



Universidade Federal da Paraíba
Centro de Ciências Sociais Aplicadas
Programa de Pós-Graduação em Administração
Curso de Mestrado

Thiago Emídio Esteves da Silva

PERCEPÇÕES DO PRINCIPAL SOBRE A QUALIDADE DOS DADOS ABERTOS
GOVERNAMENTAIS: Uma análise à luz da teoria principal-agente

João Pessoa – PB
Fevereiro/2019



Thiago Emídio Esteves da Silva

PERCEPÇÕES DO PRINCIPAL SOBRE A QUALIDADE DOS DADOS ABERTOS
GOVERNAMENTAIS: Uma análise à luz da teoria principal-agente

Dissertação apresentada ao Programa de Pós-Graduação em Administração da Universidade Federal da Paraíba, na área de concentração Administração e Sociedade, como requisito para obtenção do título de Mestre em Administração.

Orientador: Prof. Pedro Jácome de Moura Júnior

João Pessoa – PB

Fevereiro/2019

Catálogo na publicação
Seção de Catalogação e Classificação

S586p Silva, Thiago Emídio Esteves da.
PERCEPÇÕES DO PRINCIPAL SOBRE A QUALIDADE DOS DADOS
ABERTOS GOVERNAMENTAIS: Uma análise à luz da teoria
principal-agente. / Thiago Emídio Esteves da Silva. -
João Pessoa, 2019.
118 f. : il.

Orientação: Pedro Jácome de Moura Júnior.
Dissertação (Mestrado) - UFPE/CCSA.

1. Teoria Principal-Agente. 2. Qualidade de Dados. 3.
Dados Abertos Governamentais. 4. Boa Governança. I.
Moura Júnior, Pedro Jácome de. II. Título.

UFPE/CCSA

Thiago Emídio Esteves da Silva

PERCEPÇÕES DO PRINCIPAL SOBRE A QUALIDADE DOS DADOS ABERTOS
GOVERNAMENTAIS: Uma análise à luz da teoria principal-agente.

Banca apresentada em: 28 / 02 /2019

Banca examinadora

Prof. Dr. Pedro Jácome de Moura Júnior
Orientador

Prof. Dr. Francisco José da Costa
Membro Interno do Programa

Prof. Dr. Brivaldo André Marinho da Silva
Membro Externo ao Programa

“Uma decisão só é tão boa quanto os dados dos quais ela se baseia”.

Autor Desconhecido.

AGRADECIMENTOS

Agradeço a minha família, principalmente a minha querida mãe, Gláucia Esteves da Silva; a meu dedicado pai, Severino Emídio da Silva Filho; e a meu irmão, Felipe Emídio Esteves da Silva, os quais, mesmo longe, acompanharam-me de perto até aqui, ajudando-me em cada letra desta dissertação. Também agradeço a todos os meus tios, primos, avôs, mesmo aqueles que não estão mais aqui conosco (*in memoriam*), mas que contribuíram para que eu chegasse a essa conquista.

Ao meu orientador pela paciência. Obrigado, professor Pedro Jácome – sem você eu nunca teria chegado aqui.

Aos professores Brivaldo e Franzé, por estarem sempre dando um apoio quando podiam.

A todos os meus professores do programa de Pós-Graduação em Administração da Universidade Federal da Paraíba, pelo esforço constante para que eu pudesse entregar este documento assim como me aperfeiçoar como acadêmico, profissional e como pessoa.

A meus colegas, principalmente Walter, Mayara e Bruno, por estarem me ajudando nas piores horas e com esta pesquisa e sempre me estendendo a mão.

Aos meus amigos, Fillipi e Karol, e minha companheira, Maruska, por sempre me ajudarem nos momentos difíceis.

A minha amiga e revisora, Sra. Joséli, por me apoiar e ter os grandes níveis de paciência com minhas dificuldades.

E que a mochila de estudos nunca deixe de ser usada. A todos, muito obrigado.

Dedico este Mestrado aos meus pais,
Emídio e Gláucia, pelo amor, incentivo e apoio
em todas as minhas escolhas e decisões e por
estarem sempre ao meu lado.

RESUMO

Na Teoria Principal-Agente, a administração pública é o Agente dos cidadãos, ou seja, a administração pública está a serviço dos cidadãos, identificados como o Principal, os reais donos da propriedade pública em uma democracia. Para fazer seu trabalho como Agente, a administração pública fornece dados abertos governamentais aos cidadãos, principalmente através da internet, como prova de sua confiabilidade, e, assim, supostamente, consegue atenuar a assimetria de informação inerente a essa relação, além de reduzir a desconfiança dos cidadãos. Para esses dados terem o efeito desejado, eles devem, contudo, ser percebidos, pelo Principal, como tendo qualidade de dados, tanto em seu formato quanto em sua função. Essa pesquisa propõe-se a usar a Teoria Principal-Agente como orientação teórica para identificar essa percepção, pois a literatura pede mais atenção à qualidade de dados em questão de função, e, conforme visto na literatura de tecnologia da informação, requisita-se a utilização dessa teoria na avaliação de sua usabilidade de tais tecnologias. Para atingir tal propósito, foram entrevistados 12 indivíduos que trabalharam recentemente com conjuntos de dados abertos governamentais, seguido da análise de conteúdo dos relatos obtidos com base na literatura sobre a Teoria Principal-Agente, Governança e Qualidade dos dados. Esse processo resultou em 19 constructos de qualidade de dados identificados – dois deles sendo emergentes foram evidenciados e discutidos – assim como a identificação da percepção de qualidade de dados como um todo para os entrevistados.

Palavras-Chave: Teoria Principal-Agente. Qualidade de Dados. Dados Abertos Governamentais. Boa Governança.

ABSTRACT

In the Principal-Agent Theory, the public administration is the Agent of the citizens, that is, the public administration is at the service of the citizens, identified as Principals, as they are the real owners of public property in a democracy. To do its work as an Agent, the public administration provides open government data to the citizens, mainly through the internet, as evidence of its reliability, and thus, supposedly mitigate the asymmetry of information inherent to the Principal-Agent relationship and also to reduce distrust of the citizens. However, for these data to have the desired effect, they must be perceived by the Principal as having data quality, both in the form and in the function. This research uses the Principal-Agent Theory as a theoretical orientation to identify this perception because the literature calls for more attention to the quality of data in relation to its function, and in the literature of information technology to request the use of this theory in the evaluation of information technology usability. To achieve this, we interviewed 12 individuals who have recently worked with open government data sets, followed by the content analysis of the obtained reports based on the literature on principal-agent theory, governance, and data quality. This process resulted in 19 identified data quality constructs - two of which were emergent constructs - that were evidenced and discussed, as well as the identification of the data quality perception as a whole, for the interviewed individuals.

Keywords: Principal-Agent Theory. Data Quality. Open Government Data. Good Governance.

Índice de Figuras

Figura 1 – Representação Gráfica da Conexão das Ideias do Referencial Teórico.....	21
Figura 2 – Representação Gráfica da Relação Principal-Agente	23
Figura 3 – Modelo Conceitual do Referencial Teórico	40
Figura 4 – Aplicativos feitos pelos discentes	43
Figura 5 – Representação dos Constructos e suas Definições.....	54
Figura 6 – Processo de Análise dos Dados.....	57

Índice de Quadros

Quadro 1 – Definições de Ferramentas para a Tomada de Decisão	35
Quadro 2 – Exemplo de Literatura Sobre Métodos de Mensuração de Dados.....	36
Quadro 3 – Questões de Percepções Sobre os Bancos de Dados	46
Quadro 4 – Metodologias de Mensuração de Qualidade de Dados e Suas Dimensões.....	48
Quadro 5 – Resultados sobre os Respondentes e os Bancos de Dados	58
Quadro 6 – Constructos Únicos por Ordem de Entrevistados.....	59
Quadro 7 – Sentenças e Constructos por Entrevistados	59
Quadro 9 – Natureza da Resposta por Entrevistado em Relação à Primeira Sentença	60
Quadro 10 – Acessibilidade.....	62
Quadro 11 – Clareza.....	64
Quadro 12 – Completude.....	65
Quadro 13 – Compreensibilidade	67
Quadro 14 – Concisão	69
Quadro 15 – Consistência.....	70
Quadro 16 – Conveniência	72
Quadro 17 – Credibilidade	74
Quadro 22 – Exatidão	75
Quadro 19 – Facilidade de Operação	76
Quadro 20 – Interatividade	78
Quadro 21 – Interpretabilidade.....	79
Quadro 23 – Prestação de Contas	80
Quadro 24 – Quantidade adequada de dados.....	82
Quadro 27 –Transparência.....	83
Quadro 26 – Singularidade	84
Quadro 18 – Experiência	85
Quadro 25 –Responsabilidade	86
Quadro 27 – Sentenças por Constructo Expandida	87
Quadro 28 – Sentenças por Conjunto de Dados	88

Índice de Abreviaturas

AIMQ	A methodology for information quality assessment
AMEQ	Activity-based Measuring and Evaluating of product
ANAC	Agencia Nacional de Aviação Civil
BDQA	Big Data Quality Assessment
BDQAFSO	Big Data Quality Antecedents in Financial Services Organizations
CDQ	Comprehensive methodology for Data Quality management
CGU	Controladoria Geral da União
CIHI	Canadian Institute for Health Information methodology
COLDQ	Loshin methodology of Cost–effect of Low data quality
Conj	Conjunto de dados fornecido ao entrevistado
DaQuinCIS	Data Quality in Cooperative Information Systems
DDDC	Data Disclosure Desired Characteristics
DQA	Data quality assessment
DQFMPS	Data Quality Framework for Malaysia’s Public Sector Open Data Initiative
DWQ	Data Warehouse Quality Methodology
ENADE	Exame Nacional de Desempenho de Estudantes
ENCCEJA	Exame Nacional para Certificação de Competências de Jovens e Adultos
Entrev.	Codiname dado ao entrevistado
ESALQ	Escola Superior de Agricultura Luiz de Queiroz
FSMBDQA	Framework for social media big data quality analysis
GTI	Governança de Tecnologia de Informação
HDQM	Heterogeneous Data Quality Methodology
ID_Sen	Número único de identificação de TSen
INDA	Infraestrutura Nacional de Dados Abertos
INDE	Infraestrutura Nacional de Dados Espaciais
IQM	Information quality measurement
ISTAT	ISTAT methodology
LDIF	Linked Data Integration Framework
MEC	Ministério da Educação e Cultura
MEHDQA	Methods of Electronic Health Record Data Quality Assessment
MEIDIQS	Methodology to Evaluate Important Dimensions of Information Quality in Systems
MFDQAI	Methodologies for Data Quality Assessment and Improvement
ODQMF	Open Data Quality Measurement Framework
OLAP	Online Analytical Processing
OLDQAF	Objective Linked Data Quality Assessment Framework
PAC	Programa de Aceleração do Crescimento

Q	Questão sendo respondida pelo entrevistado
QAD	Quantidade Adequada de Dados
QAFD	Methodology for the quality assessment of financial data
QALD	Quality assessment for linked data: A survey
QALOD	Quality Assessment of Linked Open Data
QAMF	Quality Assessment Monitoring Framework
SI&D	Sistemas de Informação e Decisão
TCon	Nome do constructo obtido a partir da TSen dos entrevistados
TCat	Descrição do tipo ou categoria de TNat observado
TDQM	Total Data Quality Management
TI	Tecnologia da Informação
TIC	Tecnologia de Informação e Comunicação
TIQM	Total Quality Information Management
TNat	Natureza encontrada em uma TSen
TPA	Teoria Principal-Agente; Teoria da Agência
TSen	Sentença com natureza única retirado da transcrição das entrevistas
TObj	Objeto ao qual a TCon se relaciona
UDADQ	Utility-driven assessment of data quality.
UFPB	Universidade Federal da Paraíba
USP	Universidade de São Paulo

Sumário

RESUMO.....	7
ABSTRACT	8
Índice de Figuras.....	9
Índice de Quadros	10
Índice de Abreviaturas	11
Sumário.....	13
1. INTRODUÇÃO.....	16
1.1. Justificativa	20
1.2. Objetivo Geral.....	20
1.3. Objetivos Específicos	20
2. REFERENCIAL TEÓRICO.....	21
2.1. Teoria Principal-Agente.....	22
2.2. Governança e Boa Governança.....	27
2.3. Qualidade dos Dados	33
2.4. TPA e Qualidade de Dados.....	38
3. METODOLOGIA.....	41
3.1. População e Amostra	41
3.2. Obtenção dos Dados	44
3.3. Roteiro de Entrevista	45
3.4. Fontes Para a Análise dos Dados	46
3.5. Análise e Tratamento dos Dados	56
4. RESULTADOS E DISCUSSÃO	58
4.1. Saturação.....	59
4.2. Geral.....	59
4.3. TCon - Acessibilidade	61
4.4. TCon - Clareza.....	63

4.5.	TCon - Completude	64
4.6.	TCon - Compreensibilidade	66
4.7.	TCon - Concisão	68
4.8.	TCon - Consistência.....	69
4.9.	TCon - Conveniência	71
4.10.	TCon - Credibilidade.....	72
4.11.	TCon - Exatidão	74
4.12.	TCon - Facilidade de Operação.....	75
4.13.	TCon - Interatividade	77
4.14.	TCon - Interpretabilidade	78
4.15.	TCon - Prestação de Contas	79
4.16.	TCon - Quantidade Adequada de Dados.....	81
4.17.	TCon - Transparência.....	83
4.18.	TCon - Singularidade	84
4.19.	TCon Emergente- Experiência.....	84
4.20.	TCon Emergente- Responsabilidade.....	85
4.21.	Percepção de Qualidade de Dados à Luz da TPA	87
4.22.	Resultados Complementares	90
5.	CONSIDERAÇÕES FINAIS	92
5.1.	Contribuições Teóricas	92
5.2.	Contribuições Práticas	93
5.3.	Limitações.....	94
5.4.	Sugestões de Pesquisas Futuras	96
6.	REFERÊNCIAS	97
7.	ANEXOS	107
7.1.	Projeto de Atividade SI&D.....	108
8.	APÊNDICES	111

8.1. Termo de Consentimento Livre e Esclarecido – TCLE.....	111
8.2. Roteiro de Entrevistas	112
8.3. Quadro de Métricas e Referências em Batini <i>et al.</i> (2009).....	113
8.4. Primeira Lista de Conceitos Obtidos a partir da Literatura	118

1. INTRODUÇÃO

A administração pública brasileira, e, por consequência, o gestor público brasileiro, tem vários princípios que norteiam seu comportamento. Alguns desses princípios são: transparência e prestação de contas, como bem exemplificado no princípio da “Publicidade” da Constituição Federal (BRASIL, 1988) e outras leis que buscam liberar informações ao público (BRASIL, 2000, 2011, 2009). A esse esforço se dá o nome de Governança. Governança é um método usado há séculos para manter confiança e melhorar a relação entre Principal e Agente, ou seja, entre cidadãos e a administração pública (MORCK; STEIER, 2005; FORTESCUE-ALLAND, 1714).

A Teoria Principal-Agente (TPA), também chamada de Teoria da Agência, é caracterizada pela relação entre um indivíduo ou grupo – conhecido como Principal – que delega uma atividade e certo poder de decisão a outro indivíduo ou grupo – conhecido como Agente. Assim, o Principal é o proprietário dos recursos, mas quem administra tais recursos é o Agente (MITNICK, 1973; ROSS, 1973; JENSEN; MECKLING, 1976). No caso brasileiro, o agente público e administração como um todo são apenas Agentes do real proprietário do estado democrático brasileiro, os cidadãos.

Ross (1973) afirma que o Agente tende a colocar seus interesses pessoais em primeiro plano; o Principal deve, então, ficar observante do comportamento do Agente e oferecer incentivos motivadores, de forma a alinhar os comportamentos do Agente com os interesses de que o Principal necessita. O Principal deve ser observante, pois, ao delegar a ação e poder de decisão ao Agente, favorece uma assimetria informacional, já que toda informação sobre o trabalho e o que esperar dele naturalmente se centralizam no Agente, pois é esse indivíduo que está lidando com o serviço diariamente, não o Principal (MITNICK, 1973).

Para evitar desvios de comportamento, é utilizada a Governança, que busca fornecer dados confiáveis. Ela permite que o Principal tenha poder de modificar desvios de comportamento do Agente e reduz, assim, a assimetria de informação, gerando confiança em seu trabalho. (MITNICK, 1973; JENSEN; MECKLING, 1976; BANCO MUNDIAL, 1992; BANCO MUNDIAL, 2017; IBGC, 2018).

Há de se observar que Governança na literatura e em práticas organizacionais tem várias facetas e tipologias. Esforços de prestação de contas (*accountability*) e transparência, no entanto, se mantêm, como formas de mitigar a assimetria natural de informações entre

Principal e Agente (MORCK; STEIER, 2005; FORTESCUE-ALLAND, 1714; BANCO MUNDIAL, 1992; BANCO MUNDIAL, 2017; IBGC, 2018).

Para atender à demanda de Governança, são utilizadas as Tecnologias de Informação e Comunicação (TICs) e a internet para disponibilizar dados abertos governamentais. Como exemplo, têm-se o Portal da Transparência (2018), o Portal de Dados Abertos Governamentais (2018) e o Portal da Controladoria Geral da União (CGU) (2016). Há, porém, um problema na qualidade desses dados (FABRINI, 2018; ANTUNES, 2018).

Dito de outra forma, os dados devem ter qualidade, para que sejam efetivos como esforço de Governança e favoreçam a relação entre o Principal e o Agente. E como definem Cai e Zhu (2015), Batini e Scannapieco (2016) e Wang e Strong (1996), tais dados devem estar adequados ao uso do consumidor de dados e à sua intenção original. Wang e Strong (1996) chamam consumidor de dados aquele que vai usar este recurso como ferramenta de tomada de decisão.

A confiança nos dados e dados confiáveis são indispensáveis a Governança para poder mitigar conflitos de interesse entre o cidadão e administração pública, como afirmam Welch e Hinnant (2003). Isso permite ao cidadão obter informações suficientes para cobrar um serviço público de qualidade (ALENEZI, TARHINI, MASA'DEH, 2015; ANTUNES, 2018).

Qualidade de dados também é importante como ferramenta de ajuda em tomada de decisão. Segundo Kwon, Lee e Shin (2014) e Hazen *et al.* (2014), métodos e áreas de análises de dados, por exemplo, *big data*, *business intelligence & analytics*, e *data analytics*, são dependentes dessa qualidade. A *Big data* é reconhecida como método de análise de conjuntos de dados volumosos e complexos demais para sistemas tradicionais de análise (AGARWAL; DHAR, 2014; PROVOST; FAWCETT, 2013); *Business intelligence & analytics*, definido como a área maior que engloba as formas de análise e obtenção de informação usada por negócios (CHEN *et al.*, 2012); e *Data analytics*, definido como a área que engloba todas as formas de analisar dados, incluindo os de TICs (HAZEN *et al.*, 2014). Essas ferramentas podem promover a geração de valor entre atores institucionais, por meio da confiança.

Para Hazen *et al.* (2014) e para Wang, Ziad e Lee (2006), a qualidade da informação está intimamente ligada à qualidade dos dados em que são baseados. Davenport e Patil (2012) enfatizam que o futuro dependerá dos indivíduos e das organizações que melhor souberem lidar com informações. Isso inclui informações na relação entre cidadão e administração pública. Assim, a tomada de decisão baseada em qualidade de dados transforma os dados em uma nova “tecnologia de gestão”, como diz Montanholi (2017).

Instrumentos de mensuração de qualidade de dados abertos governamentais, como os de Francisco *et al.* (2017), Behkamal *et al.* (2014), Ijab *et al.* (2017) e Vertrò *et al.* (2016), assim como as leis de gestão de informação (BRASIL, 1988, 2000, 2001, 2009, 2011), contudo, focam em metodologias mais objetivas e restritas de mensuração da qualidade de dados, exigindo que estes tenham o formato adequado, mas não mencionando as qualidades subjetivas, como a utilidade e a intenção dos dados mostrados. Isso não garante uma qualidade total, mesmo que todos os dados estejam presentes da forma correta. O que importa à TPA e ao Principal, portanto, é a capacidade dos dados de seguir sua função original de informar e a sua utilidade na relação entre Principal e Agente.

Uma característica definidora de qualidade de dados é quanto estes estão relacionados aos seus usos e intenções originais, ainda mais quando se trata de qualidade de dados abertos governamentais (CAI; ZHU, 2015; WANG; STRONG, 1996; BATINI; SCANNAPIECO, 2016).

Os Frameworks de qualidade de dados abertos governamentais de Vertrò *et al.* (2016) e de Ijab *et al.* (2017) fazem menção à qualidade de dados definida como o uso das informações nos bancos de dados e, destes, só Ijab *et al.* (2017) consideraram utilidade em sua avaliação de qualidade, a saber, na dimensão relevância. Mesmo assim, não especificaram como chegaram a essa dimensão, ou esclarecendo os meios de mensuração de utilidade e intenção na relação administração pública e cidadão.

É neste ponto que a Teoria Principal-Agente se encaixa. Pesquisas em TI, como as de Bhattacharjee (1998); Bøe e Gulbrandsen (2016); Bøe, Gulbrandsen e Sørenbø, (2015); Bøe, (2016); Karanja (2017) e Coser (2017), se destacam neste caráter subjetivo de utilidade e intenção dos dados que carecem de explicação em metodologias de qualidade de dados. Bøe e Gulbrandsen (2016); Bøe, Gulbrandsen e Sørenbø, (2015); Bøe, (2016); e Coser (2017) recomendam, inclusive, que pesquisas futuras utilizem, para a TPA, tecnologias da informação em sua intenção original. Esta proposta de pesquisa atende a esta demanda adotando a TPA como teoria base para analisar qualidade de dados abertos governamentais, algo ainda não explorado. Tal proposta também pretende atender à demanda de Antunes (2018) e Moura (2018), que recomendam analisar a efetividade dos portais brasileiros na capacidade de informar o cidadão.

Os cidadãos e a administração pública são atores fundamentais nesse novo processo de garantir transparência do setor público nos seus dados abertos governamentais, para aumentar a efetividade em sua relação como agente dos cidadãos. Disto deriva a questão desta pesquisa:

Quais seriam as percepções do principal sobre a qualidade de dados abertos governamentais, à luz da Teoria Principal-Agente?

1.1. Justificativa

Essa pesquisa se justifica teoricamente pelas várias sugestões na literatura de TICs (BØE; GULBRANDSEN, 2016; BØE; GULBRANDSEN; SØREBØ, 2015; BØE, 2016; COSER, 2017) para usar a TPA, no intuito de alinhar o uso de suas ferramentas com a sua intenção original. Até o presente momento, os métodos de mensuração de qualidade de dados pouco se focaram em sua qualidade quanto à função (CAI; ZHU, 2015; WANG; STRONG, 1996), portanto teriam a ganhar em testar a TPA como orientação teórica para análise da qualidade de dados abertos governamentais. E, por fim, a literatura também pede para observar se os dados abertos governamentais realmente estão servindo a sua função em Governança (MOURA, 2018; ANTUNES, 2018).

Como formas adicionais de justificativa, tem-se:

- I. A análise da qualidade de dados, do ponto de vista da TPA, pode dar poder ao cidadão (Principal);
- II. Uma metodologia de mensuração de qualidade de dados na função e forma, na perspectiva da TPA, talvez surja como repercussão desta pesquisa e dê um novo norteamento aos estudos sobre tal; e
- III. Esta pesquisa pode expandir a frameworks sobre mensuração da qualidade de dados.

1.2. Objetivo Geral

Identificar as percepções sobre qualidade de dados abertos governamentais, à luz da TPA.

1.3. Objetivos Específicos

- Descrever a percepção do Principal sobre a Qualidade de dados abertos governamentais.
- Comparar as descrições encontradas com constructos de mensuração de Qualidade de dados abertos governamentais e de TPA.

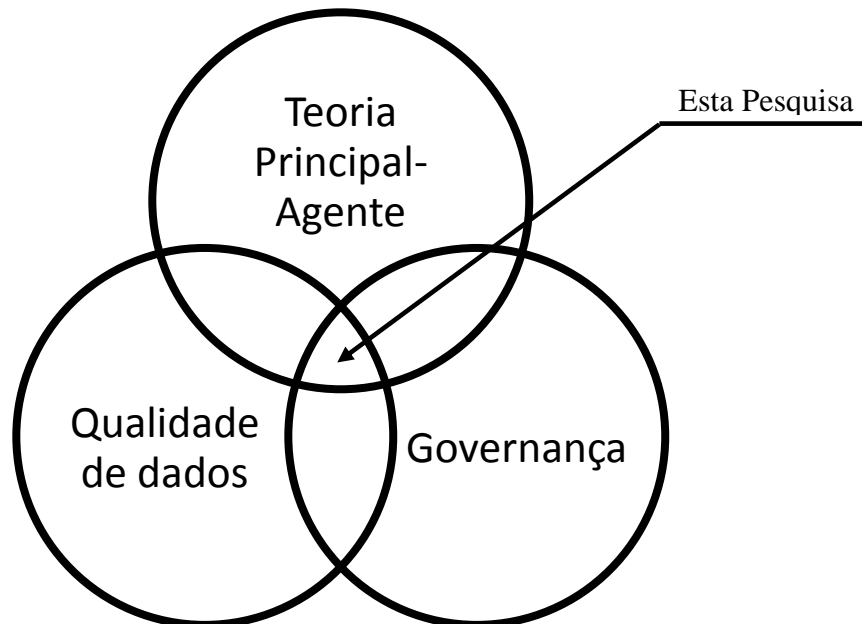
2. REFERENCIAL TEÓRICO

O desenvolvimento das TICs, nos últimos anos, vem trazendo novas possibilidades para participação cidadã, transparência e controle social das ações governamentais.

A confiabilidade e a segurança na Qualidade de Dados Abertos Governamentais (DAG) são utilizadas para fundamentar formas de avaliar a boa governança pública, que emprega esforços para mitigar um problema da relação Principal-Agente (MONTANHOLI, 2017)

O foco teórico desta pesquisa é conectar a Teoria Principal-Agente (TPA) com a Qualidade de Dados Abertos Governamentais (DAG), por intermédio do conceito de Governança, como mostra a figura abaixo:

Figura 1 – Representação Gráfica da Conexão das Ideias do Referencial Teórico.



Fonte: Autor.

Para explicar mais sobre essa ligação entre as ideias no referencial teórico, se decidiu seguir um método específico composto de 4 partes. As 3 primeiras delas seriam a explicação das áreas que dão base para a pesquisa: Qualidade de dados, Governança e TPA. Cada uma seguindo uma metodologia parecida, com, primeiramente, um contexto histórico da ideia, seguido de uma explicação do contexto atual de cada área, com definições e outras informações pertinentes ao trabalho. Essas partes finalizam com um resumo ligando as áreas umas às outras, principalmente à área que a procede. A 4ª e última parte fazem uma síntese,

remetendo à relação entre TPA e Qualidade de dados que se pretende nesta pesquisa, aos objetivos e justificativas e aos métodos.

2.1. Teoria Principal-Agente.

Alguns dos primeiros a descrever, academicamente, a TPA foram Berle e Means (1932), os quais definiram dois grupos distintos: O primeiro, formado pelos proprietários ou acionistas (o que viria a ser chamado Principal), que delega o controle de uma organização mais complexa ao segundo grupo, o de gestores, legalmente nomeados para tal e remunerados por essa responsabilidade (o que viria a ser chamado Agente). Esses gestores não tinham a propriedade do negócio, mas ainda assim o dirigiam, acumulando conhecimento que o Principal não detinha e com base em confiança. Agentes então eram obrigados a prestar contas aos proprietários para garantir a confiança e a continuação da relação.

Em 1973, Ross (1973) transformou a teoria em um problema econômico, o chamado “Problema Principal-Agente”. O texto de Ross se debruça sobre o problema de incentivo e restrições que acontece entre o proprietário (O Principal) e o gestor (O Agente). Para garantir o alinhamento do comportamento do Agente para com as necessidades ótimas do gestor, um incentivo seria necessário. Nesse contexto, Ross tenta formular uma forma de calcular qual seria essa compensação. Como base, ele utiliza as teorias sobre racionalidade limitada e parte do princípio de que indivíduos tendem a agir por interesse próprio, com liberdade de relação – algo baseado na teoria econômica neoliberal, como sugere Friedman (1968).

Já Mitnick (1973) descreve, cataloga e nomeia a relação Agente e Principal posta em prática na administração pública, assim como delimita os itens para a teoria da agência, seu funcionamento, as premissas, seus problemas e as contradições que acontecem nesse comportamento, basicamente descrevendo como instituições lidavam para modificar o comportamento do Agente para algo mais desejável ao Principal, não só seus incentivos mas todo o funcionamento institucional. Um dos pontos a notar foi a existência de uma assimetria de informações entre Agente e Proprietário, pois, uma vez que o Agente estava mais próximo da atividade, este teria mais informações. Mitnick (1976) volta, em outro momento, a falar do servidor e o serviço públicos como agentes do povo numa democracia.

Num outro momento, Mitnick (2006) chama a teoria de sua obra de 1973 de “Teoria institucional da agência” enquanto a de Ross de “Teoria econômica da agência”. É interessante notar as similaridades das teorias. Embora ambas tenham sido construídas de forma independente e concomitante, parecem complementar uma à outra. Mitnick e Ross não

se conheciam até 1974, porém ambos publicaram suas obras no mesmo ano e cada um baseado em áreas diferentes.

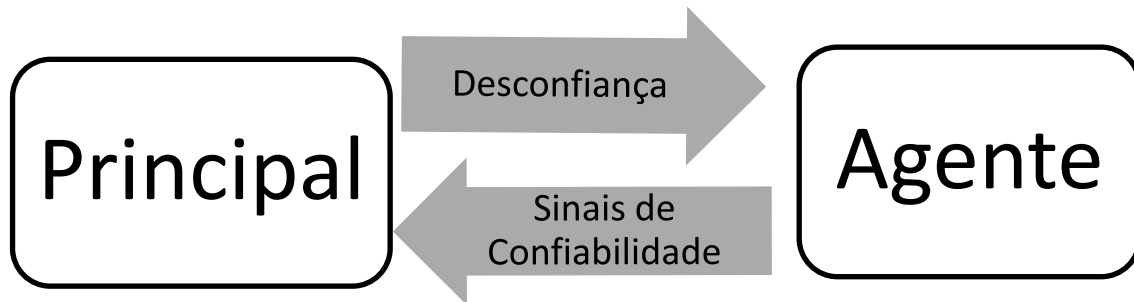
O texto mais relevante, academicamente, sobre teoria da agência, segundo o número de citações no Google Scholar (2019), contudo, é o texto de Jensen e Meckling (1976). Nele, os autores explicam a Teoria da Firma, e, entre outros assuntos, o comportamento de gestão, custo de agência e estrutura de propriedade. O texto de Jensen e Meckling se tornou tão academicamente relevante que viria a ser republicado pela Revista de Administração de Empresas, em versão portuguesa, na sua coleção de Clássicos da Administração (JENSEN, MECKLING, 2008), ofuscando trabalhos anteriores como o de Mitnick (1973) e Ross (1973).

O autor Fama (1980) complementa a teoria ao retratar que, mesmo que a organização tenha controles institucionais, o Agente – no caso gerentes comerciais – não sofrem pressão apenas de dentro da organização através de expectativas, recursos e controles mas também interage com o que há externamente à organização, o que pode influenciar suas ações para o que houver de melhor para a organização da qual faz parte.

Outro ponto de vista ainda da relação Principal e Agente é o que remete às relações contratuais, como Williamson (1985) descreve. Nessa perspectiva, incentivos e controles são fundamentais na TPA de maneira não econômica. Tais incentivos e controles são decorrentes de um modelo de gestão organizacional que estabelece a forma como a organização vai ser administrada, a partir de prêmios ou punições a seus gestores em um processo de gestão que, fundamentalmente, retrata quais os resultados são esperados e como devem ser obtidos, pressupondo a existência de uma área de controladoria que atue com neutralidade no monitoramento desse processo.

Em outra vertente, para Milgron e Roberts (1992), a assimetria de informação na TPA caracteriza-se como uma situação na qual uma das partes da transação não possui toda a informação necessária para averiguar se os termos do contrato são mutuamente aceitáveis, e se serão implementados, nem como avaliar a performance individual dos envolvidos. Entende-se que há uma relação desigual na TPA, governada por interesse próprio, de caráter racional (INGRAMS, 2018). O Principal desconfia do Agente, porém disponibiliza recursos, incentivos monetários e não monetários ao Agente, e este retribui na forma de resultados de seu serviço e dados estabelecidos com o gestor, ou seja, devolve o trabalho e sinais de confiabilidade. Essa situação é delineada na figura 2 a seguir.

