



ARQUITETURA PARA PUBLICAÇÃO DE DADOS ABERTOS EM SISTEMAS DE GOVERNO: O EXEMPLO DO SIASG

Xiao Yuan Kong

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Engenharia de Sistemas e Computação, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre em Engenharia de Sistemas e Computação.

Orientador: Jano Moreira de Souza

Rio de Janeiro

Março de 2015

ARQUITETURA PARA PUBLICAÇÃO DE DADOS ABERTOS EM SISTEMAS DE
GOVERNO: O EXEMPLO DO SIASG

Xiao Yuan Kong

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO
LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA (COPPE)
DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS
REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM
CIÊNCIAS EM ENGENHARIA DE SISTEMAS E COMPUTAÇÃO.

Examinada por:

Prof. Jano Moreira de Souza, Ph.D.

Prof. Geraldo Bonorino Xexéo, D.Sc.

Prof. Asterio Kiyoshi Tanaka, Ph.D.

RIO DE JANEIRO, RJ - BRASIL

MARÇO DE 2015

Kong, Xiao Yuan

Arquitetura para publicação de dados abertos em sistemas de governo: o exemplo do SIASG / Xiao Yuan Kong. – Rio de Janeiro: UFRJ/COPPE, 2015.

XVI, 90 p.: il.; 29,7 cm.

Orientador: Jano Moreira de Souza

Dissertação (mestrado) – UFRJ/ COPPE/ Programa de Engenharia de Sistemas e Computação, 2015.

Referências Bibliográficas: p. 87-90.

1. Dados Abertos Governamentais. 2. Governo Eletrônico. 3. ETL. I. Souza, Jano Moreira de. II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia de Sistemas e Computação. III. Título.

Agradecimentos

Primeiramente agradeço a minha família, aos meus pais pela educação e bons valores passados, e aos meus irmãos pelas conversas e discussões do dia a dia. E por estarem sempre ao meu lado, ajudando e apoiando em todas as minhas decisões.

Ao Professor Jano, meu orientador, pela oportunidade de compartilhar suas ideias que permitiram expandir meus horizontes, pela paciência e pelas valiosas sugestões.

Aos professores, Asterio Tanaka e Geraldo Xexéo por aceitarem fazer parte da banca examinadora.

Aos meus professores de graduação do DCC/UFRJ, pela boa formação em matemática e computação.

Aos professores da COPPE, em especial, os da linha de Banco de dados, Alexandre, Marta, Jano, Xexéo e Zimbrão, que foram de imensa importância no meu processo de aprendizagem.

À CAPES, pela ajuda financeira em forma de bolsa de estudos.

Ao CAPGOV, pelos projetos, que me fizeram evoluir tanto academicamente quanto profissionalmente.

À equipe do CAPGOV, em especial, Ana Cog, Daiane, Débora, Romulo, Sérgio e Vanessa por toda a ajuda, pelo incentivo e companhia, nesta longa jornada.

Aos futuros doutores, Márcio Antélio e Rogério Borba, pelas reuniões e discussões para definir o tema de trabalho.

Ao corpo administrativo, Ana Paula, Patrícia, Eliah, Solange, Claudia, Guty e Roberto Rodrigues pelo excelente trabalho prestado ao PESC que me ajudaram em diversas ocasiões.

Ao time de dados abertos da SLTI/MPOG, Augusto, Christian e Nitai, pela avaliação da API e diversas correções sugeridas.

Aos meus amigos de longa data, simplesmente pela amizade.

A todas as pessoas que de alguma forma contribuíram para realização deste trabalho e que não foram citadas aqui.

Resumo da Dissertação apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

ARQUITETURA PARA PUBLICAÇÃO DE DADOS ABERTOS EM SISTEMAS DE GOVERNO: O EXEMPLO DO SIASG

Xiao Yuan Kong

Março/2015

Orientador: Jano Moreira de Souza

Programa: Engenharia de Sistemas e Computação

O advento da tecnologia e os avanços na implementação do governo eletrônico no Brasil têm estimulado a criação de novas políticas de integração entre o governo e a sociedade. Ambos os fatores, também motivaram a implementação da Lei do Acesso à Informação (LAI) que gerou um crescente aumento do interesse público ao acesso aos dados governamentais, e a adesão à Parceria para Governo Aberto (OGP), onde o governo brasileiro assume um compromisso internacional em disponibilizar informações públicas aos cidadãos. Todos esses fatores contribuíram para uma política ficar em destaque atualmente, que é a de Dados Abertos Governamentais (DAG). Esta política tem como objetivo disponibilizar dados e informações governamentais de domínio público para a sociedade. Dentro desse contexto, este trabalho propõe uma solução para abertura de dados voltada a interoperabilidade, definindo uma arquitetura orientada a serviço seguindo padrões abertos. Dessa forma, o governo promove, ao mesmo tempo, a transparência pública, o incentivo à inovação (desenvolvimento de aplicativos, visualizações) e uma sociedade mais participativa.

Abstract of Dissertation presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

ARCHITECTURE FOR PUBLICATION OF OPEN DATA IN GOVERNMENT
SYSTEMS: THE EXAMPLE OF SIASG

Xiao Yuan Kong

March/2015

Advisor: Jano Moreira de Souza

Department: System and Computer Engineering

The advent of technology and the progress of the implementation of e-government in Brazil has been stimulating the creation of new integration policies between the government and the society. Both factors also motivated the implementation of Brazilian Access to Information Law, which generated an increasing public interest in accessing the government data, and the membership of the Open Government Partnership (OGP), where the Brazilian government has assumed an international commitment to provide public information to citizens. These factors contributed to highlight the policy of Open Government Data (OGD), which aims to provide governmental data to society. In this context, this work proposes an open data solution designed to be interoperable, by defining a service-oriented architecture following open standards. Thus, the government promotes, at the same time, public transparency, encouraging innovation (application development, new forms of data visualization), and a more engaging society.

Índice

1.	Introdução.....	1
1.1	Contextualização	1
1.2	Objetivos.....	2
1.3	Organização do Trabalho.....	3
2.	Fundamentação Teórica	4
2.1	Governo Eletrônico.....	4
2.1.1	Governo Eletrônico no Brasil	7
2.2	Dados Abertos Governamentais	11
2.2.1	Dados Abertos no Brasil.....	14
2.2.1.1	Lei de Acesso a Informação	14
2.2.1.2	Parceira do Governo Aberto.....	15
2.2.1.3	Infraestrutura Nacional de Dados Abertos	16
2.3	Qualidade de Dados.....	17
2.3.1	Introdução.....	17
2.3.2	Conceito.....	18
2.3.3	Metodologia de Qualidade de Dados.....	21
2.3.3.1	Metodologia TDQM.....	22
2.3.3.2	Metodologia TQdM.....	23
2.4	Web Service.....	23
2.4.1	SOAP	24

2.4.2	REST	25
2.4.2.1	HATEOAS	27
2.5	ETL.....	28
2.5.1	<i>Pentaho Data Integration</i>	30
2.5.2	<i>Talend Open Studio for Data Integration</i>	31
2.5.3	Comparativo de ferramentas ETL	32
3.	Proposta.....	35
3.1	Cenário Atual.....	35
3.2	Motivação	37
3.3	Arquitetura Proposta.....	39
4.	Estudo de caso	47
4.1	Sistema Estruturadores	48
4.2	SIASG.....	49
4.3	Padrões e Tecnologias adotados	52
4.3.1	Considerações	52
4.3.2	Camada de Qualidade	53
4.3.3	Camada de Serviço	54
4.4	Arquitetura da Implementação	58
4.5	Implementação.....	60
4.5.1	Camada de Qualidade de Dados	60
4.5.1.1	Carga de Dados em <i>stage</i>	60
4.5.1.2	Análise de Qualidade de Dados	62

4.5.2	Camada de Serviço	69
4.5.2.1	Carga no Banco de Dados da API de Compras.....	70
4.5.2.2	Desenvolvimento da API de Compras	71
5.	Conclusão	82
5.1	Considerações Finais	82
5.2	Resultado e Contribuições	83
5.3	Limitações	85
5.4	Trabalhos Futuros	86
	Referências Bibliográficas.....	87

Lista de Figuras

Figura 1 – Proporção de indivíduos que acessaram a <i>Internet</i> (Elaborado pelo autor, dados colhidos de (CETIC, 2013))	9
Figura 2 – Proporção da população com mais de 15 anos que acessaram os serviços do Governo Eletrônico (Elaborado pelo autor, dados colhidos de (CETIC, 2013))	10
Figura 3 – Utilização de serviços públicos pela população com mais de 15 anos (Elaborado pelo autor, dados colhidos de (CETIC, 2013))	10
Figura 4 – Categorias e dimensões de qualidade de dados (Adaptado de (WANG & STRONG, 1996)).....	21
Figura 5 – Fases da metodologia TDQM (Adaptado de (WANG, 1998))	22
Figura 6 – Exemplo de HATEOAS (Elaborado pelo autor).....	27
Figura 7 – Modelo de Maturidade de Richardson (Adaptado de (RICHARDSON, LEONARD, 2010))	28
Figura 8 – Processo ETL (Elaborado pelo autor)	29
Figura 9 – Tela do <i>Pentaho Data Integration</i> (Elaborado pelo autor)	31
Figura 10 – Tela do <i>Talend Open Studio for Data Integration</i> (Elaborado pelo autor).	32
Figura 11 – Processo de solicitação de dado pelo e-SIC (Elaborada pelo autor).....	37
Figura 12 - Número de pedidos de acesso à informação pelo e-SIC (Adaptado de (CGU, 2015b)).....	37
Figura 13 – Arquitetura proposta pelo autor (Elaborada pelo autor)	40
Figura 14 – Diagrama de Casos de usos do Provedor de Dados (Elaborada pelo autor)	42
Figura 15 – Diagrama de Casos de uso do Catalogador de Dados	43
Figura 16 – Diagrama de Casos de uso do Analista de Qualidade de Dados (Elaborada pelo autor).....	44
Figura 17 – Diagrama de Casos de uso do Desenvolvedor ETL (Elaborada pelo autor)	45

Figura 18 – Diagrama de Casos de uso do Analista de Negócio (Elaborada pelo autor)	45
Figura 19 – Diagrama de Casos de Uso do Desenvolvedor de API (Elaborada pelo autor)	
.....	45
Figura 20 - Diagrama de Casos de Uso do Cidadão (Elaborada pelo autor).....	46
Figura 21 – Diagramas de atividades da arquitetura proposta (Elaborada pelo autor)...	47
Figura 22 - Estrutura do SIASG (Elaborada pelo autor)	51
Figura 23 – Estrutura de um Recurso HAL (Adaptado de (KELLY, MIKE, 2013)).....	56
Figura 24 – Funcionamento do <i>Varnish Cache</i> (Elaborado pelo autor).....	57
Figura 25 – Implementação da Arquitetura (Elaborada pelo autor)	58
Figura 26 – Infraestrutura da API (Elaborada pelo autor).....	59
Figura 27 – Processo ETL de carga dos dados para <i>stage</i> (Elaborada pelo autor).....	62
Figura 28 - Apresenta detalhes base dados do SIASG (Elaborado pelo autor).....	64
Figura 29 - Gráfico de Estatísticas simples sobre o atributo <i>it_no_basico</i> de <i>siasg_identificacao_basica</i> (Elaborado pelo autor)	65
Figura 30 - Tabela de Estatísticas simples sobre o atributo <i>it_no_basico</i> de <i>siasg_identificacao_basica</i> (Elaborado pelo autor)	65
Figura 31 - Gráfico de Estatísticas simples sobre o atributo <i>it_co_iden_basica</i> de <i>siasg_identificacao_basica</i> (Elaborado pelo autor)	66
Figura 32 - Gráfico de Estatísticas simples sobre o atributo <i>it_co_iden_basica</i> de <i>siasg_identificacao_basica</i> (Elaborado pelo autor)	66
Figura 33 - Gráfico para os padrões encontrados para o atributo <i>it_co_iden_basica</i> da tabela <i>siasg_identificacao_basica</i> (Elaborado pelo autor).....	68
Figura 34 - Tabela para os padrões encontrados para o atributo <i>it_co_iden_basica</i> da tabela <i>siasg_identificacao_basica</i> (Elaborado pelo autor).....	68
Figura 35 - Resultado da aplicação da regra de negócio (Elaborado pelo autor).....	69

Figura 36 - Linha que não está em conformidade com a regra de negócio (Elaborado pelo autor).....	69
Figura 37 – Estrutura da API (Elaborado pelo autor).....	72
Figura 38 - Página inicial da API de dados abertos (Elaborada pelo autor).....	72
Figura 39 – Modelo de relacionamento entre os módulos (Elaborada pelo autor).....	73
Figura 40 – Modelagem do módulo de fornecedores (Elaborado pelo autor).....	74
Figura 41 – Modelagem do módulo de materiais (Elaborada pelo autor).....	74
Figura 42 – Modelagem do módulo de serviço (Elaborada pelo autor)	75
Figura 43 – Modelagem das entidades do módulo de Licitação (Elaborado pelo autor)	75
Figura 44 – Modelagem do módulo de contrato (Elaborada pelo autor).....	76
Figura 45 – Documentação da consulta básica do método Materiais (Elaborado pelo autor)	77
Figura 46 – Documentação da consulta detalhada do método Material (Elaborado pelo autor).....	77
Figura 47 – Documentação do método de Município gerado pela Swagger (Elaborado pelo autor).....	78
Figura 48 – Documentação do método de consulta de Municípios (Elaborada pelo autor)	79
Figura 49 – Dados da consulta de municípios disponibilizados em HTML (Elaborada pelo autor).....	80
Figura 50 - Dados da consulta de municípios disponibilizados em XML (Elaborada pelo autor).....	80
Figura 51 - Dados da consulta de municípios disponibilizados em JSON (Elaborada pelo autor).....	81

Figura 52 - Dados da consulta de municípios disponibilizados em CSV (Elaborada pelo autor)..... 81

Lista de Tabelas

Tabela 1 – Dimensões de qualidade de dados (Traduzido de (PIPINO <i>et al.</i> , 2002))....	20
Tabela 2 – Função de cada operação do HTTP (Elaborado pelo autor).....	26
Tabela 3 - Comparação entre as duas principais ferramentas ETL de código aberto (Adaptado de (NETO, TRAJANO CARLOS MONTASSIER, 2012))	33
Tabela 4 – Quantitativo de arquivos disponibilizados (Elaborado pelo autor)	61
Tabela 5 – Nível de maturidade da API seguindo o modelo de Richardson (Elaborado pelo autor).....	84
Tabela 6 – Comparativo da API desenvolvida com os princípios definidos pela OpenGovData (Elaborado pelo autor).....	84

Lista de abreviaturas, siglas, símbolos

API - *Application Programming Interface*

CATMAT - Catálogo de Material

CATSER - Catálogo de Serviço

CETIC - Centro de Estudos sobre Tecnologias da Informação e da Comunicação

CSV - *Comma-Separated Values*

DAG - Dados Abertos Governamentais

EJB - *Enterprise JavaBeans*

E-gov - Governo Eletrônico

e-SIC - Sistema Eletrônico do Serviço de Informações ao Cidadão

ETL - *Extract, Transform and Load*

Java EE - *Java Platform, Enterprise Edition*

JAX-RS - *Java API for RESTful Web Services*

JPA - *Java Persistence API*

JSON - *JavaScript Object Notation*

G2B – *Government to Businesses* (Governo para Empresa)

G2C – *Government to Citizens* (Governo para Cidadão)

G2E – *Government to Employees* (Governo para Funcionário)

G2G – *Government to Governments* (Governo para Governo)

HAL – *Hypertext Application Language*

HATEOAS - *Hypermedia as the Engine of Application State*

HTML - *HyperText Markup Language*

HTTP - *Hypertext Transfer Protocol*

JSON - *JavaScript Object Notation*

INDA - Infraestrutura Nacional de Dados Abertos

MIT - *Massachusetts Institute of Technology*

MPOG - Ministério do Planejamento, Orçamento e Gestão

OGP - *Open Government Partnership* (Parceria para Governo Aberto)

OKF - *Open Knowledge Foundation*

PDI - *Pentaho Data Integration*

PDM – Padrão Descritivo de Material

RDF - *Resource Description Framework*

REST - *Representational State Transfer*

SCDP - Sistema de Concessão de Diárias e Passagens
SLTI - Secretaria de Logística e Tecnologia da Informação
SERPRO - Serviço Federal de Processamento de Dados
SICONV - Sistema de Gestão de Convênios
SIC - Serviço de Informações ao Cidadão
SICAF - Sistema de Cadastramento Unificado de Fornecedores
SIAFI - Sistema Integrado de Administração Financeira
SIAPE - Sistema Integrado de Administração de Pessoas
SIASG - Sistema Integrado de Administração de Serviços Gerais
SISG - Sistema de Serviços Gerais
SIDOR - Sistema Integrado de Dados Orçamentários
SIGPLAN - Sistema de Informações Gerenciais de Planejamento
SIORG - Sistema de Informações Organizacionais
SOAP - *Simple Object Access Protocol*
TIC - Tecnologias de Informação e Comunicação
TDQM - *Total Data Quality Management*
TOS - *Talend Open Studio*
TOSDI - *Talend Open Studio for Data Integration*
TOSDQ - *Talend Open Studio for Data Quality*
TQdM - *Total Quality data Management*
TQM - *Total Quality Management*
W3C - *World Wide Web Consortium*
URI - *Uniform Resource Identifier*
XML - *eXtensible Markup Language*

1. Introdução

Neste capítulo, será apresentada uma breve contextualização acerca do tema desta dissertação. Em seguida, serão enumerados os objetivos almejados e ao final a organização do texto.

1.1 Contextualização

A globalização aliada a programas de inclusão digital tem permitido, nos últimos anos, uma rápida evolução das Tecnologias de Informação e Comunicação (TIC), principalmente da Internet. Esse cenário tem possibilitado o crescimento e a disseminação do Governo Eletrônico. Neste contexto, é esperado que as inovações tecnológicas permitam que o governo desempenhe melhor suas funções oferecendo serviços de melhor qualidade.

Em paralelo, no cenário internacional diversas parcerias foram firmadas envolvendo países ao redor do mundo. Dentre elas, cabe ressaltar a Parceria para Governo Aberto – OGP, uma aliança internacional com participação do Brasil, onde os membros assumem um compromisso para a promoção da transparência. A adesão do governo brasileiro nesta parceria foi um marco histórico.

Para atender as metas desta parceria, o governo brasileiro criou a Infraestrutura Nacional de Dados Abertos – INDA, onde se estabeleceu diversos marcos distribuídos em dois planos de ação.

Além disso, a Lei de Acesso à Informação (LAI) foi promulgada, obrigando o Governo a disponibilizar seus dados públicos e permitindo que qualquer pessoa possa solicitar dados públicos do governo. Com isso, o cidadão acaba ganhando um papel de

maior destaque no governo, contribuindo para exercício da democracia. Aliado a isso, o incentivo à transparência pública tem estimulado o desenvolvimento e a utilização de novas tecnologias.

Outro ponto a considerar é o rápido crescimento da quantidade de informações armazenadas. Cada vez mais as organizações governamentais armazenam e distribuem os dados. Para conseguir que estas informações sejam disponibilizadas e principalmente reutilizadas, é necessário garantir ao menos a qualidade dos dados. Para isso, o importante é ser sempre transparente, de forma que a abertura de dados possibilite aos consumidores de dados, a descoberta de problemas e dessa forma, estimular a aplicação de técnicas de qualidade de dados no sistema de origem.

Neste sentido, este trabalho propõe uma arquitetura para abertura de dados em sistemas estruturadores de governo que visa facilitar a publicação dos dados, garantir a qualidade dos dados publicados e seguir padrões comumente utilizados.

1.2 Objetivos

Dentro do contexto apresentado na seção anterior, os objetivos deste trabalho são:

1. Criar uma arquitetura de software que facilite a publicação de dados contidos em sistemas estruturantes em dados abertos;
2. Garantir o nível de qualidade dos dados publicados;
3. Implementar uma solução utilizando tecnologias alinhadas com padrões governamentais e internacionais, quando aplicáveis;
4. Ser um padrão referencial de governo para futuras iniciativas de publicação em dados abertos.

1.3 Organização do Trabalho

Este trabalho está organizado em cinco capítulos.

O Capítulo 1 – Introdução, que contextualiza o cenário em que se insere este trabalho e expõe as motivações, objetivos, e organização da estrutura do texto acerca desta dissertação.

O Capítulo 2 – Fundamentação Teórica, que apresenta uma breve revisão da literatura acerca dos temas que o autor considera importante para o entendimento deste trabalho: Governo Eletrônico, Dados Abertos Governamentais, Qualidade de dados, *Web Service*, e ETL. Serão apresentadas definições, classificações e modelos defendidos por estudiosos desses temas.

O Capítulo 3 – Proposta, que apresenta o cenário atual, a motivação e a proposta de uma arquitetura de software de suporte para disponibilização de dados abertos governamentais em sistema estruturadores do governo.

O Capítulo 4 – Estudo de caso, que apresenta a descrição do estudo de caso com uma breve introdução à Sistemas Estruturadores do Governo, uma visão geral sobre o SIASG, os padrões, recursos tecnológicos adotados, e, ao final, a implementação da arquitetura de *software* proposta como prova de conceito do modelo proposto.

O Capítulo 5 – Conclusão, que apresenta as conclusões do trabalho, analisando os resultados obtidos, limitações do modelo proposto, assim como possibilidades de trabalhos futuros.

Ao final, são apresentadas as Referências Bibliográficas deste trabalho. É importante ressaltar que a geração das referências foi feita usando a ferramenta Zotero¹ e que a formatação seguiu a Norma para a Elaboração Gráfica de Teses/Dissertações da COPPE/UFRJ.

2. Fundamentação Teórica

Neste capítulo, será dada uma breve fundamentação teórica acerca dos principais conceitos relacionados a este trabalho. São eles: Governo Eletrônico, Dados Abertos Governamentais, Qualidade de Dados, *Web Service* e ETL.

2.1 Governo Eletrônico

O conceito de Governo Eletrônico é multidisciplinar e bastante abrangente, ou seja, com diferentes definições na literatura. Uma definição mais genérica foi dada por (AGUNE & CARLOS, 2005) que caracteriza Governo Eletrônico como “[...] conjunto de ações modernizadoras vinculadas ao setor público que começam a ganhar visibilidade a partir de 1997”. Entretanto, para efeito deste trabalho, a definição que mais se adequa e que será abordada é relacionada a utilização de Tecnologias de Informação e Comunicação (TIC) na administração pública.

Neste sentido, (HACHIGIAN, 2002) define: “[...] Governo Eletrônico é o uso das TIC para promover governos mais eficientes e efetivos, facilitar o acesso aos serviços

¹ Disponível em <https://www.zotero.org/>

públicos, permitir maior acesso público à informação, e fazer o governo mais *accountable* para os cidadãos" ².

Em uma visão mais idealista, é possível imaginar que o Governo Eletrônico vai além da inclusão das TIC para automatização, modernização dos processos e dos serviços do governo. Pode-se pensar como uma transformação do estado, que impacta na maneira como este desempenha seu papel, e de suas relações com a sociedade. E desta forma, garantir uma maior eficiência e qualidade dos serviços prestados.

Um ponto importante a considerar é que cidadãos não são os únicos beneficiados pelo Governo Eletrônico. Dentro dessa análise de relacionamentos do governo, especialistas comumente identificam três ou quatro grupos de interações. Abaixo, segue uma classificação em quatro grupos na visão de (GRIFFIN, DAVID *et al.*, 2013):

- a) Governo para o Cidadão (G2C): envolve a relação entre o governo e cidadãos. O G2C permite que as agências do governo conversem, ouçam, se relacionem e se comuniquem continuamente com os seus cidadãos, apoiando, desta forma, a prestação de contas, a democracia e as melhorias para os serviços públicos. G2C permite que os clientes acessem informações e serviços governamentais instantaneamente, de forma conveniente e de qualquer lugar.
- b) Governo para Empresas (G2B): consiste de interações entre agências governamentais e empresas privadas. Permite iniciativas de transações eletrônicas, tal como contratações públicas e desenvolvimento de um mercado eletrônico para o governo tornando as empresas mais competitivas [pois, a burocrática é simplificada].
- c) Governo para Governo (G2G): envolve a relação entre organizações governamentais e também com outras organizações governamentais estrangeiras. Existe uma forte dependência entre os diversos departamentos do governo para entregar efetivamente serviços e atribuir responsabilidades. G2C foca na comunicação e cooperação on-line entre as agências e departamentos governamentais para compartilhar bases de dados, recursos, habilidades e capacidades sob

² Tradução do autor para: “[...] e-government is the use of ICT to promote more efficient and effective government, facilitate more accessible government services, allow greater public access to information, and make government more accountable to citizens. E-government might involve delivering services via the Internet, telephone, community centers (self-service or facilitated by others), wireless devices or other communications systems.”

o ponto de vista de evitar duplicação, aumentando a eficiência e efetividade dos processos.

- d) Governo para Funcionários (G2E): envolve a relação entre governo e seus funcionários. Ele dá aos funcionários a possibilidade de acessar informações relevantes sobre: políticas de renumeração e benefícios, treinamento e oportunidades de aprendizagem, leis, etc. G2E refere-se também aos mecanismos estratégicos e táticos para incentivar a implementação de metas e programas governamentais, bem como gestão de recursos humanos, orçamento e contabilidade.

Além disso, ainda é possível classificar o Governo Eletrônico em quatro perspectivas conforme (LENK & TRAUNMULLER, 2000) propõe:

1. A perspectiva do destinatário onde a interface do cidadão com o trabalho administrativo é o mais importante;
2. A perspectiva de processo em que reorganização de processos, que faz uso de todos os tipos de sinergias entre homem e máquina, é primordial;
3. A perspectiva de cooperação, que complementa a perspectiva de processo especialmente através da insistência na cooperação em esforços colaborativos como reuniões, negociações ou deliberações que não seguem um modelo de processo bem definido e que não pode ser totalmente padronizado;
4. A perspectiva do conhecimento que destaca a gestão de informação e conhecimento como o maior bem em diversas situações de trabalho no setor público.

Mesmo com toda essa análise acerca deste conceito, um dos maiores desafios do governo eletrônico ainda é a sua implementação. Dessa forma, (UNIES & MANAGEMENT, 2008) propõe um modelo de maturidade para o Governo Eletrônico dividida em cinco estágios:

1. Emergindo³: A presença online de um governo é composta principalmente de uma página *web* e/ou um sítio oficial; *links* para ministérios ou departamento de educação, saúde, assistência social, trabalho e finança podem ou não existir. Muitas das informações são estáticas e há uma pequena interação com os cidadãos.

³ Tradução do autor para: Emerging.

2. Aprimorado⁴: Governos provem mais informações em política pública e governança. Eles criam *links* para informações arquivadas que é facilmente acessível pelos cidadãos, como por exemplo, documentos, formulários, leis, regulamentos, relatórios, boletins informativos.
3. Interativo⁵: Governos entregam serviços *online* como formulários para pagamento de taxas e renovação de licenças. Além disso, o início de um portal interativo ou sítio com serviços para melhorar a conveniência dos cidadãos são evidentes.
4. Transacional⁶: Governos começam a transformar a si mesmo introduzindo interações de duas vias entre cidadãos e governo. Isto inclui opções para pagamento de taxas, solicitação de carteira de identidade, certidão de aniversário, passaportes e renovação de licenças, assim como, outras interações G2C, e permite que cidadãos acessem esses serviços *online* 24 horas por dia Todas as transações são conduzidas *online*.
5. Conectado⁷: Governo transformam a si mesmos em uma entidade conectada que responde pelas necessidades dos cidadãos desenvolvendo uma infraestrutura de *back office* integrada. Este é o nível mais sofisticado para iniciativas de governo eletrônico *online*.

2.1.1 Governo Eletrônico no Brasil

No cenário brasileiro, o Governo Eletrônico surgiu no ano de 2000, através do Decreto Presidencial de 3 de abril de 2000, que institui o "[...] Grupo de Trabalho Interministerial para examinar e propor políticas, diretrizes e normas relacionadas com as novas formas eletrônicas de interação." (BRASIL, 2010).

Dentre as principais ações deste grupo, está a criação do Comitê Executivo do Governo Eletrônico (CEGE) em outubro de 2000, com o objetivo de formular políticas, estabelecer diretrizes, coordenar e articular as ações de implantação do Governo Eletrônico.

