTH ORGANIZATION. Epidemic prone disease surveillance and response after the tsunami in aceh province indonesia. May 2005.
- YANG, H. *et al.* Clinical distinction and evaluation of leptospirosis in taiwa: a case control study. *J Nephrol*, [S. l.], v. 18, p. 45-53, Jan./Feb. 2005.
- YUE, P. *et al.* Sharing geospatial provenance in a service-oriented environment. *Computers, Environment and Urban Systems*, [S. l.], v. 35, n. 4, p. 333-343, 2011.
- ZEUNER, T. *Landwirtschaftliche Schaderregerprognose mit Hilfe von geographischen Informationssystemen*. Mainz: Johannes Gutenberg Universität, 2008.