Figura 2 – Representação Gráfica da Relação Principal-Agente



Fonte: Autor.

Quando se fala da TPA aplicada a um governo democrático, Pereira (2017) diz que a teoria estabelece uma interdependência sistêmica de três espécies de relação Principal-Agente, em respectiva ordem: entre cidadão e servidor público; entre políticos eleitos e servidor público e; entre cidadãos e políticos eleitos. Para questões desta pesquisa, se versará sobre a primeira relação: O cidadão como Principal e o servidor público e seus órgãos como o Agente do cidadão.

Sabe-se que a relação entre Principal e Agente, pelos meios de TI, nem sempre vão bem. Como diz Benaroch (2017), às vezes, os interesses divergentes afetam a efetividade da relação do ponto de vista do Principal. Xue, Yang e Yao (2018) informam que há desalinhamento de esforços de TI com o objetivo do Principal em relação à eficiência desses esforços nas empresas.

A Teoria Principal-Agente também ajuda, porém, a investigar intenção e utilidade entre cidadão e administração pública. As seguintes pesquisas expressam isso: Choi (2018); Bhattacharjee (1998); Bøe e Gulbrandsen (2016); Bøe, Gulbrandsen e Sørebo (2015); Bøe (2016); Karanja (2017); Coser (2017); Park e Jayakar (2013); Ingrams (2018); Butter, Liu e Tan (2012); Guillamon *et al.* (2017); Peng *et al.* (2016); e Dawson *et al.* (2017).

Um pequeno resumo dos achados dessas pesquisas será dado a seguir.

Choi (2018) analisou, com auxílio da TPA, os problemas que nascem do pedido de abertura de dados por meio de TICs. Nesse texto, o maior empecilho foi que agentes públicos tendem a se omitir da culpa pela falta de transparência ao invés de resolvê-la. Choi explica que este e outros comportamentos contra a transparência eram de se esperar pela TPA e como isso afeta (aumentando) a desconfiança do cidadão.

Bhattacharjee (1998) destaca-se por apresentar uma das poucas pesquisas experimentais em laboratório que demonstram a linha causal entre os princípios da Teoria Principal-Agente e o uso de TICs.

Em suas pesquisas, Bøe e Gulbrandsen (2016) e Bøe, Gulbrandsen e Sørebo (2015) demonstram como um Framework já utilizado no uso continuado de sistema de informação tem a ganhar com os princípios de divergência de objetivos e aversão ao risco da Teoria Principal-Agente. Eles também defendem que levar em consideração as intenções de ambos é importante para TICs, sendo o alinhamento de objetivos, ao longo do tempo, um desafio constante.

Bøe (2016) retrata como a TPA explica e também demonstra a correlação entre a competência percebida de uso de uma TIC, a harmonia de objetivos, a utilidade percebida e efetividade de uso. Ele também faz menção à necessidade da qualidade da comunicação entre Principal e Agente.

Karanja (2017), por outro lado, fala de como desalinhamentos tendem à perda de efetividade de uso de TICs em empresas, pois, como diz a Teoria Principal-Agente, os Agentes tendem a se desprender do trabalho ótimo do ponto de vista do Principal, por questões de objetivos desalinhados. No caso, os CIOs (*Chiefs of Information Officers*) focavam-se mais em questões de fraternidade com fornecedores ao invés de priorizarem o objetivo organizacional, custos baixos e homogeneidade em TICs.

Em seu trabalho, Tiago Coser (2017) fala de como Governança de Tecnologia de Informação (GTI) contribui para diminuir problemas na relação Principal-Agente. Ele escreve que as principais contribuições do modelo de GTI e, por consequência, da Teoria Principal-Agente, são o alinhamento de objetivos, a capacidade de monitoramento por Governança e padronização dos processos.

O alinhamento entre os esforços do Programa de Oportunidades em Tecnologia de Banda-larga (*Broadband Technology Opportunities Program*), com seus objetivos, foi avaliado por Park e Jayakar (2013), também à luz da TPA. Essa avaliação mostrou que, em muito, o planejamento público carecia de alinhamento com o objetivo original do BTOP e que órgãos governamentais, apesar de terem a maioria e as maiores concessões para centros públicos de computadores, não comprovaram ter relatórios de fundos melhores ou completaram os objetivos mais rapidamente.

Butter, Liu e Tan (2012) demonstraram que a assimetria de informação entre o Governo e as empresas, natural relação entre Principal-Agente, causava ineficiência para ambos, uma vez que o Governo não sabia o que deveria exigir para melhor controle nem as

empresas que trabalham com essa instituição sabiam as regulações e questões públicas o suficiente para serem competentes. Os autores recomendam, então, o que chamam de Certificação de Operador Econômico Autorizado e um sistema baseado em confiança, como formas mais eficientes para alcançar objetivos em TICs, ao invés de um conjunto extenso de controles.

Os esforços de transparência e prestação de contas por meio de mídias sociais (a saber, Facebook), entre cidadãos e governo, foram analisados por Guillamon *et al.* (2016) na perspectiva da Teoria Principal-Agente. Segundo eles, este caminho se tornou mais efetivo aos objetivos originais de tais esforços. Algo próximo ao que se quer fazer nesta pesquisa.

Peng *et al.* (2016) adicionam ainda que a teoria Principal-Agente também pode ser usada em empresas públicas, mesmo não sendo subordinadas somente ao cidadão, e que, por vezes, algumas empresas públicas são mais eficientes e melhores que os próprios governos.

Uma visão um tanto diferente da Governança no Setor Público e na relação de Principal e Agente foi dada por Dawson *et al.* (2017). Nesse texto, na relação entre órgãos públicos (Principais), que controlariam os departamentos de TI (Agentes), a Governança sairia de um papel controlador para um papel mediador. O departamento de TI usaria a Governança mais como uma forma de alinhar os objetivos de ambas as partes. Ele adiciona que a esse recurso (se assim podemos dizer) dá o nome de Teoria da Agência Legal, em que organizações inteiras, e não apenas grupos, podem ser Principal e Agente; também pode ser usada para explicar o comportamento público.

A literatura relata que o uso de TPA em TICs é interessante para alinhar objetivos ou seja, alinhar a intenção do Principal àquela dada pelo Agente. Sob tal prisma, as passagens transcritas buscam definir a TPA: “O modelo principal-agente é uma ferramenta econômica genérica que pode ser aplicada [...] potencialmente a [outros] problemas caracterizados por incongruência de objetivos, aversão ao risco e assimetria de informação” (BHATTACHERJEE, 1998); “o Framework de replicação foi ampliado utilizando a TPA. Tais estudos têm sido solicitados na literatura e os resultados do nosso estudo mostram a relevância desta perspectiva integrativa na intenção de continuidade de sistemas de informação.” (BØE; GULBRANDSEN, 2016); “ Este estudo empregou apenas uma possível 'lente' teórica da TPA, no entanto, existem outras alternativas teóricas para a modelagem dos constructos derivados da TPA. ” (BØE; GULBRANDSEN; SØREBØ, 2015); “...aqui também pode ter argumentos teóricos para incluir TPA em constructos derivados [de TICs]” (BØE, 2016); e “Recomenda-se, para novas pesquisas, que se contemplem fundamentos do problema

de agência não tratados nesta pesquisa, como aversão ao risco e conflito de interesse.” (COSER, 2017).

Resumindo, há várias ferramentas que o Principal usa para eliminar, ou reduzir, o problema na relação com seu Agente. Mitnick (1973) fala que há mecanismos institucionais para garantir a transparência, a prestação de contas e métodos de modificar a ação do Agente, ou seja, Métodos de Governança Pública, segundo o Banco Mundial (1992, 2017). No caso da Administração Pública Federal Brasileira, o Ministério da Transparência, Fiscalização e Controladoria Geral da União (CGU) seria o órgão para garantir que isso aconteça (CGU, 2016), porém o próprio Ministério alerta que o cidadão é parte integrante do processo de instauração de confiança entre Principal e Agente.

Enfim, como coloca Montanholi (2017), uma das ferramentas para mitigar problemas na relação Principal-Agente, na administração pública é a Governança. Uma vez que o Principal na democracia é o cidadão, a Governança se torna a ferramenta para controlar os Agentes delegados de autoridade, caso tomem decisões desalinhadas do interesse do Principal.

2.2. Governança e Boa Governança

A qualidade de dados e a TPA convergem quando o assunto é boa governança, por isto, é necessário entender e discutir o que é a Governança, para conectar essas duas áreas solidamente.

Primeiro, há de se separar governança de governar, mesmo que essas palavras tenham nascido da mesma origem, “κυβερνάω”, conforme registro no livro VI de “A República”, por volta de 380 antes de Cristo (PLATÃO, 1987, p. 273). Traduzida literalmente, essa palavra significa “comandar ou guiar um navio”, porém, como, na mesma passagem, governança é retratado como “poder de comandar” e “atos que evitem conflito de interesse em partes”, seu sentido ainda se confundia com governar. Nesse texto que também apresenta a ligação entre o que viria a ser chamado governança e democracia, Platão critica o sistema democrático exatamente por exigir grande quantidade de garantias de que o governante saiba o que faz.

Governança, porém, recebeu um significado diferente de governar, no manuscrito de Sir John Fortescue, aproximadamente em 1474 (FORTESCUE-ALLAND, 1714). Nesse manuscrito, governança significa um esforço organizacional, não somente de comandar mas também de, expressamente, evitar conflitos de interesse, com trocas entre comandante e comandados. No texto, o autor fala da necessidade de restringir o poder monárquico, após

vários exemplos de abuso. Para evitar tais exemplos, Sir Fortescue exigia prestação de contas, a ética e a legalidade por meio da então “nova” monarquia constitucionalista.

Em 1922, segundo Morck e Steier (2005), devido a situações de abuso de poder, parecidas formas de boa governança aconteciam também fora da esfera pública – nascia a governança corporativa. Naquele ano, um caso judicial caía sobre os gestores da Companhia Holandesa das Índias Orientais que enriqueceram em detrimento dos acionistas e em benefício próprio ou de próximos. Na ocasião, os acionistas, como Principal da relação, exigiam formas de averiguar o uso de seus investimentos, criando o que seria chamado de governança corporativa. Desse caso resultaram várias cláusulas de governança, implantadas tanto na companhia quanto na bolsa de valores, para evitar futuros problemas. Esse comportamento de abuso de poder e seguinte exigência de governança foi visto em várias partes do mundo, tanto em organizações privadas como em instituições públicas, o que traz definição atual de governança:

A governança refere-se, portanto, a todos os processos de governar, seja por um governo, mercado ou rede, seja sobre uma família, tribo, organização formal ou informal, ou território, através de leis, normas, poder ou idioma. A governança difere do governar na medida em que se concentra menos no estado e em suas instituições e mais em práticas e atividades sociais (BEVIR, 2012, p. 06).

E, para definir a melhor forma de ter governança, houve várias publicações que diziam como esses processos podiam melhorar. Um desses esforços foi o livretinho *Governance and Development*, do Banco Mundial (1992). Nele, governança é a forma em que o poder é exercido tanto em instituições públicas quanto privadas, para aumentar a solidez das relações entre atores também. O Banco Mundial estudou como governos e economias falharam e chega à definição de um método que beneficiaria os envolvidos, chamado de *good governance* ou “boa governança”. Este conceito é composto pelos seguintes esforços dos atores: Prestação de contas (chamada de *Accountability*, que representa o processo de contínua demonstração, por parte do Agente, de que sua gestão está alinhada às diretrizes previamente fixadas pelo Principal); Estrutura legal/regulatória sólida, voltada à resolução de conflitos entre partes e; Transparência de informações entre atores. Esse artigo retira dessa definição “esforços de governança” como sendo referenciado a estes três esforços.

O processo de melhoramento de governança considerado pelo Banco Mundial é uma das ferramentas à disposição do Principal, do ponto de vista da Teoria Principal-Agente, em caso de desvio da conduta deste último, pois permite solver um dos maiores problemas da relação, a assimetria de informação (SILVEIRA, 2015), principalmente em formas de

governo mais democráticas (BANCO MUNDIAL, 1992; MORCK; STEIER, 2005). Na democracia de fato, o cidadão é que é o Principal e a administração pública seu Agente. A ferramenta do cidadão, como Principal, para extinguir conflitos com o governo é, portanto, o uso desses processos de Governança, explicados pelo Banco Mundial (MONTANHOLI, 2017). Em resumo, a Governança mostra as informações necessárias ao Principal, funcionando também como o meio de alerta para que o Agente seja responsável pelos seus atos.

Os princípios da Boa Governança, como regulações efetivas, prestação de contas e transparência, seriam necessários para se ter uma governança próspera, que reduziria corrupção e desvios de conduta de todos os atores – esse efeito não se restringindo a governos democráticos apenas (BANCO MUNDIAL, 1992, 2017, IBGC, 2018). Esforços de liberação de dados abertos governamentais também seriam ligados a essa prosperidade (MACHADO FILHO, 2011).

Segundo Elbahnasawy (2014), há uma forte relação negativa entre dados abertos e serviços governamentais, vindos da internet (conhecido como *e-government*), e corrupção, oriundos da Teoria Principal-Agente. Já Ingrams (2018) corrobora a ideia da relação de governança com democracia; mais especificamente, ele correlaciona leis de acesso à informação com países em progresso democrático, a saber, com pouca competição política, sociedade civil forte, pressão regional e aceitação de tratados internacionais.

A administração pública brasileira, como mostra o Plano Diretor da Reforma do Estado (BRASIL, 1995), poderia utilizar tais princípios para melhorar o serviço público e evitar conflitos e assimetrias entre Principal e Agente em seu sistema.

Isso é refletido nos primeiros quatro princípios da administração pública: “legalidade, impessoalidade, moralidade, publicidade” (BRASIL, 1988), no caput do artigo 37 da Constituição, e se assemelham aos princípios da governança corporativa do Instituto Brasileiro de Governança Corporativa de “prestação de contas (*Accountability*), equidade, ética e transparência” (IBGC, 2018).

Governança também é expressa na Constituição Federal, como expresso no art. 5º, inciso XXXIII: “todos têm direito a receber dos órgãos públicos informações de seu interesse particular, ou de interesse coletivo ou geral”; também no inciso XXXIV, referente a “a obtenção de certidões em repartições públicas, para defesa de direitos e esclarecimento de situações de interesse pessoal”; e, por fim, no inciso LXXII em que se concede *habeas-data* “para assegurar o conhecimento de informações relativas à pessoa do impetrante, constantes de registros ou bancos de dados de entidades governamentais ou de caráter público”. Tudo

isso com restrições à intimidade e ao interesse social, tal como estabelecido no inciso LX, assim expresso: “a lei só poderá restringir a publicidade dos atos processuais quando a defesa da intimidade ou o interesse social o exigirem” (BRASIL, 1988).

Outras leis relacionadas à boa governança seriam depois adicionadas. Essas leis usam TICs como o meio pelo qual a boa governança seria expressa (Exemplo: O portal da transparência). Trata-se da Lei nº 12.527 de 2011, conhecida como a Lei de Acesso à Informação – LAI (BRASIL, 2011); Lei Complementar nº 131 de 2009, conhecida como a lei Capiberibe (BRASIL, 2009); e Lei Complementar de nº 101 de 2000 (BRASIL, 2000), conhecida como a Lei de Responsabilidade Fiscal.

Na LAI (BRASIL, 2011), a transparência ativa é dever da administração pública, logo esta deve divulgar, de forma espontânea, informações de interesse público, produzidas ou mantidas por órgãos e entidades públicas, em local físico de fácil acesso e via internet, principalmente pelo www.portaltransparencia.gov.br. O Artigo 8º da LAI, no qual se baseia a avaliação da Transparência Ativa, afirma que é “dever dos órgãos e entidades públicas promover, independente de requerimentos, a divulgação em local de fácil acesso, no âmbito de suas competências, de informações de interesse coletivo ou geral por eles produzidas e custodiadas”.

O Decreto nº 7.724, de 16 de maio de 2012, regulamentou a Lei no 12.527 (Lei de Acesso à Informação), de 18 de novembro de 2011, que dispõe sobre o acesso a informações previsto no inciso XXXIII do caput do Art. 5º, no inciso II do § 3º do Art. 37 e no § 2º do Art. 216 da Constituição. Nela, há a transparência ativa, ou seja, a transparência que se dá sem requisição direta do cidadão, e a transparência passiva, que se dá sob esse tipo de requisição.

O Ministério da Transparência, Fiscalização e a Controladoria Geral da União, segundo a CGU (2016), são as principais ferramentas institucionais que a administração pública usa para apurar os princípios de Governança. O Ministério da Transparência e Controladoria-Geral da União (CGU) está estruturado em quatro unidades finalísticas: Secretaria de Transparência e Prevenção da Corrupção; Secretaria Federal de Controle Interno; Corregedoria-Geral da União; e Ouvidoria-Geral da União. Segundo a CGU, a qualidade de dados abertos governamentais ainda pode melhorar. Assumindo seu papel de fiscalizador e controlador, a própria CGU relata que, “em 2016, avaliou detalhadamente 18 órgãos, enviando relatório em que se cobrava (sic) ajustes específicos nas respostas” (p. 33), sendo exigências externas como denúncias e representações uma ferramenta essencial para seu funcionamento. Conforme relata o TCU (2014), “No Brasil, o crescente interesse pelo tema não é diferente. Tanto no setor privado quanto no público, existem iniciativas de

melhoria da governança”. Mesmo assim, o cenário atual revela que há grandes falhas na governança brasileira que possibilitam a ocorrência de fraudes, desvios e perdas financeiras e de reputação em organizações governamentais em detrimento do Principal (CGU 2016; FABRINI, 2018). Um grande problema enfrentado no Brasil é quanto à falta de Transparência Municipal, com muitas informações não só erradas, ou tardias, como também incompletas. Isso é evidenciado na literatura acadêmica, nas pesquisas de Cruz *et al.* (2012), sobre municípios do estado de São Paulo; Vieira e Kronbauer (2017), nos municípios de Rio Grande do Sul com mais de 50 mil habitantes; Nobre (2017) e Lacerda (2017), sobre os municípios da Paraíba; e Vale (2017), nos municípios do Maranhão.

O mesmo pode ser dito de Órgãos e Repartições Públicas maiores, como dizem Gomes *et al.* (2018). Segundo eles, há dificuldades de transparência para uma grande parte do governo. Almada (2017) avaliou e comparou, por meio da internet, a transparência, dos governos brasileiro, americano e britânico. Cruz e Afonso (2018) também acharam muitas falhas. Como nota, é interessante falar da repetida correlação entre receita e transparência, o que pode ser um dos fatores em questão para a falta de dados, mas que não é desculpa para a omissão do Agente, segundo a lei (Brasil, 2011).

Essas constantes falhas de transparência e de sua qualidade afetam a forma como o cidadão brasileiro reage à administração pública. Como diz Antunes (2018, p. 175), as falhas nos portais de dados abertos ao público “não fomentam o interesse do cidadão em participar do processo decisório governamental, sendo essa a intenção central do processo informacional”.

Mesmo assim, não é como se falta de boa governança fosse um problema exclusivamente brasileiro. Choi (2018) afirma que conseguir informações através de mecanismos institucionais, na China, por exemplo, é difícil, principalmente por causa de burocratas que tentam se evadir da culpa de falta de informações em detrimento do cidadão.

Órgãos que não seguem esses esforços de governança são tão comuns que Nahon, Peled e Shkabatur (2015) criaram uma metodologia para classificar órgãos americanos como aqueles que seguem as diretrizes de dados abertos governamentais e aqueles que não o fazem. Luna-Reyes, Bertot e Mellouli (2014), em suas pesquisas, dizem que a literatura em dados abertos governamentais passa de esforços de presença de dados para dados de qualidade mais amigáveis para programas e, assim, busca-se melhorar tanto a tomada de decisão privada como manter “direitos” de controle da administração pública pelos cidadãos.

Wu *et al.* (2015) relacionaram estratégias de governança de TI com desempenho organizacional e alinhamento estratégico, mostrando uma correlação positiva para as

empresas; tal procedimento explicou como estes esforços diminuíram o receio dos Principais de investir na empresa e, portanto, aumentaram seu rendimento e confiança. Nessa pesquisa, os autores até desdobram seus efeitos na máquina pública.

Para que problemas de governança pública não ocorram, as pesquisas em Dados Abertos Governamentais (*Open Government Data*) e Governo eletrônico (*e-Government*) têm se focado em garantir não só a presença como a qualidade de dados abertos governamentais, para que estes possam ser utilizados em suas funções originais descritas na Teoria Principal-Agente (ALENEZI, TARHINI, MASA'DEH, 2015; GONZALEZ-ZAPATA; HEEKS, 2015). Também incluído aí o Brasil (BRAGA *et al.*, 2008; MONTANHOLI, 2017) – Dados Abertos Governamentais, aqui definidos como dados abertos a qualquer um e que podem ser manipulados ou reaproveitados para outras áreas (UMBRICH; NEUMAIER; POLLERES, 2015; CLARKE; MARGETTS, 2014) e Governo Eletrônico definido por uso da internet e de Web, para disponibilizar informação e serviços aos cidadãos, de forma aberta fácil e rápida (ONU, 2002; 2018)

Esse percurso nos leva à *Data Literacy*. Este conceito é definido como a capacidade de “acessar, entender, avaliar, manipular, resumir e comunicar” (SCHIELD,2004) que está ligada não só à capacidade do publicador e produtor de dados de disponibilizar dados de qualidade mas também da capacidade do consumidor de dados de tirar o de que necessita do que lhe foi disponibilizado, sendo amplamente reconhecidas como essenciais para o uso de dados como vetor de informações segundo a *International Association for Social Science Information Services and Technology (IASSIST)* e a *Association of Public Data Users (APDU)*, ambas organizações que lidam exatamente com essa relação (SCHIELD,2004; KHAN; KIM; CHANG, 2018; GRAY; GERLITZ; BOUNEGRU, 2018).

Em resumo, tanto aquele que publica e produz os dados quanto aqueles que os consomem necessitam saber como usar dados como vetor de informação, caso contrário o valor da informação é desperdiçado e pode levar ao erro (SCHIELD,2004; KHAN; KIM; CHANG, 2018; GRAY; GERLITZ; BOUNEGRU, 2018). Dados valem pouco como governança se a informação não é transmitida ou se o é com pouca qualidade. É essencial, portanto, que o cidadão, como Principal, tenha o mínimo de *Data Literacy* para poder entender os dados, assim como a administração detenha quantidades suficientes de *Data Literacy* para passar informações e poder reduzir conflito e desconfiança na relação Principal-Agente com maior facilidade. A administração pública, então, tem o papel de ensinar o cidadão o suficiente para poder obter essa informação no sistema educacional e, também, em bancos de dados, caso a dúvida cause dificuldade ao consumidor de dados de obter

informações, visto que essa dificuldade pode ser um problema para a relação Principal-Agente.

É com base neste assunto que se chega ao último assunto da pesquisa: a Qualidade de dados. Para serem efetivos em informar o cidadão, é necessário que os dados sejam de qualidade, isto é, bons geradores de informação, o que nem sempre ocorre no Brasil, como dizem Silva e Miranda (2017) e Nazário, Silva e Rover (2012). É por este aspecto que esta pesquisa se interessa.

2.3. Qualidade dos Dados

Antes de falar o que é qualidade de dados, há primeiro de se definir o que são “dados”, ou, o mais importante, no que “dados” difere de “informação”. Vulgarmente essas palavras, às vezes, são usadas como tendo o mesmo significado (VIGO, 2013).

“Dados” tem várias definições, desde “símbolos” (BELLINER; CASTRO; MILLS, 2004); “série de símbolos”(VIGO, 2013); “representação de fatos, conceitos ou instruções, expressos de maneira formalizada, passível de comunicação, interpretação ou processamento por meios humanos ou automáticos” (NBR-11515, 2007) a “resultado da transformação de signos (letras, números e outros signos), por meio de regras de ordenação de código ou sintaxe; os dados são símbolos não interpretados.” (SANTOS, 2016). Em resumo, dados são não subjetivos, tendendo a lidar diretamente com as representações da realidade.

Informação, em outra vertente, é ainda mais difícil de definir, sendo definida como: “Dados” vulgarmente (VIGO, 2013); “Dados que receberam significado por meio de conexão relacional” (BELLINER; CASTRO; MILLS, 2004); “Dados processados” (ACKOFF, 1989), “Dados organizados por padrões explicativos” (SANTOS, 2016); e até como “um ou mais dados bem formados com significado” (FLORIDI, 2009). O conceito de informação é, portanto, mais abstrato e “mais subjetivo em natureza” (VIGO, 2013) já que lida com a capacidade do sujeito de conseguir retirar significado subjetivo.

Essa definição traz o porquê da pesquisa se focar na qualidade de dados e não em qualidade de informação. Primeiramente, o termo “dados” tem um conceito mais objetivo, portanto, mais fácil de trabalhar, pois foca as representações da realidade. Segundo, Dados Abertos Governamentais são isso mesmo, dados. Mesmo se esses dados fossem processados (Exemplo: média), satisfazendo à definição de informação de Ackoff (1989), essa “informação” de Dados Abertos Governamentais chega ao consumidor através de simbologia, ou seja, chega ao Principal simplesmente como “dados”. Esta pesquisa está mais interessada

na qualidade dos dados como veículo ou plataforma para criação de informações sob a égide da TPA do que somente na qualidade das informações obtidas a partir dos dados. Por esse motivo é que se fala de qualidade de dados em questão de Função, Uso e Intenção, e não somente Qualidade de Informação.

Finalizada esta exposição, há de se explicar Qualidade de dados em si. Qualidade de dados é antiga, na verdade; quase tão antiga quanto a Estatística. Segundo Rubin (1971), um dos primeiros a fazer menção à qualidade de dados foi Tucídides, o Filósofo Atenense, no século IV antes de Cristo, pois ele, duvidando dos dados que conseguia sobre a quantidade de tijolos em um muro, utilizou um método para garantir essa qualidade, simples repetição de mensuração. A teoria sobre Qualidade de dados em Banco de dados completamente eletrônicos começa em Ivanov (1972). Em seu trabalho, o autor relaciona Precisão com a Qualidade de dados e sugere Qualidade de dados como o grau em que o formato dos dados corresponde a certas características de precisão e padronização. O grande marco moderno da qualidade de dados, contudo, é a pesquisa de Wang e Strong (1996). Nela, Qualidade de dados vai além de precisão e padronização. Qualidade de dados, para eles, deve permitir que os dados, além de apresentar precisão e padronização, sejam também melhores portadores de informação, ou seja, nesta definição, qualidade de dados é qualidade em formato e função.

Hoje, qualidade de dados é definida em duas grandes vertentes. A primeira vertente é mais restrita e objetiva, por focar a condição e forma dos dados, conforme discutido nas pesquisas de Francisco *et al.* (2017), Behkamal *et al.* (2014) e Vertrò *et al.* (2016). Tal vertente vem diretamente das teorias de Ivanov (1972) que vê Qualidade como dados precisos e que refletem a realidade.

A segunda vertente é mais “ampla e subjetiva”, segundo Cai e Zhu (2015). Ela define qualidade não só quanto à forma em que os dados se apresentam mas também quanto à sua utilidade e intenção. Como dizem Cai e Zhu (2015), Wang, Ziad e Lee (2006) e Wang e Strong (1996), qualidade de dados, nesse contexto, é “o grau em que um conjunto de dados está adequado”, tanto ao que Wang e Strong (1996) chamam de “consumidor de dados”, isto é, aquele que vai usar os dados, quanto às intenções daquele que os colocou em exposição em primeira instância. Esse fato influi pertinentemente nos Esforços de Governança (YILDIZ, 2007).

Qualidade de dados ganhou força no mundo acadêmico por sua ligação a processos de TICs, principalmente os que contribuem para tomada de decisão baseada em informações, como *Big Data* (AGARWAL; DHAR, 2014), *Business Intelligence & Analytics* (CHEN *et al.*, 2012), *Data Mining* (CAI; ZHU, 2015) e *Data Analytics* (HAZEN *et al.*, 2014), ferramentas

que podem e são usadas para analisar o Governo (ALENEZI *et al.*, 2015), incluindo órgãos do próprio Governo (CGU, 2016). Esses processos são importantes hoje e se tornarão cada vez mais necessários, uma vez que, no futuro, processos de TI, como os acima citados, se tornarão parte indispensável da tomada de decisão (DAVENPORT; PATIL, 2012).

Para facilidade de entendimento, uma tabela foi feita expondo os conceitos aqui tratados.

Quadro 1 – Definições de Ferramentas para a Tomada de Decisão

Termo	Definição
<i>Big data</i>	Conjunto de dados massivos, rápidos e complexos demais para sistemas tradicionais de análise de dados, sendo por vezes confundido com os processos usados para analisar estes conjuntos de dados (AGARWAL; DHAR, 2014; PROVOST; FAWCETT, 2013)
<i>Business Intelligence & Analytics</i>	Área maior que engloba as formas de análise e obtenção de informação e conhecimento, usada por negócios, para tomada de decisão baseada em informações. (CHEN <i>ET AL.</i> , 2012)
<i>Data analytics</i>	Processos e técnicas que quando aplicados a dados brutos os transforma em informações úteis (HAZEN <i>ET AL.</i> , 2014; CHEN <i>ET AL.</i> , 2012; PROVOST; FAWCETT, 2013)
<i>Data Mining</i>	Processo de obtenção, compreensão, e avaliação de dados para implementação em tomada de decisão ou atividade (CAI; ZHU, 2015).

Fonte: Autor.

Definir um conceito não é, entretanto, o bastante; também há de se fazer sua mensuração para estudá-lo, o que não é diferente para a Qualidade de dados. No intuito de se proceder à mensuração de Qualidade de dados, várias metodologias foram feitas, contudo a maior parte se preocupa mais com a forma ou formato que os dados apresentam do que com sua intenção ou utilidade, como registram Vetrò *et al.* (2016) ou Zhu e Gauch (2000), que usam o formato dos dados para definirem a Qualidade dos dados. Ao tentar se manter em “qualidades intrínsecas” (VETRÒ *et al.*, 2016) e “objetividade” (ZHU; GAUCH, 2000), apresentam-se várias metodologias de mensuração de qualidade de dados, porém não se foca na função dos dados de passar informação, o que convergiria a duas outras funções que se relacionam à TPA: a função para o consumidor dos dados, ou seja, o uso dos dados, e a função para o publicador/autor dos dados, ou seja, a sua intenção.

Entre as metodologias de mensuração de Qualidade de dados, várias não chegam a mensurá-la em questão de função, e, mesmo quando a mensuram, o fazem superficialmente. Isto se explica, por exemplo, como a capacidade de pedir informações mais úteis, situação discutida em Ijab *et al.* (2017), ou como uma simples passagem que subentende a função dos dados, explicada por Torodan *et al.* (2015). O quadro 2 a seguir expõe várias destas

metodologias, incluindo aquelas que não levam em consideração a função dos dados para o consumidor ou publicador dos dados ou mal o fazem.

Quadro 2 – Exemplo de Literatura Sobre Métodos de Mensuração de Dados

Acrônimo	Nome	Referência
TDQM	Total Data Quality Management	Wang (1998)
DWQ	Data Warehouse Quality Methodology	Jeusfeld et al. (1998)
AIMQ	A methodology for information quality assessment	Lee et al. (2002)
DQA	Data quality assessment	Pipino et al. (2002)
IQM	Information quality measurement	Eppler e Munzenmaier (2002)
TIQM	Total Quality Information Management	English (2003)
ISTAT	ISTAT methodology	Falorsi et al (2003)
AMEQ	Activity-based Measuring and Evaluating of product	Su e Jin (2004)
COLDQ	Loshin methodology of Cost-effect of Low data quality	Loshin (2004)
DaQuinCIS	Data Quality in Cooperative Information Systems	Scannapieco et al. (2004)
QAFD	Methodology for the quality assessment of financial data	De Amicis and Batini (2004)
CIHI	Canadian Institute for Health Information methodology	Long e Seko (2005)
CDQ	Comprehensive methodology for Data Quality management	Batini and Scannapieco (2006)
UDADQ	Utility-driven assessment of data quality.	Even e Shankaranarayanan, (2007)
MFDQAI	Methodologies for Data Quality Assessment and Improvement	Batini et al. (2009)
HDQM	Heterogeneous Data Quality Methodology	Carlo et al. (2011)
LDIF	Linked Data Integration Framework	Mendes., Mühleisen, & Bizer (2012)
MEHDQA	Methods of Electronic Health Record Data Quality Assessment	Weiskoph e Weng (2013)
QALOD	Quality Assessment of Linked Open Data	Behkamal et al. (2014)
DDDC	Data Disclosure Desired Characteristics	Lourenço (2015)
QAMF	Quality Assessment Monitoring Framework	Umbrich; Neumaie e Polleres (2015)
BDQA	Big Data Quality Assessment	Cai e Zhu (2015)
MEIDIQS	Methodology to Evaluate Important Dimensions of Information Quality in Systems	Torodan et al. (2015)
FSMBDQA	Framework for social media big data quality analysis	Jaffar et al. (2015)
ODQMF	Open Data Quality Measurement Framework	Vertrò et al. (2016)
BDQAFSO	Big Data Quality Antecedents in Financial Services Organizations	Haryadi et al. (2016)
QALD	Quality assessment for linked data: A survey	Zaveri et al. (2016)
DQFMPS	Data Quality Framework for Malaysia's Public Sector Open Data Initiative	Ijab et al. (2017)
OLDQAF	Objective Linked Data Quality Assessment Framework	Assaf, Senart, & Troncy (2018)

Fonte: Autor.