⁴ Tradução do autor para: Enhanced.

⁵ Tradução do autor para: Interactive.

⁶ Tradução do autor para: Transactional.

⁷ Tradução do autor para: Connected.

Dentre outros fatos relevantes, pode-se citar a definição do Padrão de Interoperabilidade em Governo Eletrônico (e-PING) em 2004 e do Modelo de Acessibilidade de Governo Eletrônico (e-MAG) em 2005.

O e-PING “define um conjunto mínimo de premissas, políticas e especificações técnicas que regulamentam a utilização da Tecnologia de Informação e Comunicação (TIC) na interoperabilidade de serviços de Governo Eletrônico, estabelecendo as condições de interação com os demais Poderes e esferas de governo e com a sociedade em geral.” (BRASIL, 2015).

Enquanto o e-MAG define uma série de recomendações, aderente à padrões internacionais, para facilitar e padronizar a implementação da acessibilidade digital de sítios do governo federal.

A política de Governo Eletrônico do Brasil segue um conjunto de diretrizes que atuam em três frentes fundamentais:

- a) Junto ao cidadão;
- b) Na melhoria da sua própria gestão interna;
- c) Na integração com parceiros e fornecedores.

(PORTAL DE GOVERNO ELETRÔNICO DO BRASIL, 2015).

E sendo determinado por sete diretrizes gerais de implantação e operação do Governo Eletrônico:

1. A prioridade do Governo Eletrônico é a promoção da cidadania;
2. A Inclusão Digital é indissociável do Governo Eletrônico;
3. O Software Livre é um recurso estratégico para a implementação do Governo Eletrônico;
4. A gestão do conhecimento é um instrumento estratégico de articulação e gestão das políticas públicas do Governo Eletrônico;
5. O Governo Eletrônico deve racionalizar o uso de recursos;

6. O Governo Eletrônico deve contar com um arcabouço integrado de políticas, sistemas, padrões e normas;
7. Integração das ações de Governo Eletrônico com outros níveis de governo e outros poderes. (PORTAL DE GOVERNO ELETRÔNICO DO BRASIL, 2015).

Um aspecto importante da implementação do Governo Eletrônico é a sua relação com o aumento da utilização das TIC. Segundo dados do Centro de Estudos sobre as Tecnologias da Informação e da Comunicação (CETIC), entidade responsável por monitorar a adoção das TIC no Brasil, a proporção de indivíduos que acessaram a *Internet* no Brasil tem aumentado nos últimos anos, conforme ilustra a Figura 1.

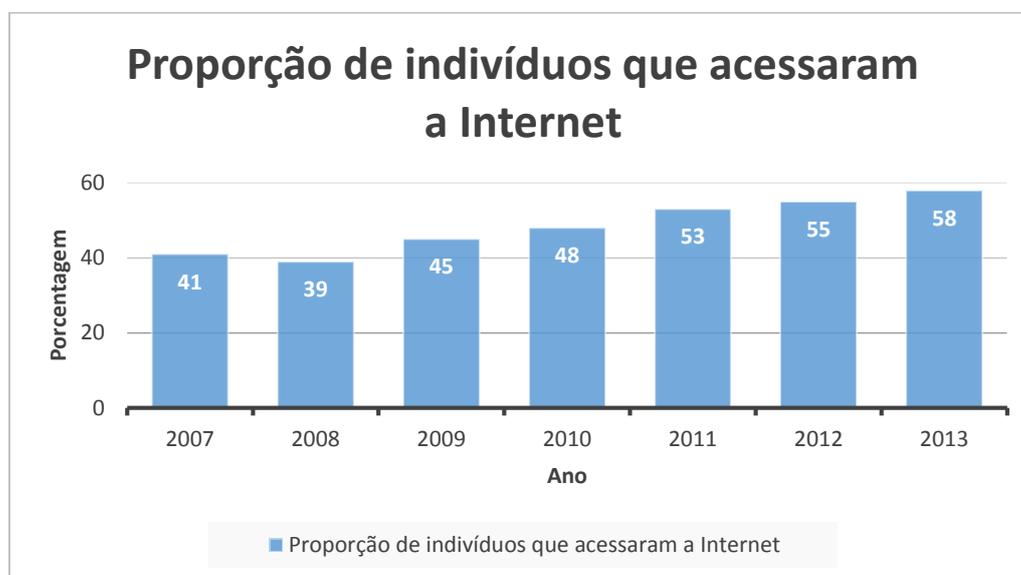


Figura 1 – Proporção de indivíduos que acessaram a *Internet* (Elaborado pelo autor, dados colhidos de (CETIC, 2013))

A Figura 2 mostra a evolução da utilização do Governo Eletrônico pela população com mais de 15 anos no Brasil. É possível perceber que os dados mais que duplicaram nos anos de 2012 e 2013, em relação aos anos anteriores. Indicando uma maior procura nos serviços eletrônicos governamentais.

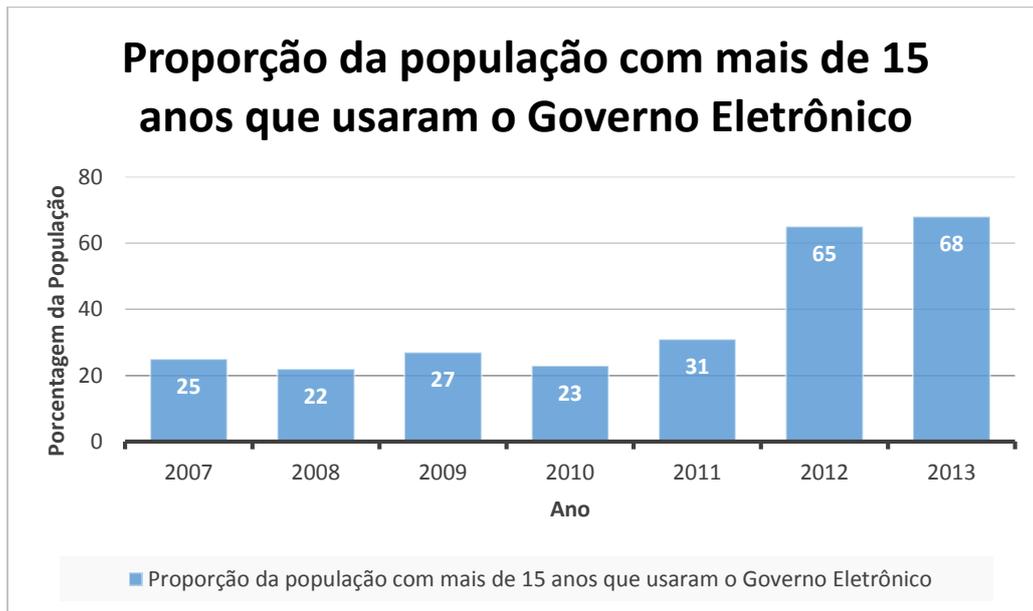


Figura 2 – Proporção da população com mais de 15 anos que acessaram os serviços do Governo Eletrônico (Elaborado pelo autor, dados colhidos de (CETIC, 2013))

A Figura 3 ilustra a utilização dos diversos serviços de Governo Eletrônico pela população com mais de 15 anos. É importante ressaltar que o cidadão pode utilizar mais de um serviço.

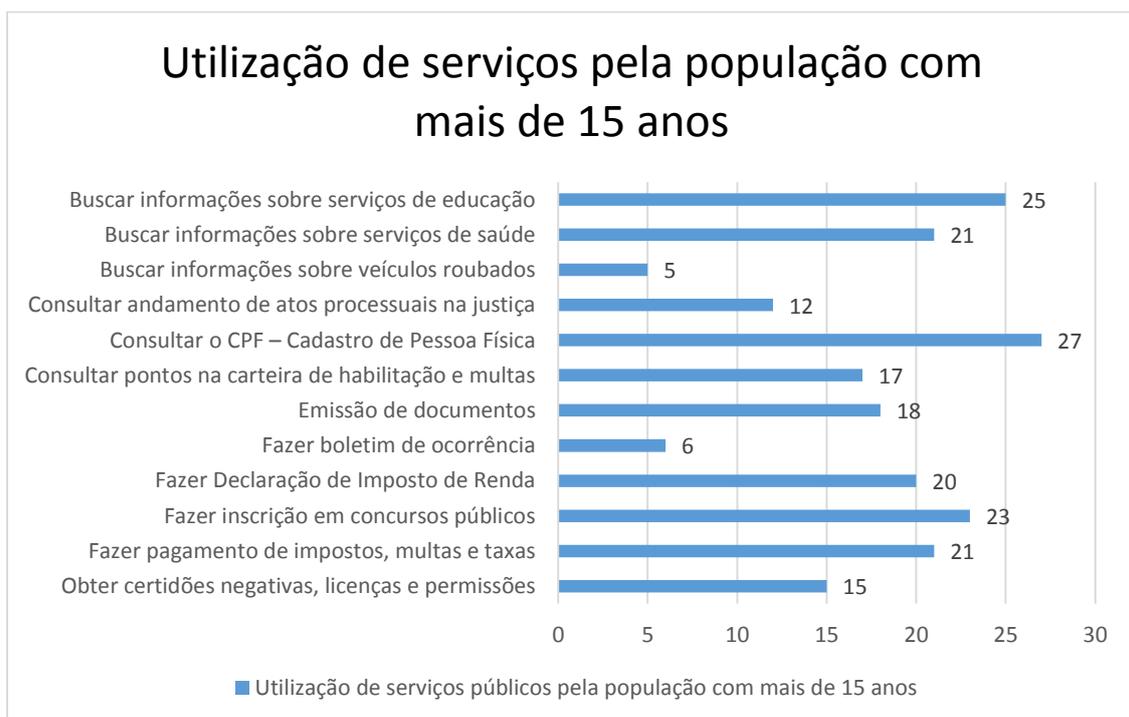


Figura 3 – Utilização de serviços públicos pela população com mais de 15 anos (Elaborado pelo autor, dados colhidos de (CETIC, 2013))

2.2 Dados Abertos Governamentais

“Em 2009, [o tema] Dados Abertos começou a aparecer nas grandes mídias, com governos de vários países (como Estados Unidos, Reino Unido, Canadá e Nova Zelândia) anunciando novas iniciativas voltadas a abertura de informação pública.”⁸ (OPEN KNOWLEDGE FOUNDATION (OKF), 2012, p.3).

Segundo a Fundação de Conhecimento Aberto, (OPEN KNOWLEDGE FOUNDATION (OKF), 2012, p.6): “Dados Abertos são dados que podem ser usados livremente, reutilizados e redistribuídos por qualquer pessoa estando sujeito a, no máximo, a exigência de creditar a sua autoria e compartilhar pela mesma licença.”⁹

No entanto, apenas exibir os dados não é o suficiente para os considerar como dados abertos. Para isso o especialista em políticas públicas David Eaves propôs as leis¹⁰ para Dados Abertos Governamentais (DAG) baseadas em três princípios: ser facilmente encontrado, manipulado e compartilhado.

1. Se o dado não puder ser encontrado ou indexado por nenhuma ferramenta de busca, então ele não existe;
2. Se não estiver disponível em um formato aberto e compreensível por máquinas, ele não pode ser aproveitado;
3. Se algum arcabouço¹¹ válido [para manipular o dado] não permitir que seja compartilhado, ele não é útil. (EAVES, DAVID, 2009).

⁸ Tradução do autor para: “In 2009 open data started to become visible in the mainstream, with various governments (such as the USA, UK, Canada and New Zealand) announcing new initiatives towards opening up their public information.”

⁹ Tradução do autor para: “Open data is data that can be freely used, re-used and redistributed by anyone - subject only, at most, to the requirement to attribute and sharealike”.

¹⁰ Traduzido e adaptado pelo autor.

¹¹ Tradução do autor para: *framework*.

Neste sentido, o grupo de especialistas em dados abertos do OpenGovData¹², definiu oito princípios¹³ para os DAG que são:

1. Completo - Todos os dados públicos estão disponíveis. Dado público é aquele que não está sujeito à limitações de privacidade, segurança ou privilégios.
2. Primário - Os dados são disponibilizados como coletados na fonte, com o maior nível de granularidade, e sem agregação ou modificação.
3. Atual - Os dados são disponibilizados tão rápido quanto o necessário para preservar o seu valor.
4. Acessível - Os dados são disponibilizados para o maior número possível de usuários e para o maior número de finalidades.
5. Compreensível por máquina - Os dados são suficientemente estruturados para permitir o processamento automatizado.
6. Não discriminatório - Os dados são disponíveis para todos, sem exigência de cadastro para acessá-los.
7. Não proprietário - Os dados são disponibilizados em formatos sobre o qual nenhuma entidade detenha o controle exclusivo (formato não proprietário).
8. Livre de Licença - Os dados não estão sujeitos a nenhuma restrição de direito autoral, patente, marca registrada ou segredo comercial. Restrições sensatas de privacidade, segurança ou privilégios são permitidas. (MALAMUD, CARL *et al.*, 2007).

Seguindo orientações da cartilha técnica para publicação de dados abertos publicada pela (SECRETARIA DE LOGÍSTICA E TECNOLOGIA DA INFORMAÇÃO (SLTI) & MINISTÉRIO DO PLANEJAMENTO ORÇAMENTO E GESTÃO (MPOG), 2012), os dados abertos precisam ser disponibilizados em formato livre, dentre os quais vale a pena destacar:

- CSV - Significa *Comma-Separated Values*, ou valores separados por vírgula, e é um formato para armazenamento de dados tabulares em texto. A codificação é muito simples: cada linha do arquivo representa uma linha na tabela, e as colunas são separadas por vírgula. Campos que podem conter vírgula devem ser delimitados por aspas. CSV é recomendado para representação de estrutura de dados mais simples, de natureza tabular, onde não existem subpropriedades ou listas, gerando um arquivo menor e mais leve para processamento. Arquivos CSV

¹² Acessível em <http://opengovdata.org/>

¹³ Traduzido e adaptado pelo autor.

são processáveis diretamente por editores de planilhas, como o OpenOffice e o MS Excel.

- JSON - É um acrônimo para *JavaScript Object Notation*. É um padrão aberto de estruturação de dados baseado em texto e legível por humano. A especificação é a RFC 4627. JSON ganhou maior utilização com o advento do Ajax. A serialização em JSON é muito simples e resulta em uma estrutura pouco verbosa o que se mostra uma ótima alternativa para o XML. JSON possibilita serialização de estrutura de objetos complexos, como listas e subpropriedades. JSON está se tornando o padrão mais utilizado para integração de dados entre repositórios e *frameworks*, também está se tornando o padrão nativo de armazenamento em alguns bancos de dados modernos.

- RDF - Significa *Resource Description Framework*, é um modelo de dados estruturado em grafos e possui diversos formatos de serialização, tais como RDF/XML, *Notation 3* e *Turtle*. Os formatos baseados em RDF têm seus dados descritos em vocabulários disponíveis na Web. Apesar da grande qualidade dos dados disponibilizados em RDF, a construção de vocabulários para seu uso não é trivial. Numa escala de níveis de qualidade/complexidade de dados abertos, o RDF está no último nível, onde se constituirá a Web semântica.

- XML - Significa *Extensible Markup Language*. É um conjunto de regras para codificar documentos com estrutura hierárquica e em um formato legível por máquina. É baseado em texto e tem como principais objetivos simplicidade, extensibilidade e usabilidade. XML é largamente utilizado como formato de troca de dados nos clássicos *Web Services* SOAP. Possui uma ampla gama de ferramentas associadas, tais como o padrão XSLT que permite transformar para outra estrutura XML ou outro formato. Apesar de sua ampla utilização, tem sido menos encorajada a utilização desse formato para integração de aplicações na Web, por utilizar mais recursos para transmissão e para o processamento dos dados. Em substituição, recomenda-se utilizar JSON.

A publicação de dados abertos traz diversas vantagens para o governo, dentre eles, vale a pena destacar:

- a) Aumento da participação social - Maior transparência do governo e aumento da democracia que impacta positivamente na maior participação da sociedade;
- b) Estímulo a inovação – A utilização de dados abertos estimulará a sociedade a elaborar diferentes formas de agregar e visualizar os dados;
- c) Economia de tempo e dinheiro - uma vez disponibilizados os dados, eles ficariam disponíveis para qualquer pessoa interessada, e dessa forma evitaria ações (de solicitação de dados) duplicadas.

- d) Melhora nos serviços governamentais – muitos dos problemas que acontecem no âmbito de governo, decorrem da falta de comunicação entre os diferentes setores de governo e da impossibilidade de sistemas conversarem entre si.

2.2.1 Dados Abertos no Brasil

Nesta seção, serão abordadas algumas das principais iniciativas relacionadas (direta ou indiretamente) à política de dados abertos no Brasil. Dentre essas iniciativas, vale destacar: Lei de Acesso a Informação, Parceria do Governo Aberto, Infraestrutura Nacional de Dados Abertos.

2.2.1.1 Lei de Acesso a Informação

A Lei de Acesso a Informação (LAI) foi regulamentada pelo Decreto 7.724/2012 que entrou em vigor no Brasil em 16 de maio de 2012. Esta lei define mecanismos que possibilitam, a qualquer pessoa, física ou jurídica, o recebimento de informações públicas dos órgãos e entidades de governo. Ficam subordinados a essa lei:

I - os órgãos públicos integrantes da administração direta dos Poderes Executivo, Legislativo, incluindo as Cortes de Contas, e Judiciário e do Ministério Público;

II - as autarquias, as fundações públicas, as empresas públicas, as sociedades de economia mista e demais entidades controladas direta ou indiretamente pela União, Estados, Distrito Federal e Municípios. (BRASIL, 2011).

A publicação desta lei é um marco importante para a democracia no Brasil, pois estabelece como regra o livre acesso a informação pública e tratando o sigilo como exceção. Dessa forma, contribui para gerar uma sociedade mais participativa, onde o cidadão deixa de ser um mero observador, e passa a atuar como protagonista de um país mais transparente.

Com finalidade de atuar como um canal de contato entre a sociedade e o setor público, foram criados postos físicos de Serviço de Informações ao Cidadão (SIC) que tem como funções:

- a) atender e orientar o público quanto ao acesso a informações;
 - b) informar sobre a tramitação de documentos nas suas respectivas unidades;
 - c) protocolizar documentos e requerimentos de acesso a informações;
- (BRASIL, 2011).

Neste contexto, foi criado também o Sistema Eletrônico do Serviço de Informação ao Cidadão (e-SIC). Este sistema “[...] centraliza as entradas e saídas de todos os pedidos de acesso dirigidos ao Poder Executivo Federal. O objetivo do e-SIC é organizar e facilitar os procedimentos de acesso à informação tanto para os cidadãos quanto para a Administração Pública.” (CGU, 2015b).

A partir do pedido de disponibilização de um dado, o órgão ou entidade tem até 20 dias corridos para atender ao pedido. Ao final deste prazo, com devida justificativa, é possível prorrogar o prazo por mais dez dias.

2.2.1.2 Parceira do Governo Aberto

A Parceria do Governo Aberto - *Open Government Partnership* (OGP) - é uma iniciativa internacional de adesão voluntária lançada em 20 de setembro de 2011, com participação inicial de oito países: África do Sul, Brasil, Estados Unidos, Filipinas, Indonésia, México, Noruega e Reino Unido. Os países participantes se comprometem “[...] a adotar medidas concretas para o fortalecimento da transparência das informações e atos governamentais, combate à corrupção, fomento à participação cidadã, gestão dos recursos públicos, integridade nos setores público e privados, entre outros objetivos.” (OGP, 2015)

A OGP definiu cinco grandes desafios, para os quais, os compromissos devem ser enquadrados, são eles: melhoria dos serviços públicos; aumento da integridade pública; gestão mais efetiva dos recursos públicos; criação de comunidades mais seguras; e aumento da responsabilidade corporativa.

Segundo a (OGP, 2015) atualmente 65 países participam desta parceria. E até o momento, cerca de mil compromissos, no total, foram assumidos por esses países para tornar seus governos mais transparentes.

2.2.1.3 Infraestrutura Nacional de Dados Abertos

Como consequência da participação do Brasil na OGP, surge o Decreto s/ºn de 15 de setembro de 2011 que institui o Plano de Ação Nacional sobre o Governo Aberto, o qual estabelece o compromisso do governo de implantar a Infraestrutura Nacional de Dados Abertos (INDA).

Neste sentido, em 12 de abril de 2012, a Instrução Normativa SLTI nº 4 (BRASIL, 2012) institui a INDA, que tem como objetivos:

- I – definir, estruturar e coordenar a política de dados abertos, bem como estabelecer o seu modelo de funcionamento;
- II – promover o ordenamento na geração, armazenamento, acesso, e compartilhamento de dados para uso do Poder Executivo federal e da sociedade;
- III- definir e disciplinar os padrões e os aspectos técnicos referentes à disponibilização e disseminação de dados para uso do Poder Executivo federal e da sociedade;
- IV – promover o compartilhamento de recursos de tecnologia da informação e evitar a duplicidade de ações e o desperdício de recursos na disseminação de dados e informações pelos órgãos e entidades do Poder Executivo federal;
- V - apoiar, capacitar e fornecer suporte para a publicação de dados abertos aos órgãos e entidades do Poder Executivo federal ou que aderirem à INDA que não possuem prática, cultura e atribuições finalísticas de disseminação de dados;
- VI – buscar a melhoria contínua da publicação de dados abertos, baseando-se nas melhores práticas concebidas nos cenários nacional e internacional;

VII – promover a colaboração entre governos dos os diferentes níveis da federação e entre o Poder Executivo federal e a sociedade, por meio da publicação e do reuso de dados abertos;

VIII – promover e apoiar o desenvolvimento da cultura da publicidade de dados e informações na gestão pública;

IX – disponibilizar tecnologias e apoiar as ações dos órgãos e entidades do Poder Executivo federal ou que aderirem à INDA na implementação da transparência ativa por meios digitais;

X – promover a participação social na construção de um ecossistema de reuso e de agregação de valor dos dados públicos.

Para conseguir alcançar os seus objetivos, foram definidos dois planos de ação pela INDA: no 1º Plano de Ação do Brasil, cinco órgãos do Governo Federal assumiram 32 compromissos, onde 25 foram totalmente cumpridas, cinco foram parcialmente implementados e dois não foram iniciados; atualmente, no 2º Plano de Ação do Brasil, 45 compromissos foram definidos, dentre os quais três deles são continuidade de compromissos que não foram integralmente finalizados no Plano de Ação anterior.

2.3 Qualidade de Dados

Esta seção tem como objetivo apresentar uma breve introdução sobre qualidade de dados, bem como apresentar os seus principais conceitos e algumas metodologias que têm como função orientar e definir técnicas para o uso da informação sobre os dados.

2.3.1 Introdução

O rápido crescimento do uso das TIC tem contribuído para o aumento da quantidade de dados produzidos. Todos os dias, volumes imensos de dados são gerados e armazenados pelas organizações. E isso acabou criando uma preocupação acerca da qualidade dessas informações. Tal cenário recorre à aplicação de Qualidade de dados, que

surgiu com a necessidade de integrar dados. Este tema tem ganhado cada vez mais destaque, principalmente nas organizações.

Segundo (ECKERSON, 2002), no final da década de 1990, muitos dos projetos de TI atrasaram ou fracassaram, devido a imprevistos com problemas de qualidade de dados. Isso acabou despertando a preocupação dos executivos para o custo real dos dados de baixa qualidade. Dados ruins podem significar decisões equivocadas, causando assim possíveis perdas para as organizações (REDMAN, 1998). Neste sentido, em pesquisa feita pelo *Data Warehousing Institute* (TDWI), os problemas de qualidade de dados causam um prejuízo de mais 600 bilhões de dólares por ano às empresas americanas (ECKERSON, 2002).

No entanto, é importante ressaltar que os benefícios de investir em qualidade de dados vão além dos motivos financeiros. A qualidade de dados gera informações mais precisas e oportunas que garantem uma melhor utilização dos recursos, trazendo uma maior confiança e satisfação por parte dos ‘clientes’. (TGC, 2002).

Neste contexto, Richard Wang, professor e diretor do programa de qualidade da informação do *Massachusetts Institute of Technology* (MIT), acredita que não somente as empresas devem investir para melhoria da qualidade dos dados. Mas também, o governo deve ter um envolvimento maior no assunto, investindo e divulgando o conceito. (CLIENTESA, 2012).

2.3.2 Conceito

Na literatura há algumas definições para o termo Qualidade de Dados, alguns autores o referem como “adequado para uso”¹⁴ outros preferem “para atender às

¹⁴ Tradução do autor para: appropriate for use.

necessidades do usuário”¹⁵. Baseado nisso, (WANG & STRONG, 1996) definem Qualidade de Dados como “[...] dados que estão prontos para uso pelos consumidores de dados”¹⁶.

Uma característica da qualidade de dados é a sua forte relação com suas diversas dimensões. Dimensão é definida por (WANG & STRONG, 1996) como: “[...] um conjunto de atributos de qualidade de dados que representa aspectos e construtos simples”¹⁷.

Dessa forma é possível concluir que qualidade de dados é multidimensional, de forma que cada dimensão mostra um lado importante da qualidade de dados. Diversos autores propõem conjuntos de dimensões diferentes. (WANG & STRONG, 1996) após longo estudo, propõe quinze dimensões para qualidade de dados. A Tabela 1 apresenta essas dimensões e as suas respectivas definições.

Dimensões	Definições
Acurácia	Os dados são corretos e confiáveis.
Atualidade	O dado está suficientemente atualizado para a tarefa a ser executada.
Compleitude	Não há ausência de dados, além de serem suficientes em tamanho e profundidade para a realização das tarefas.
Credibilidade	O dado é considerado como verdadeiro e crível.
Facilidade de acesso	O dado já está disponível ou é fácil e rapidamente recuperável.
Facilidade de Entendimento	O dado é fácil de compreender.
Facilidade de Interpretação	O dado possui linguagem, símbolos e unidades apropriadas, além de possuir definições claras.
Objetividade	O dado é imparcial, sem preconceitos e livre de interpretações tendenciosas.
Quantidade de Informação	O volume de dados é apropriado para as tarefas a serem executadas.

¹⁵ Tradução do autor para: to meet end user needs.

¹⁶ Tradução do autor para: as data that are fit for use by data consumer.

¹⁷ Tradução do autor para: as a set of data quality attributes that represent a single aspect or construct of data quality.

Relevância	O dado é aplicável e útil para as tarefas as serem executadas.
Representação concisa	Os dados são representados de forma compacta.
Representação consistente	Os dados são apresentados em um mesmo formato.
Reputação	O dado tem alta reputação quando avaliado em relação à sua fonte e conteúdo.
Segurança de acesso	O acesso ao dado é restrito de maneira apropriada, de forma a manter a segurança.
Valor Agregado	O dado é benéfico e fornece vantagens ao ser utilizado.

Tabela 1 – Dimensões de qualidade de dados (Traduzido de (PIPINO *et al.*, 2002))

Além das dimensões, (WANG & STRONG, 1996) propõe também quatro categorias para elas. As categorias são:

- a) Acessibilidade¹⁸ - As dimensões desta categoria levam em conta aspectos relativos à disponibilidade e segurança dos dados. As dimensões associadas são: Facilidade de acesso e Segurança de acesso.
- b) Contextual¹⁹ - As dimensões desta categoria agem sobre os dados dependendo do contexto. Isso significa que dependendo de quando, onde e de quem manipula os dados, resultados diferentes podem ocorrer. As dimensões associadas são: Relevância, Valor Agregado, Atualidade, Completude e Quantidade de Informação.
- c) Intrínseca²⁰ - As dimensões desta categoria atuam sobre os dados independentes do seu contexto ou aplicação. As dimensões associadas são: Acurácia, Objetividade, Credibilidade e Reputação.
- d) Representacional²¹ - As dimensões desta categoria dizem respeito a aspectos relativos ao formato dos dados. A forma como os dados são apresentados

¹⁸ Tradução do autor para: Accessibility DQ.

¹⁹ Tradução do autor para: Contextual DQ.

²⁰ Tradução do autor para: Intrinsic DQ.

²¹ Tradução do autor para: Representational DQ.

influência diretamente na clareza, coerência e facilidade de entendimento. As dimensões associadas são: Facilidade de Interpretação, Facilidade de Entendimento, Representação Concisa e Representação Consistente.

A Figura 4 abaixo representa bem essas quatro categorias apresentadas e suas respectivas dimensões.

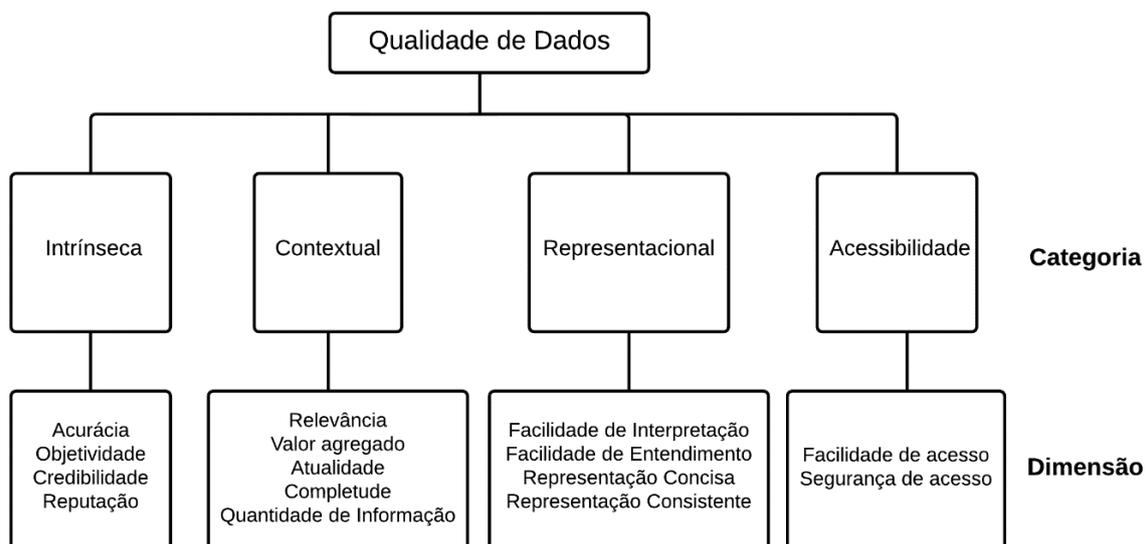


Figura 4 – Categorias e dimensões de qualidade de dados (Adaptado de (WANG & STRONG, 1996))

2.3.3 Metodologia de Qualidade de Dados

Uma metodologia de qualidade de dados é um conjunto de orientações e técnicas que ajudam a definir um processo razoável para medir e melhorar a qualidade dos dados de uma organização. Isso acontece por meio de fases e momentos de decisão, que orientam e regulam aquilo que se faz dentro de um projeto.

Aqui serão apresentadas duas metodologias que, na literatura, são classificadas como de propósito geral. Elas são chamadas assim porque são aplicadas para as mais diversas realidades, ou seja, não focam em uma atividade específica (BATINI & SCANNAPIECA, 2006).

A primeira, TDQM, define por meio de quatro fases um processo para a solução e prevenção de problemas, sendo amplamente difundida hoje por um programa do MIT nessa área. Já a segunda, TQdM é mais voltada para uma melhoria contínua, análises de custo benefício e visão gerencial.

2.3.3.1 Metodologia TDQM

Total Data Quality Management (TDQM) é uma metodologia que foi amplamente disseminada e incrementada pelo *Massachusetts Institute of Technology* (MIT) e é uma extensão para dados do programa *Total Quality Management* (TQM) que atua sobre produtos de fabricação.

O processo TDQM é dividido em quatro fases: definição, medição, análise e melhoria. Estas fases são executadas de forma iterativa, formando um ciclo. Conforme está representado na Figura 5.

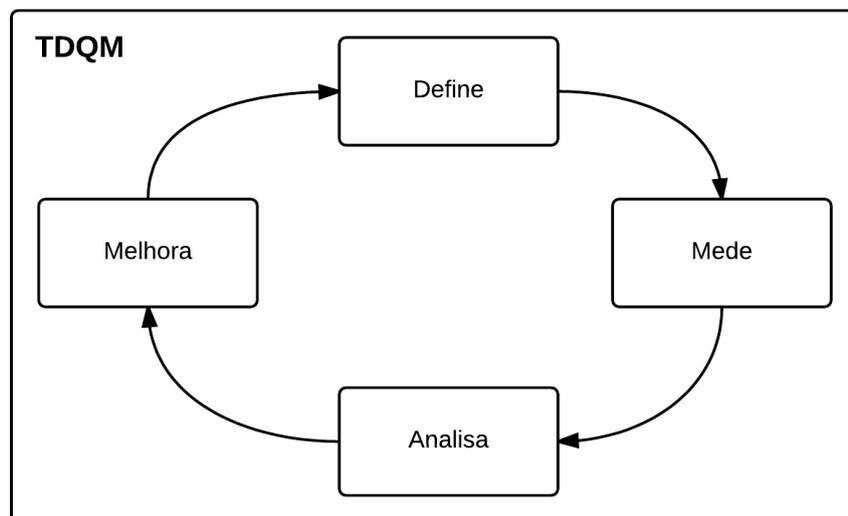


Figura 5 – Fases da metodologia TDQM (Adaptado de (WANG, 1998))

Nesta metodologia, a fase de definição inclui a identificação das dimensões de qualidade de dados e requisitos relacionados. Já a fase de medição produz métricas de qualidade que fornecem um *feedback* e permitem a comparação de qualidade com os

requisitos de qualidade predefinidos. A fase de análise é onde os possíveis problemas são identificados por meio do retorno obtidos pelas métricas. Por último, descreve-se e implementam-se as melhorias para corrigir e evitar potenciais falhas (BATINI & SCANNAPIECA, 2006).

2.3.3.2 Metodologia TQdM

TQdM é a sigla em inglês para *Total Quality data Management* e foi inicialmente projetado para *Data Warehouse*. A metodologia trabalha através de processos iterativos com o objetivo de obter uma melhoria constante da qualidade dos dados até chegar a um nível aceitável de carência de erros.

Em relação a TDQM, essa metodologia faz análises de custo benefício onde é possível avaliar os custos causados pelas perdas de qualidade, os custos gerados pelos próprios processos de melhorias e ainda os benefícios causados pela melhoria da qualidade dos dados. Por fim, a metodologia ainda prevê técnicas que verificam a efetividade das melhorias implementadas. (BATINI & SCANNAPIECA, 2006)

2.4 Web Service

Segundo a *World Wide Web Consortium (W3C)*, *Web Service* pode ser definido como um “sistema de software designado para suportar interações entre máquinas na rede”²².

Enquanto para o (E-PING, 2014), *Web service* é uma “aplicação lógica, programável que torna compatíveis entre si os mais diferentes aplicativos,

²² Tradução do autor para: “a software system designed to support interoperable machine-to-machine interaction over a Network.”

independentemente do sistema operacional, permitindo a comunicação e intercâmbio de dados entre diferentes redes”.

Dentre os principais tipos de *Web service*, há dois que são bastante difundidos e utilizados: SOAP e REST.

2.4.1 SOAP

Segundo a (W3C, 2007), *Simple Object Access Protocol* (SOAP) é um “[...] protocolo que usa o formato XML para trocar mensagens estruturadas em ambientes distribuídos e descentralizados”²³. Esse protocolo tem uma estrutura independente de modelos de programação e por isso permite a interação entre diversas plataformas e sistemas. Sendo assim, ele é muito usado para implementação de serviços *web* em ambientes altamente padronizados, onde as mensagens SOAP são descritas, por exemplo, em WSDL, que é uma linguagem de descrição em formato XML.

A grande vantagem da estrutura padronizada do SOAP é a possibilidade de criar ferramentas ou bibliotecas para poder manipular facilmente as mensagens recebidas. Porém, manter uma estrutura assim significa esforços de verificação para manter o padrão, e isso gera um *overhead*.

Devido a sua natureza, o SOAP pode ser utilizado com qualquer protocolo da camada de aplicação. Sendo assim, não seria possível tirar vantagens de um protocolo específico, como o REST faz com o HTTP. Dessa forma, ao utilizar o SOAP juntamente ao protocolo HTTP, este faz uso somente da operação POST. Com isso, o protocolo funciona apenas como meio de transporte para transportar o pacote SOAP, não podendo

²³ Tradução do autor para: “protocol intended for exchanging structured information in a decentralized, distributed environment.”

incluir qualquer característica à mensagem, pois tudo isso já está encapsulado no pacote SOAP.

2.4.2 REST

Representational State Transfer (REST) é um estilo de arquitetura de software para sistema multimídia distribuídos. Assim como o SOAP, REST também é um protocolo usado para troca de mensagens, porém faz somente uso de HTTP para carregar suas mensagens abertamente, diferentemente do SOAP, que em teoria pode usar qualquer outro protocolo da camada de aplicação.

Os sistemas que seguem esse padrão são chamados *RESTful*. A abordagem usada pelo REST não coloca restrições ao formato da mensagem a ser trocada, podendo ser usado, na prática, qualquer formato, como por exemplo: CSV, JSON e XML. Ficando a critério do desenvolvedor escolher o mais adequado. Isso permite uma maior flexibilidade, torna as aplicações mais rápidas, leves e não exige conhecimento específico para padronizar o acesso e as mensagens que deverão ser trocadas.

Além disso, por ser implementado através do protocolo HTTP, uma das suas características é a utilização das principais operações deste protocolo para a comunicação com os recursos. Dessa forma, os métodos PUT, GET, POST e DELETE são utilizados. Cada um dos métodos tem um propósito definido: GET para recuperar a representação de um recurso, POST para criar um novo recurso, PUT para modificar um recurso existente, DELETE para remover um recurso, conforme está relacionada na Tabela 2.

HTTP	Função
GET	Consultar
POST	Incluir
PUT	Alterar
DELETE	Apagar

Tabela 2 – Função de cada operação do HTTP (Elaborado pelo autor).

Por esse motivo, ele é ideal para aplicações que recebem um alto número de requisições e um grande volume de dados. Contudo, o REST, por não ter uma estrutura bem definida para troca de mensagens, dificulta a comunicação e o compartilhamento de dados entre sistemas ou plataformas, ou seja, a interoperabilidade.

O responsável por criar este padrão, Roy Fielding, definiu um conjunto de restrições para sistemas *RESTful*:

1. Cliente e Servidor – separação entre aplicações clientes e servidores.
2. Servidor sem estado – cada requisição de um cliente contém todas as informações necessárias para o entendimento e processamento da requisição, e não pode se aproveitar de nenhum contexto armazenado no servidor.
3. Cache – Aplicações clientes podem usar o cache para armazenar as respostas. A resposta deve indicar se isto é permitido ou não.
4. Interface uniforme - Define quatro regras que o serviço deve implementar para facilitar futura refatoração. São elas: identificação de recurso por meio de URI; representação do *resource* em JSON, HTML, XML, etc.; Respostas autoexplicativas por meio de metadados nas requisições e resposta; e Acréscimo de Hipermissão.
5. Arquitetura em camadas - Esta restrição diz que o cliente nunca deve chamar diretamente o servidor remoto, o recomendado é que passe por um intermediário, que pode ser um balanceado de carga, qualquer outra máquina que sirva para este propósito.
6. Código sob demanda (opcional) - É um paradigma de sistemas distribuídos que define a possibilidade de mover um código existente no servidor para a execução no cliente. No entanto, o seu uso é opcional. (FIELDING, 2000).

Dentre outros princípios, definidos por Fielding, está o uso de formato hipermissão para representar os dados. Dessa forma, o conceito de HATEOAS foi definido.

2.4.2.1 HATEOAS

HATEOAS é uma abreviação para *Hypermedia as the Engine of Application State*. Um sistema orientado a hipermídia “provê informações para navegar pelas interfaces REST do *site* dinamicamente, pois inclui nas respostas *links* hipermídia”²⁴.

O cliente inicializa a comunicação com a API REST através de uma URI fixa. E a partir do recurso recebido como resposta, será possível descobrir todas as futuras ações e a URIs para manipular o recurso. A Figura 6 ilustra esse conceito.

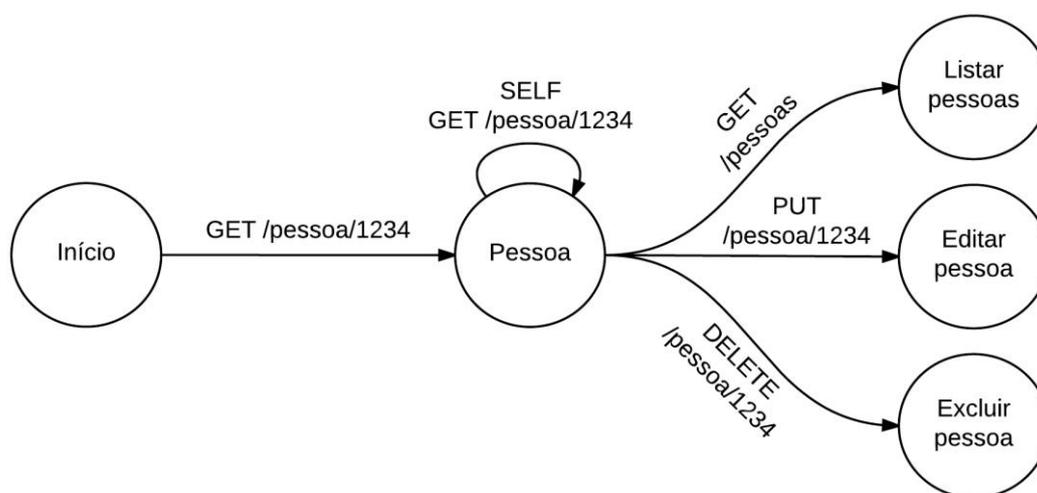


Figura 6 – Exemplo de HATEOAS (Elaborado pelo autor)

A grande vantagem desta abordagem é a facilidade do servidor em mudar as URIs dos recursos, sem causar incompatibilidades com nenhum cliente que a consome. Além de simplificar o desenvolvimento, uma vez que é a partir da requisição a um determinado recurso, é sabido todas as possíveis ações que podem ser executadas a partir dela.

²⁴ Tradução do autor para: “provides information to navigate the site’s REST interfaces dynamically by including hypermedia links with the responses.”

(RICHARDSON, LEONARD, 2010) contribuiu elaborando um Modelo de Maturidade para APIs *RESTful*, classificando-o em quatro níveis, como mostra a Figura 7.

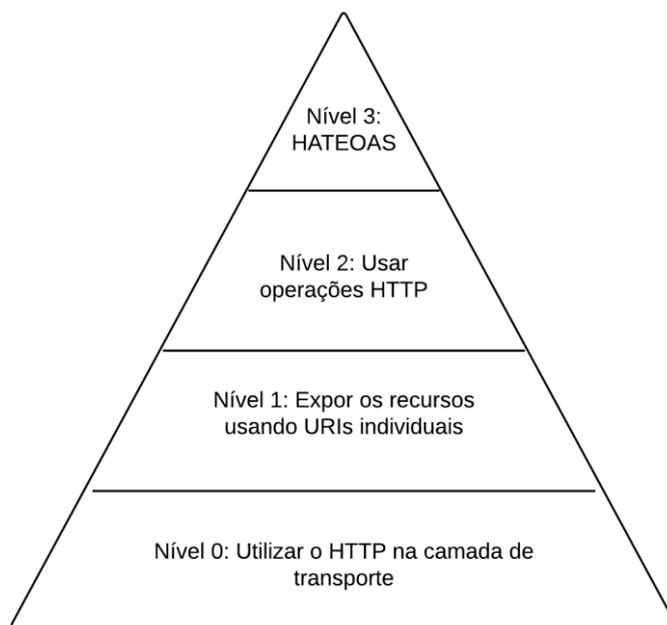


Figura 7 – Modelo de Maturidade de Richardson (Adaptado de (RICHARDSON, LEONARD, 2010))

2.5 ETL

Extraí, Transforma e Carrega²⁵ (ETL) é um processo que visa facilitar a migração e integração de dados. Esta técnica é bastante utilizada principalmente na área de *Data Warehouse*. É dividido em três fases:

- a) Extraí – consiste em fazer a extração dos dados que se deseja carregar. A extração pode ser feita a partir de um arquivo ou banco de dados.
- b) Transforma – consiste em aplicar transformações nos dados obtidos na etapa anterior, dentre as técnicas pode-se citar: codificação, derivação e agregação de

²⁵ Tradução do autor para: Extract, Transform and Load.

valores; padronização, higienização e junção de dados; conversão de tipos; aplicação de regras de negócio e outros. É importante ressaltar que esta etapa é opcional.

- c) Carrega – consiste em carregar os dados obtidos nas fases anteriores em um arquivo ou banco de dados.

Abaixo, a Figura 8 ilustra o processo ETL.

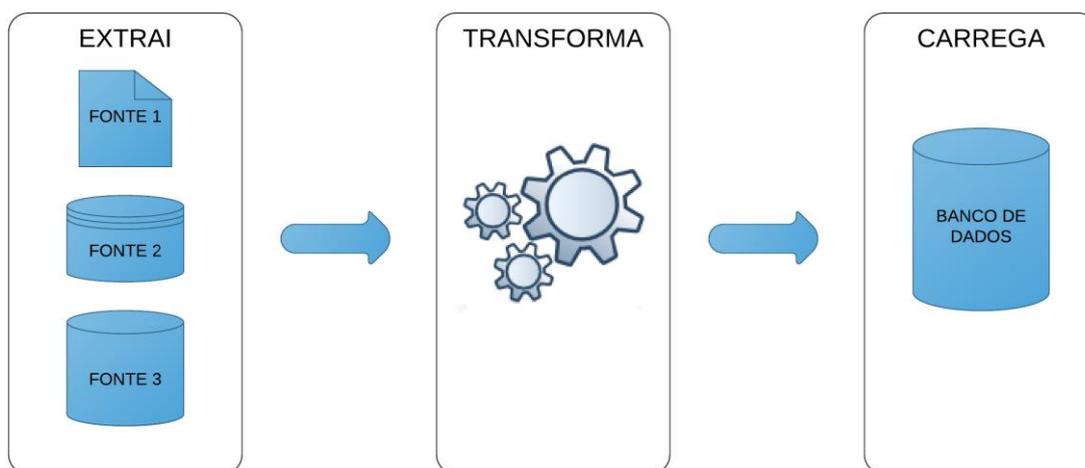


Figura 8 – Processo ETL (Elaborado pelo autor)

Para facilitar este processo, a partir da década de 90, diversas ferramentas ETL foram criadas. Essas ferramentas trazem uma série de facilidades, dentre as mais gerais pode-se citar: o acesso rápido a bases de dados e operações através de componentes específicos, visualização do fluxo de dados e monitoramento de cargas.

Embora o uso dessas ferramentas simplifique o processo de desenvolvimento, o alto custo para adquirir uma licença ainda continua sendo uma das principais desvantagens. Nesse contexto, ferramentas de código aberto passam a ser alternativas interessantes. Elas comumente fornecem os mesmos componentes das versões pagas, sendo restritivas apenas em funcionalidades avançadas como: suporte a versionamento, gerenciamento e monitoramento avançado, testes, *debugger* e desenvolvimento

colaborativo. Sendo assim, ainda seria possível alcançar grandes resultados com os recursos fornecidos pelas ferramentas gratuitas.

Atualmente, existem duas ferramentas ETL de código aberto que se destacam por oferecer soluções de qualidade para o desenvolvimento. São elas: *Talend Open Studio for Data Integration* e *Pentaho Data Integration*.

A seguir, serão apresentadas essas duas ferramentas, e ao final um breve comparativo com análise nos aspectos de infraestrutura, usabilidade e funcionalidades dessas ferramentas.

2.5.1 *Pentaho Data Integration*

A ferramenta *Pentaho Data Integration*²⁶ (PDI) ou simplesmente *Kettle* é uma excelente ferramenta ETL para a integração de dados. Ela traz um vasto recurso para suporte a extração, transformação e carregamento de dados. Além de prover grande facilidade de uso com a sua interface gráfica, sendo fácil e intuitivo, privilegiando a curva de aprendizado.

Existem dois conceitos importantes no *Kettle*: Transformação e *Job*. O primeiro é uma rotina que contém uma coleção de passos para a realização de uma tarefa: ler um arquivo CSV, remover uma coluna e exportar os dados em formato XML. Já o segundo, é uma coleção de rotinas, onde é possível executar várias transformações e até mesmo outro *Jobs*.

Uma característica do *Kettle* é que Transformações e *Jobs* são salvos em arquivos XML. Isso significa que para executar os procedimentos que estão contidos nesses

²⁶ Disponível em <http://www.pentaho.com/product/data-integration>

arquivos é preciso um interpretador. Tal interpretador já vem com a ferramenta, porém, para executar os procedimentos XML em uma plataforma externa ao *Pentaho* é preciso ter um ambiente previamente configurado.

E por fim, é importante destacar que a *Pentaho* possui uma comunidade bastante ativa, e por ter o código aberto, há um grande número de colaboradores.

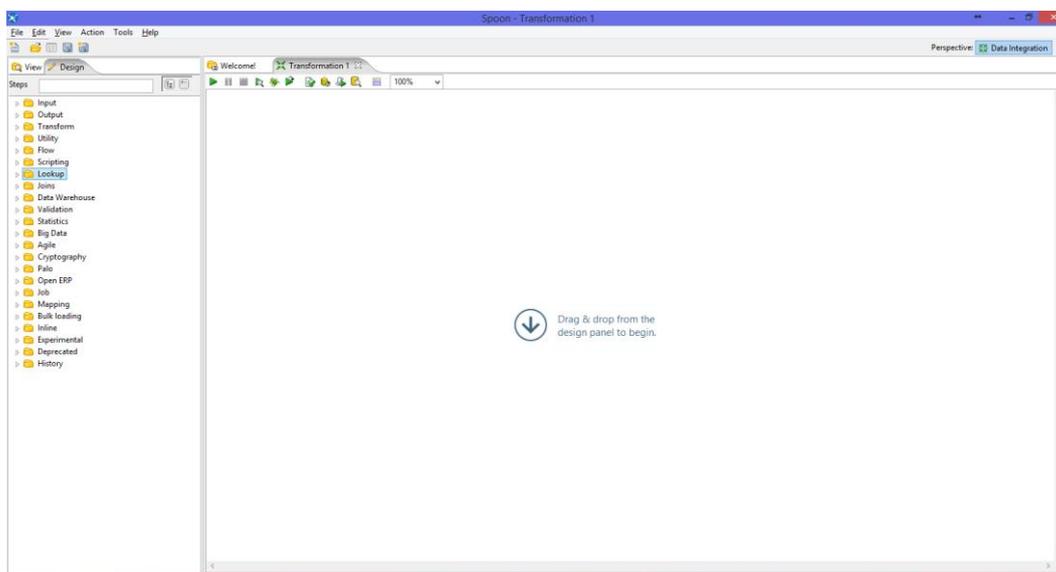


Figura 9 – Tela do *Pentaho Data Integration* (Elaborado pelo autor)

2.5.2 *Talend Open Studio for Data Integration*

Assim como o *Kettle*, o *Talend Open Studio for Data Integration* (TOSDI) é uma ferramenta ETL de código aberto para integração de dados. Criado pela empresa *Talend*²⁷, o TOSDI contém mais de 800 componentes e conectores, que permitem fazer as mais diversas extrações, transformações e cargas, desde acesso à simples base de dados até extração e carregamento em complexos *Web services*.

Um aspecto interessante, é que diferentemente do *Pentaho*, o conceito de *Job* no *Talend* engloba os conceitos de Transformação e *Job* do *Kettle*, ou seja, ele pode conter

²⁷ Disponível em <http://www.talend.com/>

uma sequência de passos para a realização de uma tarefa ou até mesmo executar outros *Jobs*.

Outro ponto interessante, o ambiente gráfico da ferramenta é baseado na interface de desenvolvimento *Eclipse*, mas voltado para a integração de dados. Dessa forma, é possível modelar diagramas de negócios, arrastar e soltar componentes e geração documentação.

Por último, é importante destacar a maneira com que a ferramenta trabalha, ela é uma geradora de código *Java*, e isso traz uma vantagem sobre outras ferramentas, que é a de suas aplicações poderem ser executadas praticamente em qualquer plataforma, já que atualmente *Java* é uma tecnologia bastante difundida.

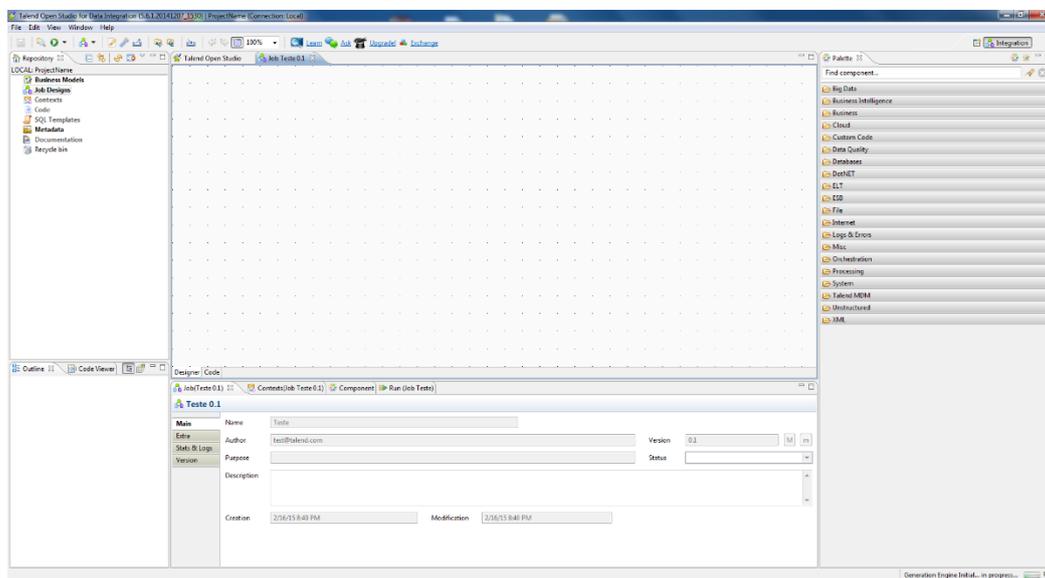


Figura 10 – Tela do *Talend Open Studio for Data Integration* (Elaborado pelo autor)

2.5.3 Comparativo de ferramentas ETL

(NETO, TRAJANO CARLOS MONTASSIER, 2012) em seu trabalho de conclusão de curso cria uma metodologia interessante para comparar ferramentas ETL. Para tal, ele se baseia nas necessidades de uma empresa de pequeno porte. O trabalho é

bem minucioso e nele são avaliadas nove categorias de requisitos. Cada requisito dentro dessas categorias é explicado detalhadamente e recebe uma nota de avaliação e um peso relativo a sua importância.

Neste contexto, o autor faz um levantamento das ferramentas ETL: TOSDI, PDI e *CloverETL*. Ele explica as principais características das ferramentas e as soluções ETL propostas por cada uma. No entanto, no que se refere aos requisitos somente são avaliadas as ferramentas *Talend* e *Pentaho*, pois o autor justifica que ao verificar a documentação da versão de código aberto da ferramenta da *Clover* encontrou deficiências de componente, o que força o usuário a utilizar linguagem de programação ou de *scripts* para suprir necessidades. A Tabela 3 apresenta o quadro resumo com o comparativo.

Critérios	Pentaho	Talend
Arquitetura e Escalabilidade	2	4
Funcionalidades ETL	25	27
Facilidade de uso	21	20
Reutilização	6	8
Depuração	13	12
Mecanismo de processamento	-	-
Conectividade	16	16
Garantia de Qualidade dos Dados	-	-
Características Gerais Ferramenta ETL	12	12
Total	95	99

Tabela 3 - Comparação entre as duas principais ferramentas ETL de código aberto (Adaptado de (NETO, TRAJANO CARLOS MONTASSIER, 2012))

O resultado da avaliação mostra que a ferramenta *Talend*, leva vantagem nos critérios de Arquitetura e Escalabilidade, Funcionalidades e Reutilização. Essa vantagem a favor do TOSDI foi devido aos recursos de versionamento dentro de um projeto *Talend*, componentes para carga de tabelas dimensionais e documentação automática. E leva uma pequena desvantagem nos critérios de Facilidade de uso e Depuração. O autor justifica

que o PDI apresenta uma curva de aprendizado melhor devido a sua facilidade de uso. Para mais detalhes do comparativo das ferramentas, ver (NETO, TRAJANO CARLOS MONTASSIER, 2012).