Se a função dos dados é informar, metodologias de mensuração de qualidade de informação seriam, então, a resposta? Não exatamente. Métodos de mensuração de qualidade de informação como a *Total Quality Information Management*, de Francisco *et al.* (2017), focam-se na usabilidade e na intenção original das informações, ou seja, quer saber da qualidade de informações já obtidas, não exatamente da qualidade dos dados como vetor

dessas informações. A metodologia *Data Disclosure Desired Characteristics*, de Lourenço (2015), é a que mais se aproxima de um método avaliativo da qualidade de dados como vetor de informação, levando em consideração sua intenção e uso. Esse autor, no entanto, se limita ao contexto de utilidade na métrica de solicitar uma forma de ranquear os bancos de dados e ter um canal de feedback aberto para pedido de dados a mais, baseando-se no conceito ao qual ele chama de “relevância”. Isso é uma forma de atender a necessidades de TPA e faz referência à qualidade de dados relacionada a sua função de informar, embora, certamente, não a única.

Como, entretanto, mostram Batini *et al.* (2009), métodos de mensuração de qualidade de dados têm a ganhar com o uso de metodologias de mensuração de qualidade de informação. Afinal, informação é uma das características da qualidade de função dos dados e dados são o veículo de obtenção de informação.

Com relação aos instrumentos de mensuração de dados abertos do governo, os instrumentos como os de Umbrich; Neumaie e Polleres, (2015), Behkamal *et al.* (2014), Ijab *et al.* (2017) e Vetrò *et al.* (2016) são especialmente interessantes exatamente por se especializarem em tratar de dados abertos governamentais – tema deste estudo. Esses grupos de dados, como é expresso na Teoria Principal-Agente, têm uma utilidade e uma intenção esperada. Como dito antes, porém, apenas Ijab *et al.* (2017) utilizam a ideia de utilidade e intenção original, em seu framework de qualidade de dados, para avaliar bancos de dados malaios, no conceito a que ele chama de “relevância”.

Há vários pontos contingentes que também afetam a TPA em Qualidade de Dados Abertos Governamentais. Dificuldades em ligar os dados de bancos de dados abertos governamentais e achar responsáveis pela transparência, por exemplo, já são um desafio bem documentado, segundo Behkamal *et al.* (2014) e Canova *et al.* (2015). Esse desafio também afeta a relação de confiança e o comportamento entre cidadão e governo (NAHON; PELED; SHKABATUR, 2015), comprometendo outras áreas da vida do primeiro, como inovação, (ZUIDERWIJK; JANSSEN; DAVIS, 2014) e sua participação em decisões governamentais (ZUIDERWIJK, JANSSEN, 2015). Essa dificuldade de achar responsáveis e ligar os vários dados presentes permite casos como o relatado por Choi (2018) em que desvio de conduta e até completa falta de dados abertos pelo Agente é ocorrência comum e que permite Agentes se omitirem da sua responsabilidade para com o Cidadão. Esse tipo de perda de qualidade de dados em Dados Abertos Governamentais cria textos como o de Weber e Yan (2017), nos quais estes expõem um programa que dá feedback automático sobre Qualidade dos Dados Abertos Governamentais, ao menos em formato de dados. Lourenço (2015) é outro estudioso

que demonstrou que a qualidade de dados não é suficientemente satisfatória nos sites de dados abertos governamentais, nomeadamente nos sites <Data.Gov> de França, Austrália, Canadá, Reino Unido, Estados Unidos da América e Nova Zelândia.

Esta falta de qualidade também é reportada na Administração Pública Brasileira, que recentemente passou por uma crise de transparência e governança, principalmente em órgãos menores como prefeituras, onde há completa ausência de acervos de dados exigidos por lei (CRUZ *ET AL.*, 2012; PAULO; VIEIRA; KRONBAUER, 2017; NOBRE, 2017; LACERDA, 2017; VALE, 2017; FABRINI, 2018; ANTUNES, 2018), ainda que, infelizmente, este problema não se restrinja a dados de órgãos menores (FABRINI, 2018; NAZÁRIO; SILVA; ROVER, 2012; SILVA; MIRANDA, 2017, LOURENÇO, 2015).

Mesmo quando estes dados estão presentes, pesquisas como a de Silva e Miranda (2017) relatam que estes têm problemas com credibilidade, integridade e erros (ALMADA, 2017), ou seja, de qualidades mais básicas. Isso afeta a usabilidade de ferramentas pelo cidadão como Principal, para atingir seus objetivos (FAGUNDES; MACEDO; FREUND, 2017), e atinge Esforços de Governança exigidos pela população (BRAGA *et al.*, 2008; MONTANHOLI, 2017). E mesmo quando os dados estão presentes, e são de qualidade, geralmente só se retrata a qualidade de dados no sentido mais restrito, tendendo a ignorar o ponto da intenção original ou o uso pelo cidadão (NAZÁRIO; SILVA; ROVER, 2012).

Vale dizer que, apesar de tudo, há esforços de Qualidade de Dados do Governo Brasileiro. O Portal Brasileiro de Dados Abertos – PBDA (2018), por exemplo, é o ambiente virtual para onde todos os dados abertos governamentais convergem (O Portal da Transparência é utilizado apenas para questões fiscais e financeiras). Nesse portal, entre outros, segundo a própria PBDA (2018), se utilizam dois recursos: a) o que se chama de Infraestrutura Nacional de Dados Abertos (INDA – 2018) e b) Infraestrutura Nacional de Dados Espaciais (INDE – 2018). Ambos apresentando um conjunto de padrões, tecnologias, procedimentos e mecanismos que seguem os Padrões de Interoperabilidade de Governo Eletrônico (e-PING), oferecendo entendimento mútuo de protocolos e arquivamento, para garantir uma “interação, dentro do governo federal e com a sociedade em geral” (PING, 2018).

2.4. TPA e Qualidade de Dados

Segundo a Teoria Principal-Agente, os dados abertos governamentais brasileiros têm uma intenção original e um uso: mostrar ao cidadão brasileiro (Principal) que o serviço prestado está indo como acordado e que se pode confiar na administração pública (Agente),

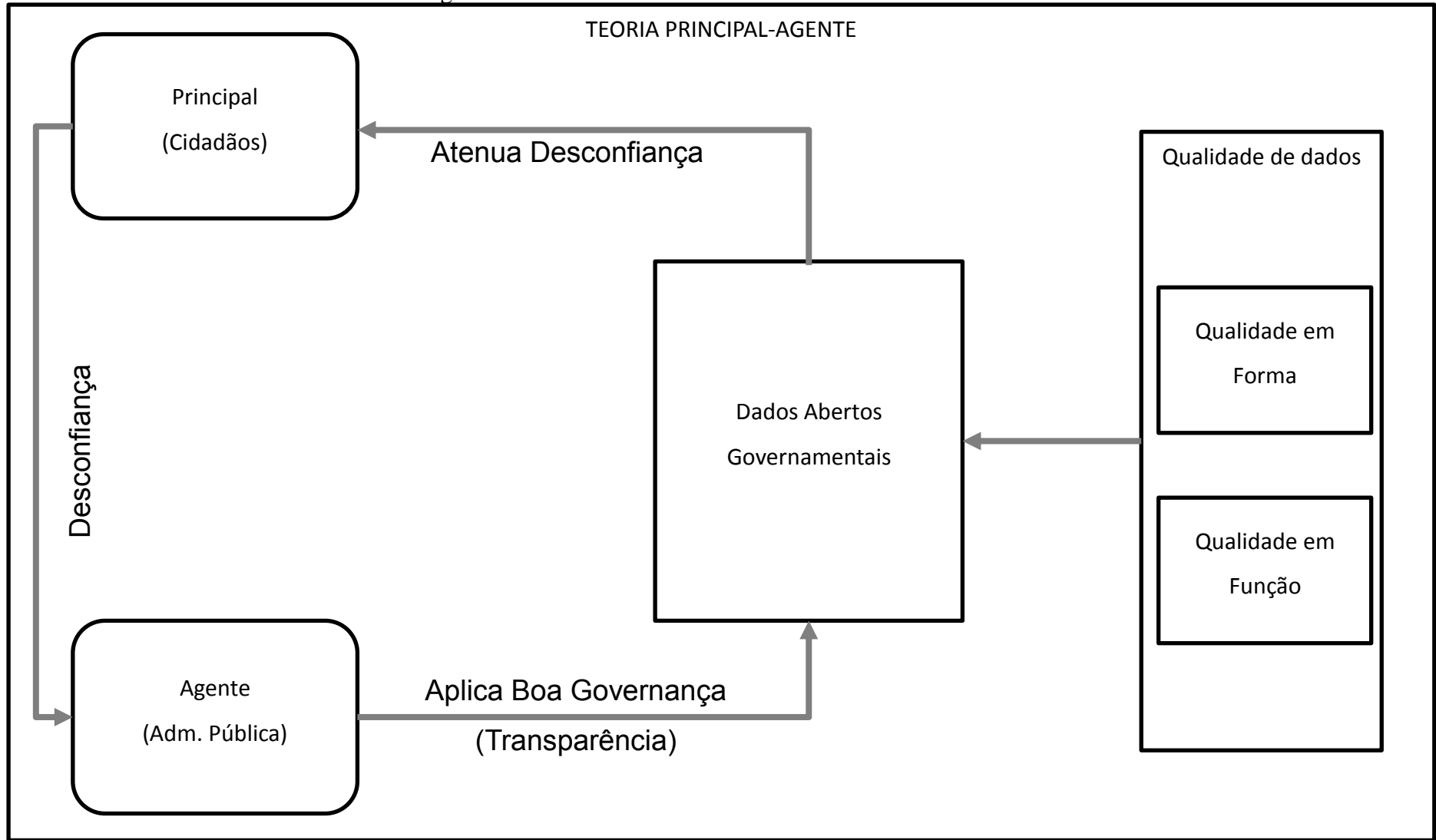
de modo que a administração pública, resolvendo a assimetria de informação e o conflito de interesses inerente à relação, possibilite atenuar a desconfiança que o Principal tem em relação ao Agente.

Para esses dados abertos governamentais servirem a este propósito, eles devem apresentar qualidade, na definição de Wang e Strong (1996), tanto de forma quanto de função, ou seja, devem servir à intenção proposta na Teoria Principal-Agente, de “ser útil” quanto à intenção original: atenuar a assimetria de informação. As metodologias de qualidade de dados atualmente relacionadas a dados abertos governamentais está, contudo, mais focada em qualidade de forma que de função, o que demonstra uma lacuna da literatura, uma vez que qualidade de dados não é algo relativo apenas à forma que os dados se apresentam, mas também como veículo para a informação, ou seja, deve-se focar a qualidade dos dados na definição de sua intenção e utilidade, conforme a visão proposta pela TPA. Essa lacuna é exatamente o que esta pesquisa tenta preencher, ao observar percepções sobre a qualidade de dados, não por meio de um método quantitativo de qualidade de dados ou informação, mas sim com aplicação de um formato mais aberto e qualitativo.

A TPA é a teoria mais interessante para isso, não só pelas recomendações da literatura de TI que a tem como a teoria base para alinhamento entre comportamento do Agente e necessidades do Principal mas também pelo fato de ela ser ligada intimamente ao conceito de governo e governança.

Para maior compreensão, a Figura 3 resume o modelo conceitual deste texto.

Figura 3 – Modelo Conceitual do Referencial Teórico



Fonte:

Autor.

3. METODOLOGIA

Esta pesquisa, segundo Zanella (2009), é de natureza qualitativa, com objetivo descritivo-exploratório. O procedimento para obtenção dos dados escolhido foi a entrevista semiestruturada porque, segundo Bryman e Bell (2011) e Mattos (2011), o método qualitativo é mais interessante em territórios ainda inexplorados e subjetivos, permitindo o estudo sem suposições. Especificamente, utilizou-se a entrevista, pois Zanella (2009) explica que, com este instrumento, se evita a contaminação de dados pela presença de outros indivíduos e permite-se evitar que o contexto social de pesquisa em grupos afete o resultado. O subtipo semiestruturada foi escolhido por tentar obter dados relativos a um único objetivo, com parte das perguntas já estruturada e outra parte mais aberta, sendo um dos subtipos mais interessantes para avaliar algo que ainda é pouco explorado – como qualidade de dados à luz da TPA, neste caso. A liberdade de ser aberto é essencial para avaliar este aspecto, enquanto a parte mais estruturada dá mais solidez ao método que se quer passar. A pesquisa semiestruturada é, portanto, o melhor formato para a coleta de dados e exploração do tema escolhido.

3.1. População e Amostra

Para obtenção dos dados, se utilizou da participação voluntária de alunos da Universidade Federal da Paraíba (UFPB) que fizeram parte de uma atividade valendo crédito da Disciplina Sistemas de Informação e Decisão, 2018.1, CCSA/DA/UFPB, (abreviada como SI&D) dos cursos de graduação em Ciência da Computação e Administração.

Essa atividade pedia que os alunos fossem separados em grupos aleatórios de até 3 integrantes. Cada grupo deveria criar um aplicativo numa plataforma OLAP (*Online Analytical Processing*, ou, Processamento Analítico Online), para estudar um conjunto de dados entre vários previamente escolhidos pelo professor da disciplina. Para decidir qual grupo ficaria com qual conjunto, foi feita uma escolha aleatória entre os grupos. A atividade tinha como intuito criar um aplicativo que possibilitasse análises descritivas, previsões e diagnósticos dos dados dos órgãos e responder a questões de negócio criadas pelos discentes (e acordadas previamente com o docente). Cada grupo deveria apresentar seus aplicativos, nas últimas semanas da matéria, em até 30 minutos. Durante cada apresentação, o grupo teria de informar quais dados foram tratados; como os dados foram obtidos; que etapas foram necessárias; qual script de carga de dados foi produzido e quantos registros foram carregados;

qual modelo de dados foi produzido; quais questões de negócio foram elaboradas a partir dos dados; e como essas questões eram respondidas pelo aplicativo.

O docente da disciplina separou, por conveniência, os seguintes órgãos para serem parte da atividade: (a) Agência Nacional de Aviação Civil, (b) Agência Nacional de Águas, (c) Secretaria de Desenvolvimento da Infraestrutura do Ministério de Planejamento (d) Ministério da Indústria, Comércio Exterior e Serviços, (e) Ministério da Defesa, (f) Ministério do Turismo,

(g) Ministério da Educação e Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, (h) Controladoria Geral da União, (i) Banco de dados do futebol europeu, (h) Banco de dados da Fórmula 1. Devido a certas dificuldades de acesso a alguns destes dados e número de alunos maior que esperado, outros conjuntos de dados, contudo, foram adicionados *a posteriori*.

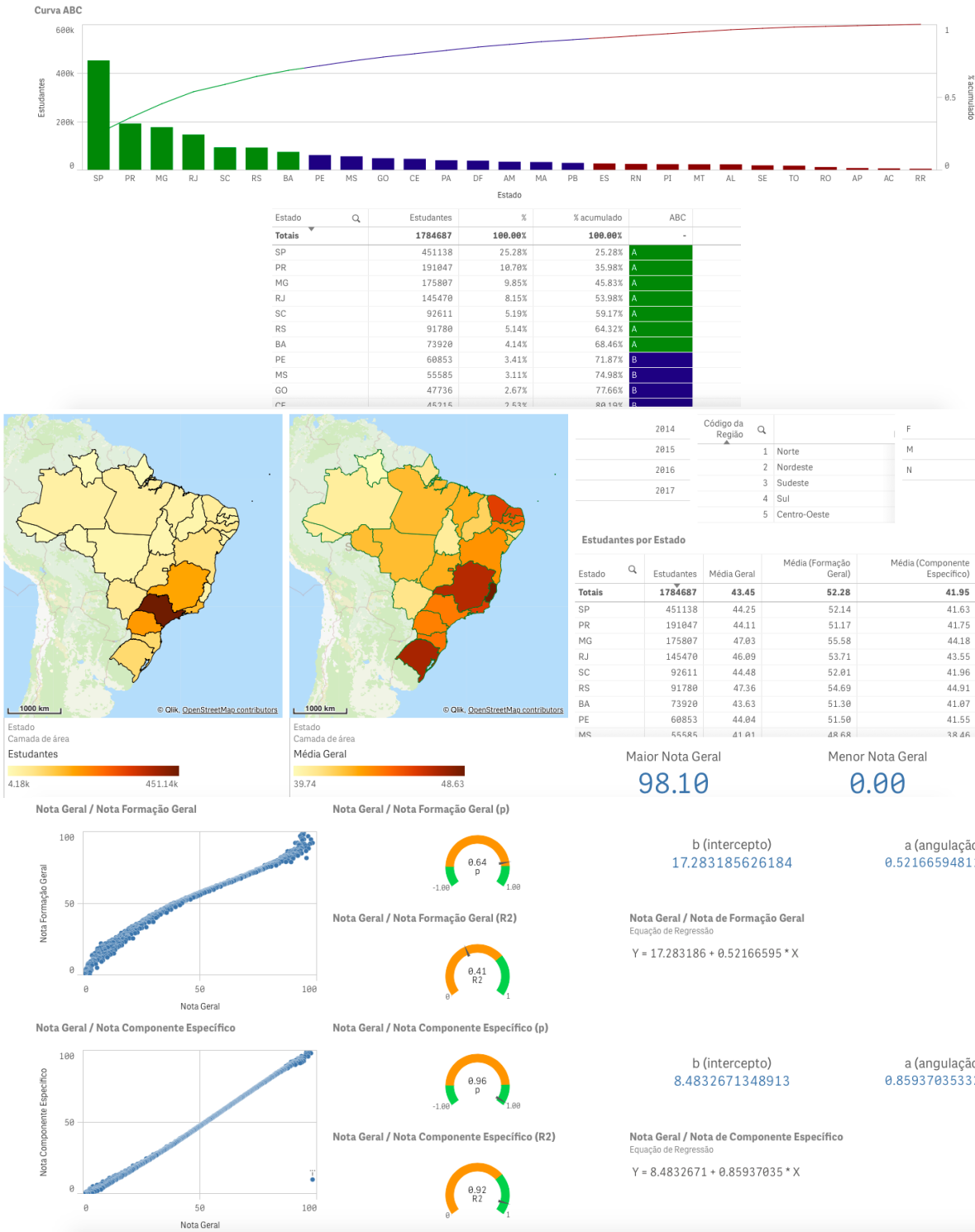
Para avaliação da pontuação relativa à atividade, se levou em consideração os seguintes tópicos do documento da matéria: (a) elaboração das questões de negócio (argumentação sobre coerência com os dados e relevância para a tomada de decisão); (b) implementação de procedimentos ETL (descrição de passos e apresentação de evidências); (c) modelagem dimensional (apresentação do modelo de dados); (d) implementação OLAP para análises descritivas (apresentação de pasta específicas do Aplicativo); (e) sugestão de análises diagnósticas (apresentação de explicações para o comportamento dos dados, conforme descritos no item anterior); (f) implementação OLAP para análises preditivas (apresentação de pasta específica do Aplicativo); (g) apresentação do grupo (coesão, comunicação, postura, capacidade de envolvimento da plateia); (h) criatividade, pró-atividade, inovação. Outros aspectos como o envolvimento na discussão de trabalhos de outros grupos e a presença nessas apresentações também fizeram, todavia, parte dos créditos para a disciplina de formas outras. Um sistema de “consultoria”, “contratações” e “demissões” com outros integrantes de outros grupos assim como uma classificação diferenciada para os primeiros colocados foram postos em prática (Para maiores informações, ver o documento da atividade nos anexos).

Essa população foi escolhida, em detrimento de outras, por vários motivos. Primeiramente, a atividade que esses alunos estavam fazendo se demonstrou uma oportunidade única, pois esses tinham certas características necessárias para a obtenção de dados sobre qualidade de dados como: ter o conhecimento necessário para poderem entender os dados; ter o conhecimento de como reconhecer qualidade dos dados; ter conhecimento de métodos para obtenção dessas informações de dentro dos dados; e ter motivação para estudar profundamente os dados expostos e não somente dar uma resposta sem base formal, analítica.

Segundamente, se fossem escolhidas outras populações sem a devida motivação ou conhecimento, mas que representariam mais o Principal, os dados obtidos poderiam não ser enviesados devido ao fato de estes não terem nem motivação de fazer uma boa análise dos dados, como realmente saber o que procurar ou como expressar isso em questão de qualidade de dados.

A atividade foi apresentada e teve várias apresentações por grupo, resultando em vários aplicativos sobre os dados, sendo alguns dos exemplos os da Figura 4 abaixo.

Figura 4 – Aplicativos feitos pelos discentes



Fonte: Discentes estudados.

3.2. Obtenção dos Dados

Durante a apresentação das atividades acima citadas, especificamente entre 22 de outubro de 2018 e 02 de novembro de 2018, foram obtidos os contatos de 31 alunos das turmas matutina e noturna. Destes 31 alunos, 23 responderam positivamente a chamadas para

fazer parte da pesquisa, entre as datas de 02 de novembro de 2018 e 14 de novembro de 2018. Entre os dias 14 de novembro e 21 de dezembro de 2018, contudo, apenas 12 respondentes realmente participaram da pesquisa, ou seja, 52,17%. A coleta de mais entrevistas foi encerrada neste nível por motivos múltiplos: restrições de tempo, indisponibilidade dos voluntários de participar realmente da pesquisa, proximidade com as festividades de fim de ano e a escassez de constructos novos, aspecto observável nas falas dos entrevistados (saturação).

As entrevistas foram feitas em dois métodos: presencial e por meio eletrônico. Originalmente, se planejou somente contar apenas com entrevistas presenciais, mas, devido a restrições de tempo, as últimas entrevistas foram feitas por meio eletrônico. Para registrar as entrevistas presenciais, se utilizou o Smartphone Samsung Galaxy J5, com microfone de lapela. Já, no método eletrônico, o programa OBS Studio foi utilizado para o registro em formato mediado por internet. Após gravação, o programa Xulia Beta 0.1 fez a primeira transcrição para formato textual que foi corrigido/revisado à mão, *a posteriori*.

3.3. Roteiro de Entrevista

O roteiro de pesquisa teve 4 perguntas referentes aos temas Circunstâncias do entrevistado e ao Banco de dados estudado. Este é seguido de um pequeno texto, solicitando o efeito *priming*, para reavivar a memória dos entrevistados, e finalizado com 8 perguntas quanto à qualidade dos dados à luz da TPA (mais informações, ver apêndices).

As 4 perguntas referentes aos temas Circunstâncias do entrevistado e aos Bancos de dados estudados foram: I) “Qual foi o órgão que você estudou para a matéria? ”; II) “Qual seu curso de graduação? ”; III) “Como foi seu papel ao fazer este trabalho? ”; e IV) “Qual foi o banco de dados que vocês estudaram?”. O intuito de ter posto estas questões foi de firmar exatamente a qual conjunto de dados os entrevistados se referiam e qual grau de entendimento esses entrevistados supostamente tinham dos dados.

O efeito *priming* é definido por Moura Jr. (2015) como “efeito temporário da ativação de estado mental relativo a uma categoria (um construto como “animal feroz”, por exemplo), a partir da exposição do indivíduo a um exemplar dessa categoria (imagem de um tubarão, por exemplo)”. Este efeito é utilizado como recurso de pesquisa neste texto, pois *priming* pode ser aplicado para situações “em que se quer estimular determinado estado mental” (MOURA Jr., 2015). Esse recurso foi utilizado para criar um diálogo que estimulasse as memórias dos dados brutos nos entrevistados. Isso foi obtido ao se proceder à leitura da seguinte passagem: “A partir de agora, tente se imaginar no processo de desenvolvimento do aplicativo para a

disciplina do Prof. Você está começando a lidar com os dados abertos brutos, do jeito que estão no site indicado pelo professor. Tudo que nós vamos discutir envolve esse momento específico em que você desenvolvia o aplicativo".

Para discutir sobre a qualidade de dados à luz da TPA, resolveu-se usar as seguintes perguntas:

Quadro 3 – Questões de Percepções Sobre os Bancos de Dados

Nº	Questão	Fundamentação
Q1	Foi enfrentada dificuldades ou facilidades para acessar os dados referente a atividade?	Explora características/atributos de Qualidade de dados.
Q2	Foi enfrentada dificuldades ou facilidades para entender esses dados?	Explora características/atributos de Qualidade de dados.
Q3	Qual sua percepção sobre a confiabilidade destes dados?	Explora características/atributos de TPA.
Q4	Qual sua percepção sobre o formato em que foi apresentado estes dados?	Explora características/atributos de Qualidade de dados.
Q5	Qual sua percepção sobre a efetividade destes dados em informar?	Explora características/atributos de Qualidade de dados.
Q6	Se tivesse a oportunidade, Como você mudaria nos dados de modo a melhorar sua efetividade?	Explora características/atributos de TPA.
Q7	Imagine-se como um cidadão ou organização atendido por este órgão. Em que você mudaria sua perspectiva?	Explora características/atributos de TPA.
Q8	Qual sua percepção sobre estes dados no papel de prova de que a organização está realizando um serviço que beneficia o cidadão?	Explora características/atributos de TPA.

Fonte: Autor

As questões foram escolhidas por referenciar as características de qualidade de dados e TPA. Mais especificamente as descritas em Mitnick (1973, 1976, 2006), para TPA, e Cai e Zhu (2015), Batini e Scannapieco (2016), Wang, Ziad e Lee (2006) e Wang e Strong (1996), para qualidade de dados. Também foram consideradas situações hipotéticas que possibilitaram o máximo de informação obtido por entrevistado. Para ajudar a investigar e analisar a capacidade dessas perguntas de promover a reflexão sobre aquilo que se pesquisa neste trabalho, dois mestrandos foram solicitados. E, para ajudar a garantir que as perguntas faziam sentido para a população dos entrevistados, 2 graduandos serviram como colaboradores na resolução de pré-testes das perguntas, os quais viriam a ser recolocados como os respondentes do roteiro completo identificados como B01 e B02.

3.4. Fontes Para a Análise dos Dados

Para obter os conceitos com quais comparar os dados obtidos sobre constructos e métricas de qualidade de dados, serão usadas, como base, revisões da literatura sobre metodologias de mensuração de qualidade de dados e informações (dados sendo veículo para criação e transporte de informação), como, por exemplo, Lourenço (2015); Vertrò *et al.*

(2016); Umbrich; Neumaie e Polleres (2015); Fagundes, Macedo e Freund (2017); Alenezi, Tarhini e Masa'deh (2015); e Batini *et al.* (2009), que dão acesso aos constructos que compõem cada tipo de mensuração, mostrando onde os conceitos se enquadram, se relacionam, quais conceitos foram mais relevantes e, principalmente, para onde convergem.

Este método foi o escolhido, pois oferece um mapa de quais artigos se deve ler em sequência, para se ter mais detalhes, e evita perder-se o foco relendo-se múltiplos artigos sobre o mesmo método ou mesmo constructo, além de possibilitar encontrar outros métodos mais condensados de obter os constructos como outras revisões da literatura. Também se procurou obter os métodos mais relevantes na sua publicação.

Para esta pesquisa, se levou em conta métodos mais amplos de qualidade de dados, para não estudar, equivocadamente, métodos feitos somente para uma pequena área. Métodos de qualidade de informação que avaliassem o papel dos dados como veículo de informação também foram estudados, pois a função, o uso e a intenção dos dados, na TPA, é, sobretudo, informar.

Daí, obtiveram-se os métodos de mensuração de qualidade de dados e informação contidos no quadro 4.