3. Proposta

Este capítulo apresenta um modelo para disponibilização de dados abertos governamentais. Será apresentado o cenário atual, a motivação para a proposta que gerou esta dissertação e a arquitetura elaborada pelo autor.

3.1 Cenário Atual

Segundo a constituição federal, "todos têm direito a receber dos órgãos públicos informações de seu interesse particular, ou de interesse coletivo". A lei que dispõe sobre os procedimentos para atender esta exigência da constituição e que devem ser observadas pelos órgãos públicos e demais órgãos controlados, de forma direta ou indiretamente pela União, Estados, Distrito Federal e Municípios é a Lei de Acesso a Informação (LAI) nº 12.527.

A LAI prevê a transparência ativa, que consiste na publicação espontânea de dados públicos por parte do governo. Caso isso não atenda ao cidadão, o mesmo poderá fazer uso da transparência passiva, nesse cenário existem duas maneiras de solicitá-la, a primeira é dirigir-se pessoalmente ao órgão no qual está a informação. Os órgãos deverão divulgar o endereço de seus SICs (Serviço de Informação ao Cidadão) nos *sites*; a outra forma é pela *internet*, através do e-SIC (Sistema Eletrônico do Serviço de Informações ao Cidadão) que engloba apenas os órgãos e entidades do poder Executivo Federal e exige que o usuário faça cadastro no sistema caso não o tenha.

O próximo passo depois que o cidadão escolhe a forma que deseja obter a informação é preencher um formulário para solicitar a informação. O prazo de resposta que o órgão tem é de 20 dias prorrogáveis por mais 10 dias mediante justificativa expressa.

A resposta do órgão pode ser um pedido negado, uma informação satisfatória ou uma informação não satisfatória. Nesse caso o cidadão entende que sua solicitação não foi atendida, assim como no caso de pedido negado cabe recurso, sendo assim, o requerente tem 10 dias para entrar com recurso a partir da data de resposta do órgão.

Dessa forma, a Figura 11 ilustra o fluxograma de atividades relacionada a solicitação de abertura de dados por parte do cidadão no e-SIC.

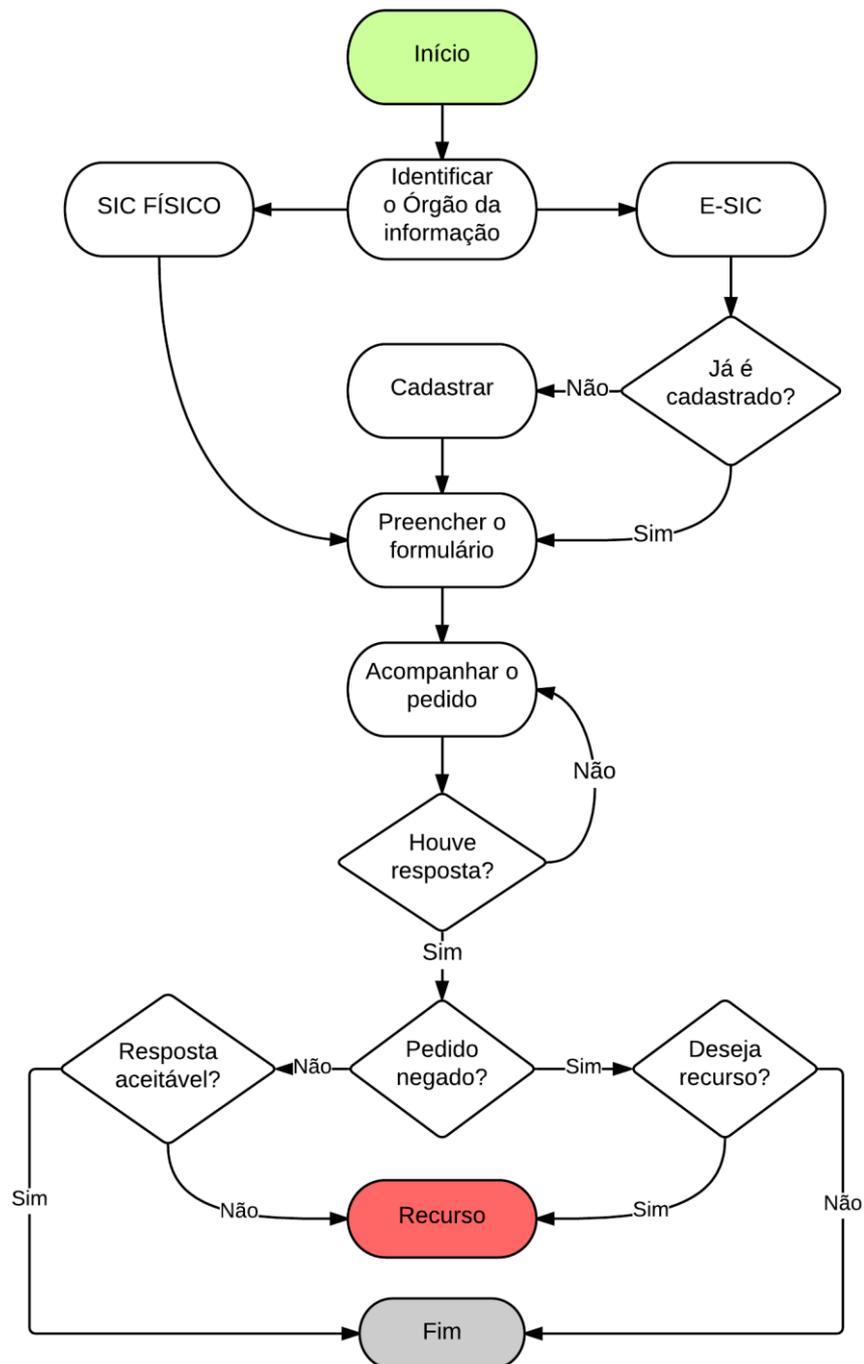


Figura 11 – Processo de solicitação de dado pelo e-SIC (Elaborada pelo autor)

Segundo dados do próprio e-SIC, a quantidade de solicitações de acesso à informação feito por pessoas físicas e jurídicas, desde que a lei entrou em vigor, entre o período de maio de 2012 a até janeiro de 2015 tem crescido a cada ano, como demonstra a Figura 12, significando um maior interesse da população em dados governamentais.

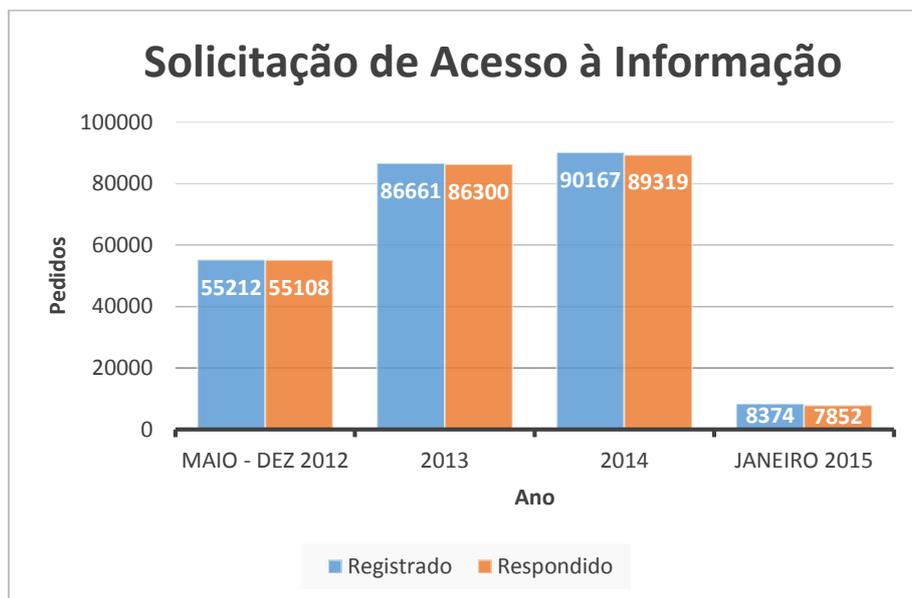


Figura 12 - Número de pedidos de acesso à informação pelo e-SIC (Adaptado de (CGU, 2015b))

3.2 Motivação

Com base no cenário apresentado, para a elaboração da proposta deste trabalho, diversas foram as motivações que fizeram o autor definir o escopo deste projeto, dentre os quais se pode citar.

Em primeiro lugar o aspecto desafiador, entusiasmante e inovador de criar uma solução para disseminar a política de dados abertos governamentais no Brasil. Isso é necessário para que o país avance nesta política.

A criação do e-SIC para registrar as solicitações de dados do cidadão foi um grande passo dado pelo Brasil, que ao mesmo tempo, permite uma maior interação entre governo e a sociedade, e também estimula a sociedade a se interessar mais pelo governo. No entanto, este tipo de solução não agrega muito valor econômico, uma vez que dificulta a reutilização dos dados e além de não ser uma solução muito eficiente em situação de dados que haja grande número de pedido de informação, pois acaba gerando um retrabalho para o governo, causando prejuízo de tempo e dinheiro.

Os compromissos assumidos pelo Brasil frente à OGP, e posteriormente a criação do Plano de Dados Abertos (PDA) em que foi estabelecido como meta a abertura de dados dos diversos sistemas da Administração Pública Federal (SIOP, SICONV, SIORG, SISPAC, SIAPE, SIAPA, SPIUNET, SIGS).

E, por fim, a citação de (MARTANO & CRAVEIRO, 2014) referente aos dados governamentais: “o esforço despendido dentro da máquina pública para disponibilizá-los é ainda uma face pouco pesquisada da mesma questão. [...] Antes dos dados poderem ser disponibilizados, é necessário que haja uma série de processos internos no órgão público para: descobrir quais bases ele possui (exploração), extrair essas bases dos sistemas internos (extração) e, por fim, tratá-las de forma a se tornarem genéricas o bastante para serem usáveis por pessoas externas ao órgão”.

Baseado nestas motivações é esperado que exista uma solução para o processo de disponibilização de dados públicos para as entidades governamentais. Dessa forma, essa é a proposta deste trabalho.

3.3 Arquitetura Proposta

O processo de abertura de dados é bastante complicado, visto que é necessária uma infraestrutura bem definida e o estabelecimento prévio dos papéis dos usuários. Para este fim, este trabalho propõe uma arquitetura de *software* que visa dar suporte ao processo de abertura de dados de sistemas estruturadores, que poderá ser reutilizado por qualquer órgão da esfera federal, estadual ou municipal. Esta proposta foi definida adotando uma arquitetura orientada a serviço, ou seja, disponibilizando os dados por meio de um *Web service* ou API, a justificava disto segue abaixo.

“Há muitas formas de disponibilizar os dados: eles podem ser publicados em páginas da web, podem ser expostos via uma interface de consulta em um *website*, ou podem ser acessados diretamente por sistemas eletrônicos via uma API (interface de programação de aplicativo).” (ALVARENGA, EVERTON ZANELLA *et al.*, 2011).

Neste sentido, o (E-PING, 2014) afirma: "A tecnologia de *Web Services* é recomendada como solução de interoperabilidade da e-PING. De maneira que, independente das tecnologias em que foram implementados, possa-se adotar um padrão de interoperabilidade que garanta escalabilidade, facilidade de uso, além de possibilitar atualização de forma simultânea e em tempo real".

Além das justificativas apresentadas, pode-se citar os diversos benefícios apontados ao publicar os dados em uma API, dentre eles: a economia de tempo e custo dos pedidos de acesso à informação do SIC; a possibilidade de cruzar dados de diferentes órgãos gerando novas agregações; aumento da participação social; estímulo a inovação; economia de tempo e dinheiro; e melhora nos serviços governamentais. É possível citar também o estímulo a economia em diversos setores do governo, devido a geração de valor e a criação de novos empregos proporcionados por esta política.

Dessa forma, Neelie Kroes, a vice-Presidente da comissão europeia e responsável pela Agenda Digital para a Europa, afirma que a política de dados abertos gera uma atividade econômica estimada de cerca de 38 bilhões de euros por ano. (NEELIE KROES, 2011).

Com base nestes motivos, este trabalho optou utilizar uma arquitetura baseada em serviços como interface para a abertura de dados abertos e que ao mesmo tempo proporciona interoperabilidade. A Figura 13 representa a arquitetura proposta.

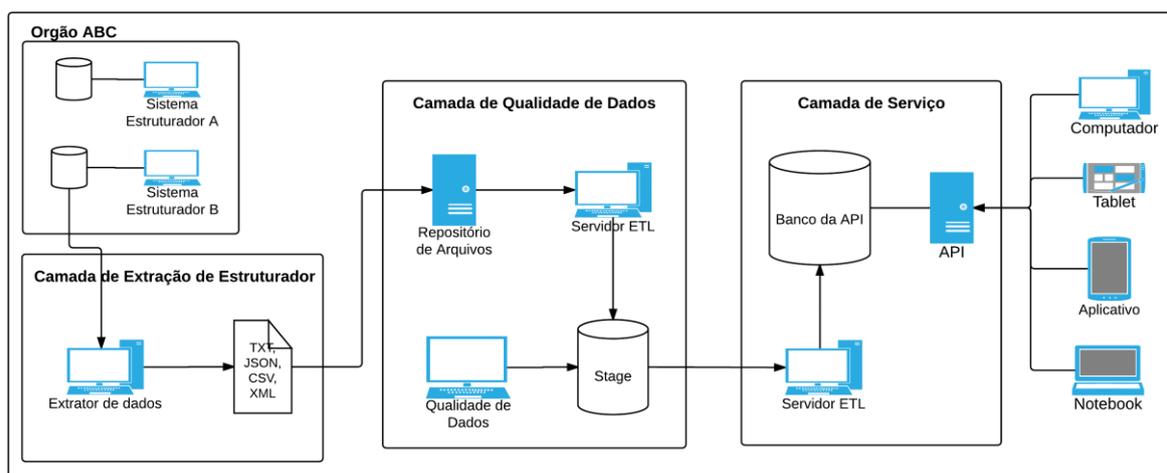


Figura 13 – Arquitetura proposta pelo autor (Elaborada pelo autor)

Na figura acima é possível ver que três camadas foram definidas: camada de extração de estruturador, camada de qualidade de dados e camada de serviço. A seguir serão explicadas cada uma destas camadas:

- Camada de extração de sistemas estruturadores – Esta camada é onde acontecerá a extração do sistema estruturador de onde os dados serão disponibilizados. Nesta camada, um extrator de dados será desenvolvido. O extrator fará a extração dos dados do banco de dados, necessárias para dar início ao processo de abertura, podendo ser o banco de dados inteiro ou apenas um conjunto de dados. Estes dados serão gerados em algum formato não proprietário, como XML, JSON, CSV,

TXT, e disponibilizados em um repositório de arquivos, e caso necessário, junto com arquivos auxiliares, como o arquivo de leiaute dos dados, devido ao formato dos arquivos não estruturados.

- Camada de qualidade de dados – esta camada é onde acontecerá o processo de qualidade de dados. Com os arquivos com os dados disponibilizados no repositório de dados, um processo ETL é desenvolvido para a extração dos dados destes arquivos e carregados em um banco de dados na área de *stage*²⁸. Esses dados devem ser carregados em sua forma bruta, ou seja, sem aplicação de nenhuma transformação. Em seguida, uma análise de qualidade dos dados é feita para garantir a qualidade dos dados disponibilizados. E, por fim, caso necessário, o processo de enriquecimento dos dados será executado.
- Camada de serviços – Esta camada é onde será desenvolvida a API de dados abertos. Esta camada conterá o Servidor ETL, o banco de dados da API e o servidor da API propriamente dito. O processo ETL é desenvolvido e executado no servidor, para extrair os dados da *stage*, aplicar as transformações necessárias e gerar o banco com os dados necessários a API. Este será usado como o banco de dados da API. Em seguida, dá-se início o desenvolvimento da API e ao final, esta API construída será usada como interface de acesso a dados para a sociedade que poderá desenvolver aplicativos consumindo-a.

Para implementar esta arquitetura é necessário que os seguintes atores estejam previamente definidos. São eles: Provedor de Dados, Catalogador de Dados,

²⁸ *Stage* é um conceito bastante usado na área de *Data Warehouse*, que representa uma área de trabalho temporária onde os dados de um sistema são copiados, para, em seguida, aplicar técnicas de qualidade, higienização, normalização de dados, etc.

Desenvolvedor de ETL, Analista de Qualidade de Dados, Analista de Negócio, Desenvolvedor de API e Cidadão.

O Provedor de dados é o ator encarregado em disponibilizar os dados. Ele é o responsável por prover uma forma de extrair os dados do sistema estruturador e disponibilizá-la no servidor de dados. Devido à dificuldade de formalizar um padrão para exportação dos dados e que este seja largamente utilizado, permitiu-se a livre escolha do formato a ser usado, podendo ser: XML, CSV, JSON, TXT. Cabe ressaltar que normalmente este ator pertence ao Órgão responsável pelo sistema. A Figura 14 ilustra esse ator com os seus casos de usos.

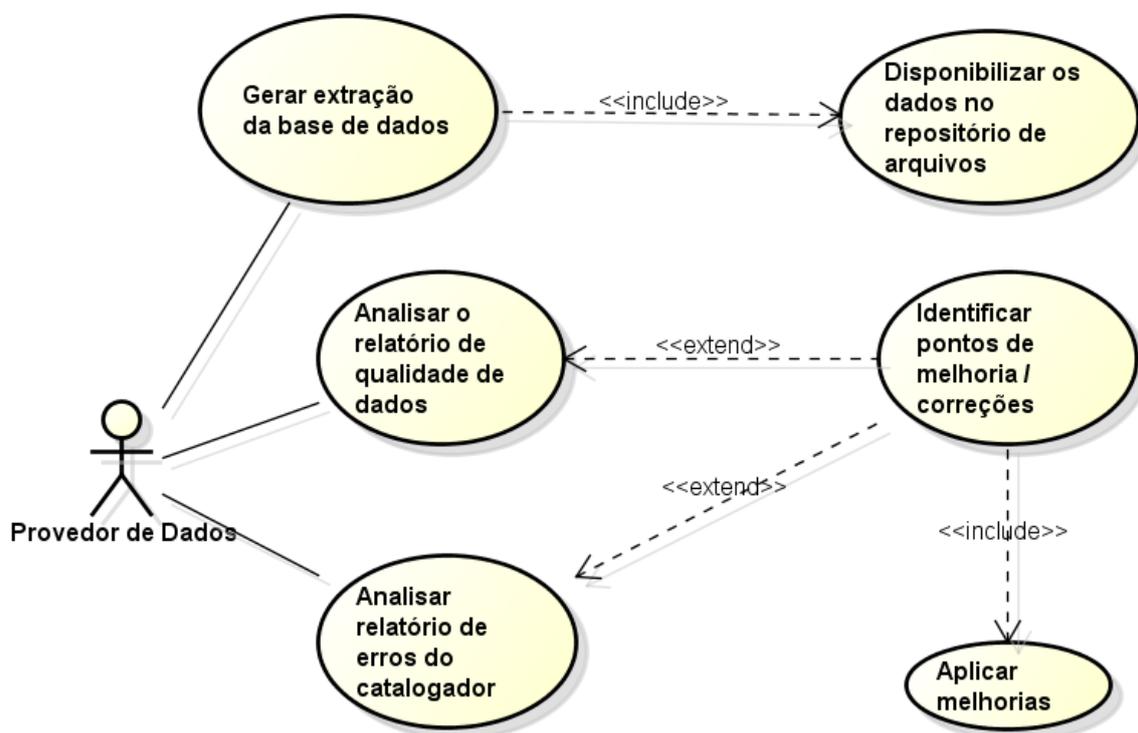


Figura 14 – Diagrama de Casos de usos do Provedor de Dados (Elaborada pelo autor)

O Catalogador de Dados é o ator responsável pela infraestrutura do repositório de dados. Este ator gerencia e define a estrutura do repositório de dados, onde serão disponibilizados os arquivos de extração com dados dos sistemas estruturadores. Dentre

suas ações estão a análise dos arquivos de extrações e elaboração do relatório de erros, caso necessário. A Figura 15 representa os casos de uso do Catalogador de Dados.

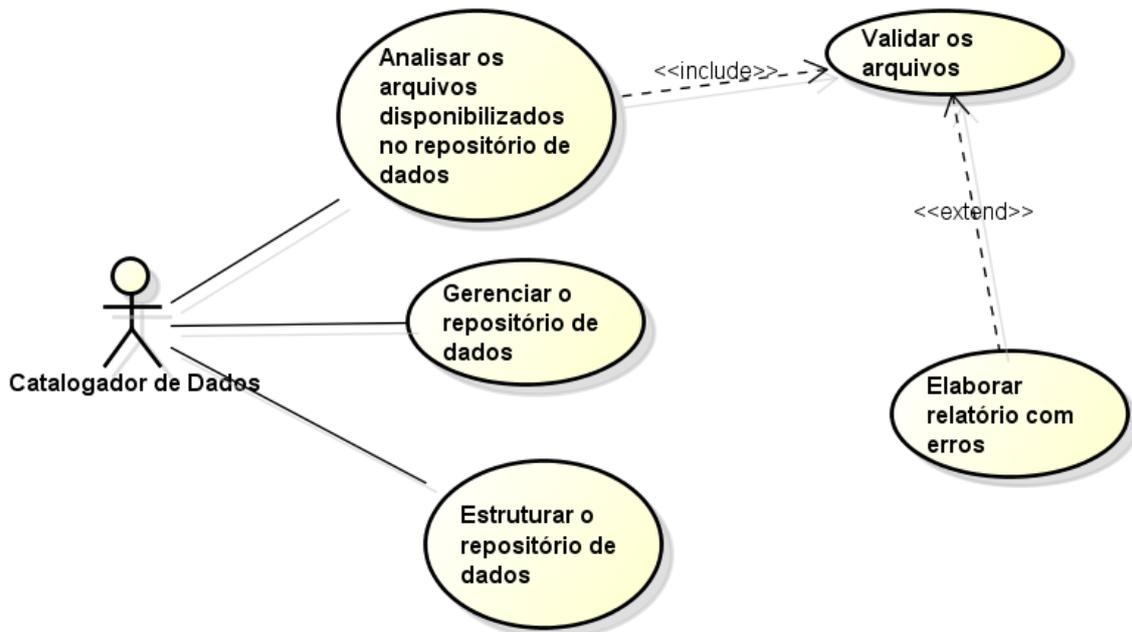


Figura 15 – Diagrama de Casos de uso do Catalogador de Dados

O Analista de Qualidade de Dados é o ator responsável por garantir que os dados publicados estejam em um grau de qualidade aceitável. Este ator utiliza técnicas de qualidade de dados para atingir os seus objetivos. Dentre as suas responsabilidades estão: definição das dimensões e métricas de qualidade, análise dos dados com base nas métricas e identificar pontos de melhorias, caso necessário. Na Figura 16, é mostrado os casos de uso referente a este papel.

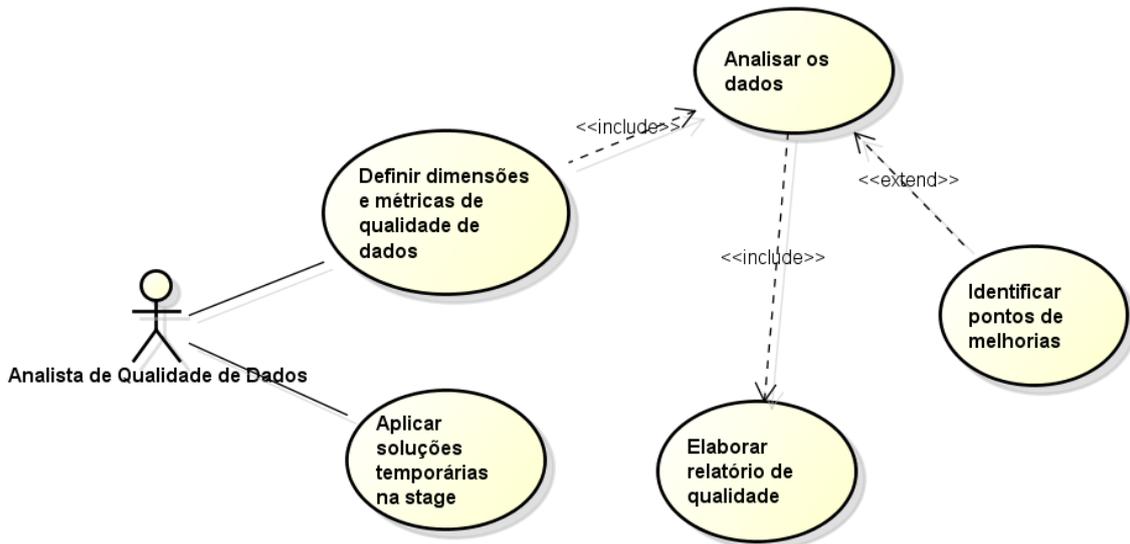


Figura 16 – Diagrama de Casos de uso do Analista de Qualidade de Dados (Elaborada pelo autor)

O Desenvolvedor ETL é o ator responsável por desenvolver todos os processos ETL necessários para a publicação de dados abertos. Este ator será requerido em dois cenários, para extrair os dados dos arquivos do repositório e carregá-los no banco de dados da área de *stage*, e para carregar o banco de dados da API logo após o processo de análise de qualidade de dados. A Figura 17 mostra os casos de uso referente a este tipo de usuário.

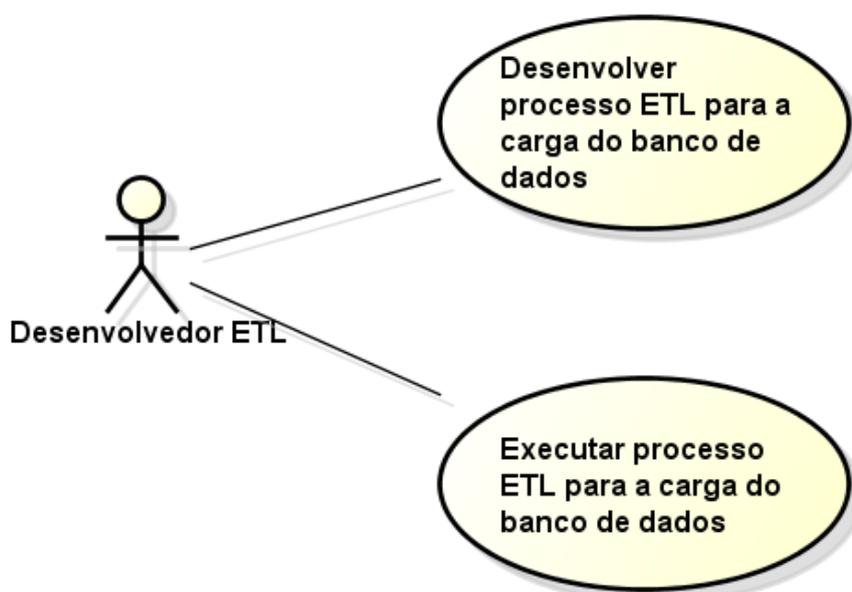


Figura 17 – Diagrama de Casos de uso do Desenvolvedor ETL (Elaborada pelo autor)

O Analista de Negócio é a pessoa que possui conhecimento sobre as regras de negócio do Sistema Estruturador. É de sua competência auxiliar os analistas e desenvolvedores nas dúvidas referentes ao negócio em todas as etapas do desenvolvimento. Além de ser responsável em propor métodos a serem desenvolvidos para a API. A Figura 18 ilustra os casos de uso deste ator.

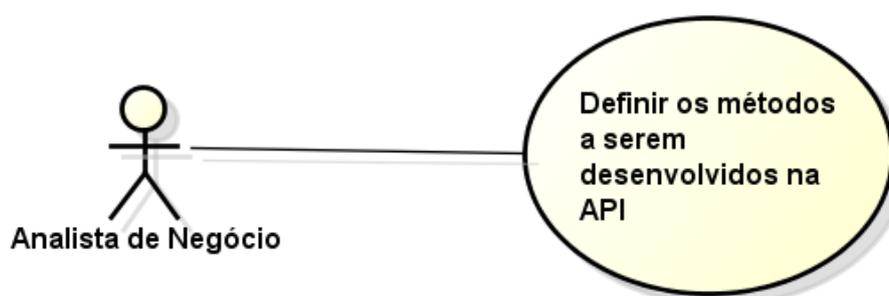


Figura 18 – Diagrama de Casos de uso do Analista de Negócio (Elaborada pelo autor)

O Desenvolvedor da API é o responsável por desenvolver a API para publicação dos dados, a sua função é desenvolver a API para dados abertos seguindo os padrões e tecnologias, implementando os métodos definidos pelo Analista de Negócio. Na Figura 19, é mostrado o diagrama de casos referente a este papel.