Quadro 4 – Metodologias de Mensuração de Qualidade de Dados e Suas Dimensões

Acrônimo (Inglês)	Nome (inglês)	Nome (português)	Dimensões	Referência Base
TDQM	<i>Total Data Quality Management</i>	Gerenciamento Total de Qualidade de Dados	acessibilidade, adequação, credibilidade, completude, concisa / consistente apresentação, facilidade de manipulação, valor agregado, livre de erros, interpretabilidade, objetividade, relevância, reputação, segurança, pontualidade, compreensibilidade	Wang (1998)
DWQ	<i>Data Warehouse Quality Methodology</i>	Metodologia de Qualidade de Data para Warehouses	exatidão, integralidade, minimalidade, rastreabilidade, interpretabilidade, evolução de metadados, acessibilidade (sistema, transacional, segurança), utilidade (interpretabilidade), pontualidade (atualidade, volatilidade), capacidade de resposta, completude, credibilidade, precisão, consistência, interpretabilidade	Jeusfeld, Quix e Jarker (1998)
AIMQ	<i>A methodology for information quality assessment</i>	Uma metodologia para avaliação da qualidade da informação	acessibilidade, adequabilidade, credibilidade, completude, concisa / consistente apresentação, facilidade de operação, livre de erros, interpretabilidade, objetividade, relevância, reputação, segurança, pontualidade, compreensibilidade	Lee <i>et al.</i> (2002)
DQA	<i>Data quality assessment</i>	Avaliação de qualidade de dados	acessibilidade, quantidade adequada de dados, credibilidade, completude, liberdade de erros, consistência, representação concisa, relevância, facilidade de manipulação, interpretabilidade, objetividade, reputação, segurança, pontualidade, compreensibilidade, valor adicionado.	Pipino <i>et al.</i> (2002)
IQM	<i>Information quality measurement</i>	Medição de qualidade da informação	acessibilidade, consistência, pontualidade, concisão, manutenibilidade, atualidade, aplicabilidade, conveniência, velocidade, abrangência, clareza, precisão, rastreabilidade, segurança, exatidão, interatividade.	Eppler e Munzenmaier (2002)

Acrônimo (Inglês)	Nome (inglês)	Nome (português)	Dimensões	Referência Base
TIQM	<i>Total Quality Information Management</i>	Gerenciamento de Informações de Qualidade Total	dimensões inerentes: conformidade de definições (consistência), completude, conformidade com os regulamentos de negócios, precisão (para fonte substituta), precisão (para realidade), precisão, não duplicação, equivalência de dados redundantes, concorrência de dados redundantes. dimensões pragmáticas: acessibilidade, pontualidade, clareza contextual, integração da derivação, usabilidade, <i>rightness</i> (completude dos fatos), custo.	Inglês (2003)
ISTAT	<i>ISTAT methodology</i>	Metodologia ISTAT	precisão, completude, consistência	Falorsi <i>et al.</i> (2003)
AMEQ	<i>Activity-based Measuring and Evaluating of product</i>	Medição e avaliação com base em atividades do produto	representação consistente, interpretabilidade, caso de compreensão, representação concisa, pontualidade, completude, valor agregado, relevância, adequação, significado, falta de confusão, disposição, legível, razoabilidade, precisão, confiabilidade, liberdade de parcialidade, deficiência de dados, deficiência de design, operação, deficiências, precisão, custo, objetividade, credibilidade, reputação, acessibilidade, exatidão, sem ambiguidade, consistência	Su e Jin (2004)
COLDQ	<i>Loshin methodology of Cost-effect of Low data quality</i>	Metodologia Loshin de custo-efeito de qualidade de dados baixa	esquema: clareza de definição, compreensividade, flexibilidade, robustez, essencialidade, granularidade do atributo, precisão de domínios, homogeneidade, identificabilidade, obtenção, relevância, simplicidade / complexidade, semântica, consistência, consistência sintática. dados: exatidão, valores nulos, completude, consistência, atualidade, pontualidade, acordo de uso, administração, <i>ubiquity</i> , apresentação: adequação, correto, interpretação, flexibilidade, formato preciso, portabilidade, consistência, uso de armazenamento, política de informação: acessibilidade, metadados, privacidade, segurança, redundância, custo.	Loshin (2004)
DaQuinCIS	<i>Data Quality in Cooperative Information Systems</i>	Qualidade de Dados em Sistemas Cooperativos de Informação	precisão, completude, consistência, atualidade, confiabilidade	Scannapieco <i>et al.</i> (2004)

Acrônimo (Inglês)	Nome (inglês)	Nome (português)	Dimensões	Referência Base
QAFD	<i>Methodology for the quality assessment of financial data</i>	Metodologia para a avaliação da qualidade de dados financeiros	precisão sintática / semântica, consistência interna / externa, completude, atualidade, unicidade.	De Amicis e Batini (2004)
CIHI	<i>Canadian Institute for Health Information methodology</i>	Metodologia do Canadian Institute for Health Information	dimensões: precisão, comparabilidade da pontualidade, usabilidade, características de relevância: sobrecobertura, subcobertura, variação de resposta simples / correlacionada, confiabilidade, cobrança e captura, unidade / item não-respondido, edição e imputação, processamento, estimativa, pontualidade, abrangência, integração, padronização, equivalência, capacidade de ligação, comparabilidade entre produtos e histórico, acessibilidade, documentação, interpretabilidade, adaptabilidade, valor.	Long e Seko (2005)
CDQ	<i>Comprehensive methodology for Data Quality management</i>	Metodologia abrangente para gerenciamento de qualidade de dados	schema: exatidão em relação ao modelo, exatidão em relação a requisitos, completude, pertinência, legibilidade, normalização. dados: precisão sintática / semântica, precisão semântica, completude, consistência, atualidade, pontualidade, volatilidade, completabilidade, reputação, acessibilidade, custo.	Batini e Scannapieco (2006)
UDADQ	<i>Utility-driven assessment of data quality</i>	Avaliação de qualidade de dados orientada a utilidade	completude, validade, precisão, atualidade	Mesmo e Shankaranarayanan, (2007)
MFDQAI	<i>Methodologies for Data Quality Assessment and Improvement</i>	Metodologias para Avaliação e Melhoria da Qualidade de Dados	precisão, integralidade, consistência, pontualidade, atualidade, volatilidade, exclusividade, quantidade adequada de dados, acessibilidade, credibilidade, usabilidade, integridade de derivação, concisão, velocidade, abrangência, clareza, rastreabilidade, segurança, exatidão, objetividade, relevância, reputação, facilidade de operação, interatividade	Batini et al. (2009)
HDQM	<i>Heterogeneous Data Quality Methodology</i>	Metodologia de Qualidade de Dados Heterogêneos	precisão, atualidade	Carlo et al. (2011)
LDIF	<i>Linked Data Integration Framework</i>	Framework de Integração de Linked Data	antes: recência, reputação. depois: exatidão, completude e atualidade	Mendes., Mühleisen, & Bizer (2012)

Acrônimo (Inglês)	Nome (inglês)	Nome (português)	Dimensões	Referência Base
MEHDQA	<i>Methods of Electronic Health Record Data Quality Assessment</i>	Métodos de Avaliação da Qualidade de Dados do Registro de Saúde Eletrônico	completude, concisão e consistência	Weiskoph e Weng (2013)
QALOD	<i>Quality Assessment of Linked Open Data</i>	Avaliação de Qualidade de Dados Abertos Vinculados	disponibilidade, licenciamento, interligação, segurança, desempenho	Behkamal <i>et al.</i> (2014)
DDDC	<i>Data Disclosure Desired Characteristics</i>	Divulgação de dados características desejadas	qualidade, completude, acesso e visibilidade, usabilidade e compreensibilidade, pontualidade, valor e utilidade, granularidade	Lourenço (2015)
QAMF	<i>Quality Assessment Monitoring Framework</i>	Framework de Monitoramento de Avaliação de Qualidade	recuperabilidade, uso e completude, abertura, capacidade de contato, precisão	Umbrich; Neumaie e Polleres (2015)
BDQA	<i>Big Data Quality Assessment</i>	Avaliação de Qualidade de Big Data	acessibilidade, pontualidade, autorização, credibilidade, definição / documentação, metadados, precisão, consistência, integridade, integridade, aptidão, legibilidade, estrutura	Cai e Zhu (2015)
MEIDIQS	<i>Methodology to Evaluate Important Dimensions of Information Quality in Systems</i>	Metodologia para Avaliar Dimensões Importantes da Qualidade da Informação em Sistemas	entrada: precisão, integralidade, atualidade. saída: exatidão, completude, confiabilidade	Torodan <i>et al.</i> (2015)
FSMBDQA	<i>Framework for social media big data quality analysis</i>	Framework para análise de qualidade de big data de mídia social	escalabilidade, performance, confiabilidade, operações rápidas, backup, acessibilidade, sustentabilidade, proteção contra perda de dados, disponibilidade, pontualidade, utilidade, consistência, compreensão, segurança, exatidão, eficiência.	Jaffar <i>et al.</i> (2015)
ODQMF	<i>Open Data Quality Measurement Framework</i>	Framework de Medição de Qualidade de Dados Abertos	rastreabilidade, atualidade, expiração, completude, conformidade, compreensão, exatidão	Vertrò <i>et al.</i> (2016)

Acrônimo (Inglês)	Nome (inglês)	Nome (português)	Dimensões	Referência Base
BDQAFSO	<i>Big Data Quality Antecedents in Financial Services Organizations</i>	Antecedentes de Qualidade de Big Data em Organizações de Serviços Financeiros	características: precisão, credibilidade, relevância, atualidade, abrangência, consistência, singularidade, oportunidade, validade e rastreabilidade. antecedentes: quatro características de big data v, clareza de metadados, credibilidade da fonte de dados e conteúdo de dados, ferramentas de big data, técnica e tecnologia, autenticidade do método de coleta de dados, clareza de política sobre uso de dados, habilidades analíticas e equipe multidisciplinar, repositório de dados, capacidade ágil, <i>business case</i> convincente	Haryadi <i>et al.</i> (2016)
QALD	<i>Quality Assessment for Linked Data</i>	Avaliação de qualidade para Linked Data	disponibilidade, licenciamento, interligação, segurança, desempenho, validade sintática, precisão semântica, consistência, concisão, completude, relevância, confiabilidade, compreensibilidade, pontualidade, concisão-concisão, interoperabilidade, interpretabilidade, versatilidade	Zaveri <i>et al.</i> (2016)
DQFMPS	<i>Data Quality Framework for Malaysia's Public Sector Open Data Initiative</i>	Framework de Qualidade de Dados para a Iniciativa de Dados Abertos do Setor Público da Malásia	acessibilidade, pontualidade, autorização, credibilidade, clareza, exatidão, autenticidade, integridade, consistência, integridade, auditoria, aptidão ao uso, legibilidade, estrutura	Ijab <i>et al.</i> (2017)
OLDQAF	<i>Objective Linked Data Quality Assessment Framework</i>	Framework de Avaliação Objetiva de Qualidade de Linked data a Objetivos	completude, disponibilidade, licenciamento, frescura, exatidão, compreensibilidade, proveniência, coerência, consistência, segurança.	Assaf, Senart e Troncy (2018)

Fonte: Autor.

A maioria dos métodos de mensuração não tem foco nos dados abertos do governo nem na função dos dados para avaliar a Qualidade de dados. Por isso a fonte principal de comparação em Qualidade de dados provém da revisão de Batini *et al.* (2009) para constructos de Qualidade de dados.

Esta decisão é resultante do fato de Batini *et al.* (2009) terem uma revisão literária ampla e definirem Qualidade de dados em sua função de informar, não se restringindo tão somente ao formato, se alinhando, portanto, à visão deste trabalho. Esse texto também não se restringe à qualidade de dados em uma área específica (como, por exemplo, em sistemas médicos ou financeiros) e contém o maior número de métricas e constructos entre as metodologias estudadas. O trabalho de Batini *et al.* (2009) também foi publicado na *ACM Computer Survey*, uma revista de alto impacto (*Impact factor: 7.669*) fazendo-o pertencer ao top 3.3% de todas as revistas do *Journal Citation Reports* (2018).

Embora sua relevância, Batini *et al.* (2009) fazem apenas uma revisão da literatura, mas não uma Metodologia de qualidade de dados. A lista completa de métricas para avaliar qualidade de dados, desses autores não tem muita conexão, uma vez que engloba vários métodos diferentes e com focos diferentes (para a lista completa de métricas, ver Apêndice 8.4), além de muitas métricas completamente quantitativas. Para poder entender o que cada métrica e dimensão foca e poder identificar de modo qualitativo cada uma delas, foi feita uma pesquisa de cada um dos métodos listados em nesses últimos autores e se retirou o significado aproximado de cada métrica e dimensão para buscá-los nas falas dos entrevistados.

Como fonte principal base, para obtenção de constructos de governança, se utilizaram os constructos de “Boa governança” (*Good Governance*) do Banco Mundial (1992; 2017), por ser a literatura que considera governos para seu constructo e por ser uma das principais literaturas modernas sobre o assunto.

Por fim, a fonte principal para obtenção de constructos de TPA foram os trabalhos de Mitnick (1973;1976; 2006) e Jensen & Meckling (1976; 2008), por lidarem com a relação Cidadão-Administração Pública e serem das mais referenciadas fontes.

Com base nesses autores, foi feita uma lista prima contendo cada TCon que se supunha aparecer com base nas literaturas de fontes principais descritas acima. Nessa lista estava associado o nome da TCon, ao seu conceito obtido a partir das fontes principais de constructos, e ao autor que foi a base para a TCon. Nessa lista, que se encontra nos apêndices, também se colocou uma descrição reduzida do que se procurou.

Para seguir o princípio da parcimônia, contudo, decidiu-se reduzir o número de constructos uma vez que, dos 44 constructos desta primeira lista, vários apresentavam proximidade ou estariam presentes em outros constructos.

Por este motivo, se retiraram os conceitos de Disponibilidade, Backup, Sustentabilidade, da lista por serem similares demais ao conceito de Acessibilidade e por não fazerem distinção suficiente para a relação Principal-Agente estudada. Eficiência, da mesma forma, não foi suficientemente dissimilar de Custo, nem Versatilidade de Usabilidade, nem Autorização de Licenciamento.

Os conceitos de TPA de Autodidatismo já estão presentes no conceito de Acessibilidade. Não foi possível dissimilar Confiança o suficiente de Credibilidade, sendo todos os conceitos relativos ao primeiro constructo observáveis no segundo, Credibilidade de Qualidade de dados. O conceito de Segurança foi retirado por não fazer parte de Qualidade de dados como vetor de informação. Por fim, Estrutura Legal, um dos conceitos integrantes de Governança não se mostrou útil averiguar em primeira instância, pois os alunos estudados provavelmente não detinham conhecimentos sobre bases legais necessários e pelas partes interessantes para este constructo serem abarcadas em outras áreas, como a de Prestação de Contas, por exemplo.

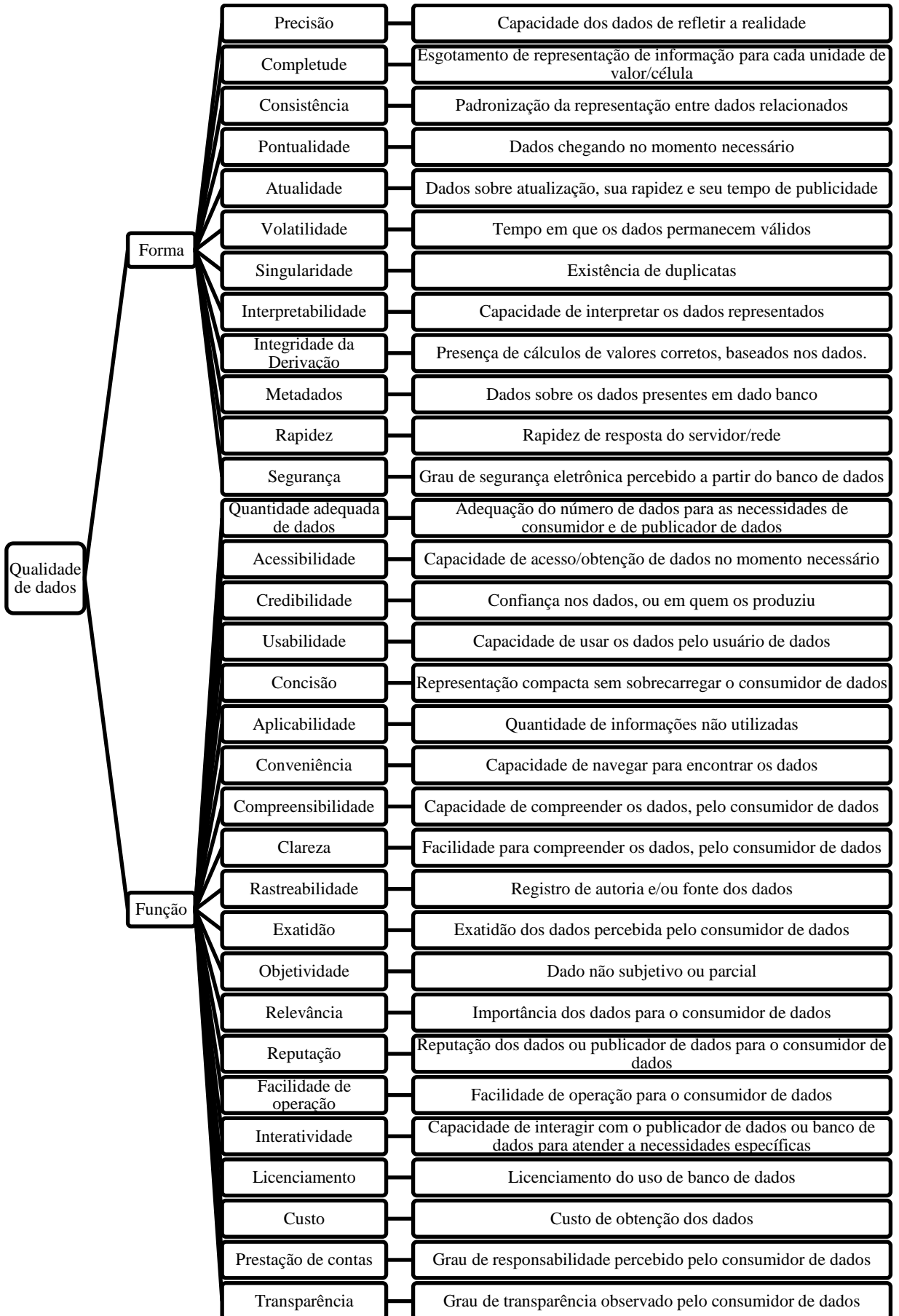
Para garantir uma melhor análise, também se separaram os conceitos em dois grandes grupos: conceitos de Qualidade de dados em forma e de Qualidade de dados em função.

Qualidade de dados em função é aqui relacionada a conceitos que avaliam como os dados estão segundo sua função, seja esta baseada na intenção original do publicador/produtor de dados, seja no uso que o consumidor de dados tem para aqueles dados. Este grupo, por causa dessa natureza, tem características mais subjetivas em suas métricas.

Qualidade de dados em forma, por outro lado, tem conceitos preocupados com o formato que o dado detém, ou seja, se é possível obter algum nível de informação, não interessado explicitamente no seu uso ou percepção perante o consumidor, produtor ou publicador de dados. Da mesma forma, este grupo, por causa de sua natureza, tem métricas mais objetivas.

Essa classificação resultou na Figura 5, a seguir exposta, que classifica os constructos de acordo com o grupo e sua definição:

Figura 5 – Representação dos Constructos e suas Definições



3.5. Análise e Tratamento dos Dados

Para análise de dados, foi utilizada a técnica de análise de conteúdo descrita em Bardin (1977) e Rocha & Deusdará (2005), por meio do programa Microsoft Excel. A análise de conteúdo é bem resumida, conforme se vê na seguinte passagem de Moura:

Análise de dados de maneira objetiva e sistemática, com potenciais resultados qualitativos e quantitativos. É, portanto, um conjunto de técnicas que têm como intenção a inferência, a partir da percepção do pesquisador, sobre o conteúdo das mensagens analisadas para extração de conhecimento, considerando o contexto no qual as mensagens foram produzidas e recebidas (MOURA, 2015, p.).

Para fazer a análise de conteúdo, é necessário fazer um tratamento dos dados obtidos. O áudio das entrevistas inicialmente foi transcrito em formato discursivo, depois dividido em sentenças (unidade de registro da fala contendo significado para o pesquisador), produzindo uma coluna de dados (TSen) no Excel, que serviu de base para as análises decorrentes.

A cada sentença TSen atribuiu-se, em outras colunas, o identificador numérico sequencial único (ID_Sen); o codinome do respondente (Entrev.); a questão precedente (Q); a natureza principal da sentença (TNat); a categoria do tipo de natureza (TCat); o objeto ao qual TSen se refere (TObj); e o tipo de constructo, ideia, conceito, dimensão ou métrica do entrevistado referenciado (TCon), se houver. Por exemplo, se uma sentença trata somente de uma descrição ou nomeação do conceito de transparência com elogios e ou ênfase de percepção positiva, referindo-se a uma planilha, esta TSen é classificada como contendo TNat com o texto “Constructo”, TCat com o texto “Positiva”, TObj com o texto “Planilha” e TCon com o texto “Transparência”. Se fosse sobre um desejo ao invés de constructo sendo retratados os valores, estes seriam “Desejo”, “Positivo”, “Planilha”, e “Transparência”, respectivamente, e assim por diante.

Em seguida, para focar nos pontos mais relevantes, a tabela foi filtrada pelas TNats relevantes para a pesquisa. As TNats escolhidas, suas TCats e seus significados foram:

- **Constructo:** Natureza que retrata de definição, nomeação ou menção direta a um conceito, constructo, ideia, dimensão ou métrica referenciados na literatura.
 - **Positiva:** Constructo de categoria positiva para o entrevistado. Retrata elogios, referências positivas e louvores à boa qualidade do constructo.
 - **Negativa:** Constructo de categoria negativo para o entrevistado. Retrata críticas, descrédito e referências negativas à qualidade do constructo.

- **Neutra:** Constructo de categoria neutra, indefinida ou indefinível. Retrata a presença ou a nomeação do constructo em relação ao banco de dados.
- **Desejo:** Natureza que retrata o desejo do entrevistado referente aos dados, para que estes sejam mais efetivos e ou eficientes.
 - **Desejo:** Natureza de desejo de melhora dos dados.

O seguinte passo foi avaliar cada uma das TSen com estes TNat, para conferir a ligação do TCon com a literatura, assim como adicionando outras informações na coluna “Obs”. O último passo foi descrever os achados no corpo da dissertação, evidenciando emergências, convergências e divergências, assim como o tipo de natureza observado, sempre comparando o descrito na literatura e o obtido a partir dos dados.

O processo é bem resumido na figura 6 a seguir:

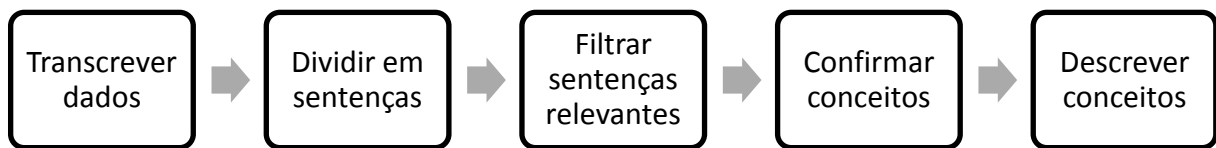


Figura 6 – Processo de Análise dos Dados

Fonte: Autor.

4. RESULTADOS E DISCUSSÃO

Cada áudio da entrevista durou em média 31 minutos; as transcrições tinham em média 849 palavras relevantes. O áudio de todas as entrevistas totaliza 6 horas e 14 minutos e 10.189 palavras relevantes.

Os resultados quanto a questões demográficas, suas representações e os dados sobre os quais se trabalhou estão representados no Quadro 5 a seguir:

Quadro 5 – Resultados sobre os Respondentes e os Bancos de Dados

Entrevista	Organização	Banco de Dados	Abreviação (Conj)	Curso de graduação
A01	Ministério da Educação	ENADE - Exame Nacional de Desempenho de Estudantes	ENADE	Administração
A02	Ministério da Educação	ENCCEJA - Exame Nacional para Certificação de Competências de Jovens e Adultos	ENCCEJA	Ciência da Computação
A03	CEPEA - Centro de Estudos Avançados em Economia Aplicada	Vendas internacionais	CEPEA	Administração
A04	ANAC - Agência Nacional de Aviação Civil	Voos internacionais	ANAC	Administração
A05	Ministério do Planejamento	PAC – Programa de Aceleração do Crescimento	PAC	Ciência da Computação
B01	Ministério do Planejamento	PAC – Programa de Aceleração do Crescimento	PAC	Ciência da Computação
B02	Ministério da Educação	ENADE - Exame Nacional de Desempenho de Estudantes	ENADE	Administração
B03	Ministério do Turismo	Investimentos e gastos	M.TUR	Ciência da Computação
B04	Ministério do Turismo	Investimentos e gastos	M.TUR	Ciência da Computação
B05	ANAC - Agência Nacional de Aviação Civil	Voos Internacionais	ANAC	Administração
C01	CEPEA - Centro de Estudos Avançados em Economia Aplicada	Vendas internacionais	CEPEA	Administração
C02	Kaggle Inc.	Dados de futebol europeu	KAGGLE	Ciência da Computação

Fonte: Autor.

Este resultado é interessante, pois possibilita equiparar as percepções de múltiplos indivíduos sobre o mesmo banco de dados, ou a mesma organização, sem ter sido a intenção original. Houve 4 bancos de dados repetidos, ou a taxa de repetição de 33%. A saber, os bancos compartilhados foram entre A01 e B02; A03 e C01; A05 e B01; e A04 e B05. Uma possível causa para isso seria que os alunos compartilharam a experiência da entrevista, o que possibilitou a participação conjunta das entrevistas, ou pura chance aleatória.

4.1. Saturação

Quadro 6 –
por Ordem de

Entrevistados	Constructos Únicos
A01	10
A02	2
A03	2
A04	1
A05	1
B01	0
B02	1
B03	2
B04	0
B05	0
C01	0
C02	0

Constructos Únicos
Entrevistados

Fonte: Autor.

A saturação da pesquisa foi atingida, levando em consideração os escritos de Bardin (1977). Como se pode ver no quadro 6, houve uma redução drástica de novos constructos antes do final da pesquisa. Originalmente se planejava entrevistar ao menos um aluno por banco de dados descrito na atividade (Ver Projeto de Atividade, nos Anexos), o que não foi possível devido à restrição de tempo, à aparente saturação e à não cooperação dos voluntários. Esta é a explicação de as entrevistas terem sido encerradas no número 12.

4.2. Geral

Quadro 7 – Sentenças e Constructos por Entrevistados

Entrevistados	Número de TSen	Número de TCon
A01	92	27
A02	64	22
A03	70	24
A04	39	19
A05	28	10
B01	47	16
B02	32	11
B03	71	25
B04	30	14
B05	82	22
C01	44	15
C02	47	10

Total	646	215
--------------	------------	------------

Fonte: Autor.

Quando se fala do número de sentenças e de constructos obtidos, num geral, se verifica que houve 646 sentenças e 214 Constructos observados entre as suas categorias: Sentenças e TCon. Os mais prolíficos em sentenças foram o A01(91) e B05(82) e os mais prolíficos em TCon foram A01 e B03. Uma nota interessante é que, pelo timbre de voz, essas pessoas são do sexo feminino, enquanto os menos prolíficos, A05(28) e B02(32), são pessoas do sexo masculino, demonstrando que, aparentemente, mulheres têm mais facilidade de expressar sentenças e conceitos que homens.

Há de se falar da tipologia de natureza das sentenças logo após cada questão. Isso é relevante, pois é possível saber a percepção dos entrevistados sobre o que a questão fala exatamente, mesmo que, no decorrer das falas, alguns entrevistados mudem a natureza da sua fala para ir de “positiva” para “neutra” e para “negativa”, contudo, devido à estrutura de Q7 e Q8, não foi possível observar uma tipologia exata para essas naturezas.

Quadro 8 – Natureza da Resposta por Entrevistado em Relação à Primeira Sentença

	A01	A02	A03	A04	A05	B01	B02	B03	B04	B05	C01	C02
Q1	+	+	+	-	+	+	+	-	+	+	-	+
Q2	-	+	+	-	+	+	+	-	-	+	+	-
Q3	+	+	?	+	+	-	+	-	-	+	+	+
Q4	-	+	+	-	-	-	+	-	-	+	+	-
Q5	+	+	+	+	-	-	+	-	-	-	-	+
Q8	?	+	+	+	+	-	-	-	-	-	+	+

Fonte: Autor.

Legenda: “+”= positivo; “-” = negativo; “?” = Dúvida

O fato mais interessante é que A3, na Q3, e A1, na Q8, responderam com uma dúvida da sua capacidade de retratar os dados como deveriam. As falas foram: “Acho que não posso dizer com certeza com os dados que eu vi, né? ” (Entrevistado: A1, TSen: 67) e “Bem... ainda estou pensando hehehe... Não sou da área, então não posso dizer” (A3, 179). Isso demonstra os problemas de pesquisar a percepção. A percepção é falha e pode ser duvidosa, inclusive para o próprio pesquisado – nem tudo o que é expresso é a realidade (Bardin, 1977).

Para os resultados, a partir daqui, será feita uma análise com base em cada um dos TCon obtidos, explicando-se a definição de cada um mais profundamente, como estes TCon se relacionam a outros e mencionando-se TSen relevantes, além de como cada um se comporta na relação Principal-Agente e afeta essa relação, à luz da teoria em estudo.

4.3. TCon - Acessibilidade

A acessibilidade é descrita, por Wang e Strong (1996), como uma das “quatro categorias maiores” de Qualidade de dados e está em implícita ou explícita em todas as metodologias de mensuração de qualidade pesquisadas, seja de qualidade de informação ou de qualidade de dados. Wang (1998); Jeusfeld *et al.* (1998); Lee *et al.* (2002); Long e Seko (2005); Pipino *et al.* (2002); Eppler e Munzenmaier (2002); Falorsi *et al.* (2003); Su e Jin (2004); Todoran *et al.* (2015); Behkamal *et al.* (2014); Vertrò *et al.* (2016); Umbrich; Neumaie e Polleres (2015); e Ijab *et al.* (2017) mencionam acessibilidade de forma direta e muitos outros, como Vertrò *et al.* (2016), já tinham a acessibilidade pressuposta para sua avaliação de Qualidade de dados.

As métricas de avaliação de acessibilidade variam. As encontradas foram: a diferença entre número de links quebrados e o número total de links (BATINI *et al.* 2009), o número total de tabelas/dados não acessíveis (FALORSI *et al.*, 2003), a demora ou a não resposta ao “fazer uma requisição de dados” (EPPLER; MUNZENMAIER, 2002), ou o grau de acessibilidade ou esforço para acessar percebido pelo consumidor de dados (ENGLISH, 2003).

O fator Acessibilidade é interrogado diretamente na Q1 e foi mencionado 22 vezes por quase todos os entrevistados. Este fator é especial e esperado. As sentenças de B01 foram as únicas a não o mencionarem diretamente – em seu lugar menciona-se uma dificuldade de Completude, embora a presença de acessibilidade seja subentendida.

Nas TCat positivas, foi considerada toda sentença que expressa positividade quanto a um constructo, como em “[...] acessar é fácilimo” (A01;6) e “[...] o acesso destes dados foi simples assim. Que o site na hora que a gente abre, é o que ele mostra. ” (A03; 159). A presença do constructo de acessibilidade foi, todavia, referenciado como neutra “[dificuldade] para acessar, nada” (A01, 2).

Um dos achados desta parte foi que os respondentes com visão mais negativa foram os que avaliaram os bancos de dados referentes ao Ministério do Turismo. Uma das sentenças dizia que a inacessibilidade “era 3 ou 4 em 15 ou 10 [dos links]” (B03,389). E havia “pagina faltando, [os dados estavam] muito ruim” (B04, 448). B03, B04 e C01 também foram os únicos a descrever acessibilidade pela métrica de links quebrados; as sentenças negativas de outros respondentes geralmente tinham mais relação com a dificuldade de entender o site em que os bancos residiam.

Houve uma prevalência de falta de acessibilidade em quase todos os bancos de dados. Os únicos que não mostraram qualquer dificuldade de acessibilidade foram os dados

referentes ao Ministério da Educação - MEC. Pelo contrário, os dados referentes ao MEC foram elogiados pela facilidade de acesso, em “[...] acessar é fácilimo” (A01,6) “eu gostei muito do nosso... do MEC, particularmente, porque era tudo bem dividido em vários exames, e dentro dos exames tinham vários anos, sem problemas” (A02,128) e “[...] para acessar ele foi tranquilo[...]” (B03,342). Algo que seria repetido em outros constructos de qualidade de dados, governança e TPA.

Este constructo é também descrito na TPA, uma vez que é o requisito mínimo para ser um sinal de confiança entre Agente e Principal. No momento em que a acessibilidade não está presente nos dados, o valor dos dados como mitigador de desconfiança diminui e é possível causar um aumento por suspeita desse comportamento que beneficiaria o Agente, permitindo que ele não siga as exigências do Principal.

Quadro 9 – Acessibilidade

ID_Sen	Conj	Entrev.	Q	TNat	TObj	TCat			
						Negativa	Positiva	Neutra	Total
2	ENADE	A01	Q1	Constructo	Dados			1	1
3	ENADE	A01	Q1	Constructo	Dados			1	1
6	ENADE	A01	Q1	Constructo	Dados		1		1
108	ENCCEJA	A02	Q1	Constructo	Dados			1	1
128	ENCCEJA	A02	Q5	Constructo	Banco		1		1
159	ENCCEJA	A03	Q1	Constructo	Dados		1		1
161	CEPEA	A03	Q1	Constructo	Site	1			1
229	ANAC	A04	Q1	Constructo	Site	1			1
267	PAC	A05	Q1	Constructo	Dados			1	1
342	ENADE	B02	Q1	Constructo	Dados			1	1
374	M.TUR	B03	Q1	Constructo	Site	1			1
375	M.TUR	B03	Q1	Constructo	Site	1			1
383	M.TUR	B03	Q1	Constructo	Dados	1			1
385	M.TUR	B03	Q1	Constructo	Site	1			1
389	M.TUR	B03	Q1	Constructo	Site	1			1
415	M.TUR	B03	Q5	Constructo	Dados	1			1
447	M.TUR	B04	Q1	Constructo	Paginas	1			1
465	M.TUR	B04	Q6	Constructo	Banco	1			1
475	ANAC	B05	Q1	Constructo	Site	1			1
557	CEPEA	C01	Q1	Constructo	Dados	1			1

						TCat			
ID_Sen	Conj	Entrev.	Q	TNat	TObj	Negativa	Positiva	Neutra	Total
561	CEPEA	C01	Q1	Constructo	Site	1			1
602	KAGGLE	C02	Q1	Constructo	Dados	1			1
Total						14	3	5	22

Fonte: Autor.

4.4. TCon - Clareza

Clareza, podendo ser também traduzido como claridade, este constructo é definido em Batini *et al.* (2009) como sendo o que apresenta a métrica de quanto os usuários consideram os dados claros. No texto original de Eppler e Munzenmaier (2002), do qual esta métrica nasceu, se entende o real conceito por trás desse constructo. Clareza é a facilidade de compreender. Ser capaz de compreender um dado é claridade, ser capaz de fazer isso facilmente é clareza.

As sentenças com esse conceito subjacente são as seguintes: “[...] como organização, [eles] poderiam trabalhar com os dados mais claros, né?”(A03, 208), “Eles deviam colocar dados mais claros, como os auditores ainda não viram isso?”(B04, 455) e “Certo, acho que foi fácil de entender porque tinha cada a descrição de cada coluna dos dados, então, do lado dizia: o nome, "o apelido", o nome da coluna, o que aquilo seria, qual o tipo de conteúdo numérico, "string", o que cada valor queria dizer” (B02,346).

Desses constructos, apenas a sentença número 346 pedia Clareza pelo conceito, ao invés de nomeá-lo, mas, considerando o resto das sentenças, se pode inferir que, implicitamente, o significado era o mesmo. Clareza é indispensável à TPA. O Principal, ao não ter toda a informação do agente, deixa de saber até como facilmente compreender. Para compensar isso, os dados devem ser os mais interpretáveis e claros possíveis, pois o principal não tem como entender tal atributo.

Considerando as falas sobre interpretabilidade e compreensibilidade de A03 e, em parte, C01, é possível entender que a percepção da clareza dos dados do CEPEA poderia melhorar. Eles dão todas as informações de modo rápido e fácil, mas aparentemente de modo não intuitivo, o que causou dificuldade ao grupo.