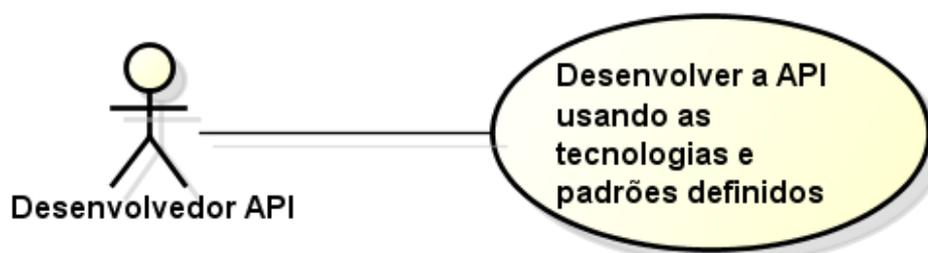


Figura 19 – Diagrama de Casos de Uso do Desenvolvedor de API (Elaborada pelo autor)

E por fim o cidadão, este ator possui papel fundamental para a política de dados abertos, pois ele é o principal interessado em acessar os dados do governo. Dentre as suas

ações estão: solicitar novo conjunto de dados a ser publicado, desenvolver aplicativos usando a API e utilização da API como instrumento de fiscalização, por meio de desenvolvimento de aplicativos. A Figura 20 demonstra as ações deste usuário.



Figura 20 - Diagrama de Casos de Uso do Cidadão (Elaborada pelo autor)

Na Figura 21 consta o diagrama de atividades, com as ações destes usuários e como eles se interagem.

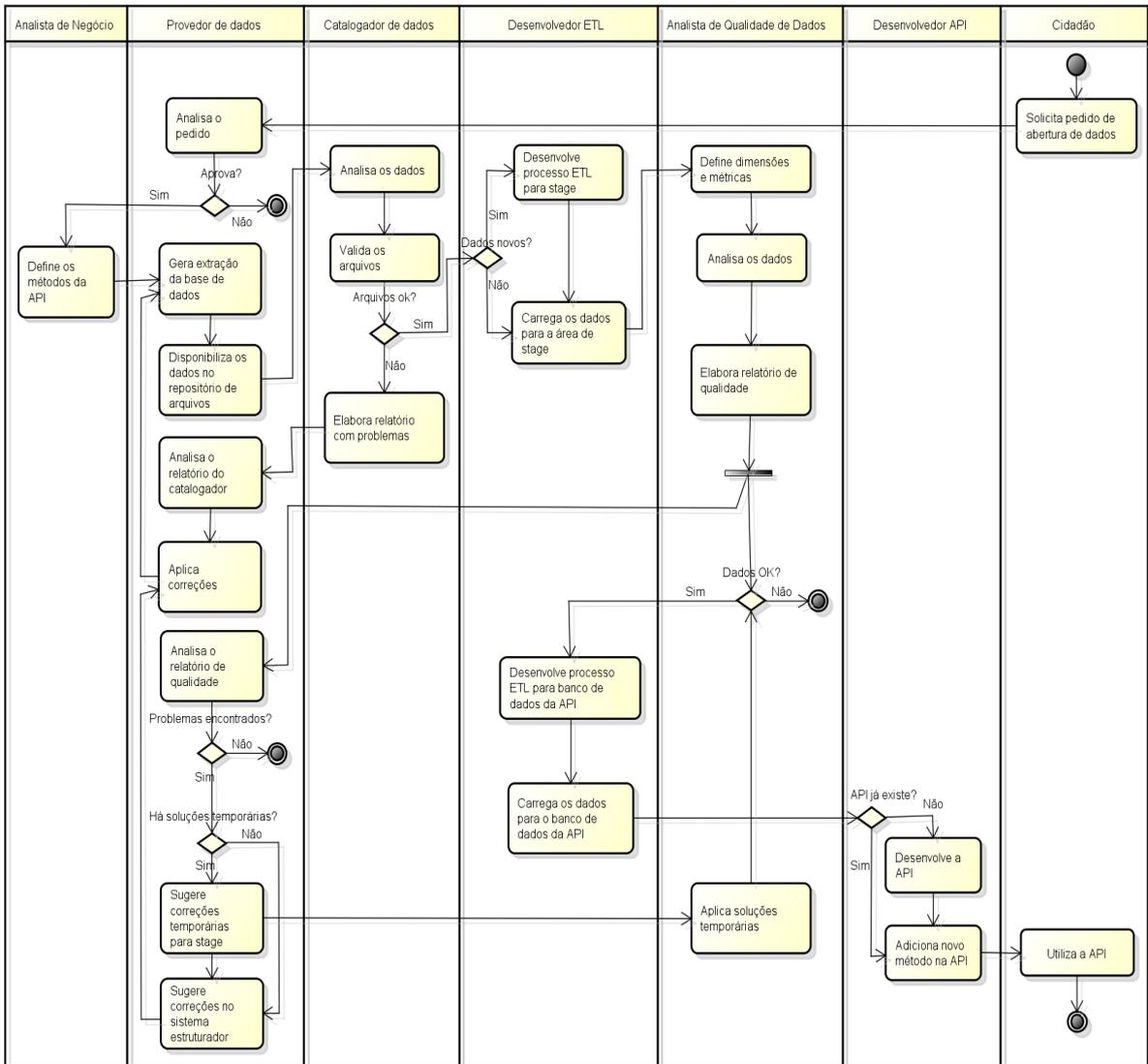


Figura 21 – Diagramas de atividades da arquitetura proposta (Elaborada pelo autor)

4. Estudo de caso

Este capítulo apresentará o estudo de caso com base da arquitetura de *software* apresentado na seção 3.3, que servirá como prova de conceito do modelo em questão. Neste sentido, será dada uma breve introdução a Sistemas Estruturadores, SIASG, Padrões e Tecnologias adotados, Arquitetura da Implementação, e a detalhes da implementação da Camada de Qualidade e da Camada de Serviço.

4.1 Sistema Estruturadores

Sistemas Estruturadores de Governo ou Sistemas de Gestão Administrativa são sistemas de informações que dão suporte aos principais processos administrativos da Administração Pública Federal. Embora alguns desses sistemas estejam sendo reconstruídos, grande parte deles foram desenvolvidos na década de 80 com base em mainframes, em antigas linguagens de programação e em padrões proprietários (FRANZOSI *et al.*, 2009).

Neste sentido, dentro do âmbito federal há diversos Sistemas Estruturadores, dentre eles, vale a pena destacar:

- Sistema de Concessão de Diárias e Passagens (SCDP) - Sistema criado para simplificar e aperfeiçoar o processo de concessão de diárias e passagens, além de melhorar o controle e reduzir gastos. O SCDP faz o cadastramento da viagem com seus respectivos trechos, a reserva das passagens, a autorização da solicitação e a emissão do bilhete. Também faz o controle do orçamento de cada órgão para gastos com diárias e passagens.

- Sistema de Gestão de Convênios (SICONV) - Sistema para registrar a celebração, a liberação de recursos, o acompanhamento da execução e a prestação de contas dos convênios realizados com o Governo Federal.

- Sistema de Informações das Empresas Estatais (SIEST) - Sistema desenvolvido e disponibilizado pelo Departamento de Coordenação e Controle das Empresas Estatais (DEST), que trata da elaboração do Plano de Dispêndios Globais (PDG) das empresas estatais para o exercício financeiro subsequente. Ele acompanha a execução e revisão do PDG para o exercício financeiro vigente e fornece informações para o Balanço Geral da União, no capítulo investimento das empresas. O SIEST cuida ainda da manutenção de informações cadastrais (perfil das estatais), contábeis (endividamento, plano de contas, balanço patrimonial) e econômico-financeiras (política de aplicações) das empresas federais. É um instrumento em permanente atualização, sendo compatível com os níveis de informações de que dispõem as estatais, bem como incorpora métodos de informatização mais avançados, com vistas à racionalização dos trabalhos de elaboração e controle dos orçamentos. É constituído de cinco módulos: Programa de Dispêndios Globais (PDG), Orçamento de Investimento, Cadastro Geral das Empresas Estatais, Balanços Patrimoniais e Endividamento.

- Sistema Integrado de Administração de Recursos Humanos (SIAPE) - Sistema informatizado de Gestão de Recursos Humanos do Poder Executivo Federal, que controla as informações cadastrais e processa os pagamentos dos servidores da Administração Pública Federal.

- Sistema Integrado de Administração Financeira do Governo Federal (SIAFI) - Modalidade de acompanhamento das atividades relacionadas com a administração financeira dos recursos da União, que centraliza ou uniformiza o processamento da execução orçamentária, recorrendo a técnicas de elaboração eletrônica de dados, com o envolvimento das unidades executoras e setoriais, sob a supervisão do Tesouro Nacional e resultando na integração dos procedimentos concernentes, essencialmente, à programação financeira, à contabilidade e à administração orçamentária.
- Sistema Integrado de Dados Orçamentários (SIDOR) - Conjunto de procedimentos, justapostos entre si, com a incumbência de cuidar do processamento de cunho orçamentário, por meio de computação eletrônica, cabendo sua supervisão à Secretaria de Orçamento Federal (SOF). [Este sistema está em processo de desativação e será substituído pelo Sistema Integrado de Planejamento e Orçamento - SIOP].
- Sistema Integrado de Administração de Serviços Gerais (SIASG) - Sistema informatizado de apoio às atividades operacionais, utilizado pelos órgãos e pelas entidades da Administração Federal direta, autárquica e fundacional, que possui três módulos básicos: o catálogo unificado de materiais e serviços, o cadastro unificado de fornecedores e o registro de preços de bens e serviços. (Fonte: Ministério do Planejamento, Orçamento e Gestão) (CGU, 2015a).

Dentre os sistemas citados, para efeito deste trabalho, o sistema escolhido para ser o objeto de estudo de caso é o SIASG. Neste contexto, os dados deste sistema serão publicados em uma API seguindo a arquitetura proposta na seção 3.3.

4.2 SIASG

O Sistema Integrado de Administração de Serviços Gerais (SIASG) “[...] é o sistema informatizado de apoio às atividades operacionais do Sistema de Serviços Gerais [SISG]. A solução tem sido uma ferramenta importante para a modernização da área de serviços gerais na Administração Federal, em especial no cadastramento de fornecedores, catálogos de materiais e serviços e no registro de preços de bens e serviços”. Dessa forma, “[...] o sistema é capaz de agilizar os processos de compra e promover a transparência dos atos do governo ao divulgar informações sobre os processos licitatórios.” (SERPRO, 2015).

O SIASG surgiu a partir do Decreto nº 1.094, de 23 de março de 1994 que regulamenta o SISG e institui o SIASG, como ferramenta informatizada auxiliar do SISG, com “a finalidade de integrar e dotar os órgãos da administração direta, autárquica e fundacional de instrumento de modernização, em todos os níveis.” (BRASIL, 1994).

Segundo a Portaria Normativa N.º 2, de 27 de Outubro de 2000, o SIASG tem como objetivos:

a) Promover a implementação de políticas e diretrizes relativas às atividades de administração de materiais, obras e serviços, transportes, comunicações administrativas, licitações e contratos, adotados na Administração Pública Federal direta e indireta, autárquica e fundacional;

b) Gerenciar e operacionalizar o funcionamento sistemático das atividades do SISG, com a utilização dos módulos Catálogo de Materiais – CATMAT, Catálogo de Serviços – CATSER, Sistema de Divulgação Eletrônica – SIDEC, Sistema de Cadastramento Unificado de Fornecedores – SICAF, Sistema de Registro de Preços – SIREP, Sistema de Contratações – SICON e Minuta de Empenho, visando:

- i. Modernizar as atividades do Estado;
- ii. Dotar as compras do Governo de melhor qualidade e de absoluta transparência em seus procedimentos;
- iii. Integrar outros sistemas da Administração Pública.

c) Garantir a segurança dos registros efetuados em outros sistemas, para os casos de integração sistêmica. (BRASIL, 2000).

O SIASG é composto por diversos módulos:

- Catálogo de Materiais e Serviços (CATMAT / CATSER) - Tendo como base primária a *Federal Supply Classification* [sistema desenvolvido pelo Departamento de Defesa dos Estados Unidos], objetiva a formação de uma linguagem única de materiais e serviços para a Administração Pública, além de propiciar a definição de padrões de qualidade para os materiais e serviços adquiridos pelo Governo.

- Portal de Compras do Governo (Comprasnet) - Este Sistema faz parte do esforço permanente do Estado para a modernização de suas atividades, disponibilizando instrumentos eficientes que ofereçam à sociedade condições adequadas para acesso a informações relativas às compras e contratações da Administração Pública Federal.

- Sistema de Cadastramento Unificado de Fornecedores (SICAF) - O SICAF tem por finalidade cadastrar e habilitar parcialmente os interessados, pessoas físicas

ou jurídicas, em participar de licitações realizadas por órgãos/entidades da Administração Pública integrantes do SISG, bem como acompanhar o desempenho dos fornecedores cadastrados e ampliar as opções de compra do Governo.

- Sistema de Gestão de Contratos (SICON) - Possibilita o gerenciamento dos contratos firmados pela Administração, disponibilizando informações gerenciais das contratações, viabilizando redução de custos operacionais.

- Sistema de Divulgação Eletrônica de Compras e Contratações (SIDECE) - Tem por objetivo tornar as compras do Governo transparentes para toda a Sociedade, aumentando as oportunidades para os Fornecedores, através da divulgação eletrônica das informações relativas às licitações, utilizando, também, a *Internet* no endereço www.comprasnet.gov.br.

- Minuta de Empenho (SISME) - Possibilita a geração automática de minuta de Nota de Empenho e o seu envio para o SIAFI, racionalizando os procedimentos e evitando possíveis divergências nas operações que passam a ser integradas.

- Sistema de Registro de Preço (SISRP) - Possibilita o registro dos preços praticados para a Administração Pública, auxiliando os gestores nas compras, identificando os preços contratados.

- Sistema de Preços Praticados (SISPP) - Possibilita o registro dos preços praticados para a Administração Pública, auxiliando os gestores nas compras, identificando os preços contratados. Auxilia, ainda, na obtenção de PREÇOS DE REFERÊNCIA. (MPOG & SLTI, 2001, NETO, DIÓGENES LIMA, 2010).

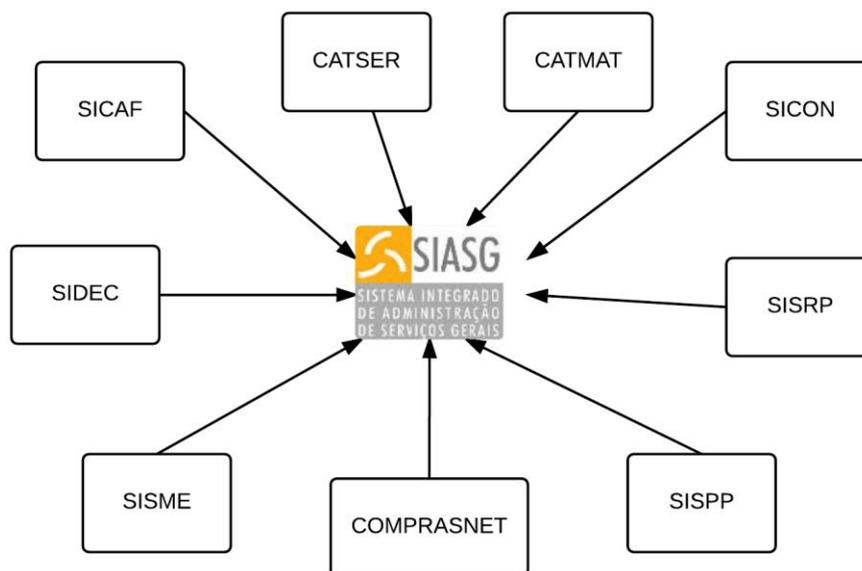


Figura 22 - Estrutura do SIASG (Elaborada pelo autor)

A seguir, serão descritos os padrões e tecnologias que foram adotados na implementação do estudo de caso.

4.3 Padrões e Tecnologias adotados

Neste capítulo, será apresentado como foi implementada a arquitetura de *software* proposta para a abertura de dados, explicitando as principais tecnologias usadas em cada etapa do processo, tal como as respectivas justificativas e opções inerentes ao desenvolvimento tecnológico.

4.3.1 Considerações

A partir do Decreto de 29 de Outubro de 2003, os Comitês Técnicos foram instituídos, no âmbito do Comitê Executivo do Governo Eletrônico (CEGE). Estes Comitês Técnicos têm como uma de suas denominações a implementação do *software* livre pelo Governo Federal. Dessa forma, o governo sempre que possível deve adotar padrões abertos no desenvolvimento de serviços e aplicativos, além de priorizar soluções, programas e serviços baseados em *software* livre.

Além disso, a partir do CEGE também foi estabelecido o e-PING, que tem entre suas políticas gerais:

- a) Adoção de padrões abertos nas especificações técnicas, sempre que possível.
- b) Priorização de *software* livre na implementação dos padrões de interoperabilidade
- c) Utilização de soluções amplamente utilizadas pelo mercado. Devido a uma maior comunidade seria possível reduzir os custos e dos riscos no desenvolvimento dos sistemas governamentais.

Além disso, a adoção de *software* livre tem sido uma opção estratégica do Governo Federal para reduzir custos, ampliar a concorrência, gerar empregos e desenvolver o conhecimento e a inteligência do país na área.

Pelos motivos citados acima, este trabalho procura dar preferência as tecnologias de *software* livre para implementar a arquitetura proposta. Vale a pena ressaltar que este trabalho é baseado fortemente na Arquitetura Orientada a Serviço utilizando uma API REST como interface de comunicação.

4.3.2 Camada de Qualidade

Certamente não existe uma ferramenta que seja a melhor para todos os cenários. Contudo, pelo comparativo apresentado na seção 2.5.3, a ferramenta *Talend* foi a escolhida, pois mostrou-se a mais adequada para os objetivos deste trabalho.

- *Talend Open Studio for Data Integration*²⁹ (TOSDI) – conforme apresentado na seção 2.5.2, TOSDI é uma versão de código aberto da plataforma *Talend*, desenvolvida para integração de base de dados. Contém mais de 800 componentes e conectores para diversos propósitos e bancos de dados.
- *Talend Open Studio for Data Quality*³⁰ (TOSDQ) – é uma versão de código aberto da plataforma *Talend*, com foco na análise da qualidade dos dados que permite análises personalizadas sobre os dados. A ferramenta também traz uma série de componentes que permitem a análise de métricas já predefinidas, tais como acurácia, completude e integridade dos dados, que oferecem ao usuário uma visão diferenciada sobre os dados que ele deseja analisar. Os resultados são apresentados em forma de gráficos e tabelas com dados detalhados, podendo ser exportados para outros formatos.

²⁹ Disponível em: <http://www.talend.com/products/talend-open-studio>

³⁰ Disponível em: <http://www.talend.com/products/talend-open-studio>

- PostgreSQL³¹ - É um sistema gerenciador de banco de dados objeto relacional de código aberto bastante utilizado. Como um servidor de banco de dados, sua função primária é armazenar dados com segurança e permitir a extração desses dados por aplicações de *softwares*.

4.3.3 Camada de Serviço

Nesta seção serão descritas todas as tecnologias utilizadas para o desenvolvimento da camada de serviço.

- Java - É uma linguagem de programação multiplataforma desenvolvida no início da década de 1990 pela empresa *Sun Microsystems*. A linguagem *Java* é orientada a objetos e compilada em *bytecode*. Programas escritos em *Java* podem ser executados em qualquer sistema operacional, desde que o interpretador *Java* esteja instalado.
- REST – Conforme apresentado na seção 2.4.2, *Representational State Transfer* (REST) é um estilo arquitetural que consiste de um conjunto de princípios e restrições que definem como HTTP e URIs devem ser utilizados.
- EJB – *Enterprise JavaBeans* (EJB) é um componente do ecossistema *Java EE*. Sua função é encapsular a lógica de negócio de uma aplicação. O container EJB é responsável por serviços como gestão de transações e autorizações de segurança. A grande vantagem deste componente é que é gerenciado pelo container que está sendo executado.
- Hibernate³² – é uma ferramenta para mapeamento objeto-relacional de código aberto. O objetivo desta biblioteca é facilitar o mapeamento dos atributos de um

³¹ Disponível em: <http://www.postgresql.org/>

³² Disponível em: <http://hibernate.org/>

banco de dados tradicional em uma classe *Java*. Esta biblioteca implementa a especificação JPA do *Java EE*.

- Resteasy³³ – é uma biblioteca desenvolvida pela JBOSS que tem o propósito de ajudar a construção de API REST. É uma implementação da especificação do JAX-RS do *Java EE*.
- Swagger³⁴ - É uma especificação e implementação de um *framework* completo para descrever, produzir, consumir e visualizar *Web services* do tipo *RESTful*. O objetivo do *Swagger* é permitir que sistemas clientes e de documentação sejam atualizados ao mesmo tempo que o servidor. A documentação dos métodos, parâmetros e modelos é integrada ao código do servidor, permitindo que APIs sempre estejam sincronizadas.
- HATEOAS - É uma abreviação de *Hypermedia as the Engine of Application State* e é uma derivação da arquitetura de aplicação REST. Seu princípio é que um cliente interage com uma aplicação de rede inteiramente através de hipermídia provida dinamicamente pelos servidores de aplicação. Clientes REST não precisam saber como interagir com qualquer aplicação em particular, somente é necessário um entendimento genérico de hipermídia.
- HAL³⁵ - É uma abreviação para *Hypertext Application Language*. HAL é um formato simples que permite de forma fácil e consistente fazer o *hyperlink* entre recursos de uma API. Além disso, HAL permite a criação de bibliotecas de uso geral que podem ser facilmente incorporadas à APIs que utilizem HAL.

³³ Disponível em: <http://resteasy.jboss.org>

³⁴ Disponível em: <http://swagger.io/>

³⁵ Disponível em: http://stateless.co/hal_specification.html

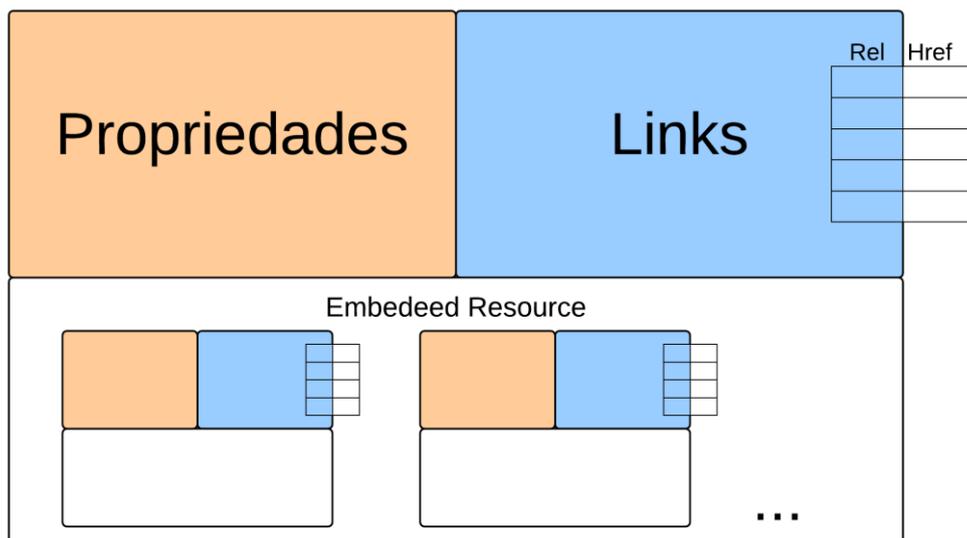


Figura 23 – Estrutura de um Recurso HAL (Adaptado de (KELLY, MIKE, 2013))

- JSON - *JavaScript Object Notation* (JSON) é uma formatação leve de troca de dados. Possui a característica de ser fácil de ler e escrever para humanos além de ser de fácil interpretação e geração por máquinas. JSON é baseado em um subconjunto da linguagem de programação *Javascript*. JSON é completamente independente da linguagem, o que o faz ser um formato ideal para troca de dados.
- XML - *eXtensible Markup Language* (XML) é uma linguagem de marcação. Ela é recomendada pela W3C para a criação de documentos que possuem dados. É uma linguagem extensível, pois permite definir os elementos de marcação. Seu maior propósito é compartilhar informações através da internet.
- CSV - *Comma-Separated Values* (CSV) é um formato de arquivo que armazena dados do tipo tabular (números e textos) em texto puro. Um arquivo CSV consiste de linhas de dados separados por quebra de linha. Cada coluna é separada por um delimitador, que pode ser uma vírgula ou tabulação.
- HTML - *HyperText Markup Language* (HTML) é uma linguagem de marcação utilizada para produzir páginas na web. Documentos HTML são interpretados por

navegadores para gerarem páginas. É a linguagem base da *internet* e foi criada para ser de fácil entendimento para pessoas e para o computador.

- *JBoss Application Server*³⁶ - É um servidor de aplicação de código aberto baseado na plataforma J2EE e desenvolvido em *Java*. O JBoss provê um ambiente completo para que outras aplicações sejam executadas a partir dele, através de serviços providos pelo servidor de aplicação.
- *Apache HTTP Server*³⁷ - Servidor *Apache* é uma ferramenta de código aberto bastante conhecida e amplamente usada devido a sua segurança, performance e compatibilidade.
- *Varnish Cache*³⁸ - *Varnish Cache* é uma ferramenta de código aberto que é usada como um acelerador de requisições HTTP. Esta ferramenta atua como um proxy HTTP reverso, que filtra as requisições que são acessadas frequentemente, guardando uma cópia em cache do conteúdo solicitado. Dessa forma, ao ser solicitado apenas retorna a página guardada em memória diminuindo a quantidade de acessos ao servidor remoto.

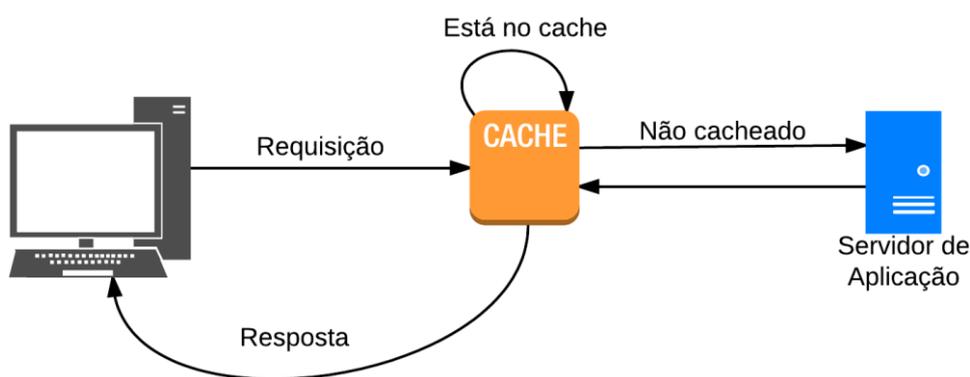


Figura 24 – Funcionamento do *Varnish Cache* (Elaborado pelo autor)

³⁶ Disponível em: <http://jbossas.jboss.org/>

³⁷ Disponível em: <http://httpd.apache.org/>

³⁸ Disponível em: <https://www.varnish-cache.org/>

A seguir, serão mostrados os padrões e tecnologias que serão utilizados na implementação, como ficou definido na Arquitetura da Implementação da proposta, fazendo um estudo de caso com o SIASG.

4.4 Arquitetura da Implementação

Com base nos padrões e tecnologias discutidas na seção anterior, a arquitetura da implementação está ilustrada conforme a Figura 25.

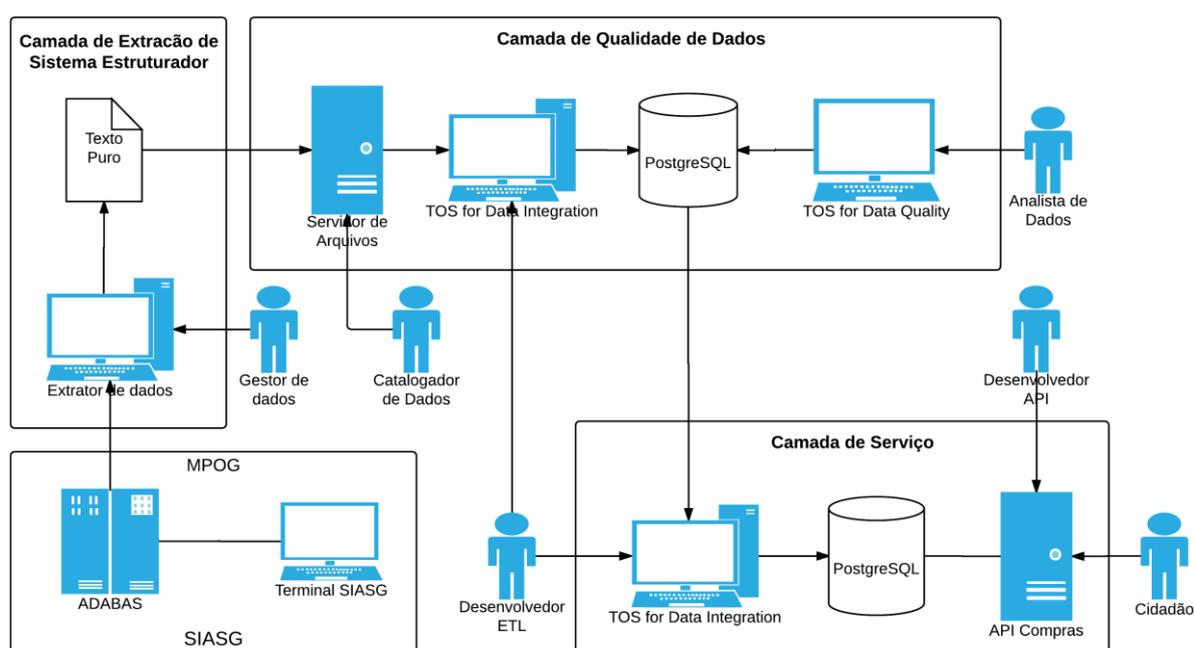


Figura 25 – Implementação da Arquitetura (Elaborada pelo autor)

Como foi detalhado na seção 3.3 a arquitetura definida contém três camadas:

- 1) Camada de Extração de Sistema Estruturador – envolve os atores e componentes necessários para a camada de extração de dados. Devido à criticidade e segurança dos acessos a esses dados, esta camada foi desenvolvida pela equipe do Serviço Federal de Processamento de Dados (SERPRO). Ele é responsável pela infraestrutura do SIASG.

- 2) Camada de Qualidade de dados – envolve os atores e componentes necessários para a implementação da camada de qualidade de dados. É necessário o envolvimento dos seguintes atores: Analista de Negócio, Catalogador de Dados, Desenvolvedor de ETL, Analista de Qualidade de Dados.
- 3) Camada de Serviço – envolve os atores e componentes necessários para a implementação da camada de serviço. São necessários os seguintes atores: Analista de Negócio, Desenvolvedor ETL, Desenvolvedor da API e o Cidadão.

Adicionalmente à implementação da arquitetura, foi decidido a utilização do *Apache Server* e do *Varnish Cache*. A adoção dessas duas tecnologias extras foi com o intuito de garantir melhor performance e segurança para o servidor JBoss e consequentemente para o API. Vale ressaltar que a utilização do *Apache Server* e *Varnish Cache* não é obrigatória. A Figura 26 descreve essa estrutura.

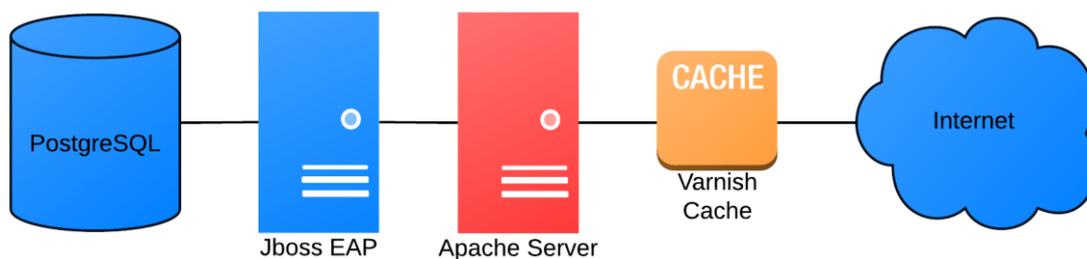


Figura 26 – Infraestrutura da API (Elaborada pelo autor)

Na próxima seção, será detalhado como foram implementadas as camadas de Qualidade de dados e de Serviço.

4.5 Implementação

4.5.1 Camada de Qualidade de Dados

Nesta seção será detalhada como foi implementada a Camada de Qualidade de Dados. Esta camada é dividida em duas etapas: a carga dos dados em *stage* e a análise de qualidade de dados.

Na primeira etapa, será feita a extração dos dados dos arquivos disponibilizados no repositório de dados e em seguida carregá-los em uma área de *stage*. Na próxima etapa será feita uma análise de qualidade dos dados usando a ferramenta TOSDQ Qualidade de Dados. Os atores envolvidos são: Analista de Dados, Analista de Negócio, Catalogador de Dados e Desenvolvedor ETL.

4.5.1.1 Carga de Dados em *stage*

Este banco será usado como uma área de *stage* para o processo de qualidade de dados. Para conseguir o acesso aos dados do SIASG, uma demanda foi solicitada junto ao SERPRO, responsável pela infraestrutura e por manter o sistema, para disponibilizar uma extração da base de dados do SIASG. O provedor de dados disponibiliza os dados juntamente com arquivos auxiliares, contendo a descrição do leiaute da estrutura dos dados no repositório de dados. Dessa forma o Catalogador de Dados pode iniciar o processo de validação da demanda.

A Tabela 4 mostra a quantidade de arquivos obtidos a partir do SIASG. É possível perceber que nos primeiros anos desta extração a quantidade de arquivos ainda é pequena. A explicação para isto é que essa época representa o início da implantação do SIASG, onde vários dos módulos do mesmo ainda estavam em desenvolvimento. Dessa forma,

foram disponibilizados 6115 arquivos contendo dados do período de março de 1997 a março de 2014, distribuídos em 83 tabelas.

Ano	Arquivos
1997	41
1998	72
1999	90
2000	96
2001	103
2002	179
2003	307
2004	324
2005	398
2006	530
2007	541
2008	544
2009	557
2010	548
2011	543
2012	537
2013	515
2014	190
Total	6115

Tabela 4 – Quantitativo de arquivos disponibilizados (Elaborado pelo autor)

O catalogador de dados analisa os arquivos recebidos à procura de potenciais problemas: arquivos corrompidos, estrutura dos arquivos fora do padrão estabelecido, ausência de arquivos, e outros. Ao final, o catalogador organizará os arquivos, montará um catálogo dos arquivos a serem carregado e um *e-mail* será enviado para sinalizar a conclusão desta etapa, dando início ao processo de carga dos dados.

Ao receber o *e-mail*, o Desenvolvedor de ETL, iniciará o desenvolvimento de um processo para fazer o ETL dos dados, com a ajuda da ferramenta TOSDI. O *job* a ser desenvolvido fará uma carga bruta dos dados para o banco de dados da *stage*. É importante ressaltar que nesta etapa os dados não serão tratados e nenhuma chave primária será criada. A justificativa disso é o processo de análise de dados na etapa

seguinte. A Figura 27 ilustra o *job* desenvolvido para a carga dos dados do SIASG dos arquivos para o banco de dados da *stage*.

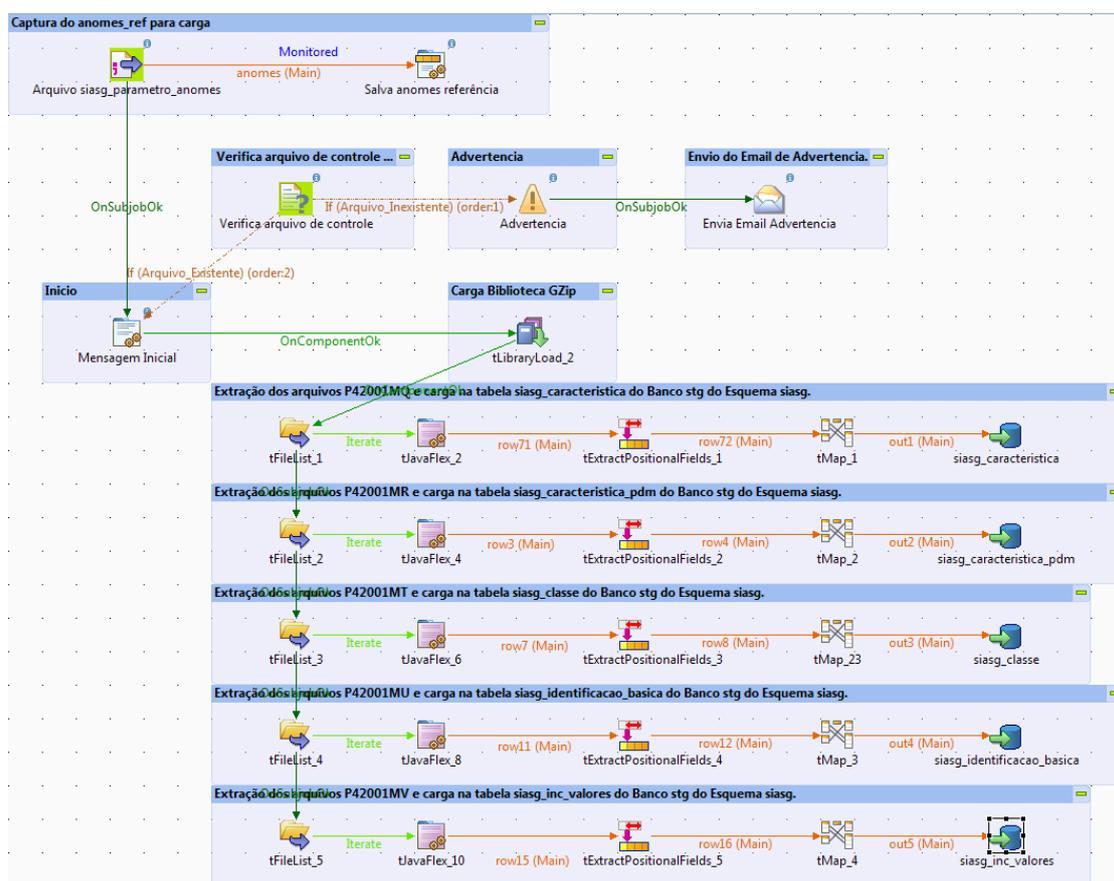


Figura 27 – Processo ETL de carga dos dados para *stage* (Elaborada pelo autor)

Com o banco de dados carregado, o desenvolvedor ETL envia um *e-mail* para o Analista de Qualidade de Dados poder iniciar o processo de análise de Qualidade de Dados.

4.5.1.2 Análise de Qualidade de Dados

Como explicitado na seção 2.3, a análise de qualidade de dados possui um papel fundamental para Dados Abertos. Sendo assim, nesta seção será mostrado como foi realizada a etapa da análise de qualidade de dados. Vale ressaltar que a análise de Qualidade de Dados não é o foco deste trabalho. Portanto, nenhuma das metodologias

consagradas na literatura será aplicado, apenas será feito uma breve avaliação usando os recursos da ferramenta TOSDQ.

Para analisar os dados do SIASG carregados na *stage* serão usadas duas métricas que a ferramenta oferece, a primeira é conhecida como análise da estrutura geral do banco de dados³⁹, e a segunda análise de coluna⁴⁰. E por fim será demonstrado o uso de uma análise personalizada fazendo uso de regras de negócio.

a) Análise da Estrutura Geral do Banco de Dados

Esse tipo de análise oferece uma visão geral do conteúdo do banco de dados. Ela calcula para cada esquema a quantidade de tabelas, registros em cada tabela, índices, chaves primárias, etc. Dessa forma, é possível verificar onde os dados estão mais concentrados, as tabelas que estão vazias, sem chaves primárias e assim por diante.

A Figura 28 mostra que o banco de dados do SIASG contém 84 tabelas e em média cada uma tem mais de 1,5 milhões de linhas. É possível ver que as principais tabelas desses sistemas são: `siasg_empenho_material`, `sidec_resultado_compra`, `siasg_empenho`. Essas três tabelas juntas concentram 70% de todos os dados do banco (em número de registros).

³⁹ Tradução do autor para: database structure overview analysis

⁴⁰ Tradução do autor para: column analysis

▼ Statistical Information									
Catálogo	#rows	#schemata	#rows/schema	#tables	#rows/table	#views	#rows/view	#keys	#indexes
stg	133996581	1	133996581,00	84	1595197,39	0	NaN	0	44

Schema	#rows	#tables	#rows/table	#views	#rows/view	#keys	#indexes
siasg	133996581	84	1595197,39	0	NaN	0	44

Tabela	#rows	#keys	#indexes
siasg_empenho_material	44392764	0	1
sidec_resultado_compra	31200777	0	3
siasg_empenho	19383379	0	1
sidec_registro_preco	6617713	0	3
sidec_compra_sem_licitacao	3912196	0	0
siasg_cronograma_doc_ateste_item	2816959	0	0
siasg_cronograma_parcela	2809176	0	1

Figura 28 - Apresenta detalhes base dados do SIASG (Elaborado pelo autor)

b) Análise de Coluna

É uma análise mais pontual que a Análise da Estrutura Geral do Banco de Dados, permitindo que o usuário trace um perfil dos dados em uma coluna. Isso pode ser feito através de indicadores, sendo que cada um fornece algum tipo específico de informação sobre o campo em questão. Nessa análise serão demonstrados os Indicadores de Estatística Simples⁴¹ e Estatística de Frequência de Padrão⁴².

b.1) Indicador de Estatística Simples

As dimensões trabalhadas por esses indicadores são as que agem sobre os dados de forma objetiva. Eles fornecem estatísticas simples sobre o número de registros, dando uma visão geral do conteúdo da coluna, dados nulos, dados distintos, duplicados, entre outras coisas que se resumem a contagem.

⁴¹ Tradução do autor para: Simple Statistics

⁴² Tradução do autor para: Pattern Frequency Statistics

A Figura 29 demonstra a aplicação desses indicadores. Foi escolhida para esta análise a tabela `siasg_identificacao_basica` do módulo CATMAT do SIASG. Nela nota-se que o campo analisado, `it_no_basico`, que representa o nome do Padrão Descrito de Material (PDM), não é candidato a ser uma chave primária, pois há valores duplicados. Também é possível ver que não existem valores brancos ou nulos para esse campo. Os mesmos resultados podem ser observados na Figura 30 na forma de tabela.

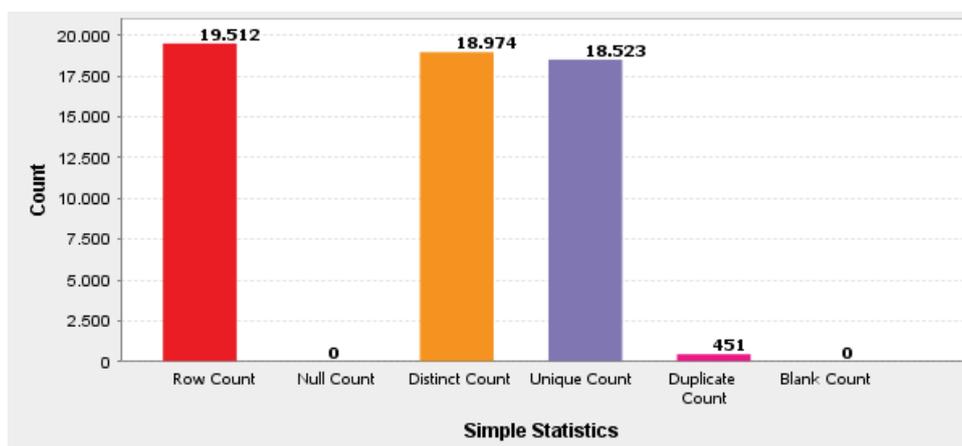


Figura 29 - Gráfico de Estatísticas simples sobre o atributo `it_no_basico` de `siasg_identificacao_basica` (Elaborado pelo autor)

Label	Contar	%
Row Count	19512.00	100.00%
Null Count	0.00	0.00%
Distinct Count	18974.00	97.24%
Unique Count	18523.00	94.93%
Duplicate Count	451.00	2.31%
Blank Count	0.00	0.00%

Figura 30 - Tabela de Estatísticas simples sobre o atributo `it_no_basico` de `siasg_identificacao_basica` (Elaborado pelo autor)

Enquanto nas Figura 31 e Figura 32, foi analisado o código do PDM. As imagens mostram que os valores desse campo são todos distintos. Dessa forma, pode-se concluir

que esse campo, `it_co_iden_basica`, é um bom identificador para as instâncias dessa entidade.

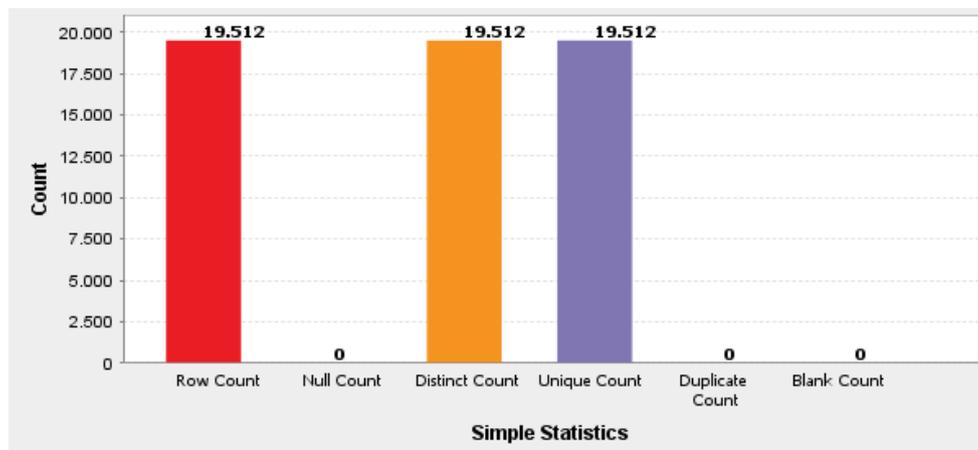


Figura 31 - Gráfico de Estatísticas simples sobre o atributo `it_co_iden_basica` de `siasg_identificacao_basica` (Elaborado pelo autor)

Label	Contar	%
Row Count	19512.00	100.00%
Null Count	0.00	0.00%
Distinct Count	19512.00	100.00%
Unique Count	19512.00	100.00%
Duplicate Count	0.00	0.00%
Blank Count	0.00	0.00%

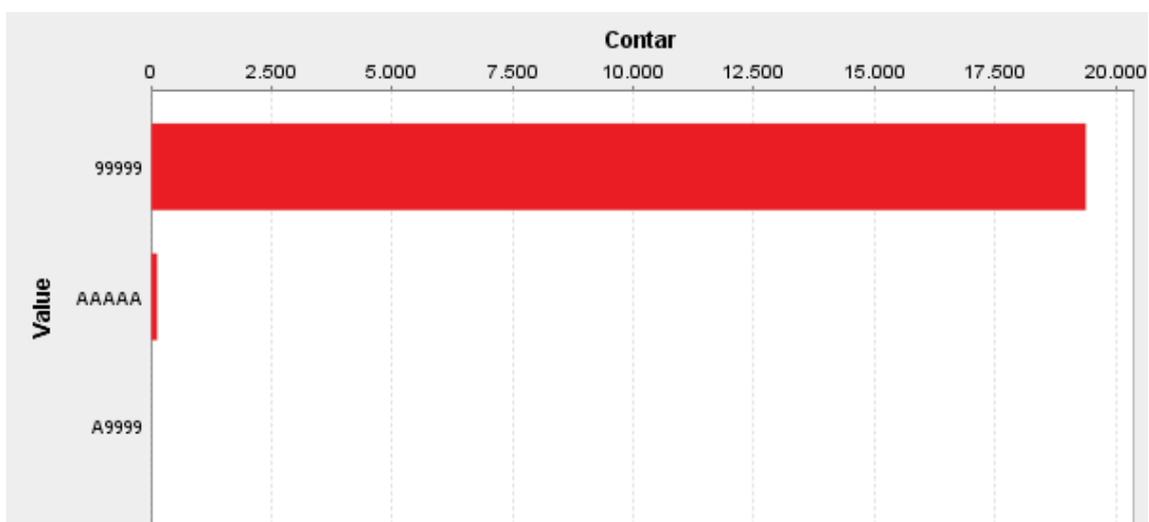
Figura 32 - Gráfico de Estatísticas simples sobre o atributo `it_co_iden_basica` de `siasg_identificacao_basica` (Elaborado pelo autor)

Após a análise desses dois campos uma série de questionamentos podem ser feitos: É possível ter PDMs com nomes idênticos, porém com o código diferente? Ao adicionar um novo item relacionado a este PDM, como será feito o relacionamento com este PDM? O código do PDM será sempre conhecido? O nome do PDM não devia ser único assim como o campo código? É possível que seja um registro duplicado que em algum momento mudou de código?

b.2) Indicador de Estatística de Frequência de Padrão

Esse grupo de indicadores é usado para determinar padrões frequentes. Com eles é possível abordar principalmente o aspecto representacional do dado. Um de seus indicadores, a Estatística de Baixa Frequência de Padrão⁴³, exibem os padrões menos frequentes, enquanto um outro, a Estatística de Frequência de Padrão⁴⁴, exhibe os mais frequentes. Eles analisam a estrutura do dado e substituem todos os caracteres alfabéticos minúsculos por “a”, todos os caracteres alfabéticos maiúsculos por “A” e todos os caracteres numéricos por “9”.

A Figura 33 e Figura 34 ilustram o resultado da aplicação do indicador Estatística de Frequência de Padrão sobre o campo `it_co_iden_basica` da mesma tabela citada anteriormente. Pode ser visto que o campo apresenta apenas três padrões. O padrão mais frequente é totalmente numérico e de tamanho cinco, ele representa 99.33% dos dados. Os outros dois padrões são somente letras e letras com números, de tamanhos cinco, e representam somente uma parcela bem pequena dos dados dessa coluna.



⁴³ Tradução do autor para: Pattern Low Frequency Statistics

⁴⁴ Tradução do autor para: Pattern Frequency Statistics

que pode ser definido no TOSDQ e assim agir sobre os dados. O resultado é apresentado por meio da Figura 35, onde é possível ver que apesar de 99,99% dos casos estarem de acordo com essa regra estabelecida, existe um único caso que não está de acordo na Figura 36.

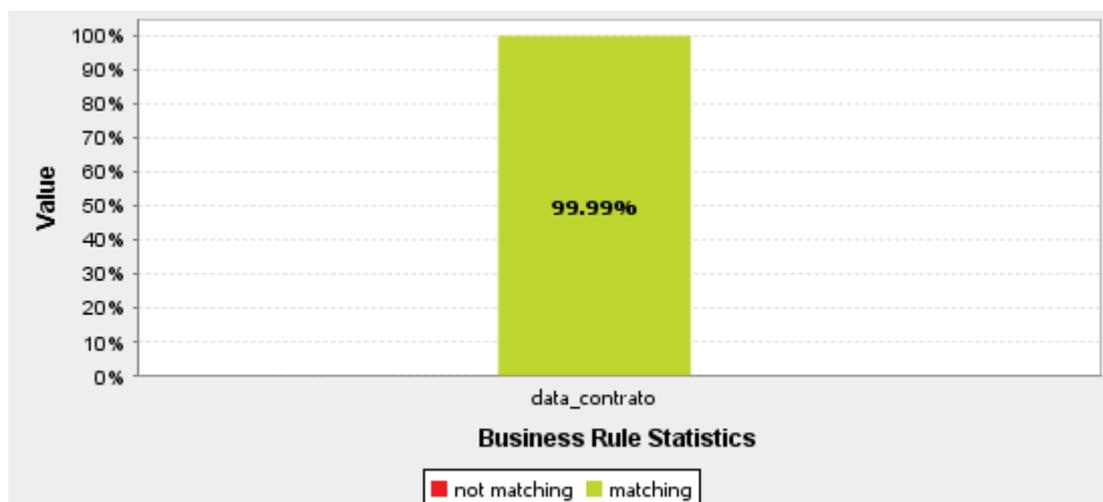


Figura 35 - Resultado da aplicação da regra de negócio (Elaborado pelo autor)

soal	it_va_total	it_va_parcela	it_da_inicio_vigencia	it_da_termino_vigencia	it_in_execuc
	000000000000313640	000000000000313640	20000717	20000131	

Figura 36 - Linha que não está em conformidade com a regra de negócio (Elaborado pelo autor)

Ao final desta etapa de análise, um relatório é elaborado com todas as análises efetuadas, e entregue ao gestor dos dados que se encarregará em avaliar. Ao final, caso nenhum problema seja relatado e nenhuma solução temporária for proposta pelo gestor, será possível avançar para a implementação da camada de serviço.

4.5.2 Camada de Serviço

Nesta seção será detalhado como foi implementada a Camada de Serviço. Esta camada é dividida em duas fases: carga do banco de dados da API e o desenvolvimento da API de compras.

Na primeira etapa será mostrado como foi feito a o ETL dos dados do *stage* para a geração do banco de dados da API. Em seguida, na próxima etapa será comentado sobre o processo de desenvolvimento da API, exibindo a aplicação desenvolvida com as respectivas telas. Os atores envolvidos são: Desenvolvedor API, Desenvolvedor ETL, e o Cidadão. Será abordado todo o procedimento para gerar o banco de dados e a API de dados abertos do SIASG.

4.5.2.1 Carga no Banco de Dados da API de Compras

Esta etapa consiste em carregar o banco de dados de produção. Foi desenvolvido com ajuda da ferramenta TOSDI. Para o desenvolvimento de processos ETL, diversas técnicas e transformações de dados foram aplicadas para a carga das tabelas. Entre as principais regras implementadas, vale destacar:

- a) Remoção de Zeros e espaços em brancos - Como os dados vieram em arquivos de texto puro, posicionais e seguindo um leiaute pré-definido, os campos numéricos são preenchidos com zeros à esquerda até completar o tamanho máximo do campo. Por exemplo: Se o campo tem tamanho de seis dígitos, o valor 1234 virá como 001234. Algo semelhante ocorre com os campos alfanuméricos, porém estes são preenchidos com espaço em branco à direita. Por exemplo: Se o campo tem tamanho de seis caracteres, a palavra 'casa' virará 'casa ').
- b) Conversão de tipo - Todos os campos foram carregados como campo textual (*varchar*), porém para gerar o banco de dados da API estes campos foram convertidos para os seus respectivos tipos: Inteiro, Numérico, Data, etc.
- c) Normalização dos dados: Devido à natureza dos bancos hierárquicos na qual se origina os dados do SIASG, o banco gerado pela extração segue a estrutura lógica

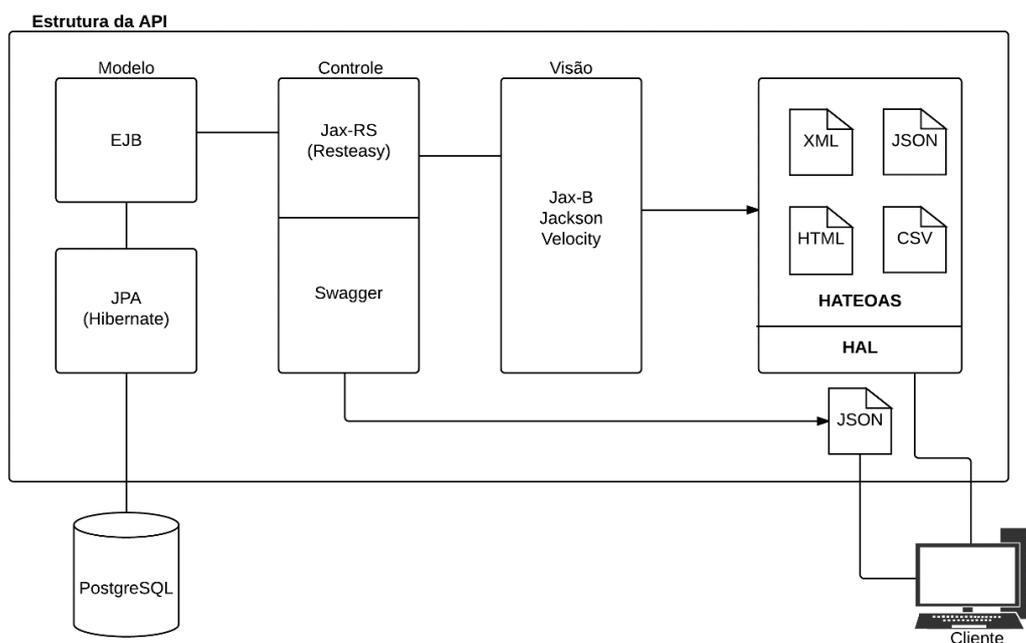
desses arquivos⁴⁵. Dessa forma, foi necessária aplicar o processo de normalização das tabelas.