O mesmo não pode ser dito dos dados do Ministério do Turismo. Para aqueles que os estudaram, ele tem quase tudo o que deveria, no requisito de forma e de atrelar as legislações,

mas a falta de completude danificou demais a compreensão e a clareza dos dados. Basicamente, é transparente que faltam dados.

Quadro 10 – Clareza

ID_Sen	Conj	Entrev.	Q	TNat	TObj	TCat		
						Desejo	Positiva	Total
208	CEPEA	A03	Q7	Desejo	Dados	1		1
346	ENADE	B02	Q2	Constructo	Dados		1	1
455	M.TUR	B04	Q3	Desejo	Dados	1		1
Total						2	1	3

Fonte: Autor.

4.5. TCon - Completude

Completude é um constructo que avalia quão completo está o conjunto de dados, ou seja, se os dados passam todos os dados que poderiam sobre o assunto daquele aglomerado de dados. As métricas de completude geralmente envolvem algum cálculo com o número de dados nulos, perdidos e ou faltando (VERTRÒ *et al.*, 2016; WANG, 1998; UMBRICH; NEUMAIE; POLLERES, 2015); o número de dados, linhas, registros presentes em relação a expectativas (JEUSFELD *et al.*, 1998; ENGLISH, 2003); a quantidade de dados presentes, em comparação com o máximo possível para cada vetor ou célula (BATINI; SCANNAPIECO, 2006) ou se os dados presentes têm profundidade suficiente para serem usados pelo consumidor de dados (ENGLISH, 2003).

Há, porém, uma diferença entre Completude e o constructo de Quantidade Adequada de Dados (QAD). Se completude trata da “profundidade dos dados”, em English (2003), tentando esgotar a possibilidade das unidades de dados de informar, então, QAD difere por focar na existência de unidades de dados necessárias para o consumidor de dados entender, isto é, se as unidades de dados englobam todas as necessidades para aquele conjunto de dados.

Para esclarecer melhor, vejamos: Se há linhas ou células faltando numa lista de estudantes, isso pode ser visto como um problema de completude, mas, se é necessário ter o número de matrícula, como coluna, por exemplo, nessa lista, para seu consumidor de dados e, por algum motivo a lista não a apresenta, então é um exemplo de problema de QAD.

Os respondentes demonstraram referenciar a completude nas entrevistas A03, B02, C01. Não se sabe ao certo o porquê, mas se pode especular que, no caso de A03 e C01, não houve qualquer dificuldade de completude, portanto, eles não comentaram positiva ou negativamente. O caso de B02 pode ser diferente: ou B02 não percebeu completude no

sentido de esgotar a capacidade de representação de uma unidade de dados, ou, como alerta Fabrini (2018), as expectativas quanto à quantidade e qualidade dos dados do governo brasileiro são baixas o suficiente para não aparecerem aqui.

Das métricas de completude de característica negativa, A04, que lidou com os dados do Ministério do trabalho, fala de valores que estavam em branco sem explicação, mais especificamente na seguinte passagem: “Algumas informações estavam faltando. Em algumas partes da planilha, mas eu creio que se deve devido a algumas peculiaridades. E se referia a avião de carga, e por essa questão. Algumas coisas têm e outras não.” (A04, 240).

A05 e B01, que tratavam dos dados do PAC, esperavam ver alguns valores aparecer, o que não ocorreu, conforme se conclui das seguintes passagens: “Teve algumas coisas faltando, as datas, não tinha todas as datas. Acho que de conclusão das obras.” (A05, 277) e “Teve o problema das datas né, que a gente achou que, eles não colocavam todas das datas, não tinha o projeto que tinha sido finalizado só que a gente não sabia se é que estava atualizado, e dizia que está concluído pois dizia estar no estágio 90, mas não tinha data técnica da obra, então, não dava para filtrar e ver quanto tempo dava para finalizar uma obra” (B01,303).

B03 e B04 tiveram uma mistura das duas métricas acima mais a falta de profundidade demonstrada e de valores em comparação ao máximo, nas passagens “Quase não tinha dado para informar. Se depender destes dados não dá para tirar nada.” (B04,462) e “[...]só não tinha as informações que deveria ter.” (B03,419) respectivamente, mostrando uma séria dificuldade de completude nos dados do Ministério do Turismo.

Quadro 11 – Completude

ID_Sen	Conj	Entrev.	Q	TNat	TObj	TCat				
						Desejo	Negativa	Positiva	Neutra	Total
49	ENADE	A01	Q5	Constructo	Dados			1		1
71	ENADE	A01	Q8	Constructo	Planilha		1			1
74	ENADE	A01	Q8	Constructo	Dados		1			1
77	ENADE	A01	Q8	Constructo	Dados				1	1
117	ENCCEJA	A02	Q3	Constructo	Planilha		1			1
231	ANAC	A04	Q1	Constructo	Planilha		1			1
240	ANAC	A04	Q4	Constructo	Dados		1			1
242	ANAC	A04	Q4	Constructo	Dados		1			1
243	ANAC	A04	Q4	Constructo	Dados		1			1
277	PAC	A05	Q4	Constructo	Dados		1			1

ID_Sen	Conj	Entrev.	Q	TNat	TObj	TCat				
						Desejo	Negativa	Positiva	Neutra	Total
297	PAC	B01	Q3	Constructo	Dados		1			1
313	PAC	B01	Q5	Constructo	Dados		1			1
402	M.TUR	B03	Q2	Constructo	Dados		1			1
404	M.TUR	B03	Q2	Constructo	Dados		1			1
419	M.TUR	B03	Q5	Constructo	Dados		1			1
421	M.TUR	B03	Q6	Desejo	Dados	1				1
446	M.TUR	B04	Q1	Constructo	Células		1			1
461	M.TUR	B04	Q5	Constructo	Dados		1			1
462	M.TUR	B04	Q5	Constructo	Células		1			1
464	M.TUR	B04	Q6	Desejo	Dados	1				1
476	ANAC	B05	Q1	Constructo	Banco			1		1
493	ANAC	B05	Q2	Constructo	Planilha				1	1
502	ANAC	B05	Q2	Constructo	Planilha				1	1
620	KAGGLE	C02	Q3	Constructo	Dados			1		1
Total						2	16	3	3	24

Fonte: Autor.

4.6. TCon - Compreensibilidade

Colocada por Batini *et al.* (2009), este constructo tem só uma métrica: a compreensão do consumidor de dados do que está sendo representado no conjunto de dados.

Este constructo está ligado à capacidade de entendimento dos dados, o que é diferente de interpretabilidade. As métricas de interpretabilidade são unidades de valor corretas e definições claras, ou seja, decifrar o que há nos dados. Compreensibilidade, por outro lado, é a extensão de quão é possível, para o consumidor de dados, compreender, segundo Lee *et al.* (20), Wang e Strong (1996) e Batini *et al.* (2009).

Assim, mesmo que uma planilha disponibilize todas as informações para interpretar que há “100” “chapas de 1/3 de polegadas” de “Poliacrilato de metila” na “localidade” “Polo Industrial Luiz Cavalcante”, isso não quer dizer que o leitor compreenda o que é uma placa de poliacrilato de metila (Placa de acrílico), o que é ou quanto é “1/3 de Polegadas” ou onde é esse polo (Maceió-AL).

No caso, as TSen marcadas negativas em compreensibilidade têm em comum o fato de que não se compreendeu o que os dados significam, tendo-se de se utilizar, muitas vezes o serviço de busca do Google para entender o que exatamente essas definições e metadados queriam dizer. Veja-se, por exemplo: “A gente fez assim, a gente teve uma amostra do trabalho anterior, daí eu entendi o que era. Com cada, cada, parte, mas, eu não saberia assim a

banana nanica “P” tipo “A” eh... o que isso significa? ... Produtor? Entendeu? É uma coisa assim, e é muito difícil para uma pessoa que não é dali identificar o que é o quê. Então isso é questão de facilidade ou de impedimento, não era fácil não [entender]”. (A03,172) e “É... Eles poderiam tentar facilitar a alguém que não é do ramo a compreender aquilo que eles estão tentando trazer de informação. Acho que eles estão mais interessados em passar as informações daqueles que já estão no ramo. Que pessoas que não estão envolvidas. Clientes finais” (A04,253-254).

A única TSen positiva tem o mesmo viés, no entanto, ela só fala de quanto “[...]foi fácil de entender. ” (B02,345), sendo talvez uma característica única do banco de dados sobre o ENCCEJA. A interpretabilidade dos dados é essencial, mas, para a relação entre Principal e Agente, a compreensibilidade é igualmente valiosa, uma vez que a incapacidade de o Principal entender o que os dados significam elimina a capacidade dos dados de ser um sinal de confiança e a crença de que os dados refletem aquilo que haveria. Isso deixaria ainda dúvidas sobre as atividades do Agente.

Quadro 12 – Compreensibilidade

ID_Sen	Conj	Entrev.	Q	TNat	TObj	TCat		
						Negativa	Positiva	Total
171	CEPEA	A03	Q2	Constructo	Planilha	1		1
172	CEPEA	A03	Q2	Constructo	Planilha	1		1
176	CEPEA	A03	Q2	Constructo	Planilha	1		1
179	CEPEA	A03	Q3	Constructo	Dados	1		1
212	CEPEA	A03	Q8	Constructo	Dados	1		1
215	CEPEA	A03	Q8	Constructo	Células	1		1
253	ANAC	A04	Q7	Constructo	Dados	1		1
254	ANAC	A04	Q7	Constructo	Dados	1		1
255	ANAC	A04	Q7	Constructo	Dados	1		1
257	ANAC	A04	Q7	Constructo	Dados	1		1
320	PAC	B01	Q4	Constructo	Dados	1		1
347	ENADE	B02	Q2	Constructo	Dados		1	1
468	M.TUR	B04	Q7	Constructo	Banco	1		1
522	ANAC	B05	Q5	Constructo	Planilha	1		1
531	ANAC	B05	Q7	Constructo	Planilha	1		1
533	ANAC	B05	Q6	Constructo	Site	1		1
565	CEPEA	C01	Q2	Constructo	Planilha	1		1
Total						16	1	17

Fonte: Autor.

4.7. TCon - Concisão

Nessa dimensão, em Batini *et al.* (2009), a métrica da dimensão consiste em mostrar uma representação concisa das informações. A essa definição some-se a seguinte passagem: “Medida em que os dados são representados de forma compacta e sem sobrecarregar [o consumidor de dados]” (WANG; STRONG, 1996), do que se pode depreender que a ideia por trás do constructo é passar a maior quantidade de informações possíveis nos dados, sem dificultar seu entendimento.

Concisão teve maiores menções negativas e pedidos dentro dos dados do Ministério da Educação com A02, A01, B02, isso porque os dados eram tão sobrecarregados e com tantas linhas que separá-los se tornava uma tarefa de programação (improvável de realizar por graduandos em Administração). As sentenças que a avaliam são as seguintes. “[...]podiam usar algo que possa condensar os dados, naquilo que realmente interessa.” (A02, 144); “[é] obvio que dados do ENADE tem que ser grande, só que [eu] acho que dá para ficar mais condensado que um giga e meio. Não tinha como tirar.” (A02, 146); “Tinha dados que exigia a inscrição do autor em tal matéria ou vetor com respostas do candidato de tal matéria, esse tipo de coisa a gente não tem como extrair. Não dá.” (A02,140); “Em vez de ter ela com a sigla do estado, PB, PE, RS e tal, é um código numérico e aí se quiser saber, qual era o estado, teria que ir no dicionário, ver qual era.”(A01,19); “A gente abriu o arquivo com um programa e dividiu a tabela, pois era muito grande, e o Qlick não aceitava aquele tamanho de tabela. A gente precisou fazer isso em python, criou o programa, dividiu a tabela em duas e salvou.” (B02,360).

A03 e B05, que trabalharam com os bancos do CEPEA e ANAC, respectivamente, mencionaram que alguns dados vieram com cabeçalhos com células mescladas, o que dificultou o trabalho com os dados: “Isso, porque ele tem um esforço de na hora de pegar os dados, organizar, a tabela ela vem toda completinha, só tem alguns detalhes, da visão os dados vindo todo juntos, e essas coisas que atrapalha entender esses dados. ” (A03,214). É interessante notar também que nem C01 nem C02 mencionaram o constructo de concisão, o que, a julgar das sentenças dos entrevistados, não foi exatamente um problema, mesmo que A03 tenha notado essa dificuldade como referenciado acima. Outro fato interessante foi que esse foi o constructo com mais natureza de “desejo” em comparação ao total, demonstrando que é essencial para o entendimento dos dados, logo o que justifica seu papel na TPA.

Na TPA, a falta de concisão dificulta a redução da assimetria de informação, não pela falta de dados, mas pelo excesso de detalhe ou prolixidade; daí se gera dificuldade de obtenção de informações, pois o cérebro do Principal só consegue aguentar um número

limitado de dados e informações ao mesmo tempo. Concisão tem um relacionamento próximo com credibilidade, e não só nas sentenças e nos motivos. Ainda assim, mesmo que o entrevistado comente dificuldade de obter informação por falta de concisão, isso não parece afetar a credibilidade.

Pode-se conjecturar que isso se deve à natureza da concisão. A falta de qualidade em concisão demonstra ao Principal que os dados estão presentes, mas não facilmente compreensíveis, portanto a desconfiança do Principal é resolvida, mesmo sem atenuar exatamente a assimetria de informação, o que permitiria esconder desvios de interesse e conduta dos órgãos/funcionários públicos mais facilmente, algo que se retratou em Choi (2018), na China, um país conhecido pela prevalência de dados quanto à forma em detrimento da função.

Quadro 13 – Concisão

ID_Sen	Conj	Entrev.	Q	TNat	TObj	TCat			
						Desejo	Negativa	Positiva	Total
19	ENADE	A01	Q2	Constructo	Dados		1		1
135	ENCCEJA	A02	Q6	Desejo	Dados	1			1
136	ENCCEJA	A02	Q6	Desejo	Dados	1			1
140	ENCCEJA	A02	Q6	Desejo	Dados	1			1
141	ENCCEJA	A02	Q6	Desejo	Dados	1			1
144	ENCCEJA	A02	Q6	Constructo	Dados		1		1
145	ENCCEJA	A02	Q6	Desejo	Dados	1			1
149	ENCCEJA	A02	Q7	Desejo	Dados	1			1
214	CEPEA	A03	Q8	Constructo	Células			1	1
360	ENADE	B02	Q6	Constructo	Dados		1		1
482	ANAC	B05	Q1	Constructo	Planilha		1		1
515	ANAC	B05	Q4	Constructo	Planilha			1	1
Total						6	4	2	12

Fonte: Autor.

4.8. TCon - Consistência

A dimensão Consistência é definida pelas seguintes métricas em Batini *et al.* (2009): a divisão dos números dos valores consistentes entre si, pelo número de valores totais (WANG, 1998); o número de tûplas que violam as restrições/números de codificação diferentes (ENGLISH, 2002); número de páginas com desvio de estilos-guia (EPPLER; MUNZENMAIER, 2002); e a percepção dos consumidores de dados sobre representação consistente (LEE *et al.*, 2002). Disso se conclui que o constructo por trás da dimensão são desvios de padrão entre as unidades de dados relacionadas, ou seja, consistência das células.

O constructo Consistência é um dos que mais apresentaram menções de T_{Nat} negativa, tendo várias críticas à uniformidade dos dados. A maior parte destas críticas foi tipificada em inconsistência temporal entre múltiplas iterações nos dados no ENADE, conforme se lê em: “e for projetar que uma coisa ruim é que eles mudavam... éh... alguns padrões com os códigos né? As tabelas, ah, em vários anos...não era... não era uma coisa muito constante, ” (A01, 37); no ENCCEJA, em: “Mas teve dificuldade com o formato que variava. ” (A02,125). O mesmo ocorre no CEPEA, em: “Então era tudo mais ou menos do mesmo jeito, só que eles não colocavam todos os dados na mesma coluna, então, você tinha que ver um por um para saber o que estava faltando nos dados.” (A03,189).

Também houve inconsistências entre células de uma mesma coluna ou linha, nos dados do PAC, em: “não é necessário você colocar um número exatamente. Contanto que você coloque uma *string* padronizada Tipo ‘saneamento’, e não ter aqui como ‘saneamento básico’ ‘saneamento A’, tem que ser bem definido isso” (B01, 315); e no M.TUR: “a maioria...a maioria não... uma boa parte assim, entre as células tinha a sigla PB e [em] outros tinha ‘Paraíba’.” (B03, 500). ANAC foi a única com um elogio à consistência: “nessa planilha final que a gente utilizou, [es]tava bem organizado [em relação a de consistência]. Muito mesmo.” (B05,512). O desejo de A01, é somente para sanar o problema acima citado.

Se conjectura que: A04 não aparece, provavelmente por não querer elogiar ou mencionar e considerar a consistência subentendida. B02 não aparece, provavelmente porque, como ele mesmo disse nas questões sobre o grupo de trabalho, ele ficou com a parte da programação, não na coleta de dados, e C02 simplesmente parece considerar os dados como consistentes.

Consistência tem também um grande valor para a relação Principal e Agente. Sendo um dos conceitos em relação ao formato de dados, ele é também estritamente necessário para a tomada de decisão e obtenção de informação pelo Principal. Um ser humano não se incomoda de ler dados um pouco inconsistentes, mesmo que seja um por um. Apesar disso, esse é um processo ineficiente, então atualmente se usam métodos computacionais. Esses métodos, segundo Eppler e Munzenmaier (2002), são sensíveis a inconsistências, fazendo a qualidade neste quesito indispensável para obtenção da informação pelo Principal, ainda mais por meio eletrônico, o formato pelo qual se disponibilizam dados abertos governamentais em maior parte.

ID_Sen	Conj	Entrev.	Q	TNat	TObj	TCat				
						Desejo	Negativa	Positiva	Neutra	Total
32	ENADE	A01	Q3	Constructo	Células				1	1
37	ENADE	A01	Q3	Constructo	Dados		1			1
53	ENADE	A01	Q6	Desejo	Dados	1				1
60	ENADE	A01	Q6	Constructo	Dados		1			1
125	ENCCEJA	A02	Q4	Constructo	Planilha		1			1
187	CEPEA	A03	Q4	Constructo	Banco		1			1
189	CEPEA	A03	Q4	Constructo	Banco		1			1
191	CEPEA	A03	Q4	Constructo	Células		1			1
193	CEPEA	A03	Q4	Constructo	Células		1			1
199	CEPEA	A03	Q6	Desejo	Dados	1				1
204	CEPEA	A03	Q5	Constructo	Dados		1			1
206	CEPEA	A03	Q5	Constructo	Dados		1			1
284	PAC	A05	Q5	Constructo	Dados		1			1
315	PAC	B01	Q5	Constructo	Dados		1			1
326	PAC	B01	Q8	Constructo	Dados		1			1
334	PAC	B01	Q7	Constructo	Dados		1			1
400	M.TUR	B03	Q2	Constructo	Dados		1			1
401	M.TUR	B03	Q2	Constructo	Dados		1			1
450	M.TUR	B04	Q2	Constructo	Células		1			1
451	M.TUR	B04	Q2	Constructo	Banco		1			1
512	ANAC	B05	Q4	Constructo	Planilha			1		1
586	CEPEA	C01	Q5	Constructo	Banco		1			1
587	CEPEA	C01	Q5	Constructo	Banco		1			1
Total						2	19	1	1	23

Fonte: Autor.

4.9. TCon - Conveniência

Essa dimensão em Batini *et al.* (2009) vem de Eppler e Munzenmaier (2002). Este constructo avalia quão difícil é, para o consumidor de dados, obter os dados. Batini *et al.* (2009) definem sua única métrica como o número de “caminhos de navegação difíceis” para o consumidor de dados, como, por exemplo, quão diretamente o consumidor de dados chega aos dados que lhe interessam, sem erros ou dificuldade de os encontrar.

O constructo obtido considerou a percepção de dificuldade ou confusão para achar planilhas e ou páginas (conjuntos de dados), como se verifica nos seguintes depoimentos: “[...] para encontrar os dados no site, a gente teve muita dificuldade em saber qual os dados

teriam base para analisar ou não. Estava tudo espalhado” (A4,228); no caso de A1, somente o fato de ter tantos dados tornou difícil navegar e encontrar um conjunto de dados específico: “A gente não viu outros por causa da própria quantidade de dados. Muitos dados. Todos separados.” (A01,83). Como parte do constructo retirado de Eppler e Munzenmaier (2002) e Lee *et al.* (2002), também há falta de indicação sobre o conteúdo das planilhas, como aqui expresso: “[...] meu único problema seria dar um formato mais claro e acessível, ainda mais se fosse um cidadão comum, que não tivesse instrução profissional.” (A02,150) e novamente na ID_Sen 228.

Conveniência tem um efeito específico na TPA. O Principal, por não saber ao certo o que procurar, ou o que tem há nas planilhas, pela assimetria natural de informação descrita por Jensen e Meckling (1976), não teria onde obter informação, “sobrecarregando” o leitor, como visto no texto de Wand e Strong (1996), falando de concisão.

Quadro 15 – Conveniência

ID_Sen	Conj	Entrev.	Q	TNat	TObj	TCat			
						Desejo	Negativa	Positiva	Total
83	ENADE	A01	Q8	Constructo	Dados		1		1
150	ENCCEJA	A02	Q7	Constructo	Dados	1			1
210	CEPEA	A03	Q7	Desejo	Dados	1			1
228	ANAC	A04	Q1	Constructo	Site		1		1
393	M.TUR	B03	Q1	Constructo	Site			1	1
478	ANAC	B05	Q1	Constructo	Site		1		1
526	ANAC	B05	Q5	Constructo	Banco		1		1
Total						2	4	1	7

Fonte: Autor.

4.10. TCon - Credibilidade

Credibilidade, na definição de Batini *et al.* (2009) e Lee *et al.* (2002), tem duas métricas, uma lidando com a extensão de quanto o consumidor de dados “acredita”, “confia” nas informações e lhe “dá credibilidade” (LEE *et al.*, 2002) e outra com relação a qual o “número de túplas com valores padrão” (BATINI *et al.*, 2009). Disso se retirou que o conceito por trás da dimensão credibilidade é a confiança nos dados, e isso foi o procurado nas falas dos entrevistados.

Assim como no caso da Acessibilidade, a Credibilidade é expressamente referenciada em uma das perguntas no roteiro de pesquisa, especificamente, na Q3. Esperou-se, portanto, que credibilidade fosse mencionada ao menos uma vez por entrevista, o que ocorreu neste constructo, mas não em Acessibilidade, como já vimos.

Um dos pontos relevantes foram as várias referências à Credibilidade como sendo positiva ou neutra (no sentido de estar presente; não houve, entretanto, característica expressa de positividade). Com tais características, especificamente, a passagem “de bom, acho que é propriamente por ser confiável acho que é um ponto bom.” (A01,39), vê-se que se trata de uma declaração que vem logo após uma série de críticas à consistência dos dados e à facilidade dos dados. O mesmo acontece na passagem “então, a da gente, o que eu gostei foi a da transparência e da confiabilidade.” (A02,154). Isso mostra que, mesmo que o entrevistado veja dificuldades de qualidade de dados, isso não demonstra ter afetado a sua percepção de confiabilidade nos dados. Isso remete a um dos textos em que Fabrini (2018) discute que o consumidor de dados abertos governamentais brasileiro já espera qualidade de dados inferior (em questão de forma, a saber completude e acessibilidade); se, por ventura, o mínimo de dados é atendido, já se cria confiança do consumidor de dados, mesmo que não se garanta atender à função de informar.

Esse efeito aparece nesta pesquisa também nos dados de B03 e B04, que descreveram vários problemas de qualidade de dados em questão de forma e, por consequência, responderam muito duramente a esse item. Isso se comprova no julgamento de B04, que considera a confiabilidade como “Péssimo. Com tanto dinheiro ali indo para tanto lugar, como se tem confiança com dados tão ruins? ” (B04,454), “Eu acho que [os dados] não era nem um pouco confiável” (B03, 406).

Há, porém, também um contraponto em B01 e B02, que estudaram os dados do PAC e do ENADE, respectivamente. Esses respondentes, mesmo verificando fontes de dados que apresentaram poucas dificuldades de qualidade em questão de forma, ainda assim demonstraram forte falta de confiança: “Não dá para ter confiabilidade” (B01, 325); e “não muito boa, por conta disso tudo, mas o suficiente para seguir essas regras eu acho.” (A05 e A01). Mesmo quando seus colegas de conjunto de dados, A05 e A01, tiveram uma resposta menos negativa, como em “bem sim, não vi qualquer motivo para duvidar dos dados ali...”, percebe-se, talvez, um viés mais como manifestação de desgosto em relação à administração pública do que uma análise real dos dados, fazendo que os respondentes B01 e B02 respondessem negativamente, do que se conclui que se faz necessário mais investigação sobre esse efeito.

A credibilidade obviamente afeta a TPA, pois, como na visualização do contexto teórico, a credibilidade é uma métrica também de desconfiança. Para todos os efeitos, credibilidade é a métrica da efetividade dos dados em atenuar a desconfiança de um Principal, no caso, o entrevistado. Seria com base nesse constructo, portanto, que se avaliaria a

efetividade dos esforços de governança, do agente e de qualidade de dados. Mostrar credibilidade é o que atenuaria a desconfiança. Com apenas 14 (56%) das sentenças retratando algum tipo de confiabilidade nos dados, é difícil, entretanto, afirmar que os dados estão servindo em sua função para atenuar desconfiança.

Quadro 16 – Credibilidade

ID_Sen	Conj	Entrev.	Q	TNat	TObj	TCat				
						Desejo	Negativa	Positiva	Neutra	Total
39	ENADE	A01	Q3	Constructo	Dados			1		1
116	ENCCEJA	A02	Q3	Constructo	Planilha				1	1
130	ENCCEJA	A02	Q5	Constructo	Dados			1		1
154	ENCCEJA	A02	Q8	Constructo	Dados				1	1
155	ENCCEJA	A02	Q8	Constructo	Dados				1	1
180	CEPEA	A03	Q3	Constructo	Dados				1	1
209	CEPEA	A03	Q7	Constructo	Dados				1	1
236	ANAC	A04	Q3	Constructo	Dados				1	1
273	PAC	A05	Q3	Constructo	Dados				1	1
299	PAC	B01	Q3	Constructo	Dados		1			1
325	PAC	B01	Q8	Constructo	Dados		1			1
333	PAC	B01	Q7	Constructo	Dados		1			1
336	PAC	B01	Q7	Constructo	Dados		1			1
371	ENADE	B02	Q3	Constructo	Dados		1			1
406	M.TUR	B03	Q3	Constructo	Dados		1			1
416	M.TUR	B03	Q5	Constructo	Dados		1			1
432	M.TUR	B03	Q7	Desejo	Dados	1				1
454	M.TUR	B04	Q3	Constructo	Dados		1			1
469	M.TUR	B04	Q7	Constructo	Banco		1			1
471	M.TUR	B04	Q8	Constructo	Dados		1			1
504	ANAC	B05	Q3	Constructo	Planilha				1	1
570	CEPEA	C01	Q3	Constructo	Planilha				1	1
619	KAGGLE	C02	Q3	Constructo	Dados			1		1
635	KAGGLE	C02	Q6	Constructo	Dados			1		1
644	KAGGLE	C02	Q8	Constructo	Dados			1		1
Total						1	10	5	9	25

Fonte: Autor.

4.11. TCon - Exatidão

Precisão é um dos constructos mais repetidos em métodos de mensuração de qualidade, pois é essencial para trabalhar com os dados (WANG; STRONG, 1996), e dados

não precisos não refletem a realidade, portanto não têm significado (ENGLISH, 2003). Há, contudo, uma diferença em como mensurar Precisão e Exatidão, uma vez que esses constructos não são iguais. Enquanto Precisão avalia se os dados presentes na tabela refletem os reais valores na realidade, a qual se tem que mensurar os valores dos dados para ver se se igualam aos presentes no conjunto de dados, Exatidão, por outro lado, foca na percepção que o usuário tem da precisão dos dados. A Exatidão não se importa exatamente com a margem de erro em si, mas sim se o consumidor de dados aceita e ou percebe esse erro, como dizem Batini *et al.* (2009) e Lee *et al.* (2002).

Um exemplo seria que os dados sobre as finanças de um projeto ao arredondar os valores dos gastos por centenas de reais ao invés de refletir até os centavos, e/ou estes dados terem alguns erros na casa das dezenas de reais, esses dados não tem precisão, mas para o gestor que está lendo estes gastos esses dados podem ter exatidão, uma vez que ele sabe que foi arredondado e erros refletem o suficiente para ser aceito. Pode haver erros, também, por parte do consumidor de dados ao não reconhecer que há dados erros ou arredondamentos.

Como nenhum dos estudantes conhecia o valor real do que era representado pelos dados, não foi possível dizer com certeza a Precisão dos dados, nem seria possível sem algum deles ter relação com o órgão estudado, embora as respondentes A04 e B05 reconhecessem que os dados pareciam condizer com o que elas conheciam da realidade, a saber: “o que esperar de tráfego aéreo a partir dos estados” (A04, 237). Estas percepções podem não ser corretas, mas ainda assim refletem exatidão com a métrica “*This information is reliable*” de Lee *et al.* (2002) para Exatidão.

Quadro 17 – Exatidão

						TCat	
ID_Sen	Conj	Entrev.	Q	TNat	TObj	Neutra	Total
237	ANAC	A04	Q3	Constructo	Dados	1	1
507	ANAC	B05	Q3	Constructo	Planilha	1	1
Total						2	2

Fonte: Autor.

4.12. TCon - Facilidade de Operação

Facilidade de operação é definido em Batini *et al.* (2009) e Lee *et al.* (2002) como sendo a extensão em que os dados podem ser manipulados para atender às necessidades do consumidor de dados.

Dados do Ministério da Educação, ou seja, do ENADE e ENCCEJA, tiveram uma grande falha de facilidade de operação. Eles eram extremamente extensos, o que não é

exatamente uma falha de Facilidade de Operação, segundo em *Batini et al.* (2009), mas uma falha dos alunos por não usarem os programas certos para a tarefa, Como esta falta de qualidade foi, entretanto, percebida pelos alunos, se adicionou essa característica ao constructo Facilidade de Operação, devido à proximidade com o conceito e se estar estudando exatamente a percepção de qualidade de dados.

Devido a sua enorme dimensão, não foi possível abrir os dados acima citados em Microsoft Excel ou na plataforma OLAP escolhida, como relatado na sentença de A02 “[...] os dados eram em torno de um giga[byte], um giga[byte] e meio, e só em tentar tratar, e converter no excel, já travava tudo. Já era coisa muito complexa. E ficou inviável.” (87). Também se comprova tal dimensão em A01: “muitas linhas foi... hum... um milhão e setecentas mil linhas” (89). Uma situação parecida ocorreu com os respondentes que analisaram os Conj CEPEA, PAC e MTUR.

Os respondentes do PAC e CEPEA falaram de dificuldades de trabalhar com conjuntos de dados que tinham informações com células mescladas: “Tinha também que ter a localização onde foi feita, quando tinha mais de um estado, ficava tudo colado no mesmo espaço, então tivemos que separar à mão.” (A05, 282); sobre o banco de dados do M.TUR, verificou-se um simples pedido: para que acertassem outras falhas em qualidade para evitar mais problemas. Assim se expressou A05, 282: “[...]eu pediria para acertarem isso. Não tem condições de trabalhar ou investir assim”.

A facilidade de operação tem relação forte com a TPA e com a qualidade de dados. Dados que não são facilmente operacionalizáveis são tão inúteis quanto os dados não compreensíveis, uma vez que, sem operacionalização, a informação fica presa atrás de um método ineficiente de obtenção de dados, o que pode não afetar alguns dos Principais que não se debruçam sobre os dados, mas que afeta amargamente o trabalho daqueles que o fazem. A dificuldade de operação também abre chances para erro humano, e, portanto, também ao equívoco, algo que afeta mais a confiança do Principal que a simples falta de esforço para produzir os dados.

Quadro 18 – Facilidade de Operação

ID_Sen	Conj	Entrev.	Q	TNat	TObj	TCat			
						Negativa	Positiva	Neutra	Total
22	ENADE	A01	Q2	Constructo	Dados			1	1
38	ENADE	A01	Q3	Constructo	Dados	1			1
87	ENADE	A01	Q8	Constructo	Planilha	1			1
89	ENADE	A01	Q8	Constructo	Planilha	1			1

ID_Sen	Conj	Entrev.	Q	TNat	TObj	TCat			
						Negativa	Positiva	Neutra	Total
124	ENCCEJA	A02	Q4	Constructo	Dados			1	1
143	ENCCEJA	A02	Q6	Constructo	Dados	1			1
170	CEPEA	A03	Q2	Constructo	Planilha	1			1
281	PAC	A05	Q4	Constructo	Células	1			1
286	PAC	A05	Q6	Constructo	Dados	1			1
352	ENADE	B02	Q4	Constructo	Dados			1	1
354	ENADE	B02	Q4	Constructo	Dados		1		1
362	ENADE	B02	Q6	Constructo	Dados	1			1
409	M.TUR	B03	Q4	Constructo	Dados			1	1
430	M.TUR	B03	Q7	Constructo	Dados	1			1
Total						9	1	4	14

Fonte: Autor.