- d) Mapeamento de valores: alguns tipos de dados normalmente presente na maioria dos bancos relacionais não existe dentro do *Adabas*, como é o caso do tipo “*Booleano*”, para estes casos a representação costuma utilizar campos do tipo texto. Como por exemplo: ‘S’, ‘SIM’, ‘N’, todos esses casos foram convertidos para booleano no banco gerado.

Aplicando-se essas técnicas, foi possível gerar o banco de dados com os dados necessários para a API. Dessa forma, é possível iniciar o Desenvolvimento da API de Compras.

4.5.2.2 Desenvolvimento da API de Compras

Para o desenvolvimento da API, foram utilizadas as tecnologias descritas na seção 4.3.3. Dessa forma, a estrutura da API como ilustrado na Figura 37.



⁴⁵ O conceito de arquivos no *Adabas* é equivalente ao conceito de tabela no banco de dados relacional.

Figura 37 – Estrutura da API (Elaborado pelo autor)

Ao final, a página principal da API desenvolvida na Figura 38. Como pode ser percebida na imagem, a página está alinhada com a identidade visual vigente adotada pelo governo.

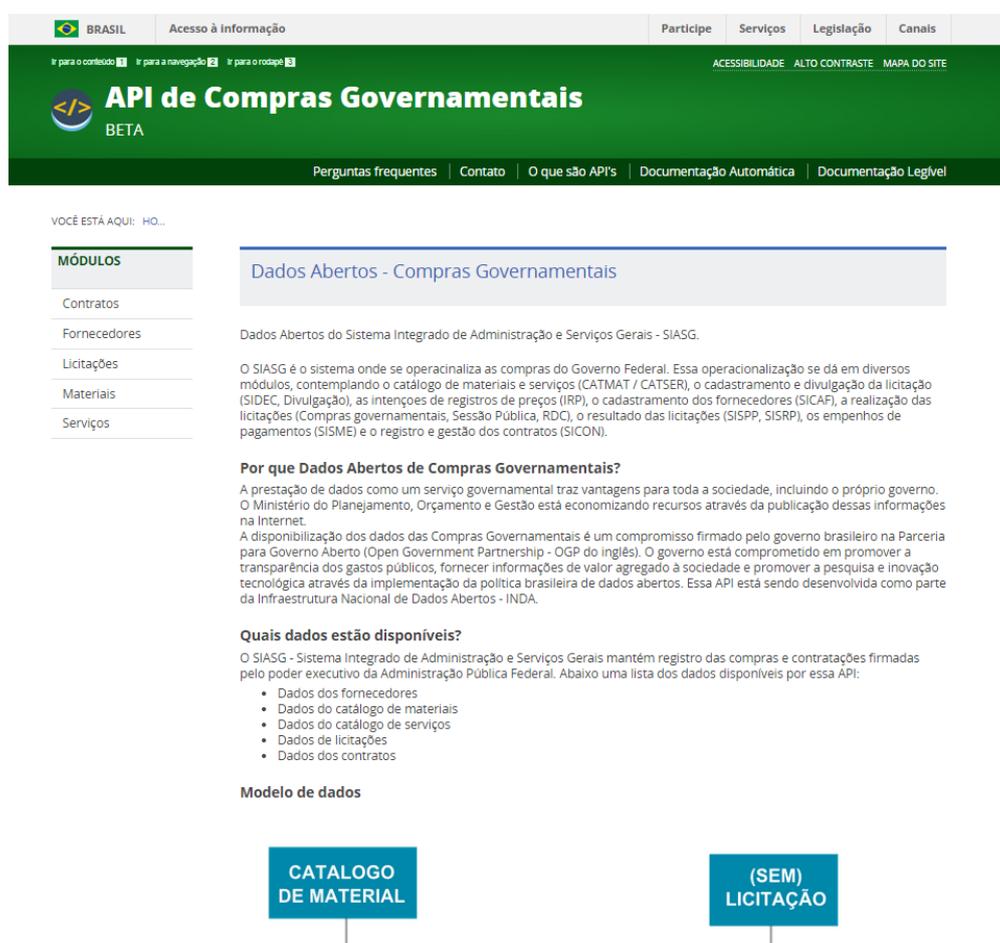


Figura 38 - Página inicial da API de dados abertos (Elaborada pelo autor)

A API de compras contém, ao todo, cinco módulos: Contratos, Fornecedores, Licitações, Materiais e Serviços. Em comparação com os módulos do SIASG, os dados foram extraídos dos módulos CATMAT, CATSER, SICAF, SICON, SIDEDEC, SISPP e SISRP. A Figura 39 apresenta uma visão geral de como esses módulos apresentados na API se relacionam.

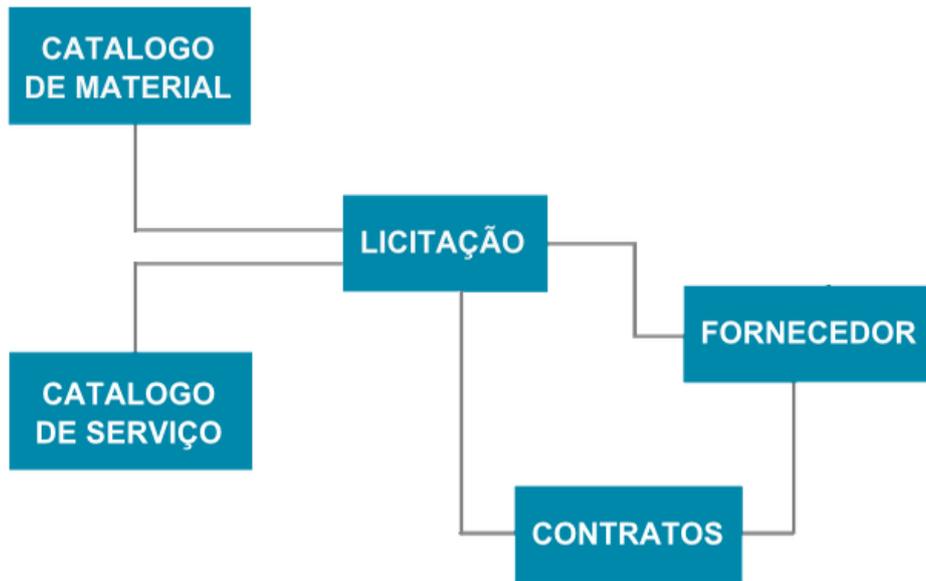


Figura 39 – Modelo de relacionamento entre os módulos (Elaborada pelo autor)

O módulo de fornecedores disponibiliza dados do módulo SICAF do SIASG. Este módulo contém dados acerca de: Fornecedores, Ocorrências, Linha de Fornecedores, Municípios, e outros. A Figura 40 representa este módulo.

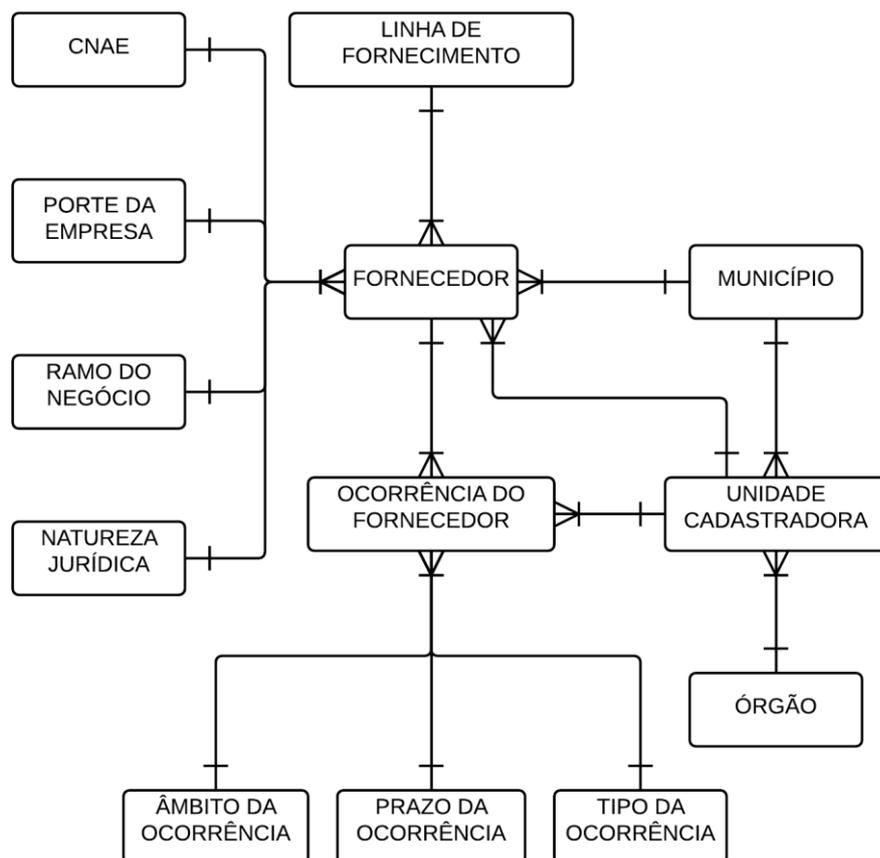


Figura 40 – Modelagem do módulo de fornecedores (Elaborado pelo autor)

O módulo de Materiais disponibiliza dados do módulo CATMAT do SIASG. Como representado na Figura 41, este modelo contém informações dos materiais catalogados pela Administração Pública, e a classificação destes itens, organizado em uma hierarquia: Grupo, Classe e Padrão Descritivo do Material (PDM).

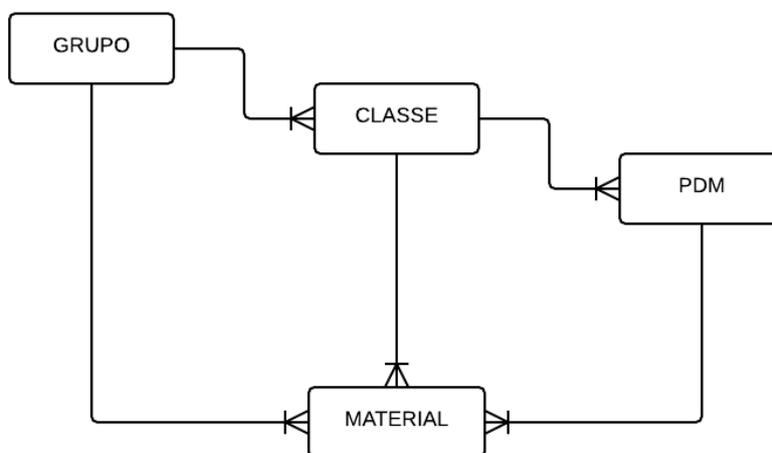


Figura 41 – Modelagem do módulo de materiais (Elaborada pelo autor)

O módulo de Serviço apresenta dados do CATSER do SIASG. O modelo contempla dados serviços catalogados pela Administração Pública e uma classificação hierárquica semelhante ao de Materiais, organizada em: Seção, Divisão, Grupo, Classe e Subclasse. A Figura 42 ilustra este módulo.

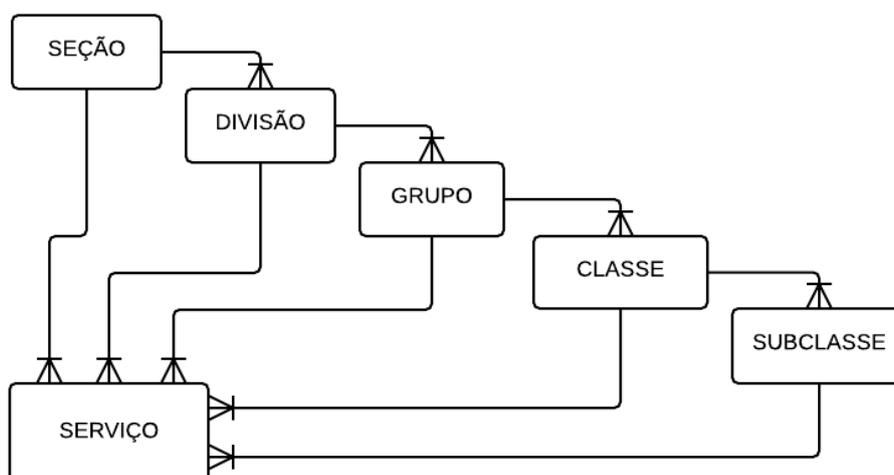


Figura 42 – Modelagem do módulo de serviço (Elaborada pelo autor)

O módulo de licitação envolve dados de três módulos do SIASG: SIDEC, SISPP, SISRP. Neste módulo é possível visualizar informações referentes a Intenção de Registro de Preço, Aviso das Licitações, Modalidades das Licitações, Licitação por Registro de Preço e Preço Praticado, além dos Órgãos e Unidades Administrativas de Serviços Gerais que integram o SIASG. A Figura 43 detalha esse relacionamento.

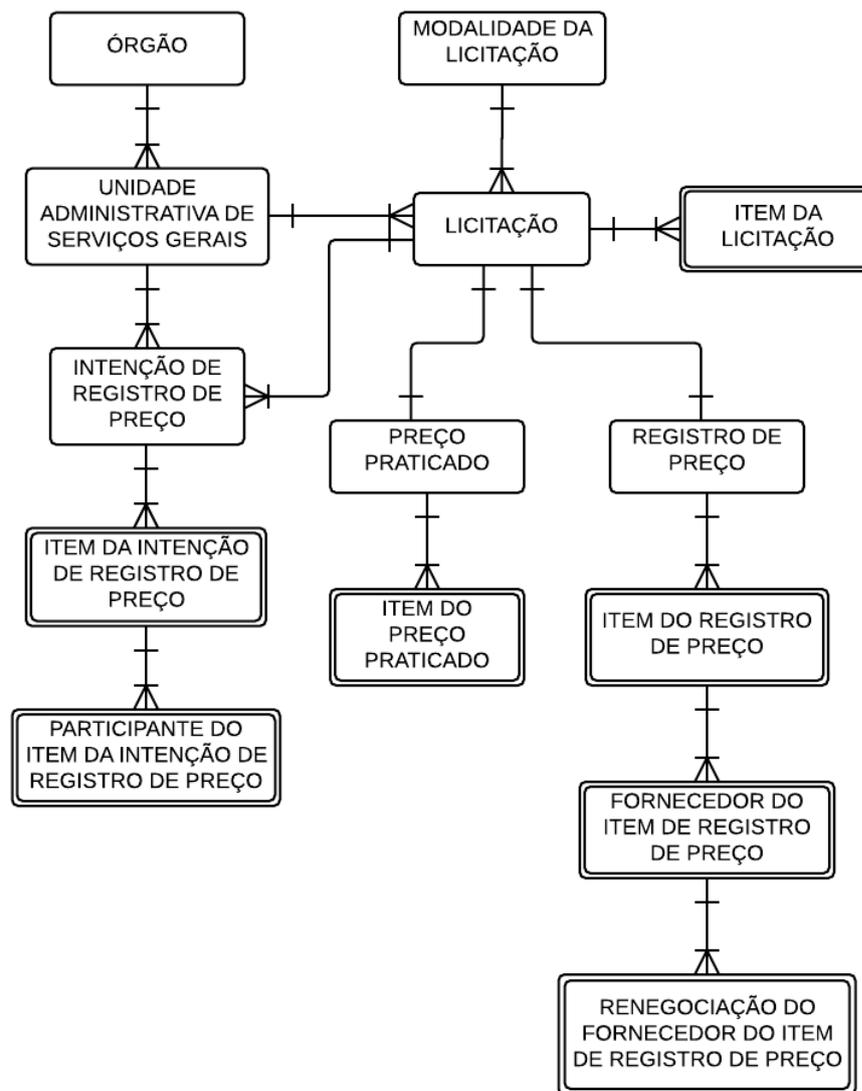


Figura 43 – Modelagem das entidades do módulo de Licitação (Elaborado pelo autor)

E por fim, a Figura 44 ilustra o módulo de contratos. Este módulo contém dados do módulo SICON do SIASG. Será disponibilizado dados referentes a Contratos, Aditivos, Apostilamento e Tipos de Contratos.

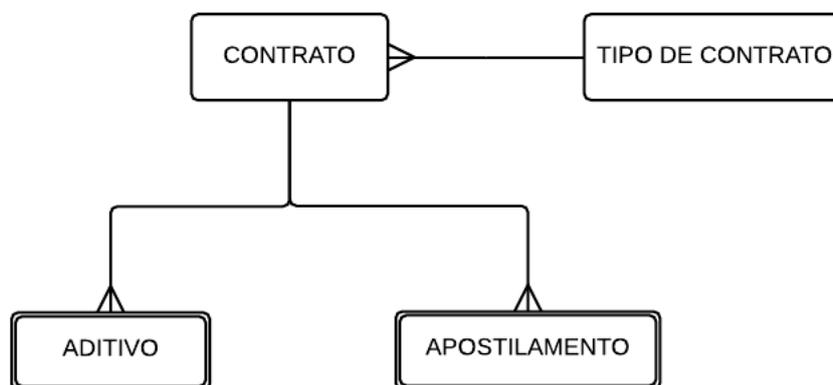


Figura 44 – Modelagem do módulo de contrato (Elaborada pelo autor)

Em cada módulo, os métodos foram classificados em dois tipos: consulta básica e consulta detalhada. No primeiro tipo, a consulta retorna uma lista dos 500 primeiros registros da consulta específica, e permite que parâmetros adicionais sejam passados e estes serão usados para fazer um filtro. Já na consulta detalhada, o único parâmetro aceito é o identificador do recurso a ser detalhado. As Figura 45 e a Figura 46 apresentam a documentação desses dois tipos de métodos, o método de consulta básica de materiais e consulta detalhada do material, respectivamente.

Materiais

Fornece uma lista das informações relacionadas aos materiais cadastrados no CATMAT.

URL:
<http://compras.dados.gov.br/materiais/v1/materiais.{formato}>

Formatos:
html, xml, json, csv

Exemplo:
<http://compras.dados.gov.br/materiais/v1/materiais.html?grupo=84>

Parâmetros da Solicitação

Campo	Tipo	Obrigatório	Descrição
classe	Texto	Não	Código da classe do material.
codigo_item_material	Texto	Não	Código do item de material.
codigo_material	Texto	Não	Código do material.
descricao_item_material	Texto	Não	Descrição do item de material.
grupo	Texto	Não	Código do grupo do material.
material	Texto	Não	Descrição do material.
offset	Inteiro	Não	Quantidade de registros ignorados a partir do início da lista de resultados ordenando pelo ID. Útil para paginar consultas que retornam mais que 500 resultados. Ex.: offset=3000, retorna até 500 registros ignorando os 3000 primeiros.
order	Texto	Não	Atributo utilizado para indicar se ordenação é crescente ou decrescente
order_by	Texto	Não	Atributo utilizado para ordenação
sustentavel	Booleano	Não	Indicador se o item é ou não sustentável.

Informações da Resposta

Campo	Descrição
codigo	Código do item de material.
descricao	Descrição do material.
id_classe	Código da classe de material.
id_grupo	Código do grupo de material
id_pdm	Código do padrão descritivo de material.
status	Indicador se o item é ou não ativo.
sustentavel	Indicador se o item é ou não sustentável.

Figura 45 – Documentação da consulta básica do método Materiais (Elaborado pelo autor)

Material

Todos os dados disponíveis de um material cadastrado no CATMAT.

URL:
<http://compras.dados.gov.br/materiais/id/material/{id}.{formato}>

Formatos:
html, xml, json, csv

Exemplo:
<http://compras.dados.gov.br/materiais/id/material/000227505.html>

Parâmetros da Solicitação

Campo	Tipo	Obrigatório	Descrição
id	Inteiro	Sim	Código do material.

Informações da Resposta

Campo	Descrição
codigo	Código do item de material.
descricao	Descrição do material.
id_classe	Código da classe de material.
id_grupo	Código do grupo de material
id_pdm	Código do padrão descritivo de material.
status	Indicador se o item é ou não ativo.
sustentavel	Indicador se o item é ou não sustentável.

Figura 46 – Documentação da consulta detalhada do método Material (Elaborado pelo autor)

Para facilitar a utilização da API, tanto para máquinas quanto para usuários, uma documentação automática foi gerada, utilizando a biblioteca *Swagger*. A Figura 47 representa a documentação sobre o método Municípios do módulo de Fornecedores em JSON voltado para máquinas e na Figura 48 apresenta a mesma documentação, porém com foco ao usuário, em HTML.

```
1 {
2   "apiVersion": "1.0.0",
3   "swaggerVersion": "1.2",
4   "basePath": "http://compras.dados.gov.br/fornecedores",
5   "resourcePath": "/municipio",
6   "apis": [
7     {
8       "path": "/v1/municipios",
9       "operations": [
10        {
11          "method": "GET",
12          "summary": "Fornece uma lista com informações relacionadas a municipios dentro do território brasileiro.",
13          "notes": "/fornecedores/v1/municipios.html?uf=PE",
14          "type": "resource",
15          "nickname": "Municipios",
16          "produces": [
17            "text/html",
18            "application/json",
19            "application/xml",
20            "text/csv"
21          ],
22          "parameters": [
23            {
24              "name": "uf",
25              "description": "Sigla da unidade federativa.",
26              "required": false,
27              "type": "string",
28              "paramType": "query",
29              "allowMultiple": false
30            },
31            {
32              "name": "nome",
33              "description": "Parte do nome do municipio.",
34              "required": false,
35              "type": "string",
36              "paramType": "query",
37              "allowMultiple": false
38            },
39            {
40              "name": "codigo_ibge",
41              "description": "Código de identificação único de acordo com o IBGE.",
42              "required": false,
43              "type": "integer",
44              "format": "int32",
45              "paramType": "query",
46              "allowMultiple": false
47            },
48            {
49              "name": "ativo",
```

Figura 47 – Documentação do método de Município gerado pela Swagger (Elaborado pelo autor)

Município

Todos os dados disponíveis de um município no SICAF.

URL:

<http://compras.dados.gov.br/fornecedores/id/municipio/{id}.{formato}>

Formatos:

html, xml, json, csv

Exemplo:

<http://compras.dados.gov.br/fornecedores/id/municipio/25313.html>

Parâmetros da Solicitação

Campo	Tipo	Obrigatório	Descrição
id	Texto	Sim	Id do Município.

Informações da Resposta

Campo	Descrição
ativo	Se o Município está ativo.
codigo_ibge	Código de identificação único de acordo com o IBGE.
id	Identificador único de município no SICAF.
nome	Nome do Município.
nome_uf	Nome da Unidade Federativa.
sigla_uf	Sigla da Unidade Federativa.

Figura 48 – Documentação do método de consulta de Municípios (Elaborada pelo autor)

Ao acessar um recurso da API uma resposta é esperada. A resposta é disponibilizada em formatos não proprietários e compreensíveis por máquinas. As Figura 49, Figura 50, Figura 51, Figura 52 representam a consulta de uma lista de municípios, que retorna a resposta nos formatos HTML, XML, JSON, CSV, respectivamente. Note que os recursos implementam o HAL e ao mesmo tempo o conceito de hipermídia do HATEOAS.

Lista de municípios filtrados por uf,

Exibindo resultados 1 - 183 de 183 Veja também em: [xml](#) [json](#) [csv](#)

Município 5444: ARACOIABA	
Id	5444
Código IBGE	2301208
Nome	ARACOIABA
Nome da UF	PERNAMBUCO
Sigla da UF	PE
Ativo	Sim
Veja mais:	Fornecedores neste município Unidades Cadastradoras neste município Licitações neste município Intenções de Registro de Preços neste município

Município 5460: CASINHAS	
Id	

**Figura 49 – Dados da consulta de municípios disponibilizados em HTML
(Elaborada pelo autor)**

```
▼<resource count="183" offset="0">
  ▼<_links>
    <link href="http://compras.dados.gov.br/fornecedores/v1/municipios.xml?uf=PE" rel="self" title="Lista de municípios filtrados por uf, "/>
    <link href="/fornecedores/v1/municipios.xml?uf=PE&offset=0" rel="first" title="Primeira página"/>
  </_links>
  ▼<_embedded>
    ▼<resource>
      ▼<_links>
        <link href="/fornecedores/id/municipio/5444" rel="self" title="Município 5444: ARACOIABA"/>
        <link href="/fornecedores/v1/fornecedores?id_municipio=5444" rel="fornecedores" title="Fornecedores neste município"/>
        <link href="/licitacoes/v1/uasgs?id_municipio=5444&unidade_cadastradora=true" rel="unidades_cadastradora" title="Unidades Cadastradoras neste município"/>
        <link href="/licitacoes/v1/licitacoes?uasg_municipio=5444" rel="licitacoes" title="Licitações neste município"/>
        <link href="/licitacoes/v1/irps?uasg_municipio=5444" rel="irps" title="Intenções de Registro de Preços neste município"/>
      </_links>
      <id>5444</id>
      <codigo_ibge>2301208</codigo_ibge>
      <nome>ARACOIABA</nome>
      <nome_uf>PERNAMBUCO</nome_uf>
      <sigla_uf>PE</sigla_uf>
      <ativo>true</ativo>
    </resource>
    ▼<resource>
      ▼<_links>
        <link href="/fornecedores/id/municipio/5460" rel="self" title="Município 5460: CASINHAS"/>
        <link href="/fornecedores/v1/fornecedores?id_municipio=5460" rel="fornecedores" title="Fornecedores neste município"/>
        <link href="/licitacoes/v1/uasgs?id_municipio=5460&unidade_cadastradora=true" rel="unidades_cadastradora" title="Unidades Cadastradoras neste município"/>
        <link href="/licitacoes/v1/licitacoes?uasg_municipio=5460" rel="licitacoes" title="Licitações neste município"/>
        <link href="/licitacoes/v1/irps?uasg_municipio=5460" rel="irps" title="Intenções de Registro de Preços neste município"/>
      </_links>
      <id>5460</id>
      <codigo_ibge>2604155</codigo_ibge>
      <nome>CASINHAS</nome>
```

**Figura 50 - Dados da consulta de municípios disponibilizados em XML
(Elaborada pelo autor)**

```

1 {
2   "links": {
3     "self": {
4       "href": "http://compras.dados.gov.br/fornecedores/v1/municipios.json?uf=PE",
5       "title": "Lista de municípios filtrados por uf, "
6     },
7     "first": {
8       "href": "/fornecedores/v1/municipios.json?uf=PE&offset=0",
9       "title": "Primeira página"
10    }
11  },
12  "_embedded": {
13    "municipios": [
14      {
15        "id": 5444,
16        "codigo_ibge": 2301208,
17        "nome": "ARACOIABA",
18        "nome_uf": "PERNAMBUCO",
19        "sigla_uf": "PE",
20        "ativo": true,
21        "links": {
22          "self": {
23            "href": "/fornecedores/id/municipio/5444",
24            "title": "Município 5444: ARACOIABA"
25          },
26          "fornecedores": {
27            "href": "/fornecedores/v1/fornecedores?id_municipio=5444",
28            "title": "Fornecedores neste município"
29          },
30          "irps": {
31            "href": "/licitacoes/v1/irps?uasg_municipio=5444",
32            "title": "Intenções de Registro de Preços neste município"
33          },
34          "licitacoes": {
35            "href": "/licitacoes/v1/licitacoes?uasg_municipio=5444",
36            "title": "Licitações neste município"
37          },
38          "unidades_cadastradora": {
39            "href": "/licitacoes/v1/uasgs?id_municipio=5444&unidade_cadastradora=true",
40            "title": "Unidades Cadastradoras neste município"
41          }
42        }
43      },
44      {
45        "id": 5460,
46        "codigo_ibge": 2604155,
47        "nome": "CASINHAS",
48        "nome_uf": "PERNAMBUCO",
49        "sigla_uf": "PE",
50        "ativo": true,