4.13. TCon - Interatividade

Interatividade tem seu conceito em Eppler e Munzenmaier (2002). Esse conceito tem como métrica o suporte a modificações nos dados, para facilidade de seu entendimento, seja por questionário ou por implantação de filtros. Os respondentes que mencionaram este constructo elogiaram sua existência para filtrar os dados, como se lê em: “Ah sim, muito útil, eles tinham um... Um programa no site que ajudava a escolher que dados você podia pegar.”(C01,576) do CEPEA; ou falaram de uma plataforma para que pudessem visualizar/filtrar os microdados, como no Desejo em: “Agora um ponto interessante que a gente encontrou, não de planilha, mas no site mesmo, foi que tinha um mapazinha lá e você clicava, numa região, numa cidade, daí ele mostrava de um jeito bem interessante. Gostaria de mais dados assim” (B05, 516) da ANAC; ou do processo de pedido de abertura de dados (cujo entrevistado falou não saber como funcionava): “Sei que eles tem um negócio assim, você pode pedir os dados, e eles tem um período para eles mandarem para você.” (B03, 431).

Os dados do Agente, segundo Mitnick (1973), não funcionam para resolver a desconfiança do Principal quando este não consegue extrair as informações dos dados. Isso inclui dados faltando, como nos constructos de Completude e Quantidade adequada de dados ou mesmo a sobrecarga de dados lidando com os constructos Concisão, Conveniência e Facilidade de operação. Uma plataforma para filtragem e requerimentos seria a chave para melhorar todas estas métricas, inclusive no próprio constructo de Interatividade. Uma plataforma online que possibilite a filtragem e pesquisa dentro de um banco de dados não é difícil de fazer (por experiência do autor e das aulas de SI&D), e já existem plataformas,

como essas descritas, disponíveis a gestores públicos no portal dados.gov.br (PPDA, 2018). Assim, não há motivo para não se mensurar e se ter um bom grau neste constructo, para quem estiver interessado na qualidade de dados do Principal, evidentemente.

Quadro 19 – Interatividade

ID_Sen	Conj	Entrev.	Q	TNat	TObj	TCat				
						Desejo	Negativa	Positiva	Neutra	Total
516	ANAC	B05	Q4	Constructo	Planilha				1	1
518	ANAC	B05	Q4	Desejo	Planilha	1				1
576	CEPEA	C01	Q4	Constructo	Site			1		1
594	CEPEA	C01	Q7	Desejo	Dados	1				1
Total						2	1	1	1	5

Fonte: Autor.

4.14. TCon - Interpretabilidade

Como dito anteriormente no TCon Compreensibilidade, a Interpretabilidade tem como conceito base que os dados sejam interpretáveis, isto é, ela avalia se os dados têm a linguagem, a documentação e o formato adequados, para que se possam decifrar os dados, e como eles se relacionam, mas não necessariamente compreender as ideias por trás desses dados (LEE *et al.*, 2002; WANG; STRONG, 1996; BATINI *et al.*, 2009).

Compreensibilidade e Clareza são dependentes de Interpretabilidade. Com baixa Interpretabilidade, é muito difícil compreender algum conjunto de dados ou deles abstrair informações. O mesmo acontece com a Clareza. Considerado isso, Interpretabilidade é indispensável para as relações descritas em TPA. Ora, se Interpretabilidade é necessário para fazer dados se tornarem informação (WANG; STRONG, 1996), é também a base para outras métricas de qualidade de dados e informações (WANG; STRONG, 1996), tendo em vista que dados são um veículo para transmissão de informações (SANTOS, 2016), e informação é o que afeta a desconfiança do Principal (MITNICK, 1973). Então, Interpretabilidade é um dos principais constructos para amenizar a desconfiança do Principal. Mesmo assim, é o que mais falha. Isso demonstra uma falta de efetividade dos dados abertos governamentais brasileiros de seguir suas funções, quais sejam: informar, reduzir ou eliminar a desconfiança do Principal e ter dados de qualidade. Em nossas pesquisas, esse constructo se mostrou, em maior parte, negativo.

Quadro 20 – Interpretabilidade

ID_Sen	Conj	Entrev.	Q	TNat	TObj	TCat				
						Desejo	Negativa	Positiva	Neutra	Total
5	ENADE	A01	Q1	Constructo	Dados				1	1
9	ENADE	A01	Q2	Constructo	Dados		1			1
10	ENADE	A01	Q2	Constructo	Dados				1	1
24	ENADE	A01	Q2	Constructo	Dados		1			1
43	ENADE	A01	Q4	Constructo	Dados		1			1
64	ENADE	A01	Q7	Constructo	Dados		1			1
169	CEPEA	A03	Q2	Constructo	Dados		1			1
233	ANAC	A04	Q2	Constructo	Planilha		1			1
259	ANAC	A04	Q6	Desejo	Dados	1				1
260	ANAC	A04	Q6	Desejo	Dados	1				1
269	PAC	A05	Q2	Constructo	Dados			1		1
303	PAC	B01	Q2	Constructo	Dados				1	1
319	PAC	B01	Q4	Constructo	Dados		1			1
343	ENADE	B02	Q1	Constructo	Dados			1		1
351	ENADE	B02	Q4	Constructo	Dados		1			1
484	ANAC	B05	Q1	Constructo	Planilha		1			1
486	ANAC	B05	Q1	Constructo	Planilha		1			1
487	ANAC	B05	Q1	Constructo	Planilha		1			1
528	ANAC	B05	Q7	Desejo	Dados	1				1
529	ANAC	B05	Q7	Constructo	Banco		1			1
530	ANAC	B05	Q7	Constructo	Planilha		1			1
572	CEPEA	C01	Q3	Constructo	Planilha		1			1
579	CEPEA	C01	Q4	Constructo	Planilha		1			1
583	CEPEA	C01	Q5	Constructo	Planilha		1			1
591	CEPEA	C01	Q6	Desejo	Dados	1				1
608	KAGGLE	C02	Q1	Constructo	Dados		1			1
615	KAGGLE	C02	Q2	Constructo	Dados		1			1
624	KAGGLE	C02	Q4	Constructo	Dados		1			1
625	KAGGLE	C02	Q4	Constructo	Dados		1			1
Total						4	20	2	3	29

Fonte: Autor.

4.15. TCon - Prestação de Contas

Prestação de contas, “*Accountability*”, ou Responsabilização, é definida da forma mais simples: como ter agentes responsáveis pelas suas próprias ações e líderes/gestores públicos sendo ultimamente responsáveis a suas populações para ações do governo (BANCO MUNDIAL, 1992). Este também é um dos três conceitos fundamentais da “Boa Governança”

(BANCO MUNDIAL, 1992), que, apesar do nome, não se trata exatamente apenas de uma governança boa, mas sim de vários processos para atingi-la.

O conceito procurado nas sentenças dos respondentes foram, portanto, as sentenças que continham constructos relacionados a essa responsabilidade e ao processo de se responsabilizar por desvios de interesse e comportamento do agente ou por desvios de seus integrantes, mais especificamente, o grau de responsabilização percebido pelo consumidor dos dados. Das respostas, 2/3 delas retrataram esforços de governança com característica positiva, mas em todas as sentenças, com exceção de uma (597), fala-se de prestação de contas por parte de subordinados do publicador de dados, mas da responsabilidade dele próprio. Como falou A05: “a gente via que ele [ANAC] realmente monitorava as atividades que essas empresas realizavam em questão aos voos.” (A05, 251). Basicamente a ANAC era o agente que mantinha outros agentes se responsabilizando, assim como colocando seus dados – dos outros agentes – para o Principal (os cidadãos). No caso, a responsabilidade pela prestação de contas é das empresas de avião e dos dados, não da ANAC, a publicadora dos dados.

A única sentença positiva que falou da prestação de contas do órgão publicador foi a seguinte: “Aparentemente, eles estão fazendo o trabalho bem deles, se responsabilizando” (C01, 597) em que o conceito foi nomeado, mas não definido, pois aqui o respondente queria dizer que eles estão transparecendo informação.

Por outro lado, as sentenças de conteúdo negativo demonstravam quanto se tenta escapar da responsabilidade de informar o Principal: “acho que não [tinham a intenção de informar], só acho que eles não tinham os dados voltados a dar informação.” (B01,338) “[se houvesse] você usaria os dados para descobrir exatamente porque eles não estão sendo usados [como papel de prova de serviço]” (B03, 434).

Sem ter como responsabilizar os agentes que não fazem seu trabalho, não é possível haver uma boa relação Principal-Agente. E sinais de Agentes tentando fugir a essa responsabilidade são piores ainda. Uma forma de avaliar a fuga à responsabilidade é necessária, assim como encontrar o responsável. No Quadro 23, vê-se o levantamento referente à responsabilização.

Quadro 21 – Prestação de Contas

ID_Sen	Conj	Entrev.	Q	TNat	TObj	TCat		
						Negativa	Positiva	Total
152	ENCCEJA	A02	Q8	Constructo	Dados		1	1
251	ANAC	A04	Q5	Constructo	Dados		1	1

ID_Sen	Conj	Entrev.	Q	TNat	TObj	TCat		
						Negativa	Positiva	Total
262	ANAC	A04	Q8	Constructo	Dados		1	1
338	PAC	B01	Q7	Constructo	Dados	1		1
436	M.TUR	B03	Q8	Constructo	Dados	1		1
597	CEPEA	C01	Q8	Constructo	Dados		1	1
Total						2	4	6

Fonte: Autor.

4.16. TCon - Quantidade Adequada de Dados

Como dito em Integralidade, a Quantidade Adequada de Dados (QAD) se volta em medir se os dados cobrem as “unidades de dados necessárias” (BATINI, *et al.* 2009) para o consumidor de dados. A dimensão original só detinha essa métrica (de modo quantitativo) e, por consequência, o conceito a essa dimensão subjacente é o mesmo.

Em casos de planilhas do tipo “.xlsx”, se representariam dados que necessitariam de mais “colunas”, a fim de se reproduzir, graficamente (??), maior quantidade adequada de dados, mas, interessantemente, os autores Batini *et al.* (2009) aparentam considerar granularidade como parte de QAD também, ou seja, aumentar o nível de detalhamento de um dado (granularidade) faz parte de uma QAD. Batini *et al.* não descrevem esse raciocínio diretamente, mas sim com auxílio das tabelas e a partir do fato de ela ser representada várias vezes nas metodologias que eles pesquisaram.

Esta pesquisa concorda com essa visão e, portanto, considerou granularidade como fazendo parte deste constructo também. As respostas pareceram se dividir quase perfeitamente em três grupos, um referente aos dados de TCat positiva e neutras, um para as negativas e um para os desejos.

Sobre os dados de TCat negativo, como ANAC, PAC e M.TUR, em todas as sentenças se percebe que houve falta de abrangência de uma ou mais áreas necessárias nos dados, como, por exemplo: “[os dados] não estavam na planilha e também a gente não encontrou nenhum banco de dados que coincidissem com o que a gente pesquisou, porque a única forma que ela mostrava era aquela planilha” (A04,245); os TNat com “ND” que eram do CEPEA apenas pediam por mais detalhamento ou granularidade dos dados como expresso em: “ehh...talvez, foi um dos pontos que o professor abarcou em relação as outras informações, eles poderiam botar essas outras informações junto à de preço. Quando fomos olhar essas informações de outras coisas, não tinha além do preço” (A03,221); e “E: eh, mas não é bom adicionar mais informação, tipo o preço por kilo ou sei lá.” O entrevistado A01, do ENADE, e o entrevistado B03, do M.TUR, na ID_Sen 418, demonstraram que os

conjuntos de dados tinham o que consideravam detalhamento mínimo. Exemplo: “Rapaz; Boa, boa, mesmo? Bem, os dados estavam extremamente completos, para cada ano, acho que a gente foi o que teve mais... é... o grupo que teve mais teve dados. Em volume de dados. A gente praticamente encheu da-do limite do upload do Qlick. Mas assim de coisa boa, diria que foi muitos dados” (A01,34-35).

Quantidade adequada de dados é essencial para quem quer abstrair informação para ter confiabilidade. Uma quantidade de dados com todo o conhecimento e atos do Agente é supostamente uma situação ideal, embora, além de ser humanamente impossível, isso sobrecarregaria o consumidor e o produtor de dados que tentasse relatar tanto. Mesmo assim, quando o agente é o Governo, essa burocracia é necessária.

Mesmo que Batini *et al.* (2009) supusessem que granularidade de dados fosse parte de QAD, a granularidade é mais importante para a qualidade de dados em TPA que em TI. De jeito nenhum a TPA espera que a assimetria de informação seja totalmente sanada, mas apenas o suficiente para evitar altos graus de desconfiança. A granularidade é, portanto, necessária. Infelizmente os alunos entrevistados não sabiam a ideia do conceito, o que não permitiu tal discussão aqui, mas esta suposição da importância de granularidade é interessante para futuras pesquisas.

Quadro 22 – Quantidade adequada de dados

ID_Sen	Conj	Entrev.	Q	TNat	TObj	TCat				
						Desejo	Negativa	Positiva	Neutra	Total
34	ENADE	A01	Q3	Constructo	Dados				1	1
35	ENADE	A01	Q3	Constructo	Dados			1		1
221	CEPEA	A03	Q8	Desejo	Dados	1				1
224	CEPEA	A03	Q8	Desejo	Dados	1				1
245	ANAC	A04	Q4	Constructo	Dados		1			1
246	ANAC	A04	Q4	Constructo	Dados		1			1
283	PAC	A05	Q5	Constructo	Dados		1			1
311	PAC	B01	Q5	Constructo	Dados		1			1
312	PAC	B01	Q5	Constructo	Dados		1			1
418	M.TUR	B03	Q5	Constructo	Dados				1	1
438	M.TUR	B03	Q8	Constructo	Planilha		1			1
441	M.TUR	B03	Q8	Constructo	Planilha		1			1
459	M.TUR	B04	Q4	Constructo	Células		1			1
589	CEPEA	C01	Q5	Desejo	Dados	1				1
Total						3	8	1	2	14

Fonte: Autor.

4.17. TCon - Transparência

O constructo transparência assim como prestação de contas é definido a partir dos constructos de boa governança descrevidas pelo Banco Mundial (1992), mais especificamente o nome completo do constructo é Transparência de Informação. No contexto desta pesquisa, tal constructo foi reconhecido pelo grau de transparência percebido no órgão pelos entrevistados.

Neste quesito, A02 e C01 falam como o ENCCEJA e o CEPEA aparentavam ser transparente para todos, incluindo, entre outros motivos, o excesso de informação, como se constata na passagem: “pois [os dados] têm uma transparência muito grande” (A02,153). O entrevistado A02, entretanto, avisa: “Ele [o MEC] falava mais em relação à prova em si, do órgão a gente não conseguiu muito, a gente conseguiu mais trabalhar no[s dados do] ENCCEJA mesmo” (A02,132). Com base nesse depoimento, portanto, não há transparência total. Os entrevistados B03 e A05 apenas relatam quão deficientes são os dados para passar transparência, conforme se verifica em: “Fraco, os dados pareciam pouco para tudo aquilo. Também que tinha esses problemas” (A05,290).

Esses dados são interessantes para a relação entre Principal e Agente, pois o sentimento de que os dados estão transparentes é um requisito atenuador de desconfiança do Principal; sem o sentimento de transparência nos dados, o Principal pode, ao contrário do que se deseja, aumentar sua desconfiança, por acreditar que o Agente está tentando esconder algo.

Quadro 23 –Transparência

ID_Sen	Conj	Entrev.	Q	TNat	TObj	TCat		
						Negativa	Positiva	Total
132	ENCCEJA	A02	Q5	Constructo	Dados	1		1
148	ENCCEJA	A02	Q7	Constructo	Dados		1	1
153	ENCCEJA	A02	Q8	Constructo	Dados		1	1
290	PAC	A05	Q8	Constructo	Dados	1		1
434	M.TUR	B03	Q8	Constructo	Dados	1		1
571	CEPEA	C01	Q3	Constructo	Planilha		1	1
Total						3	3	6

Fonte: Autor.

4.18. TCon - Singularidade

Singularidade é descrito como a não duplicação de dados, conforme explicam Batini *et al.* (2009), Wang (1998), Wang & Strong (1996) e Vertrò *et al.* A05 e B01 que fizeram o trabalho baseado nos dados relacionados ao PAC. Esse foi o único constructo em que se encontrou qualquer menção de singularidade, no caso sentenças de TCat negativas, a saber: “e também tinha que as vezes o nome de onde tinha sido feita a obra dentro do nome da obra, a sigla do estado ficava depois. E na coluna seguinte vinha a mesma informação.” (A05, 280). E também: “Por exemplo. Eles dão nome ao empreendimento, e depois eles colocam um código lá definido para obra de saneamento, como X Mil, tracinho, PB, e depois na coluna, Paraíba, aí, tipo, se já tem a identificação, para que usar o nome único se tem já a identificação?” (B01,314)

O maior efeito da não singularidade para o cidadão, como Principal, é que ficar relendo o mesmo dado, seja visualmente ou auxiliado por computador, causa retrabalho, perda de recursos e correção humana por vezes.

Quadro 24 – Singularidade

						TCat	
ID_Sen	Conj	Entrev.	Q	TNat	TObj	Negativa	Total
280	PAC	A05	Q4	Constructo	Dados	1	1
314	PAC	B01	Q5	Constructo	Dados	1	1
Total						2	2

Fonte: Autor.

4.19. TCon Emergente- Experiência

O primeiro e um dos poucos constructos emergentes ganhou TCon, pois é usado na sentença 356, em que o Principal menciona estar tentando eliminar a assimetria de informação inerente à TPA e ou à desconfiança. Só houve uma TSen com este constructo. Nela, o entrevistado explica o seguinte: “Para que esses dados informassem realmente alguém de alguma coisa, a pessoa precisaria trabalhar encima deles, fazer como a gente, ficar encima deles porque uma pessoa leiga, ‘não iria trabalhar os dados’ e pegasse a tabela não teria muita coisa informativa só teria dados realmente”. (B02, 356. sic)

Com base nisso é possível propor um constructo do tipo subjetivo, como em Lee *et al.* (2009), em que, junto de outros constructos subjetivos, o consumidor de dados evidenciaria várias métricas com respostas em Likert – as métricas para este suposto constructo poderiam ser as seguintes:

- O grau de experiência com os dados;
- O grau de experiência com pesquisas de dados num geral; e
- O grau de importância de retirar informação desses dados.

Esse constructo teria como finalidade identificar *outliers* nessas métricas e assim identificar quanto das métricas subjetivas está referenciada aos dados e quanto está relacionado à natureza da experiência, conhecimento e necessidade do Principal. Isso permitiria interpretar melhor outras métricas subjetivas.

Na aplicação da metodologia AIMQ, por exemplo, um indivíduo que assinalou “facilidade de operação” como sendo baixa pode ter mais de um motivo para essa resposta. Um desses motivos pode ser o fato de o consumidor de dados não ter instrução ou experiência para entender direito o que é “facilidade de operação” e, por não saber trabalhar com esse constructo, o acha de difícil operação. Por outro lado, um pesquisador extremamente experiente pode assinalar facilidade de operação no máximo, pois está acostumado com a qualidade dos dados presentes, inclusive com as partes ruins.

Identificar esses *outliers* e considerar seu papel é pertinente à qualidade de mensuração de quaisquer instrumentos (COSTA, 2011), ainda mais ao considerar que constructos subjetivos dependem das experiências, necessidades e conhecimentos do indivíduo (BATINI *et al.*, 2010) e, portanto, é de se supor que isso afete mensuração de qualidade de dados em questão de função também.

Este tipo de informação já é coletado, embora como informação adicional e não como parte da metodologia de mensuração de qualidade de dados ou informações. Isso diminui a capacidade dessas metodologias de mensurarem corretamente essas qualidades subjetivas, logo não se pode relegar isso a informações adicionais ou subentendidas, e daí nasce o constructo Experiência.

Quadro 25 – Experiência

						TCat	
ID_Sen	Conj	Entrev.	Q	TNat	TObj	Neutra	Total
356	ENADE	B02	Q5	Constructo	Dados	1	1
Total						1	1

Fonte: Autor.

4.20. TCon Emergente- Responsabilidade

Responsabilidade é um dos constructos emergentes, mas sua ideia advém de múltiplas áreas. Ele se assemelha aos conceitos de Rastreabilidade, Interatividade e Prestação de contas. Ele veio dos seguintes desejos de B03 “[eu iria] fazer um relatório de saber quem está fazendo

esse relatório. Quem é o responsável? ” (B03,422) e em “Sei que eles têm um negócio assim, você pode pedir os dados, e eles tem um período para eles mandarem para você de volta. ” (B03,431)

Essa descrição sobre o procedimento de identificar o responsável pelos dados e o processo de responsabilidade para com o publicador de dados é um desejo que faz muito sentido à TPA, Qualidade de dados e Boa governança. O desejo, todavia, não foi identificado como constructo nas leituras sobre essas áreas. A cartilha técnica para publicação de dados abertos (SLTI, 2019), no entanto, pede que os bancos de dados identifiquem o órgão responsável pela publicação e criação dos dados. Responsabilidade é um conceito amplo, logo esse constructo pode ser expandido para além dessa métrica.

Possíveis métricas para este constructo envolvem a checagem da presença dos seguintes critérios nos bancos de dados ou em seus metadados:

- I. Nomeação e contato do indivíduo/cargo, grupo e organização responsável pela produção dados;
- II. Se houver, nomeação e contato de indivíduo/cargo, grupo e organizações, internos ao publicador de dados, responsáveis pela qualidade dos dados;
- III. Se houver, nomeação e contato de indivíduo/cargo, grupo e organizações, externos ao publicador de dados , responsáveis pela qualidade dos dados;
- IV. Link ou detalhamento direto do processo passo a passo, e com limites de tempo de resposta, para reclamar/criticar a qualidade de dados e exigir melhoramentos.

O intuito deste constructo é que sempre se tenha como contatar o responsável, pois saber exatamente qual órgão fez um dado não quer dizer exatamente de quem é a responsabilidade; além disso, há a possibilidade de os indivíduos integrantes do publicador de dados ficarem só passando a responsabilidade ao invés de resolverem o problema. Algo semelhante a isso é descrito em Choi (2018), sobre a China. Isso pode ser evitado ao se descrever, no próprio documento, quem é o responsável, o que já foi feito, e o processo interno e externo de exigência de qualidade de dados. Para dados abertos governamentais, apenas o nome do autor não é suficiente, ainda mais se houver mais de um autor envolvido.

Quadro 26 –Responsabilidade

						TCat	
ID_Sen	Conj	Entrev.	Q	TNat	TObj	Desejo	Total
422	M.TUR	B03	Q6	Desejo	Dados	1	1
Total						1	1

Fonte: Autor.

4.21. Percepção de Qualidade de Dados à Luz da TPA

Quadro 27 – Sentenças por Constructo Expandida

TCon \ TCat	TCat				Negativa	Neutra	Positiva	Desejo	Total
	Negativa	Neutra	Positiva	Desejo					
Acessibilidade	14	5	3		63,64%	22,73%	13,64%	0,00%	22
Clareza			1	2	0,00%	0,00%	33,33%	66,67%	3
Completude	16	3	3	2	66,67%	12,50%	12,50%	8,33%	24
Compreensibilidade	16		1		94,12%	0,00%	5,88%	0,00%	17
Concisão	5		1	6	41,67%	0,00%	8,33%	50,00%	12
Consistência	19	1	1	2	82,61%	4,35%	4,35%	8,70%	23
Conveniência	5		1	1	71,43%	0,00%	14,29%	14,29%	7
Credibilidade	10	10	6	1	37,04%	37,04%	22,22%	3,70%	27
Experiência	1				100,00%	0,00%	0,00%	0,00%	1
Facilidade de operação	9	4	1		64,29%	28,57%	7,14%	0,00%	14
Interatividade	1	1	1	2	20,00%	20,00%	20,00%	40,00%	5
Interpretabilidade	20	3	2	4	68,97%	10,34%	6,90%	13,79%	29
Precisão		2			0,00%	100,00%	0,00%	0,00%	2
Prestação de contas	2		4		33,33%	0,00%	66,67%	0,00%	6
Quantidade apropriada de dados	8	2	1	3	57,14%	14,29%	7,14%	21,43%	14
Responsabilidade				1	0,00%	0,00%	0,00%	100,00%	1
Singularidade	2				100,00%	0,00%	0,00%	0,00%	2
Transparência	3		3		50,00%	0,00%	50,00%	0,00%	6
Total Geral	131	31	29	24	60,93%	14,42%	13,49%	11,16%	215
Média	8,73	3,44	2,07	2,40	52,83%	13,88%	15,13%	18,16%	4,04

Fonte: Autor.

Como houve sentenças que falavam da presença do constructo, desejos ou com característica positivas, se utilizaram mais, para a análise final, os constructos com TCat negativa em comparação ao total de constructos relatados percebidos, pois a falha de qualidade de dados é a parte mais importante na relação entre Principal e Agente. Considerando isso, a qualidade dos dados percebida pelo total de entrevistados só pode ser definida como baixa. Com 60,93% dos constructos tendo uma categoria de natureza negativa, não há como dizer que há qualidade nestes dados, do ponto de vista da TPA.

Os constructos que tiveram maiores problemas de qualidade de dados e com mais de 3 sentenças (para evitar *outsiders*) foram a Compreensibilidade, com 94% das sentenças,

seguida por Consistência e Conveniência, com 83 e 71%, respectivamente, demonstrando uma dificuldade dos entrevistados em compreender os dados, o layout das planilhas e os conteúdos delas. Considerando os constructos seguintes: Interpretabilidade (69%), Completude(67%) e Acessibilidade (64%), foram descritos quase todos os constructos considerados mais relevantes para a qualidade de dados em Wang (1998) e Wang & Strong (1996), e os considerados intrínsecos aos dados, em Vertrò *et al.* (2016) os únicos constructos não mencionados foram os de Precisão, por impossibilidade de percepção exata dada natureza metodológica da pesquisa; os relacionados à temporalidade/agilidade dos dados, pela natureza da atividade exercida pelos entrevistados e; Singularidade, por ser um dos mais fáceis de corrigir. É interessante notar, porém, que a maior dificuldade de qualidade dos dados ainda é de forma, mas o mais mencionado negativamente foi Compreensibilidade, o constructo mais próximo à função dos dados.

Os constructos com TCat negativa menor e mais de 3 sentenças, por outro lado, são Interatividade com 20%, Prestação de contas com 33% e Credibilidade com 37%. Pode-se conjecturar que esse nível observado em Interatividade se deve pela sua natureza positiva e não essencial, e, portanto, tendo mais respostas positivas e de desejo que o normal. Supõe-se o mesmo para Prestação de contas e Credibilidade, conclusão adicionada à baixa expectativa que os entrevistados demonstraram em outras sentenças.

Quadro 28 – Sentenças por Conjunto de Dados

Conj \ TCat	Negativa	Neutra	Positiva	Desejo	Negativa	Neutra	Positiva	Desejo	Total
ANAC	25	7	5	4	60,98%	17,07%	12,20%	9,76%	41
CEPEA	24	3	3	8	63,16%	7,89%	7,89%	21,05%	38
ENADE	17	12	8	1	44,74%	31,58%	21,05%	2,63%	38
ENCCEJA	5	5	6	7	21,74%	21,74%	26,09%	30,43%	23
KAGGLE	5		4	1	50,00%	0,00%	40,00%	10,00%	10
M.TUR	32	2	1	4	82,05%	5,13%	2,56%	10,26%	39
PAC	22	3	1		84,62%	11,54%	3,85%	0,00%	26
Total	130	32	28	25	60,47%	14,88%	13,49%	11,16%	215

Fonte: Autor.

As percepções por constructos, aparentemente, também não foram muito boas, porém melhor divididas. O pior conjunto de dados, considerando as TCat negativas, ficou sendo o do PAC, com 84,62%; o segundo lugar ficou com o M.TUR, com 82,05%. Em ambos, os

entrevistados trabalharam diretamente com dados financeiros, o que pode ter afetado esse nível. É possível especular que esta mensuração de qualidade de dados seja relacionada ao tipo de dados lidado, ainda mais quando se leva em conta que as duas melhores, assim consideradas por TCat negativas, foram a ENADE (44,74%) e a ENCCEJA (21,74%) que, obviamente, são bancos de dados que não lidam com o lado financeiro dos órgãos, mas sim com resultados de uma de suas atividades. Pode-se afirmar, portanto, que, com base nesses dados, houve uma disparidade entre qualidade de dados. Os dados relacionados a finanças detinham muito mais constructos negativos de qualidade de dados que os relacionados aos de resultados. Há várias possíveis suposições para esse resultado, entre elas, tanto a natureza mais difícil de lidar com dados sobre finanças quanto uma tentativa malfeita de maquiagem de desvio de recursos. Esse fato tornou os alunos entrevistados mais desconfiados. Dito de outro modo, foi contraprodutivo na função de dados abertos como atenuador de desconfiança.

Há de se relatar que ao menos algumas exigências públicas de qualidade de dados foram seguidas. Nas percepções dos estudantes com os conjuntos de dados do ENADE, ENCCEJA, KAGGLE, CEPEA e ANAC, há menções positivas quanto às normas da Infraestrutura Nacional de Dados Abertos (INDA, 2018), as quais foram seguidas, sendo a ordem desses conjuntos também representante de graus decrescentes em relação à adesão. O MEC demonstrou-se ser o que mais seguiu essas diretrizes. Se isso foi por questões de reconhecer a legitimidade da busca de qualidade de dados, por questões legais, ou pura aleatoriedade, não é possível ter certeza, mas certamente o MEC e os estudantes que fizeram estudos sobre essa instituição relataram que o órgão parece seguir não só recomendações de qualidade de dados, como a INDA (INDA, 2018) e a Cartilha técnica para publicação de dados abertos (SLTI, 2019) como também outras boas práticas de qualidade de dados internacionais, como as “*5 starts of data quality*” (VERTRÒ *et al.*, 2018), também chamadas de “As 5 estrelas de qualidade de dados”.

É possível notar também a heterogeneidade das percepções sobre dados abertos governamentais e seu papel como sinais de confiabilidade de Agente para Principal. O que levou a criação dos constructos emergentes, Responsabilidade e Experiência. Estes constructos observados são interessantes para garantir a qualidade dos dados e entender a mensuração de qualidade de dados respectivamente. Por este motivo, foram notados e discutidos, mesmo com relativa pouca menção nas sentenças dos entrevistados.

Não foram encontrados, nas sentenças dos entrevistados, vários constructos listados como possíveis de aparecer. Eles são: Pontualidade, Precisão, Atualidade, Volatilidade, Usabilidade, Integridade da derivação, Metadados, Aplicabilidade, Rapidez, Custo,

Objetividade, Relevância, Licenciamento. Para explicar tais ausências, se supõem os seguintes motivos:

- Constructos de temporalidade como Pontualidade, Atualidade, Volatilidade, Metadados e Rapidez não apareceram pelo fato de a atividade de SI&D não exigir atualização, tempo específico ou validade, fazendo-os irrelevantes para os entrevistados e, por isso, ignorados.
- Constructos relacionados à utilidade (Usabilidade, Aplicabilidade, Relevância, Objetividade), provavelmente não apareceram por serem melhor explicados por outros constructos, e também porque, na atividade, os alunos estavam livres para fazer as questões a que eles iriam responder, sendo considerado usabilidade como sinônimo de “facilidade para fazer a atividade”.
- Custos e Licenciamento não foram utilizados, provavelmente por conta de dados normais de Dados Abertos Governamentais não exigirem isso, porém pode haver algum caso, como dados militares ou dados que exijam digitalizar dados antigos, em que sejam ativados.
- No caso de Precisão, se supõe que, apesar de ser um constructo base, não foi observado, talvez pela incapacidade dos entrevistados de saberem a realidade que causou os dados.
- Finalmente, Integridade da derivação, que provavelmente não foi percebida pelos alunos ou ignorada, uma vez que não era essencial à atividade ver este tipo de informação.

4.22. Resultados Complementares

Um dos resultados mais interessantes é que houve um TNat chamado “Erro de percepção” no qual A02, que fala nas ID_Sen 95, 96, 97,109, 131 e 134 dos dados do ENCCEJA, referenciou o fato de que todos os dados que ele pegou no site por algum motivo: “[...] vieram em um arquivo separado por vírgula, esse tipo de arquivo, ele não abre direito no [Microsoft] Excel, abre todo ilegível [...]” (A02, 95) e que “[...] eles [os dados] ficavam todos condensados em uma única célula [...]” (A02, 97).

Ao se investigarem as afirmações diretamente no banco de dados, se percebeu que o arquivo do ENCCEJA tinha o formato “. cvs” do tipo “CVS separado por vírgula” considerado, por Vertrò *et al.* (2016), um bom formato de arquivo. O problema foi que o aluno não soube se expressar e, por engano, respondeu que os dados estavam “todos na mesma “[...] célula [...]” (A02, 97) quando a expressão correta seria “coluna” (em Microsoft

Excel). O aluno, mesmo sendo o curso de Ciência da Computação, também não soube como isolar automaticamente tabulações separadas por vírgula, o que aumentou seu trabalho e o de seu grupo, que teve de fazer a separação de forma manual. Este incidente demonstrou falta de *Data Literacy*, tanto pela instituição de ensino, por não ser capaz de passar esses conhecimentos, como também uma falta por parte do aluno, que teve dificuldade de extrair a informação que os dados continham. Esta falta de motivação mútua pelo boa *Data Literacy* é reconhecida também como limitação de pesquisa. Também se pode supor que houvesse uma falta de *Data Literacy* parte do publicador de dados, uma vez que não tentou explicar como os dados estavam sendo mostrados.

5. CONSIDERAÇÕES FINAIS

A pesquisa se iniciou com o objetivo geral de identificar as percepções sobre qualidade de dados abertos governamentais, à luz da TPA, e foi capaz de identificar vários constructos de qualidade de dados abertos governamentais, principalmente erros nessa qualidade, do ponto de visto da teoria aqui discutida, permitindo, após discussões e análises, várias contribuições.

5.1. Contribuições Teóricas

A contribuição principal deste trabalho foi além de responder a várias recomendações da literatura (BØE; GULBRANDSEN, 2016; BØE; GULBRANDSEN; SØREBØ, 2015; BØE, 2016; COSER, 2017; CAI; ZHU, 2015; WANG; STRONG, 1996; MOURA, 2018; ANTUNES, 2018), pois se pôs a demonstrar empiricamente que os constructos com mais características negativas de qualidade dos dados percebidas detinham mais natureza de forma que de função. Esses “piores constructos encontrados”, a saber: Interpretabilidade, Consistência, Compreensibilidade, Completude e Acessibilidade também foram os mais básicos, ou seja, seria necessário atender a estes constructos primeiro para poder avaliar a qualidade de outros, demonstrando uma hierarquia de qualidade de dados presente.

Outro ponto importante foi a disparidade na qualidade de dados entre dados referentes a finanças e os relacionados a resultados dos órgãos. Os dados pesquisados relacionados aos resultados se mostraram com uma qualidade superior aos relacionados a finanças, o que pode demonstrar não só atividade suspeita como também uma desconfiança maior dos pesquisados.

A pesquisa também contribuiu na teoria ao demonstrar a presença de dois constructos emergentes nos desejos dos respondentes, que serviriam para melhorar métricas de qualidade de dados. A esses constructos deu-se o nome de Experiência e Responsabilidade, cujas métricas e formas de uso também foram aqui discutidos. A literatura de qualidade de dados também ganha pelo fato de este trabalho complementar a lista de Batini *et al.* (2009), atualizando seus conceitos base. Isto facilitará, possivelmente, o avanço nos estudos que conceituam qualidade de dados como sendo algo considerado tanto em forma quanto em função dos dados.

Esta pesquisa também aproxima várias teorias e áreas que pouco se referenciam e trabalham em conjunto, o que lhes permite aproximação por interdisciplinaridade. A saber, este trabalho aproximou a TPA, que é ligada à área de estudo da gestão, e Qualidade de Dados, que é ligada à área de estudo da informação e computação. Principalmente essa

pesquisa demonstrou que a função dos dados também é importante, uma área por vezes pouco explorada na área de tecnologias de informação e comunicação, dando-lhe mais uma teoria com a qual pode analisar resultados – a TPA.

5.2. Contribuições Práticas

Uma das maiores contribuições práticas foi que, ao explicar os constructos que são importantes à qualidade de dados, este trabalho permite que Principais possam saber que conceitos levar em consideração para poder exigir dados de melhor qualidade, uma vez que sabem que métricas e conceitos exigir (através da metodologia desta pesquisa) ou quais pontos são mais pertinentes para atender a seus interesses (através dos resultados desta pesquisa).

Gestores públicos dos órgãos estudados podem, se interessados, verificar o impacto real de políticas de qualidade de dados e governança, por meio dos constructos desenvolvidos neste trabalho, tendo como mostrar se as exigências internas (do órgão) sobre a qualidade de dados abertos governamentais estão sendo suficientes para informar efetivamente o consumidor de dados. Não se pode fazer isso inteiramente nesta dissertação por limitações de tempo e recursos, mas isso não restringe a possibilidade para outros usos.

Gestores também podem utilizar esta dissertação como um mapa que mostre o que fazer ou o que não fazer em qualidade de dados, e como garantir uma boa governança. Este trabalho dá fontes para fazer *benchmark* para ter boa governança e qualidade de dados efetiva do ponto de vista de qualquer um, não só Principais ou cidadãos.

Esta pesquisa, pelo fato de contar com a participação de alunos voluntários, permitiu que se detectassem falhas no conhecimento dos alunos voluntários, ou seja, lhes faltava Data Literacy, o que foi relatado ao professor da disciplina e serviu para que esse docente busque (e encontre) caminhos que o levem a melhorar ainda mais suas aulas.

Se a Qualidade dos dados ou dos métodos de mensuração de qualidade de dados melhorarem em decorrência das evidências ou conceitos deste trabalho, os consumidores de dados e Principais de várias relações podem ter mais informação para tomada de decisão e evitar conflito de interesses, não se restringindo ao contexto público.

Sabendo-se, no entanto, das limitações de recursos da administração pública e para não sobrecarregar aqueles que publicam e produzem dados abertos, como diz Moura (2018), algumas sugestões se demonstraram mais pertinentes que outras por possibilitar aumentar qualidade dos dados com poucos recursos. São elas: (a) Quando mensurando a qualidade de dados, é interessante questionar ao consumidor de dados sobre sua percepção de qualidade de

dados e como melhorar, como em Weber e Yan (2017); (b) É interessante também exigir o uso de algum recurso automático disponível gratuitamente ou facilmente programável, para garantir qualidade básica de dados, principalmente nos constructos de forma, como Acessibilidade e Singularidade; (c) Colocar junto dos conjuntos de dados algumas informações de como acessar o órgão responsável pela publicação ou as reportar, caso não estejam disponíveis ou utilizáveis, dentro dos arquivos e no site, não se exigindo um contato formal, mas apenas um alarme.

5.3. Limitações

Toda pesquisa tem suas limitações e esta não foi diferente. Dois dos entrevistados A03, e B01, não conseguiram obter dados da lista proposta pelo professor (Ver Anexos) e, como alternativa, o professor da disciplina deixou que fizessem com o CEPEA, ou Centro de Pesquisa Agropecuários, uma colaboração entre a Escola Superior de Agricultura Luiz de Queiroz (ESALQ) e a Universidade de São Paulo (USP). O entrevistado C02 teve a limitação que ele fez seu trabalho sobre os dados da Kaggle Inc., uma companhia da subsidiária da Google Inc. e, portanto, não-governamental. O banco de dados escolhido para a atividade foi de jogos do futebol europeu, baseado nos registros que incluem a Fifa e companhias privadas variadas. Não se pode dizer, portanto, que este banco de dados é de dados abertos governamentais em sua completude. Estes dados no máximo são apenas dados abertos. Este detalhe só foi notado, infelizmente, quando não se podia voltar atrás, por questões de tempo.

O plano original de dissertação envolvia usar o framework de Ijab *et al.* (2017), pois este continha um contexto próximo ao que se queria pesquisar em qualidade de dados. Pela falta de acesso, contudo, às métricas do framework em língua entendível pelo autor e em tempo hábil para a pré-banca, utilizou-se, para identificar constructos, o framework de Vertrò *et al.* (2016). Assim foi decidido porque este era um dos frameworks mais novos, com boa base teórica, boa relevância acadêmica e que focava em dados abertos governamentais, embora focasse em mensuração de qualidade de dados de forma intrínseca, isso é, em questão da forma, diferentemente do que esta pesquisa se propõe a fazer. Após reflexão profunda, se decidiu utilizar a revisão da literatura de Batini *et al.* (2009) para obtenção de conceitos-base.

Houve problemas, também, com o uso de Batini *et al.* (2009) como fonte de conceitos. Esse texto é antigo (10 anos entre sua publicação e esta dissertação), tem poucas definições profundas e métricas bastante diferentes, além de ser uma revisão da literatura, não uma metodologia de mensuração de qualidade de dados. Continha, portanto, constructos por vezes conflitantes ou errados. Para sanar isso, houve a leitura de múltiplos outros trabalhos para se

obterem definições mais profundas dos conceitos e, de alguma forma, tentar atualizar Batini *et al.* (2009). Só foi possível, contudo, analisar apenas 21 artigos sobre possíveis melhoramentos do método de Batini *et al.* (2009), algo que não cobre por inteiro a literatura, e, portanto, limitando o escopo da pesquisa.

Como dito na metodologia, por questões de tempo, saturação e falta de cooperação dos entrevistados, não foi possível fazer esta pesquisa com mais pessoas ou esgotar os bancos de dados envolvidos na atividade de SI&D. O plano de dissertação planejava ter entre 20 e 30 participantes. Mesmo que se tenha chegado a números próximos ao que se espera de saturação, pela leitura de Bardin (1977), um estudo envolvendo 7 conjuntos de dados, um deles não sendo de dados abertos governamentais, restringiu-se o poder de inferência sobre esta pesquisa.

Sentenças com pedidos de esclarecimento, que tinham orações iguais ou similares a “Como assim?” (B01, 309) ou “Ahn? [...]” (A05, 205) ou que pediam esclarecimentos do que as questões representavam apareceram em 27 sentenças, as sentenças de ID_Sen: 30, 93, 101, 103, 121, 122, 147, 165, 166,167, 183, 184, 185, 201, 217, 219, 248, 271, 295, 309, 317, 332, 349, 358, 413, 457, 510, 574, 622, respectivamente. Uma possível causa dessa intercorrência pode ter sido por conta da baixa voz do entrevistador/autor. Esta limitação também retrata a dificuldade de se entender a questão Q4, uma vez que 11 sentenças (40,74%) tinham Q4 como questão.

O respondente A03 relatou não lembrar do significado do acrônimo CEPEA na ID_Sen 157, mesmo que tenha apresentado seminário sobre este mesmo órgão na semana anterior à da entrevista. Isso põe em dúvida a capacidade desse e de outros respondentes de recordar suas percepções originais sobre qualidade de dados, denunciando sua despreocupação com o assunto, uma vez que a necessidade já passou e não haveria mais utilidade em lembrar estes detalhes, mesmo que tivessem se voluntariado a participar de pesquisa sobre isso. Outra possibilidade é que o respondente A03, por usar o acrônimo o tempo todo, considerou o significado do acrônimo informação redundante e acabou esquecendo.

O fato de alunos terem demonstrado falta da *Data Literacy* para compreender certos dados permite duvidar se os entrevistados detinham *Data Literacy* suficiente para perceberem qualidade de dados efetivamente e em qualidade satisfatória.

As representações de TObj também foram limitadas, pois considerou-se o objeto como aquilo descrito nas sentenças, o que causa a perda do real objeto em meio a um grande número de sentenças, a estes casos foi atribuído o TObj “dados” por ser o mais abrangente,

mas por este motivo a coluna TObj teve, por ironia, uma limitação de qualidade de dados do tipo Completude.

E, finalizando, se percebeu que, devido ao fato de a atividade de SI&D dar links diretos aos alunos sobre onde achar os dados abertos e quais órgãos estudar exatamente, limitou-se a quantidade e qualidade de certos constructos. Uma vez que a pesquisa não se iniciou do nada, não há como dizer que os constructos Conveniência e Acessibilidade estavam corretos, pois foi o professor participante que escolheu os órgãos para pesquisar e seriam más práticas de ensino exigir de alunos uma atividade de pesquisa com dados realizada em órgãos com dificuldade de acesso ou com dificuldade de navegação.

5.4. Sugestões de Pesquisas Futuras

Para sugestões de pesquisas futuras, se recomenda atualizar e aumentar as dimensões e constructos em Batini *et al.* (2009) e nesta dissertação, pois métodos de mensuração de qualidade de dados e métodos de mensuração de qualidade de informação sempre podem melhorar, criando-se novos conceitos ou novas métricas. Um mapa teórico ou uma nova metodologia de mensuração tem relevância acadêmica.

Outra recomendação é repetir a metodologia aqui empregada em outras circunstâncias, com outros bancos de dados, outras organizações, ou com alunos de outros cursos. Isso inclui aplicar o presente método a indivíduos de outros níveis educacionais e de especialidades diferentes, para se verificar tanto o método quanto os conceitos.

Também se recomenda continuar a atender às recomendações da literatura tratadas nesta pesquisa (BØE; GULBRANDSEN, 2016; BØE; GULBRANDSEN; SØREBØ, 2015; BØE, 2016; COSER, 2017; CAI; ZHU, 2015; WANG; STRONG, 1996; MOURA, 2018; ANTUNES, 2018), para aproximar áreas teóricas que pouco se relacionam.

Por fim, se sugere a continuada de estudos de metodologias, conceitos e métricas aqui estudados, registrados e descritos, para a criação de uma nova metodologia de mensuração de qualidade de dados como vetor de informação ou para dados abertos, que englobe todas as metodologias e métricas possíveis. Um exemplo seria uma métrica que consideraria a totalidade de constructos possíveis de analisar em qualidade de dados, mas que fosse sucinta o suficiente para ser usada.

6. REFERÊNCIAS

- AGARWAL, Ritu; DHAR, Vasant. Big data, data science, and analytics: The opportunity and challenge for IS research. **IS research**, v. 25, n. 3, p. 443-448, 2014.
- ALENEZI, Hussain; TARHINI, Ali; MASA'DEH, R. Investigating the strategic relationship between information quality and e-government benefits: A literature review. **International Review of Social Sciences and Humanities**. v.9 n.1 p.33-50, 2015.
- ALMADA, M. P. **Avaliação da e-Transparência em Portais de Governos Nacionais: Uma Comparação entre Brasil, Estados Unidos e Reino Unido**. Tese (Doutorado em Comunicação Social) – Programa de Pós-Graduação em Comunicação e Cultura Contemporâneas, Universidade Federal da Bahia, 2017.
- AMICIS, F. D. BATINI A.T. A methodology for data quality assessment on financial data. **Studies in Communication Sciences**, v. 4, n. 2, p. 115-137, 2004.
- ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 11515**: Guia de práticas para segurança física relativas ao armazenamento de dados. ABNT: Rio de Janeiro, 2007
- BANCO MUNDIAL. **Governance and Development**, World Bank, 1992. 61 p.
- BANCO MUNDIAL. **Governance and the Law**, Washington, 2017. 61 p.
- BARDIN. L. **Análise de conteúdo**. Lisboa: Editora Edições 70, 1977. 229 p.
- BATINI, C.; SCANNAPIECO, M. **Data and Information Quality: Dimensions, Principles and Techniques**. Cham, Switzerland: Springer International Publishing, 2016.
- BATINI, Carlo et al. Methodologies for data quality assessment and improvement. **ACM computing surveys (CSUR)**, v. 41, n. 3, p. 16, 2009.
- BEHKAMAL, Behshid et al. A metrics-driven approach for quality assessment of linked open data. **Journal of theoretical and applied electronic commerce research**, v. 9, n. 2, p. 64-79, 2014.
- BELLINGER, Gene; CASTRO, Durval; MILLS, Anthony. **Data, information, knowledge, and wisdom**. 2004. Disponível em: https://s3.amazonaws.com/academia.edu.documents/35500195/bellinger_on_ackoff_data_info_know_wisdom.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1549764181&Signature=tCorkMovm78kOAlmPSVwVJZaOUs%3D&response-content-disposition=inline%3B%20filename%3DBellinger_on_ackoff_data_info_know_wisdo.pdf, Acesso em 20 de Jan 2019.
- BENAROCH, Michel; CHERNOBAI, Anna. Operational IT failures, IT value-destruction, and board-level IT governance changes. **MIS Quarterly**, 2017.
- BEVIR, Mark. **Governance: A very short introduction**. OUP Oxford, 2012.

BHATTACHERJEE, Anol. Managerial influences on intraorganizational information technology use: A Principal- agent model. **Decision Sciences**, v. 29, n. 1, p. 139-162, 1998.

BHATTACHERJEE, Anol; LIN, Chieh-Peng. A unified model of IT continuance: three complementary perspectives and crossover effects. **European Journal of Information Systems**, v. 24, n. 4, p. 364-373, 2015.

BØE, Tove. The Interaction Effect of Perceived Competence and Goal Harmony on Perceived Usefulness. **Creative Education**, v. 7, n. 08, p. 1136, 2016.

BØE, Tove; GULBRANDSEN, Boge. The Role of Agency Theory and Perceived Goal Divergence in IS Continuance: A Replication and Extension Study. In: **System Sciences (HICSS), 2016 49th Hawaii International Conference on**. IEEE, 2016. p. 94-103.

BØE, Tove; GULBRANDSEN, Boge; SØREBØ, Øystein. How to stimulate the continued use of ICT in higher education: Integrating information systems continuance theory and agency theory. **Computers in Human Behavior**, v. 50, p. 375-384, 2015.

BRYMAN, A.; BELL, E. **Business Research Methods 3ed**. New York: Oxford, 2011.

BRAGA, Lamartine Vieira et al. O papel do governo eletrônico no fortalecimento da Governança do setor público. **Revista do Serviço Público - RSP**, v. 59, n. 1, p. 05-21, 2008.

BRASIL. Constituição (1988). **Constituição da República Federativa do Brasil**: promulgada em 5 de outubro de 1988, Brasília, 1988. Disponível em: <https://presrepublica.jusbrasil.com.br/legislacao/91972/constituicao-da-republica-federativa-do-brasil-1988>. Acesso em: 21 jun. 2018.

_____. **Lei Complementar nº 101**, de 04 de maio de 2000. Disponível em: http://www.planalto.gov.br/ccivil_03/LEIS/LCP/Lcp101.htm. Acesso em: 22 jun. 2018.

_____. **Lei Complementar nº 131**, de 27 de maio de 2009. Disponível em: http://www.planalto.gov.br/ccivil_03/LEIS/LCP/Lcp131.htm. Acesso em: 23 jun. 2018.

_____. **Lei nº 12.527**, de 18 de novembro de 2011. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/112527.htm. Acesso em: 21 jun. 2018.

BUTTER, Frank AG; LIU, Jianwei; TAN, Yao-Hua. Using IT to engender trust in government-to-business relationships: The Authorized Economic Operator (AEO) as an example. **Government Information Quarterly**, v. 29, n. 2, p. 261-274, 2012.

CAI, Li; ZHU, Yangyong. The challenges of data quality and data quality assessment in the big data era. **Data Science Journal**, v. 14, n. 2, p. 1-10, 2015.

CANOVA, Lorenzo et al. Collaborative Open Data Versioning: A Pragmatic Approach Using Linked Data. In: **Conference for E-Democracy and Open Governement**. 2015. p. 171.

CGU – CONTROLADORIA GERAL DA UNIÃO. **Relatório de gestão exercício 2016**. Disponível em: <http://www.cgu.gov.br/sobre/auditorias/arquivos/2016/relatorio-de-gestao-cgu-2016.pdf/view> Acesso em: 29 ago. 2018.

CHEN, Hsinchun; CHIANG, Roger HL; STOREY, Veda C. Business intelligence and analytics: from big data to big impact. **Management Information Systems Quarterly**, p. 1165-1188, 2012.

CHOI, Jeong Min. Factors influencing public officials' responses to requests for information disclosure. **Government Information Quarterly**, v. 35, n. 1, p. 30-42, 2018.

CLARKE, Amanda; MARGETTS, Helen. Governments and citizens getting to know each other? Open, closed, and big data in public management reform. **Policy & Internet**, v. 6, n. 4, p. 393-417, 2014.

COSER, Tiago. Governança de tecnologia da informação para reduzir problemas de agência no nível operacional. **Revista Gestão & Tecnologia**, v. 17, n. 3, p. 112-145, 2017.

COSTA, FJ da. Mensuração e desenvolvimento de escalas: aplicações em administração. **Rio de Janeiro: Ciência Moderna**, 2011.

CRUZ, Cláudia Ferreira et al. Transparência da gestão pública municipal: um estudo a partir dos portais eletrônicos dos maiores municípios brasileiros. **Revista de Administração Pública**, v. 46, n. 1, p. 153-176, 2012.

CRUZ, Cláudia Ferreira da; AFONSO, Luís Eduardo. Gestão fiscal e pilares da Lei de Responsabilidade Fiscal: evidências em grandes municípios. **Rev. Adm. Pública**, Rio de Janeiro, v. 52, n. 1, p. 126-148, 2018. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0034-76122018000100126&lng=en&nrm=iso. Acesso em: 23 Ago. 2018.

DAWSON, Gregory S. et al. An examination of effective IT governance in the public sector using the legal view of agency theory. **Journal of Management Information Systems**, v. 33, n. 4, p. 1180-1208, 2016.

ELBAHNASAWY, Nasr G. E-government, internet adoption, and corruption: an empirical investigation. **World Development**, v. 57, p. 114-126, 2014.

ENGLISH, Larry P. **Improving data warehouse and business information quality: methods for reducing costs and increasing profits**. New York: Wiley, 1999.

EPPLER, Martin J.; MUENZENMAYER, Peter. Measuring Information Quality in the Web Context: A Survey of State-of-the-Art Instruments and an Application Methodology. In: **Proceedings...** International Conference on Information Systems(ICIQ). 2002. p. 187-196.

FABRINI, Fábio. Depois de seis anos de Lei de Acesso, estatais ainda resistem a abrir dados. **Folha de São Paulo**, São Paulo, 21 de Ago 2018. Disponível em: <https://www1.folha.uol.com.br/poder/2018/08/depois-de-seis-anos-de-lei-de-acesso-estatais-ainda-resistem-a-abrir-dados.shtml>. Acesso em: 22 ago. 2018.

FAGUNDES, Priscila Basto; MACEDO, Douglas Dyllon Jeronimo de; FREUND, Gislaine Parra. A produção científica sobre qualidade de dados em big data: um estudo na base de dados Web of Science. **RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação**, v. 16, n. 1, p. 194-210, 2017.

FAMA, E. F. Agency Problems and the Theory of the Firm, **The Journal of Political Economy**, v.88, n.2, p. 288-307, 1980.

FALORSI, P. D. et al. Improving the quality of toponymic data in the italian public administration. In: **Proceedings...** of the ICDT. Vol.3. 2003.

FLORIDI, Luciano. Philosophical conceptions of information. In: **Formal theories of information**. Springer Berlin Heidelberg, 2009. p. 13-53.

FRANCISCO, Maritza et al. Total Data Quality Management and Total Information Quality Management Applied to Customer Relationship Management. In: **Proceedings of the 9th International Conference on Information Management and Engineering**. p. 40-45. 2017.

FRIEDMAN, M. **capitalism and freedom**. Chicago: University of Chicago Press, 1962.

GOMES, Wilson; AMORIM, Paula Karini Dias Ferreira; ALMADA, Maria Paula. Novos desafios para a ideia de transparência pública. In: **E-Compós**. 2018.

GOOGLE SCHOLAR. **Google acadêmico**. 2018. Disponível em: <https://scholar.google.com.br/>. Acesso em: 16 Jan. 2019.

GONÇALVES, Alcindo. O conceito de Governança. In: ENCONTRO DO CONPEDI, 14, 2005, Fortaleza, **Anais...**, Fortaleza, 2005. p. 197.

GONZALEZ-ZAPATA, Felipe; HEEKS, Richard. The multiple meanings of open government data: Understanding different stakeholders and their perspectives. **Government Information Quarterly**, v. 32, n. 4, p. 441-452, 2015.

GRAY, Jonathan; GERLITZ, Carolin; BOUNEGRU, Liliana. Data infrastructure literacy. **Big Data & Society**, v. 5, n. 2, p. 2053951718786316, 2018.

GUILLAMÓN, María-Dolores et al. Factors influencing social media use in local governments: The case of Italy and Spain. **Government Information Quarterly**, v. 33, n. 3, p. 460-471, 2016.

HARYADI, A. F. et al. Antecedents of Big Data Quality An Empirical Examination in Financial Service Organizations. 4TH IEEE INTERNATIONAL CONFERENCE ON BIG DATA (BIG DATA), 4., 2016, Washington. **Anais eletrônicos...** Washington: 2016. Disponível em: <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7840595>. Acesso em: 07 Dez. 2018.

IBGC, **Princípios Básicos de Governança Corporativa**, Disponível em: <http://www.ibgc.org.br/governanca/governanca-corporativa/principios-basicos> Acesso em: 01 Ago. 2018.

IBGE. **Pesquisa Nacional por Amostra de Domicílios (PNAD):** Acesso à internet e à televisão e posse de telefone móvel celular para uso pessoal. 2015. Disponível em: <http://www.ibge.gov.br/home/estatistica/populacao/acessoainternet2015/default.shtm>
Acesso em: 30 Jul. 2018.

IJAB, Mohamad Taha et al. Towards Big Data Quality Framework for Malaysia's Public Sector Open Data Initiative. In: **International Visual Informatics Conference**. Springer, Cham, 2017. p. 79-87.

INFRAESTRUTURA NACIONAL DE DADOS ABERTOS – INDA. **Plano de ação da INDA**. 2018. Disponível em: <http://wiki.dados.gov.br/Plano-de-Acao-da-INDA.ashx>. Acesso em: 06 Set. 2018.

INFRAESTRUTURA NACIONAL DE DADOS ESPACIAIS – INDE. **Plano de ação da INDA**. 2018. Disponível em: <http://www.inde.gov.br/a-inde/apresentacao.html>. Acesso em: 06 Set. 2018.

INGRAMS, Alex. Democratic transition and transparency reform: An fsQCA analysis of access to information laws in twenty-three countries. **Government Information Quarterly**, v. 35 n.1 p.428-436, 2018.

IVANOV, Kristo. Quality control of information. **On the concept of accuracy of information in data—banks and in management information systems**. Tese de doutorado, Estocolmo, The Royal Institute of Technology, 1972.

JAAFAR, Nouf et al. Framework for social media big data quality analysis. In: **New Trends in Database and Information Systems II**. Springer, Cham, 2015. p. 301-314.

JENSEN, Michael C.; MECKLING, William H. Theory of the firm: Managerial behavior, agency costs and ownership structure. **Journal of financial economics**, v. 3, n. 4, p. 305-360, 1976.

JENSEN, Michael C.; MECKLING, William H. Teoria da firma: comportamento dos administradores, custos de agência e estrutura de propriedade. **Revista de Administração de Empresas**, v. 48, n. 2, p. 87-125, 2008.

JEUSFELD, Manfred A.; QUIX, Christoph; JARKE, Matthias. Design and analysis of quality information for data warehouses. In: **International Conference on Conceptual Modeling**. Springer, Berlin, Heidelberg, 1998. p. 349-362.

JOURNAL CITATION REPORTS. **Site do Journal Citation Reports**. Disponível em <https://jcr.incites.thomsonreuters.com/JCRJournalHomeAction.action?SID=A2-YTz8U9GojD740ThUIa50wFgqs4ZIohkB-18x2dQO7vT54ydjEeA24DqhpckAx3Dx31KWx2BIvNPWoTjms8fIAewcwQx3Dx3D-9vvmzcn dpRgQCGPd1c2qPQx3Dx3D-wx2BJQh9GKVmtdJw3700KssQx3Dx3D&SrcApp=IC2LS&Init=Yes>. Acesso em: 05 jan. 2019.

KARANJA, Erastus. The role of the chief information security officer in the management of IT security. **Information & Computer Security**, v. 25, n. 3, p. 300-329, 2017.

KHAN, Hammad R.; KIM, Jeonghyun; CHANG, Hsia-Ching. Toward an Understanding of Data Literacy. In: **Proceedings**. iConference, 2018.

KROGSTIE, John. Capturing enterprise data integration challenges using a semiotic data quality framework. **Business & information systems engineering**, v. 57, n. 1, p. 27-36, 2015.

KWON, Ohbyung; LEE, Namyoon; SHIN, Bongsik. Data quality management, data usage experience and acquisition intention of big data analytics. **International Journal of Information Management**, v. 34, n. 3, p. 387-394, 2014.

LACERDA, Sérvulo Mário de Paiva. **Governo aberto, transparência e governo eletrônico nas câmaras municipais paraibanas: um estudo multicaso**. 2017. 130 f. Dissertação (Mestrado em Administração) - Universidade Federal da Paraíba, João Pessoa, 2017.

LEE, Yang W. et al. AIMQ: a methodology for information quality assessment. **Information & management**, v. 40, n. 2, p. 133-146, 2002.

LONG, J. A; SEKO C. A. A cyclic-hierarchical method for database data-quality evaluation and improvement. In: **Information quality**. Routledge, 2014. p. 64-78.

LOSHIN, David. **Enterprise knowledge management: The data quality approach**. Morgan Kaufmann, 2001.

LOURENÇO, Rui Pedro. An analysis of open government portals: A perspective of transparency for accountability. **Government Information Quarterly**, v. 32, n. 3, p. 323-332, 2015.

LUCERO, Everton. **Governança da Internet: aspectos da formação de um regime global e oportunidades para a ação diplomática**. Fundação Alexandre de Gusmão, 2011.

MACHADO, Cláudio Pinheiro. **Responsabilidade social e Governança: o debate e as implicações**. São Paulo: Cengage Learning, 2011.

MAGALHÃES, R, S, P. **Governança em Organizações Públicas – Desafios para entender os fatores críticos de sucesso: O caso do Tribunal de Contas da União**. 2011. Dissertação. Fundação Getúlio Vargas. Disponível em: <http://portal2.tcu.gov.br/portal/pls/portal/docs/2297032.pdf> Acesso em: 20 de Jul. de 2018.

MAJONE, Giandomenico. Two logics of delegation: agency and fiduciary relations in EU governance. **European Union Politics**, v. 2, n. 1, p. 103-122, 2001.

MATTOS, Pedro Lincoln CL. “Os resultados desta pesquisa (qualitativa) não podem ser generalizados”: pondo os pingos nos is de tal ressalva. **Cadernos Ebape. BR**, p. 450-468, 2011.

MITNICK, B. M. Fiduciary rationality and public policy: The theory of agency and some consequences, In: **Proceedings of the 1973 Annual Meeting of the American Political**

Science Association. New Orleans, 1973. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1020859; Acesso em: 11 de Jul. de 2018.

MITNICK, Barry M. A typology of conceptions of the public interest. **Administration & Society**, v. 8, n. 1, p. 5-28, 1976.

MITNICK, Barry M. "**The Origins of Agency Theory**", 2006, Disponível em: <http://www.pitt.edu/~mitnick/agencytheory/agencytheoryoriginrev11806r.htm>, Acesso em 26 de Jun 2018.

MITRA, A. **Classifying data for successful modeling.**, 2014, Disponível em: <https://dwbi.org/data-modelling/dimensional-model/16-classifying-data-for-successful-modeling>, Acesso em 22 de Jul 2018.

MONTANHOLI, Márcia de Souza. **Princípios da Governança pública aplicada aos Tribunais de Contas do Estado e Município de São Paulo.** 2017. 127 f. Dissertação (Mestrado em Ciências Contábeis e Atuariais) - Programa de Estudos Pós-Graduados em Ciências Contábeis e Atuariais, Pontifícia Universidade Católica de São Paulo, São Paulo, 2017.

MORCK, Randall; STEIER, Lloyd. **A History of Corporate Governance around the World: Family Business Groups to Professional Managers.** University of Chicago Press, Chicago, 2005. Disponível em: <http://papers.nber.org/books/morc05-1> Acesso em: 02 de Ago de 2018.

MOURA Jr., Pedro Jácome de. **SE QUER QUE SEJA BEM FEITO, FAÇA EM EQUIPE: flow e desempenho em equipes de tecnologia da informação.** 2015. 229 p. Tese (Doutorado em Administração), Programa de Pós-Graduação em Administração - PPGA. Universidade Federal da Paraíba, 2015.