```

Figura 51 - Dados da consulta de municípios disponibilizados em JSON (Elaborada pelo autor)

```

1 Fornecedores neste municipio > uri,Unidades Cadastradoras neste municipio > uri,Licitações neste municipio > uri,Intenções de Registro de Preços neste municipio > uri,Id,Código IBGE,Nome,Nome
2 /fornecedores/v1/fornecedores?id_municipio=5444,/licitacoes/v1/uasgs?id_municipio=5444&unidade_cadastradora=true,/licitacoes/v1/licitacoes?uasg_municipio=5444,/licitacoes/v1/irps?uasg_municip
3 /fornecedores/v1/fornecedores?id_municipio=5460,/licitacoes/v1/uasgs?id_municipio=5460&unidade_cadastradora=true,/licitacoes/v1/licitacoes?uasg_municipio=5460,/licitacoes/v1/irps?uasg_municip
4 /fornecedores/v1/fornecedores?id_municipio=5509,/licitacoes/v1/uasgs?id_municipio=5509&unidade_cadastradora=true,/licitacoes/v1/licitacoes?uasg_municipio=5509,/licitacoes/v1/irps?uasg_municip
5 /fornecedores/v1/fornecedores?id_municipio=5525,/licitacoes/v1/uasgs?id_municipio=5525&unidade_cadastradora=true,/licitacoes/v1/licitacoes?uasg_municipio=5525,/licitacoes/v1/irps?uasg_municip
6 /fornecedores/v1/fornecedores?id_municipio=5568,/licitacoes/v1/uasgs?id_municipio=5568&unidade_cadastradora=true,/licitacoes/v1/licitacoes?uasg_municipio=5568,/licitacoes/v1/irps?uasg_municip
7 /fornecedores/v1/fornecedores?id_municipio=5584,/licitacoes/v1/uasgs?id_municipio=5584&unidade_cadastradora=true,/licitacoes/v1/licitacoes?uasg_municipio=5584,/licitacoes/v1/irps?uasg_municip
8 /fornecedores/v1/fornecedores?id_municipio=22896,/licitacoes/v1/uasgs?id_municipio=22896&unidade_cadastradora=true,/licitacoes/v1/licitacoes?uasg_municipio=22896,/licitacoes/v1/irps?uasg_muni
9 /fornecedores/v1/fornecedores?id_municipio=22918,/licitacoes/v1/uasgs?id_municipio=22918&unidade_cadastradora=true,/licitacoes/v1/licitacoes?uasg_municipio=22918,/licitacoes/v1/irps?uasg_muni
10 /fornecedores/v1/fornecedores?id_municipio=22934,/licitacoes/v1/uasgs?id_municipio=22934&unidade_cadastradora=true,/licitacoes/v1/licitacoes?uasg_municipio=22934,/licitacoes/v1/irps?uasg_muni
11 /fornecedores/v1/fornecedores?id_municipio=22950,/licitacoes/v1/uasgs?id_municipio=22950&unidade_cadastradora=true,/licitacoes/v1/licitacoes?uasg_municipio=22950,/licitacoes/v1/irps?uasg_muni
12 /fornecedores/v1/fornecedores?id_municipio=22977,/licitacoes/v1/uasgs?id_municipio=22977&unidade_cadastradora=true,/licitacoes/v1/licitacoes?uasg_municipio=22977,/licitacoes/v1/irps?uasg_muni
13 /fornecedores/v1/fornecedores?id_municipio=22993,/licitacoes/v1/uasgs?id_municipio=22993&unidade_cadastradora=true,/licitacoes/v1/licitacoes?uasg_municipio=22993,/licitacoes/v1/irps?uasg_muni
14 /fornecedores/v1/fornecedores?id_municipio=23019,/licitacoes/v1/uasgs?id_municipio=23019&unidade_cadastradora=true,/licitacoes/v1/licitacoes?uasg_municipio=23019,/licitacoes/v1/irps?uasg_muni
15 /fornecedores/v1/fornecedores?id_municipio=23035,/licitacoes/v1/uasgs?id_municipio=23035&unidade_cadastradora=true,/licitacoes/v1/licitacoes?uasg_municipio=23035,/licitacoes/v1/irps?uasg_muni
16 /fornecedores/v1/fornecedores?id_municipio=23051,/licitacoes/v1/uasgs?id_municipio=23051&unidade_cadastradora=true,/licitacoes/v1/licitacoes?uasg_municipio=23051,/licitacoes/v1/irps?uasg_muni
17 /fornecedores/v1/fornecedores?id_municipio=23078,/licitacoes/v1/uasgs?id_municipio=23078&unidade_cadastradora=true,/licitacoes/v1/licitacoes?uasg_municipio=23078,/licitacoes/v1/irps?uasg_muni
18 /fornecedores/v1/fornecedores?id_municipio=23094,/licitacoes/v1/uasgs?id_municipio=23094&unidade_cadastradora=true,/licitacoes/v1/licitacoes?uasg_municipio=23094,/licitacoes/v1/irps?uasg_muni
19 /fornecedores/v1/fornecedores?id_municipio=23116,/licitacoes/v1/uasgs?id_municipio=23116&unidade_cadastradora=true,/licitacoes/v1/licitacoes?uasg_municipio=23116,/licitacoes/v1/irps?uasg_muni
20 /fornecedores/v1/fornecedores?id_municipio=23132,/licitacoes/v1/uasgs?id_municipio=23132&unidade_cadastradora=true,/licitacoes/v1/licitacoes?uasg_municipio=23132,/licitacoes/v1/irps?uasg_muni
21 /fornecedores/v1/fornecedores?id_municipio=23159,/licitacoes/v1/uasgs?id_municipio=23159&unidade_cadastradora=true,/licitacoes/v1/licitacoes?uasg_municipio=23159,/licitacoes/v1/irps?uasg_muni
22 /fornecedores/v1/fornecedores?id_municipio=23175,/licitacoes/v1/uasgs?id_municipio=23175&unidade_cadastradora=true,/licitacoes/v1/licitacoes?uasg_municipio=23175,/licitacoes/v1/irps?uasg_muni
23 /fornecedores/v1/fornecedores?id_municipio=23191,/licitacoes/v1/uasgs?id_municipio=23191&unidade_cadastradora=true,/licitacoes/v1/licitacoes?uasg_municipio=23191,/licitacoes/v1/irps?uasg_muni
24 /fornecedores/v1/fornecedores?id_municipio=23213,/licitacoes/v1/uasgs?id_municipio=23213&unidade_cadastradora=true,/licitacoes/v1/licitacoes?uasg_municipio=23213,/licitacoes/v1/irps?uasg_muni
25 /fornecedores/v1/fornecedores?id_municipio=23230,/licitacoes/v1/uasgs?id_municipio=23230&unidade_cadastradora=true,/licitacoes/v1/licitacoes?uasg_municipio=23230,/licitacoes/v1/irps?uasg_muni
26 /fornecedores/v1/fornecedores?id_municipio=23256,/licitacoes/v1/uasgs?id_municipio=23256&unidade_cadastradora=true,/licitacoes/v1/licitacoes?uasg_municipio=23256,/licitacoes/v1/irps?uasg_muni
27 /fornecedores/v1/fornecedores?id_municipio=23272,/licitacoes/v1/uasgs?id_municipio=23272&unidade_cadastradora=true,/licitacoes/v1/licitacoes?uasg_municipio=23272,/licitacoes/v1/irps?uasg_muni
28 /fornecedores/v1/fornecedores?id_municipio=23289,/licitacoes/v1/uasgs?id_municipio=23289&unidade_cadastradora=true,/licitacoes/v1/licitacoes?uasg_municipio=23289,/licitacoes/v1/irps?uasg_muni
29 /fornecedores/v1/fornecedores?id_municipio=23310,/licitacoes/v1/uasgs?id_municipio=23310&unidade_cadastradora=true,/licitacoes/v1/licitacoes?uasg_municipio=23310,/licitacoes/v1/irps?uasg_muni
30 /fornecedores/v1/fornecedores?id_municipio=23337,/licitacoes/v1/uasgs?id_municipio=23337&unidade_cadastradora=true,/licitacoes/v1/licitacoes?uasg_municipio=23337,/licitacoes/v1/irps?uasg_muni
31 /fornecedores/v1/fornecedores?id_municipio=23353,/licitacoes/v1/uasgs?id_municipio=23353&unidade_cadastradora=true,/licitacoes/v1/licitacoes?uasg_municipio=23353,/licitacoes/v1/irps?uasg_muni
32 /fornecedores/v1/fornecedores?id_municipio=23370,/licitacoes/v1/uasgs?id_municipio=23370&unidade_cadastradora=true,/licitacoes/v1/licitacoes?uasg_municipio=23370,/licitacoes/v1/irps?uasg_muni
33 /fornecedores/v1/fornecedores?id_municipio=23396,/licitacoes/v1/uasgs?id_municipio=23396&unidade_cadastradora=true,/licitacoes/v1/licitacoes?uasg_municipio=23396,/licitacoes/v1/irps?uasg_muni
34 /fornecedores/v1/fornecedores?id_municipio=23418,/licitacoes/v1/uasgs?id_municipio=23418&unidade_cadastradora=true,/licitacoes/v1/licitacoes?uasg_municipio=23418,/licitacoes/v1/irps?uasg_muni
35 /fornecedores/v1/fornecedores?id_municipio=23434,/licitacoes/v1/uasgs?id_municipio=23434&unidade_cadastradora=true,/licitacoes/v1/licitacoes?uasg_municipio=23434,/licitacoes/v1/irps?uasg_muni
36 /fornecedores/v1/fornecedores?id_municipio=23450,/licitacoes/v1/uasgs?id_municipio=23450&unidade_cadastradora=true,/licitacoes/v1/licitacoes?uasg_municipio=23450,/licitacoes/v1/irps?uasg_muni
37 /fornecedores/v1/fornecedores?id_municipio=23477,/licitacoes/v1/uasgs?id_municipio=23477&unidade_cadastradora=true,/licitacoes/v1/licitacoes?uasg_municipio=23477,/licitacoes/v1/irps?uasg_muni
38 /fornecedores/v1/fornecedores?id_municipio=23493,/licitacoes/v1/uasgs?id_municipio=23493&unidade_cadastradora=true,/licitacoes/v1/licitacoes?uasg_municipio=23493,/licitacoes/v1/irps?uasg_muni
39 /fornecedores/v1/fornecedores?id_municipio=23515,/licitacoes/v1/uasgs?id_municipio=23515&unidade_cadastradora=true,/licitacoes/v1/licitacoes?uasg_municipio=23515,/licitacoes/v1/irps?uasg_muni
40 /fornecedores/v1/fornecedores?id_municipio=23531,/licitacoes/v1/uasgs?id_municipio=23531&unidade_cadastradora=true,/licitacoes/v1/licitacoes?uasg_municipio=23531,/licitacoes/v1/irps?uasg_muni

```

Figura 52 - Dados da consulta de municípios disponibilizados em CSV (Elaborada pelo autor)

5. Conclusão

Este capítulo aborda as considerações finais referentes a este trabalho, os resultados e contribuições, as limitações e perspectivas futuras.

5.1 Considerações Finais

Este trabalho buscou estudar sobre o processo de publicação em dados abertos no Brasil. Mesmo com as leis e diretrizes que obrigam entidades governamentais a publicarem os dados abertamente para a sociedade, o processo de solicitação pelo canal de atendimento ao cidadão ainda é bastante demorado, e nem sempre é possível obter a informação necessária. Muitos dos dados armazenados pelos órgãos públicos ainda são mantidos em banco de dados hierárquicos, dificultando ainda mais a abertura da informação.

Além disso, a adesão à Parceria do Governo Aberto tem estimulado alguns órgãos a publicarem seus dados para a sociedade. No entanto, são casos pontuais e carece de um modelo que ao mesmo tempo facilite a publicação e estimule o sentimento de dados abertos.

O modelo apresentado nesta dissertação está alinhado à política de *software* de código aberto adotado pelo governo federal, ao mesmo tempo em que segue as boas práticas determinadas pelo e-PING, e outras organizações especializadas no tema. O propósito deste modelo é ser uma referência em publicação de dados abertos e a partir disso possibilitar mais discussões e pesquisas relacionadas a este tema.

A inclusão da análise de qualidade é de suma importância para a política de dados abertos, devido aos benefícios trazidos. A disponibilização de dados com qualidade

produz dados com maior valor agregado, o que impacta positivamente na reutilização dos dados, devido a maior facilidade de interpretação e análise do usuário. Além de agilizar a análise, controle, fiscalização das ações governamentais.

Por outro lado, fazendo uma análise crítica, como (JANSSEN *et al.*, 2012) argumenta, a disponibilização de dados governamentais não deve ser vista como um propósito final, e sim como um meio para poder se produzir um impacto social positivo. Para isso é necessário que haja um interesse por parte da sociedade em reutilizar os dados disponibilizados. Neste sentido, (DINIZ, 2010) afirma: “Não há valor na disponibilização de dados governamentais abertos se a sociedade não tem interesse em reutilizá-los”.

5.2 Resultado e Contribuições

A principal contribuição deste trabalho é o modelo para publicação de dados abertos, que está alinhado com o tripé estabelecido pela INDA: Transparência, Colaboração e Participação. Foram listados os atores principais, um diagrama de atividades e uma arquitetura que poderá ser usada para futuros projetos de aberturas de dados.

Outra contribuição desta dissertação é a implementação da API REST proposta, que atingiu o nível máximo de maturidade seguindo o modelo de Richardson (apresentado na seção 2.4.2), conforme descrito na Tabela 5, que segundo o autor este nível de maturidade torna o protocolo HTTP mais auto documentável, permitindo uma maior “compreensão” das máquinas.

Maturidade	Status	Observação
Nível 3	OK	HATEOAS implementado na estrutura do HAL.

Nível 2	OK	API é somente leitura, então apenas o GET é usado.
Nível 1	OK	URI definidas e única.
Nível 0	OK	API REST usa por padrão HTTP.

Tabela 5 – Nível de maturidade da API seguindo o modelo de Richardson (Elaborado pelo autor)

Outro fato importante, a API desenvolvida está alinhada com a definição da OpenGovData, veja Tabela 6, em sete princípios: completo, primário, acessível, compreensível por máquina, não discriminatório, não proprietário e livre de licença, mas fere o princípio de atualidade dos dados, por falta de dados mais recentes, devido a algumas limitações que serão discutidas na próxima seção (5.3).

Princípio	Situação	Observação
Completo	OK	Todos os dados de compras considerados públicos foram disponibilizados. Devido à restrição de privacidade, dados de CPF e endereço não foram disponibilizados integralmente.
Primário	OK	Algumas adaptações tiveram que ser feitas devido a estrutura do banco hierárquico do SIASG, mas houve um esforço para garantir a primariedade do dado.
Atual	Não	Os dados que conseguimos para este trabalho são de março de 2014, portanto necessitam ser atualizados.
Acessível	OK	Sim. Inclusive já está indexado pelo Google.
Compreensível por máquina	OK	HATEOAS + HAL + Swagger.
Não discriminatório	OK	Não há restrições de pessoas nem lugar.
Não proprietário	OK	Dados são disponibilizados nos formatos: CSV, JSON, HTML e XML.
Livre de Licença	OK	Dados sob a licença <i>Open Database License</i> (ODbL) que permite os usuários compartilhar, modificar e usar os dados livremente, portanto nenhuma restrição é imposta.

Tabela 6 – Comparativo da API desenvolvida com os princípios definidos pela OpenGovData (Elaborado pelo autor)

Além disso, é possível citar também, uma contribuição indireta deste trabalho, o estímulo à adoção da política de dados abertos, que acaba trazendo benefício, “[...] para gestão interna do governo, uma vez que amplia o potencial de melhoria dos processos,

pelo retorno dado pela sociedade, e, principalmente, traz um incentivo importante para mudança de cultura do segredo” (DUTRA & LOPES, 2013).

5.3 Limitações

Este trabalho propõe uma arquitetura para publicação em dados abertos. Embora a prova de conceito tenha sido implementada com sucesso, este trabalho possui algumas limitações.

Mesmo a arquitetura sendo bastante robusta para diversos fins, incluindo o processo de atualização dos dados, não foi possível obter mais de uma extração de dados do SIASG. Sendo assim, a única extração de dados disponibilizada pelo SERPRO é referente aos dados de março de 2014.

A implantação da camada de extração de dados do sistema estruturador não pôde ser implementada, devido a problemas de segurança e criticidade inerente ao SIASG. Por este motivo, foi necessário fazer uma solicitação junto ao SERPRO para que fosse desenvolvido um extrator e os dados disponibilizados no repositório de arquivos.

O processo de análise de qualidade dos dados, por si só, teria insumos suficiente para produzir pelo menos outra dissertação. Por este motivo foi decidido incluir no escopo deste trabalho, apenas uma análise superficial usando a ferramenta de qualidade de dados TOSDQ.

5.4 Trabalhos Futuros

Dentre as possíveis extensões ao trabalho desenvolvido nesta dissertação, pode-se considerar como trabalho futuro diversas evoluções com base no modelo proposto.

Dentre as quais se pode citar:

- Evolução da abordagem da qualidade de dados - Avaliação e definição de dimensões e métricas de qualidade de dados que são mais relevantes no contexto de dados abertos governamentais.
- Evolução da API de dados abertos para dados ligados abertos - Para atingir o nível máximo (cinco estrelas) da classificação de dados ligados abertos é necessário que diversas melhorias sejam implementadas como: inclusão da serialização em RDF, definição da ontologia de compras do governo.
- Desenvolvimento de um *framework* para APIs - A partir das tecnologias e padrões adotados no desenvolvimento da API, é possível encapsular toda a complexidade da inclusão de todas essas tecnologias em um *framework*. Neste sentido, facilitaria bastante a criação de futuras APIs.
- Fomentar a inovação na sociedade quanto ao uso de APIs - Estimular a sociedade quanto a participação, pois a publicação em dados abertos somente faz sentido se a alguém a utiliza. Para isso, eventos como *Hackaton* devem ser promovidos periodicamente para divulgar e estimular o uso das API.

Referências Bibliográficas

AGUNE, R., CARLOS, J., 2005, "Governo eletrônico e novos processos de trabalho". In: *Gestão pública no Brasil contemporâneo*. São Paulo: Fundap.

ALVARENGA, EVERTON ZANELLA, COMUNIDADE TRANSPARÊNCIA HACKER, ESFERA.MOBI, 2011. "Manual dos Dados Abertos: Governo". . Abril 2011. S.l.: s.n. Acessado em: 17 Fevereiro 2015. Disponível em: <http://www.w3c.br/pub/Materiais/PublicacoesW3C/Manual_Dados_Abertos_WEB.pdf>.

BATINI, C., SCANNAPIECA, M., 2006, *Data quality concepts, methodologies and techniques*. . Berlin; New York, Springer. Acessado em: 2 Outubro 2013.

BRASIL, 1994. "DECRETO Nº 1.094, DE 23 DE MARÇO DE 1994.". . 1994. S.l.: s.n. Acessado em: 15 Fevereiro 2015. Disponível em: <http://www.planalto.gov.br/ccivil_03/decreto/Antigos/D1094.htm>.

BRASIL, 2000. "PORTARIA NORMATIVA N.º 2, DE 27 DE OUTUBRO DE 2000". . 2000. S.l.: s.n. Acessado em: 25 Fevereiro 2015. Disponível em: <http://www.comprasnet.gov.br/legislacao/portarias/p02_00.htm>.

BRASIL, 2010. "Decreto de 03.04.2000". . 2010. S.l.: s.n. Acessado em: 15 Fevereiro 2015. Disponível em: <http://www.governoeletronico.gov.br/anexos/E15_90Decreto_3_de_abril_de_2000.pdf>.

BRASIL, 2011. "LEI Nº 12.527". . 18 Novembro 2011. S.l.: s.n. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm>.

BRASIL, 2012. "Instrução Normativa da INDA". . 12 Abril 2012. S.l.: s.n. Acessado em: 17 Fevereiro 2015. Disponível em: <<http://dados.gov.br/instrucao-normativa-da-inda/>>.

BRASIL, 2015. "Padrões de Interoperabilidade de Governo Eletrônico - ePING". . 2015. S.l.: s.n. Acessado em: 15 Fevereiro 2015. Disponível em: <<http://eping.governoeletronico.gov.br/>>.

CETIC, 2013. "Indicadores TIC Domicílios". . 2013. S.l.: s.n. Acessado em: 15 Fevereiro 2015. Disponível em: <<http://www.cetic.br/pesquisa/domicilios/indicadores>>.

CGU, 2015a. "Portal de Transparência Pública". . 2015. S.l.: s.n. Acessado em: 17 Fevereiro 2015. Disponível em: <<http://www3.transparencia.gov.br/TransparenciaPublica/glossario/>>.

CGU, 2015b. "Portal do Acesso à Informação". . 2015. S.l.: s.n. Acessado em: 17 Fevereiro 2015. Disponível em: <<http://www.acessoinformacao.gov.br/assuntos/relatorios-dados/relatorios-estatisticos/relatorios-estatisticos>>.

CLIENTESA, 2012. "Boas perspectivas para o Brasil". . 2012. S.l.: s.n.

DINIZ, V., 2010. "Como conseguir dados governamentais abertos". In: *CONGRESSO CONSAD DE GESTÃO PÚBLICA, III, Brasília*. S.l.: s.n. 2010.

DUTRA, C.C., LOPES, K.M.G., 2013, "Dados abertos: Uma forma inovadora de transparência.". In: .

EAVES, DAVID, 2009. "The three laws of open government". . 30 Setembro 2009. S.l.: s.n. Acessado em: 17 Fevereiro 2015. Disponível em: <<http://eaves.ca/2009/09/30/three-law-of-open-government-data/>>.

ECKERSON, W.W., 2002, "Data quality and the bottom line". In: *TDWI Report, The Data Warehouse Institute*.

E-PING, 2014, *Padrões de Interoperabilidade de Governo Eletrônico*. . S.l., s.n. Acessado em: 17 Fevereiro 2015.

FIELDING, R.T., 2000. *Architectural styles and the design of network-based software architectures*. . S.l.: University of California, Irvine.

FRANZOSI, E.M., GARCIA, A., RODRIGUES, S.A., et al., 2009, "Uma proposta de arquitetura referencial SOA para desenvolvimento de sistemas para o governo". In: *Bento Gonçalves-RS. SBC*.

GRIFFIN, DAVID, HALPIN, EDWARD F., DISSANAYAKE, LAKSHMAN, et al., 2013, *Digital Public Administration and E-Government in Developing Nations: Policy and Practice (Advances in Electronic Government, Digital Divide, and Regional Development)*. . 1. S.l., Information Science Reference.

HACHIGIAN, N., 2002, "Roadmap for E-government in the Developing World". In: .

JANSSEN, M., CHARALABIDIS, Y., ZUIDERWIJK, A., 2012, "Benefits, Adoption Barriers and Myths of Open Data and Open Government". In: *Information Systems Management*. v. 29, pp. 258–268.

KELLY, MIKE, 2013. "HAL - Hypertext Application Language". . 13 Junho 2013. S.l.: s.n. Acessado em: 17 Fevereiro 2015. Disponível em: <http://stateless.co/hal_specification.html>.

LENK, K., TRAUNMULLER, R., 2000. "A framework for electronic government". In: *11th International Workshop on Database and Expert Systems Applications, 2000. Proceedings*. S.l.: s.n. 2000. pp. 271–277.

MALAMUD, CARL, O'REILLY, TIM, ELIN, GREG, et al., 2007. "The 8 Principles of Open Government Data". . 8 Dezembro 2007. S.l.: s.n. Acessado em: 17 Fevereiro 2015. Disponível em: <<http://opengovdata.org/>>.

MARTANO, A.M., CRAVEIRO, G.S., 2014, "Abertura e Disponibilização de Dados Abertos Governamentais: Estudos de Caso". In: .

MPOG, SLTI, 2001. "SIASG - Modernização, Transparência e Desburocratização das Compras Governamentais". . 2001. S.l.: s.n. Acessado em: 17 Fevereiro 2015. Disponível

em: <<http://www.fondcf.ufms.br/SIASG-MAIO%202001%20FORUM%20IFES%20Beethoven.ppt>>.

NEELIE KROES, 2011. "Data is the new gold". . 12 Dezembro 2011. S.l.: s.n.

NETO, DIÓGENES LIMA, 2010. "SIASG - Conceitos e Perspectivas". . 2010. S.l.: s.n. Acessado em: 17 Fevereiro 2015. Disponível em: <<http://www.scribd.com/doc/27684439/SIASG-Conceitos-e-Perspectivas-Prof-Diogenes>>.

NETO, TRAJANO CARLOS MONTASSIER, 2012. "AVALIAÇÃO DAS FERRAMENTAS ETL OPEN-SOURCE TALEND E KETTLE PARA PROJETOS DE DATA WAREHOUSE EM EMPRESAS DE PEQUENO PORTE". . 2012. S.l.: s.n. Acessado em: 17 Fevereiro 2015. Disponível em: <http://www.ambientelivre.com.br/downloads/doc_download/87-tcc-ferramentas-de-etl-open-source-talend-e-kettle.html>.

OGP, 2015. "OGP Como Funciona". . 2015. S.l.: s.n. Acessado em: 17 Fevereiro 2015. Disponível em: <http://www.governoaberto.cgu.gov.br/a-ogp/como_Funciona.asp>.

OPEN KNOWLEDGE FOUNDATION (OKF), 2012. "Open Data Handbook Documentation". . 14 Novembro 2012. S.l.: s.n. Acessado em: 17 Fevereiro 2015. Disponível em: <<http://opendatahandbook.org/pdf/OpenDataHandbook.pdf>>.

PIPINO, L.L., LEE, Y.W., WANG, R.Y., 2002, "Data Quality Assessment". In: *Commun. ACM*. v. 45, pp. 211–218.

PORTAL DE GOVERNO ELETRÔNICO DO BRASIL, 2015. "Portal de Governo Eletrônico do Brasil - Princípios". . 2015. S.l.: s.n. Acessado em: 15 Fevereiro 2015. Disponível em: <<http://www.governoeletronico.gov.br/o-gov.br/principios>>.

REDMAN, T.C., 1998, "The Impact of Poor Data Quality on the Typical Enterprise". In: *Commun. ACM*. v. 41, pp. 79–82.

RICHARDSON, LEONARD, 2010. "Richardson Maturity Model". . 2010. S.l.: s.n.

SECRETARIA DE LOGÍSTICA E TECNOLOGIA DA INFORMAÇÃO (SLTI), MINISTÉRIO DO PLANEJAMENTO ORÇAMENTO E GESTÃO (MPOG), 2012. "Cartilha Técnica para Publicação de Dados Abertos no Brasil v1.0". . 2012. S.l.: s.n. Acessado em: 17 Fevereiro 2015. Disponível em: <<http://www.dados.gov.br/cartilha-publicacao-dados-abertos/>>.

SERPRO, 2015. "Siasg - Sistema Integrado de Administração de Serviços Gerais". . 2015. S.l.: s.n. Acessado em: 15 Fevereiro 2015. Disponível em: <<https://www5.serpro.gov.br/conteudo-solucoes/produtos/administracao-federal/siasg-sistema-integrado-de-administracao-de-servicos-gerais>>.

TGC, 2002. "TDWI STUDY SHOWS THAT POOR DATA QUALITY COSTS \$600B A YEAR". . 2 Dezembro 2002. S.l.: s.n. Acessado em: 15 Fevereiro 2015. Disponível em: <<http://www.tgc.com/dsstar/02/0212/103909.html>>.

UNIES, N., MANAGEMENT, D., 2008, *United Nations e-government survey 2008: From e-government to connected governance*. . S.l., UN.

W3C, 2007. "SOAP Version 1.2 Part 1: Messaging Framework (Second Edition)". . 27 Abril 2007. S.l.: s.n. Acessado em: 17 Fevereiro 2015. Disponível em: <<http://www.w3.org/TR/soap12-part1/>>.

WANG, R.Y., 1998, "A Product Perspective on Total Data Quality Management". In: *Commun. ACM*. v. 41, pp. 58–65.

WANG, R.Y., STRONG, D.M., 1996, "Beyond accuracy: What data quality means to data consumers". In: *Journal of management information systems*. pp. 5–33.