MOURA JR, P. J. . Governança de TI na administração pública: o CIO entre a cruz e a espada. **GVCasos - Revista Brasileira de Casos de Ensino em Administração** , v. 8, p. 7, 2018..

MOURA Jr, P. J., & BELLINI, C. G. P. Repertory grid, laddering, and content analysis: a psychometric approach to understand the information technology workforce. **JISTEM- Journal of Information Systems and Technology Management**, v. 11, n. 1, p. 215-232, 2014.

NAHON, Karine; PELED, Alon; SHKABATUR, Jennifer. Cities' Open Government Data Heart Beat. In: **Conference for E-Democracy and Open Governement.** 2015. p. 209.

NAZÁRIO, Débora Cabral; SILVA, Paulo Fernando; ROVER, Aires José. Avaliação da qualidade da informação disponibilizada no Portal da Transparência do Governo Federal. **Revista Democracia Digital e Governo Eletrônico**, v. 1, n. 6, 2012.

NOBRE, Carla Janaína Ferreira. **A condição financeira governamental e sua influência na transparência da gestão pública municipal**, 2017. 87 p. Dissertação (Mestrado em Ciências Contábeis), Programa de Pós-Graduação em Ciências Contábeis - PPGCC, Universidade Federal da Paraíba, 2017.

ONU – Organização das Nações Unidas. **Benchmarking e-government: a global perspective**. New York: Division for Public Economics and Public Administration, 2002. Disponível em: <http://publicadministration.un.org/egovkb/en-us/Reports/Benchmarking-E-Government-2011>, Acesso em: 21 Ago. 2018.

ONU – Organização das Nações Unidas. **UN e-government knowledgebase: frequently asked questions**, 2018. Disponível em: <https://publicadministration.un.org/egovkb/en-us/About/UNeGovDD-Framework>, acesso em: 21 Ago. 2018.

PARKER, M. B., et al. An evaluation of Information quality frameworks for the World Wide Web. In: 8th Annual Conference on WWW Applications, Bloemfontein, Free State Province, **Anais**, 08 - 06 Sep, 2006.

PENG, Mike W. et al. Theories of the (state-owned) firm. **Asia Pacific Journal of Management**, v. 33, n. 2, p. 293-317, 2016.

PEREIRA, Luiz Carlos Bresser. **A Reforma Do Estado Dos Anos 90: Lógica e Mecanismos de Controle**. Trabalho apresentado à 2ª reunião do Círculo de Montevidéu. Barcelona, 1997.

_____. **Reforma do Estado para a cidadania: a reforma gerencial brasileira na perspectiva internacional**. Editora 34, 1998.

PEREIRA, Paulo Ricardo Maroso. **Direito de reclamação do usuário de serviço público: um instrumento de controle social no atual modelo de administração pública brasileira**. Dissertação (Mestrado em Direito), UFRGS – Universidade Federal do Rio Grande do Sul. Porto Alegre, p. 219. 2017.

PIPINO, Leo L.; LEE, Yang W.; WANG, Richard Y. Data quality assessment. **Communications of the ACM**, v. 45, n. 4, p. 211-218, 2002.

PADRÕES DE INTEROPERABILIDADE DE GOVERNO ELETRÔNICO – ePING. **ePING**. Disponível em <http://eping.governoeletronico.gov.br/#parte1>. Acesso em 06 Set. 2018

PLATÃO. **A República**. Introdução, tradução e notas de Maria Helena da Rocha Pereira. Fundação Calouste Gulbenkian, Lisboa, 1973.

PORTAL DA TRANSPARÊNCIA. **Site do portal da transparência**. Disponível em: <http://www.portaltransparencia.gov.br> . Acesso em: 29 ago. 2018.

PORTAL BRASILEIRO DE DADOS ABERTOS - PPDA. **Portal dos dados abertos**. Disponível em <http://dados.gov.br>. Acesso em 06 Set. 2018

PROVOST, Foster; FAWCETT, Tom. Data science and its relationship to big data and data-driven decision making. **Big Data**, v. 1, n. 1, p. 51-59, 2013.

ROCHA, D.; DEUSDARÁ, B. Análise de Conteúdo e Análise do Discurso: aproximações e afastamentos na (re)construção de uma trajetória. **Alea: Estudos Neolatinos**, v. 7, n. 2, p. 305-322, 2005.

ROSS, Stephen A. The economic theory of agency: The principal's problem. **The American Economic Review**, v. 63, n. 2, p. 134-139, 1973.

RUBIN, E. Questions & Answers: Quantitative Commentary in Thucydides. **The American Statistician**, v. 25, n. 5, p. 52-54, 1971.

SANTOS, Maria Isabel Araújo Silva dos et al. **A segurança do segredo: proposta de framework de aplicação dos instrumentos de proteção do segredo no ambiente de inovação da base industrial de defesa**. Tese (doutorado) - Universidade Federal de Santa Catarina, Centro Tecnológico, Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento, Florianópolis, 2016.

SCANNAPIECO, Monica et al. The DaQuinCIS architecture: a platform for exchanging and improving data quality in cooperative information systems. **Information systems**, v. 29, n. 7, p. 551-582, 2004.

SCHIELD, Milo. Information literacy, statistical literacy and data literacy. In: **IASSIST QUARTERLY (IQ)**. 2004.

SILVA, Ambrozina; MIRANDA, Junia. Portal da Transparência e qualidade da informação: um estudo do Portal da Transparência de um governo local brasileiro In: International Conference on Information Resources Management (CONF-IRM) 25., **Proceedings**, 2017.

SILVEIRA, Alexandre Di Miceli. **Governança corporativa no Brasil e no mundo: teoria e prática**. Elsevier Brasil, 2017.

SLTI - Secretaria de Logística e Tecnologia da Informação. **Cartilha Técnica para Publicação de Dados Abertos no Brasil v1.0**. Disponível em <http://wiki.gtinda.ibge.gov.br/GetFile.aspx?Page=Tecnologia&File=Cartilha%20T%c3%a9cnica%20para%20Publica%c3%a7%c3%a3o%20de%20Dados%20Abertos%20no%20Brasil%20v1.pdf> Acesso em: 02 Fev. 2019.

SU, Zhanming; JIN, Zhanming. A methodology for information quality assessment in the designing and manufacturing processes of mechanical products. In: **Proceedings... Information Quality Management: Theory and Applications**. IGI Global, 2007. p. 190-220.

TCU, Tribunal de Contas da União. **Governança Pública: referencial básico de Governança aplicável a órgãos e entidades da administração pública e ações indutoras de melhoria / Tribunal de Contas da União. – Brasília: TCU, Secretaria de Planejamento, Governança e Gestão, 2016.**

TODORAN, Ion-George et al. A methodology to evaluate important dimensions of information quality in systems. **Journal of Data and Information Quality (JDIQ)**, v. 6, n. 2-3, p. 11, 2015.

VALE, Raimundo Costa. **Portais de Transparência: um estudo sobre o cumprimento da Lei Complementar nº 131 de 27 de maio de 2009, pelos municípios do Estado do Maranhão**. Dissertação (Mestrado em Administração Pública), Instituto Politécnico da Guarda. São Luís, p. 145. 2017.

VETRÒ, Antonio et al. Open data quality measurement framework: Definition and application to Open Government Data. **Government Information Quarterly**, v. 33, n. 2, p. 325-337, 2016.

VIGO, Ronaldo. Complexity over uncertainty in generalized representational information theory (GRIT): A structure-sensitive general theory of information. **Information**, v. 4, n. 1, p. 1-30, 2013.

WEISKOPF, Nicole Gray; WENG, Chunhua. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. **Journal of the American Medical Informatics Association**, v. 20, n. 1, p. 144-151, 2013.

XUE, Ling; YANG, Ke; YAO, Yuiliang. Examining the Effects of Interfirm Managerial Social Ties on IT Components Diversity: An Agency Perspective. **Management Information Systems Quarterly**, v. 42, n. 2, p. 679-694, 2018.

WANG, Richard Y. A product perspective on total data quality management. **Communications of the ACM**, v. 41, n. 2, p. 58-65, 1998.

WANG, Richard; STRONG, Diane M. Beyond accuracy: What data quality means to data consumers. **Journal of management information systems**, v. 12, n. 4, p. 5-33, 1996.

WANG, R. Y.; ZIAD, M.; LEE, Y. W. **Data quality**. Springer Science & Business Media, 2006

WEBER, Nicholas; YAN, Ann. Integrating user feedback with open data quality models. In: **Proceedings of the Association for Information Science and Technology**, v. 54, n. 1, p. 824-826, 2017.

YILDIZ, Mete. E-government research: Reviewing the literature, limitations, and ways forward. **Government information quarterly**, v. 24, n. 3, p. 646-665, 2007.

ZANELLA, Liane Carly Hermes. **Metodologia de estudo e de pesquisa em administração**. Florianópolis: Departamento de Ciências da Administração/UFSC, 2009.

ZAVERI, Amrapali et al. Quality assessment for linked data: A survey. **Semantic Web**, v. 7, n. 1, p. 63-93, 2016.

ZHU, Xiaolan; GAUCH, Susan. Incorporating quality metrics in centralized/distributed information retrieval on the World Wide Web. In: **Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval**. ACM, 2000. p. 288-295.

ZUIDERWIJK, Anneke; JANSSEN, Marijn; DAVIS, Chris. Innovation with open data: Essential elements of open data ecosystems. **Information Polity**, v. 19, n. 1, 2, p. 17-33, 2014.

ZUIDERWIJK, Anneke; JANSSEN, Marijn. Participation and data quality in open data use: Open data infrastructures evaluated. In: **Proceedings** of the 15th European Conference on E-Government. 2015. p. 351-359.

7. ANEXOS

7.1. Projeto de Atividade SI&D

Sistemas de Informação e Decisão

Prof. Pedro Jácome

Projetos de Aplicativo – 3ª unidade da disciplina

Descrição

A disciplina “Sistemas de Informação e Decisão” (SI&D) tem como requerimento para definição da nota referente à sua 3ª unidade a elaboração e apresentação de um artefato computacional baseado em modelagem dimensional e interface OLAP, denominado aqui aplicativo (ou App). Este App deverá ser desenvolvido em grupo e apresentado à turma em data e hora definidos pelo professor da disciplina.

Regras

As seguintes regras devem ser observadas durante a realização da tarefa:

- 1) os grupos serão formados por até três integrantes, de acordo com sorteio realizado pelo professor;
- 2) a cada grupo será designado uma fonte de dados, acessível por meio de link web. Os dados poderão estar em formatos variados e deverão ser tratados pelo grupo para que sejam utilizados no App (processo ETL);
- 3) o App deverá ser desenvolvido exclusivamente na plataforma de software OLAP utilizada em sala de aula ao longo da disciplina (softwares adicionais podem ser necessários apenas para preparação dos dados, ou seja, na etapa ETL);
- 4) o App desenvolvido por cada grupo deverá utilizar, no mínimo, os recursos da plataforma de software apresentados em sala de aula, podendo expandir o uso dos recursos da plataforma de acordo com o interesse, curiosidade e iniciativa de cada grupo;
- 5) o App desenvolvido por cada grupo deverá possibilitar análises descritivas (qual o comportamento dos dados), diagnósticas (por que os dados se comportam dessa forma) e preditivas (como será o comportamento dos dados no futuro);
- 6) o App desenvolvido deverá responder a questões de negócio formuladas por cada grupo, de acordo com os dados designados ao grupo pelo professor;
- 7) a ordem de apresentação dos Apps será definida em sorteio realizado pelo professor, com no máximo 15 dias de antecedência da apresentação;
- 8) cada grupo deverá realizar a sua apresentação em até 30 minutos;
- 9) o App deverá ser compartilhado com o professor até a véspera da apresentação. A plataforma OLAP utilizada permite o compartilhamento e as instruções serão fornecidas previamente;
- 10) durante a apresentação, o grupo deverá informar (a) quais dados foram tratados, (b) como os dados foram obtidos, (c) que etapas ETL foram necessárias, (d) qual script de carga de dados foi produzido e quantos registros foram carregados, (e) qual modelo de dados foi produzido, (f) quais questões de negócio foram elaboradas a partir dos dados e (g) como essas questões são respondidas pelo App;
- 11) casos omissos serão resolvidos pelo professor da disciplina.

Dados

Os seguintes repositórios de dados abertos poderão ser usados para elaboração dos Apps (a definição de quais dados serão trabalhados por cada grupo será feita pelo professor):

- a) Agência Nacional de Aviação Civil (ANAC), dados disponíveis em <http://www.anac.gov.br/assuntos/dados-e-estatisticas>;
- b) Agência Nacional e Águas (ANA), dados disponíveis em <http://atlasirrigacao.ana.gov.br/> e <http://atlasestados.ana.gov.br/>;
- c) Secretaria de Desenvolvimento da Infraestrutura do Ministério de Planejamento (Programas PAC e Avançar), dados disponíveis em <http://dados.gov.br/dataset/obras-do-pac-programa-de-aceleracao-do-crescimento> e <http://dados.gov.br/dataset/avancar>;
- d) Ministério da Indústria, Comércio Exterior e Serviços, dados disponíveis em <http://www.mdic.gov.br/index.php/comercio-servicos/estatisticas-do-comercio-exterior-de-servicos>
- e) Ministério da Defesa, dados disponíveis em <http://www.defesa.gov.br/dados-abertos>;
- f) Ministério do Turismo, dados disponíveis em <http://dados.turismo.gov.br>, <http://dados.turismo.gov.br/index.php/chegada-de-turistas-internacionais> e <http://dados.turismo.gov.br/eventos-turisticos>;
- g) Ministério da Educação (MEC), Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep), dados disponíveis em <http://portal.inep.gov.br/microdados>;
- h) Controladoria Geral da União (CGU), dados da Copa 2014, disponíveis em <http://dados.gov.br/dataset/copa-2014#>;
- i) Banco de dados do futebol europeu, disponível em <https://www.kaggle.com/hugomathien/soccer/data>;
- j) Banco de dados da Fórmula 1, disponível em <https://www.kaggle.com/cjgdev/formula-1-race-data-19502017/data>;

Condução dos projetos, avaliação e premiações

A implementação, apresentação e dinâmica de elaboração/condução dos projetos responderá pela nota da 3ª unidade da disciplina, em escala de zero a dez pontos. A nota será resultante de dois processos distintos de obtenção de pontos, sendo o primeiro processo descrito abaixo e avaliado pelo professor da disciplina com atribuição de pontuação ao grupo, em escala de zero a dez pontos:

- a) elaboração das questões de negócio (argumentação sobre coerência com os dados e relevância para a tomada de decisão): 1,0 ponto;
- b) implementação de procedimentos ETL (descrição de passos e apresentação de evidências): 1,5 pontos;
- c) modelagem dimensional (apresentação do modelo de dados): 0,5 ponto;
- d) implementação OLAP para análises descritivas (apresentação de pasta específica do App): 2,0 pontos;
- e) sugestão de análises diagnósticas (apresentação de explicações para o comportamento dos dados, conforme descritos no item anterior): 1,0 ponto;
- f) implementação OLAP para análises preditivas (apresentação de pasta específica do App): 2,0 pontos;
- g) apresentação do grupo (coesão, comunicação, postura, capacidade de envolvimento da plateia): 1,0 ponto;
- h) criatividade, pró-atividade, inovação: 1,0 ponto.

O segundo processo de obtenção de pontos ocorre ao longo da implementação do projeto e é resultante da dinâmica de elaboração e condução do projeto em equipe:

- a) os grupos podem contratar consultoria a outros grupos. O grupo contratante deve negociar com o grupo consultor a quantidade de pontos que serão repassados em pagamento pelos serviços. A pontuação acordada (não inferior a 0,5 ponto e não superior a 1,0 ponto) será deduzida da nota

final do grupo contratante e acrescida à nota final do grupo consultor. A consultoria não significa execução de serviços, mas apenas recomendação de como executá-los. Cada grupo só poderá ser contratado uma vez com essa finalidade. Os grupos envolvidos deverão informar ao professor a sua decisão e os detalhes da “contratação”;

- b) os grupos podem alterar a sua composição por meio de “contratação” e “demissão” de integrantes, desde que respeitem a regra de composição de até três integrantes;
- c) para “demitir” algum integrante, o grupo deve informar ao professor a sua decisão, opcionalmente adicionando alguma justificativa. A “demissão” sem justificativa plausível (critério avaliado pelo professor) implica o pagamento de 1,0 ponto ao integrante demitido. Este ponto será deduzido da nota final do grupo e passa a pertencer ao integrante “demitido”;
- d) para “contratar” algum integrante pertencente a outro grupo, o grupo contratante deve negociar com o grupo de origem do “contratado” a quantidade de pontos que serão repassados em pagamento pela “contratação”. A pontuação acordada (não superior a 2,0 pontos) será deduzida da nota final do grupo contratante e acrescida à nota final do grupo de origem do “contratado”. Os grupos envolvidos deverão informar ao professor a sua decisão e os detalhes da “contratação”;
- e) para “contratar” algum integrante não pertencente a outro grupo, o grupo contratante deve convidá-lo e, havendo concordância, a eventual pontuação individual do integrante “contratado” será proporcionalmente acrescida à nota final do grupo “contratante”. O grupo deve informar ao professor a sua decisão;
- f) o integrante “demitido” que eventualmente não consiga “recolocação”, poderá executar o projeto individualmente ou poderá unir-se a outro em mesma situação, desde que respeitem a regra de composição de até três integrantes. O indivíduo (ou o grupo) deve informar ao professor a sua decisão; e
- g) os prazos para “contratação” e “demissão” encerram-se uma semana antes do início das apresentações. O prazo para consultoria encerra-se dois dias antes do início das apresentações.

Ao final de cada apresentação, o professor informará a nota obtida pelo grupo, bem como os eventuais acréscimos/decréscimos de pontos decorrentes de “contratações”, “demissões” e consultorias. Ao final de todas as apresentações, o professor apresentará o *ranking* de grupos e pontuações, com as seguintes premiações:

- 1) o primeiro colocado terá nota 10 na unidade. O segundo colocado terá nota 9,5 na unidade. O terceiro colocado terá nota 9,0 na unidade. Para tanto, devem ter obtido nota do professor maior ou igual a 7,0 pontos;
- 2) o primeiro colocado indicará um grupo concorrente para receber a pontuação especial (de 0,5 a 1,5 pontos) correspondente ao “Destaque SI&D”. O primeiro colocado deverá indicar os critérios adotados para atribuição do destaque;
- 3) os dois primeiros colocados terão prioridade na indicação de seus integrantes como monitores (concorrerão a bolsa de “Monitoria SI&D”) para a disciplina, para fins de atuação no próximo semestre;
- 4) os Apps dos três primeiros colocados serão usados como referência para as turmas seguintes e comporão a “Galeria SI&D”, que é o catálogo dos grupos vencedores e seus Apps, permanentemente exposta na comunidade virtual da disciplina;

8. APÊNDICES

8.1. Termo de Consentimento Livre e Esclarecido – TCLE

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

PESQUISA: A PERCEPÇÃO DA QUALIDADE DOS DADOS ABERTOS GOVERNAMENTAIS À LUZ DA TEORIA PRINCIPAL-AGENTE

As informações contidas nessa folha, fornecidas por Thiago Emídio Esteves da Silva, Contato (082) 98818-7163 têm por objetivo firmar acordo escrito com o(a) voluntário(a) para participação da pesquisa acima referida, autorizando sua participação com pleno conhecimento da natureza dos procedimentos a que ele(a) será submetido(a).

- **Natureza da pesquisa:** Esta pesquisa tem como finalidades: “Descrever a percepção de qualidade de dados abertos governamentais, à luz da Teoria Principal-Agente.”
- **Participantes da pesquisa:** Entre 20-40 Alunos da UFPB.
- **Envolvimento na pesquisa:** Ao participar deste estudo você tem liberdade de se recusar a participar e ainda de se recusar a continuar participando em qualquer fase da pesquisa, sem qualquer prejuízo para você. Sempre que quiser poderá pedir mais informações sobre a pesquisa através do telefone do coordenador do projeto e, se necessário, por meio do telefone do Comitê de Ética em Pesquisa.
- **Confidencialidade:** Todas as informações coletadas neste estudo são estritamente confidenciais. Os dados do(a) voluntário(a) serão identificados com um código, e não com o nome. Apenas os membros da pesquisa terão conhecimento dos dados, assegurando assim sua privacidade.
- **Benefícios:** Ao participar desta pesquisa você não terá nenhum benefício direto. Entretanto, esperamos que este estudo contribua com informações importantes que deve acrescentar elementos importantes à literatura, onde o pesquisador se compromete a divulgar os resultados obtidos.
- **Você não terá nenhum tipo de despesa ao autorizar sua participação nesta pesquisa, bem como nada será pago pela participação. Nem seu nome será gravado fora das premissas deste contrato.**

Após estes esclarecimentos, solicito o seu consentimento de forma livre para permitir sua participação nesta pesquisa. Portanto, preencha os itens que seguem:

CONSENTIMENTO LIVRE E ESCLARECIDO

Eu, _____, RG/CPF _____ após a leitura e compreensão destas informações, entendo que a participação de (escrever o nome do menor), sob minha responsabilidade, é voluntária, e que ele(a) pode sair a qualquer momento do estudo, sem prejuízo algum. Confiro que recebi cópia deste termo de consentimento, e autorizo a execução do trabalho de pesquisa e a divulgação dos dados obtidos neste estudo.

Obs: Não assine esse termo se ainda tiver dúvida a respeito.

8.2. Roteiro de Entrevistas

1. ASSINATURA DA TLCE E EXPLICAÇÃO DA ENTREVISTA
2. PERGUNTAS DEMOGRÁFICAS.
 - 2.1. Qual foi o órgão que você estudou para a matéria?
 - 2.2. Qual seu curso de graduação?
 - 2.3. Como foi seu papel ao fazer este trabalho?
 - 2.4. Qual foi o banco de dados que vocês estudaram?
3. PRIMING
 - 3.1. “A partir de agora, tente se imaginar no processo de desenvolvimento do aplicativo para a disciplina do Prof. Pedro Jácome. Você está começando a lidar com os dados abertos brutos, do jeito que estão no site indicado pelo professor. Tudo que nós vamos discutir envolve esse momento específico em que você desenvolvia o aplicativo”.
4. QUESTÕES REFERENTES AO TRABALHO.
 - 4.1. Foi enfrentada dificuldades ou facilidades para acessar os dados referente a atividade?
 - 4.2. Foi enfrentada dificuldades ou facilidades para entender esses dados?
 - 4.3. Qual sua percepção sobre a confiabilidade destes dados?
 - 4.4. Qual sua percepção sobre o formato em que foi apresentado estes dados?
 - 4.5. Qual sua percepção sobre a efetividade destes dados em informar?
 - 4.6. Se tivesse a oportunidade, Como você mudaria nos dados de modo a melhorar sua efetividade?
 - 4.7. Imagine-se como um cidadão ou organização atendido por este órgão. Em que você mudaria sua perspectiva?
 - 4.8. Qual sua percepção sobre estes dados no papel de prova de que a organização está realizando um serviço que beneficia o cidadão?

8.3. Quadro de Métricas e Referências em Batini *et al.* (2009)

Dimensões	Metodologias de mensuração de qualidade de dados e informação		Autores Base												
	Definição das Métricas em Batini <i>et al.</i> (2009)		Wang, (1998)	English (1999)	Jensfeld <i>et al.</i> (1998)	Lee <i>et al.</i> (2002)	Long e Seko (2005)	Pipino <i>et al.</i> (2002)	Eppler e Munzemaier (2002)	Falorsi <i>et al.</i> (2003)	Su e Jin (2004)	Loshin (2004)	Scannapieco <i>et al.</i> (2004)	Amicis e Batini (2004)	Batini e Scannapieco (2006)
			TDQM	DWQ	TIQM	AIMQ	CHH	DQA	IQM	ISTAT	AMEQ	COLDQ	DaQuincis	QAFD	CDQ
Precisão	Precisão sintática: é medida como a distância entre o valor armazenado no banco de dados e o correto. Precisão Sintática = Número de valores corretos / número de valores totais		X		X			X		X	X	X	X	X	X
Precisão	Número de tuplas precisas entregues			X											
Precisão	Avaliação do Usuário - Questionário				X	X			X						
Completeness	Integralidade = Número de valores não nulos / número total de valores		X		X			X			X	X	X	X	X
Completeness	Completeness = Número de tuplas entregues / Número esperado			X	X										
Completeness	Integralidade dos dados da Web = $(T_{max} - T_{atual}) * (Completeness_{Max} - Completeness_{atual}) / 2$														X
Completeness	Avaliação do Usuário - Questionário				X										
Consistência	Consistência = Número de valores consistentes / número de valores totais		X					X							

Dimensões	Definição das Métricas em Batini et al. (2009)	Metodologias de mensuração de qualidade de dados e informação												
		TDQM	DWQ	TIQM	AIMQ	CHHI	DQA	IQM	ISTAT	AMEQ	COLDQ	DaQuin CIS	QAFD	CDQ
Autores Base														
Consistência	Número de tuplas que violam restrições, número de diferenças de codificação		X											
Consistência	Número de páginas com desvio de guia de estilo							X						
Consistência	Avaliação do Usuário - Questionário				X	X								
Pontualidade	Pontualidade = (max (0; 1-Atualidade / Volatilidade)) s						X			X			X	
Pontualidade	Porcentagem de execuções de processos que podem ser executadas dentro do prazo necessário			X						X				
Pontualidade	Avaliação do Usuário - Questionário				X	X								
Atualidade	Atualidade = Hora em que os dados são armazenados no sistema - hora em que os dados são atualizados no mundo real								X		X			
Atualidade	Hora da última atualização	X						X				X		
Atualidade	Atualidade = hora do pedido - última atualização				X	X								
Atualidade	Atualidade = Idade + (Tempo de entrega - Tempo de entrada)													X
Atualidade	Avaliação do Usuário - Questionário				X	X								
Volatilidade	Tempo em que os dados permanecem válidos	X												X
Singularidade	Número de duplicatas			X								X		

Dimensões	Definição das Métricas em Batini et al. (2009)	Metodologias de mensuração de qualidade de dados e informação												
		TDQM	DWQ	TQM	AIMQ	CHHI	DQA	IQM	ISTAT	AMEQ	COLDQ	DaQuin CIS	QAFD	CDQ
Quantidade adequada de dados	Montante de dados apropriado = Mín. ((Número de unidades de dados fornecidas / Número de unidades de dados necessárias); (Número de unidades de dados necessárias / Número de unidades de dados fornecidas))						X							
Quantidade adequada de dados	Avaliação do Usuário - Questionário				X									
Acessibilidade	Acessibilidade = $\max(0; 1 - (\text{Prazo de entrega} - \text{Hora da solicitação}) / (\text{Prazo final} - \text{Hora da solicitação}))$							X						
Acessibilidade	Número de links quebrados - Número de âncoras quebradas								X					
Acessibilidade	Avaliação do Usuário - Questionário			X	X									
Credibilidade	Número de tuplas com valores padrão		X											
Credibilidade	Avaliação do Usuário - Questionário				X									
Interpretabilidade	Número de tuplas com dados interpretáveis, documentação para valores-chave		X											
Interpretabilidade	Avaliação do Usuário - Questionário				X									
Usabilidade	Avaliação do Usuário - Questionário			X										

Autores Base

Batini e Scannapieco (2006)

Amicis e Batini (2004)

Scannapieco et al. (2004)

Loshin (2004)

Su e Jin (2004)

Falorsi et al (2003)

Eppler e Munzmaier (2002)

Pipino et al. (2002)

Long e Seko (2005)

Lee et al. (2002)

Jeusfeld et al. (1998)

English (1999)

Wang,(1998)

Dimensões	Metodologias de mensuração de qualidade de dados e informação	Batini e Scannapieco (2006)	Amicis e Batini (2004)	Scannapieco et al. (2004)	Loshin (2004)	Su e Jin (2004)	Falorsi et al (2003)	Eppler e Munzmaier (2002)	Pipino et al. (2002)	Long e Seko (2005)	Lee et al. (2002)	Jeusfeld et al. (1998)	English (1999)	Wang,(1998)
		CDQ	QAFD	DaQuin CIS	COLDQ	AMEQ	ISTAT	IQM	DQA	CHI	AIMQ	TIQM	DWQ	TDQM
Relevância	Avaliação do Usuário - Questionário										X			
Reputação	Avaliação do Usuário - Questionário										X			
Facilidade de operação	Avaliação do Usuário - Questionário										X			
Interatividade	Número de formulários - Número de páginas personalizáveis							X						

Fonte: Autor, Baseado em Batini et al. (2009)

8.4. Primeira Lista de Conceitos Obtidos a partir da Literatura

TCon	Conceito obtido através da fonte principal	Autor Base
Qualidade de dados		
Precisão	Capacidade dos dados de refletir a realidade	Batini et al(2009)
Compleitude	Esgotamento de representação de informação para cada unidade de valor/célula	Batini et al(2009)
Consistência	Padronização da representação entre dados relacionados	Batini et al(2009)
Pontualidade	Dados chegando no momento necessário	Batini et al(2009)
Atualidade	Dados sobre atualização, sua rapidez e seu tempo de publicidade	Batini et al(2009)
Volatilidade	Tempo em que os dados permanecem válidos	Batini et al(2009)
Singularidade	Existência de duplicatas	Batini et al(2009)
Quantidade adequada de dados	Adequação do número de dados para as necessidades de consumidor e de publicador de dados	Batini et al(2009)
Acessibilidade	Capacidade de acesso/obter dados no momento necessário	Batini et al(2009)
Credibilidade	Confiança nos dados, ou em quem os produziu.	Batini et al(2009)
Interpretabilidade	Capacidade de interpretar os dados representados	Batini et al(2009)
Usabilidade	Capacidade de usar os dados pelo usuário de dados	Batini et al(2009)
Integridade da Derivação	Presença de cálculos de valores corretos, baseados nos dados.	Batini et al(2009)
Concisão	Representação compacta sem sobrecarregar o consumidor de dados	Batini et al(2009)
Metadados	Dados sobre os dados presentes em dado banco	Batini et al(2009)
Aplicabilidade	Quantidade de informações não utilizadas	Batini et al(2009)
Conveniência	Capacidade de navegar para encontrar os dados	Batini et al(2009)
Rapidez	Rapidez de resposta do servidor/rede	Batini et al(2009)
Compreensibilidade	Capacidade de compreender os dados, pelo consumidor de dados	Batini et al(2009)
Clareza	Facilidade para compreender dos dados, pelo consumidor de dados	Batini et al(2009)
Rastreabilidade	Registro de autoria e/ou fonte dos dados	Batini et al(2009)
Segurança	Grau de segurança eletrônica percebido a partir do banco de dados	Batini et al(2009)
Exatidão	Exatidão dos dados percebida pelo consumidor de dados	Batini et al(2009)
Objetividade	Dado não subjetivo ou parcial	Batini et al(2009)
Relevância	Importância dos dados para o consumidor	Batini et al(2009)
Reputação	Reputação dos dados ou publicador de dados	Batini et al(2009)
Facilidade de operação	Facilidade de operação para o consumidor de dados	Batini et al(2009)
Interatividade	Capacidade de interagir com o publicador de dados ou banco de dados para atender necessidades específicas.	Batini et al(2009)
Licenciamento	Licenciamento do uso de banco de dados	Behkamal et al. (2014)
Interligação	Capacidade de interligar bancos de dados	Behkamal et al. (2014)
Disponibilidade	A disponibilidade para o consumidor de dados	Behkamal et al. (2014)
Custo	Custo de obter os dados	Loshin (2004)
Backup	Proteção quanto a perda de dados	Jaffar et al. (2015)
Sustentabilidade	Sustentabilidade da conexão com os dados (quando necessária, inclui APIs)	Jaffar et al. (2015)
Eficiência	Eficiência (em recursos) em conexão de dados	Jaffar et al. (2015)
Versatilidade	Versatilidade dos dados para o consumidor de dados	Zaveri et al. (2016)

	(não exato, mas necessário observar)	
Autorização	Autorização para o uso/acesso para o consumidor de dados	Ijab (2017)
TPA		
Autodidatismo	Presença de formas com quais o principal pode aprender como obter dados sozinho	Mitnich (1973, 1976, 2006)
Confiança	Grau de desconfiança atenuada pela presença de dados abertos	Mitnich (1973, 1976, 2006)
Governança		
Prestação de contas	Grau de responsabilidade percebido pelo consumidor de dados	Banco Mundial (1992,2017)
Transparência	Grau de transparência observado pelo consumidor de dados	Banco Mundial (1992,2017)
Estrutura legal	Leis voltadas ao desenvolvimento e governança	Banco Mundial (1992,2017)

Fonte: Autor