



UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO TECNOLÓGICO
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA E GESTÃO DO
CONHECIMENTO

Rafael Bassegio Caumo

Indicadores socioeconômicos produzidos a partir de Big Data:
Um *framework* para avaliação da qualidade estatística aplicado ao turismo

FLORIANÓPOLIS
2021

Rafael Bassegio Caumo

Indicadores socioeconômicos produzidos a partir de Big Data:

Um *framework* para avaliação da qualidade estatística aplicado ao turismo

Tese submetida ao Programa de Pós-graduação em Engenharia e Gestão do Conhecimento da Universidade Federal de Santa Catarina para a obtenção do título de Doutor em Engenharia e Gestão do Conhecimento.

Orientador: Prof. Dr. João Artur de Souza

Coorientador: Prof. Dr. Alexandre Augusto Biz

Coorientadora externa: Profa. Dra. Denise Britz do Nascimento Silva

Florianópolis

2021

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Caumo, Rafael Bassegio

Indicadores socioeconômicos produzidos a partir de Big Data: Um framework para avaliação da qualidade estatística aplicado ao turismo / Rafael Bassegio Caumo ; orientador, João Artur de Souza, coorientador, Alexandre Augusto Biz, coorientadora, Denise Britz do Nascimento Silva, 2021.

240 p.

Tese (doutorado) - Universidade Federal de Santa Catarina, Centro Tecnológico, Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento, Florianópolis, 2021.

Inclui referências.

1. Engenharia e Gestão do Conhecimento. 2. Indicadores socioeconômicos. 3. Big Data. 4. Qualidade estatística. 5. Turismo. I. de Souza, João Artur. II. Biz, Alexandre Augusto. III. do Nascimento Silva, Denise Britz IV. Universidade Federal de Santa Catarina. Programa de Pós Graduação em Engenharia e Gestão do Conhecimento. V. Título.

Rafael Bassegio Caumo

Indicadores socioeconômicos produzidos a partir de Big Data:

Um *framework* para avaliação da qualidade estatística aplicado ao turismo

O presente trabalho em nível de doutorado foi avaliado e aprovado por banca examinadora composta pelos seguintes membros:

Prof. Alexandre Leopoldo Gonçalves, Dr.
Universidade Federal de Santa Catarina

Prof. José Leomar Todesco, Dr.
Universidade Federal de Santa Catarina

Prof. Paulo de Martino Jannuzzi, Dr.
Escola Nacional de Ciências Estatísticas

Prof. Cristiano Ferraz, Dr.
Universidade Federal de Pernambuco

Certificamos que esta é a **versão original e final** do trabalho de conclusão que foi julgado adequado para obtenção do título de doutor em Engenharia e Gestão do Conhecimento.

Prof. Dr. Roberto Carlos dos Santos Pacheco
Coordenador do Programa

Prof. Dr. João Artur de Souza
Orientador

Prof. Dr. Alexandre Augusto Biz – Coorientador
Profa. Dra. Denise Britz do Nascimento Silva – Coorientadora externa

Florianópolis, 07 de julho de 2021.

AGRADECIMENTOS

Os agradecimentos vão para todos que de alguma forma me apoiaram e/ou estiveram presentes, direta ou indiretamente, nesta longínqua caminhada.

A todos os familiares, em especial ao meu pai, Jatir, e à minha mãe, Santana, assim como ao meu irmão, Lucas. Por estarem sempre ao meu lado, assim como por toda base e sustento de vital importância.

Ao meu amor e companheira, Bruna. Sempre presente, me deu muita força para continuar. Devo muito a ela que esteve e segue incondicionalmente ao meu lado.

A todos os amigos que a vida trouxe. Fontes de alento e energia para seguir. Confraria da Cachoeira, Corre Atrás, Estela, CS, Meia Nuca e todos os outros. Estarão sempre muito bem guardados, cada um com muita consideração.

Ao corpo docente e aos colegas de PPGE/GC e IGTI, principalmente aos professores orientadores Dr. João Artur de Souza e Dr. Alexandre Augusto Biz, por todos os ensinamentos e contribuições, e aos colegas Alessandro Ribeiro, sobretudo pelo auxílio com a programação em Python, e Aline Valdati, especialmente pelo conhecimento de metodologia compartilhado.

À coorientadora Profa. Dra. Denise Britz do Nascimento Silva por toda a atenção e dedicação despendida, contribuindo expressivamente e de forma crucial para a evolução deste trabalho. Foi de extrema satisfação poder ter aprendido ainda mais com tão destacada e competente profissional.

Aos professores avaliadores componentes da banca, Dr. Paulo de Martino Jannuzzi, Dr. José Leomar Todesco, Dr. Alexandre Leopoldo Gonçalves e Dr. Cristiano Ferraz, e aos profissionais especialistas participantes do painel Delphi, pela disponibilidade, contribuições e ensinamentos.

Aos colegas professores da ESAG/Udesc, minha segunda casa nesses quatro anos, por todos os momentos vividos.

Muitíssimo obrigado.

RESUMO

Representações de conhecimento explícito sob a forma de indicadores socioeconômicos são tradicionalmente construídas por produtores de estatísticas públicas do setor público, pela academia e por instituições do terceiro setor, usualmente se utilizando de censos, pesquisas amostrais e registros administrativos. Apesar de reconhecidos pelo rigor científico, tais processos de produção de informação são por vezes caros, demorados e de operacionalização complexa, uma vez que consideram predominantemente levantamentos em campo conduzidos por pessoas. Com o advento da Revolução dos Dados e do Big Data, a utilização de fontes alternativas de dados traz oportunidades de contorno a algumas das problemáticas dos processos tradicionais de produção de indicadores socioeconômicos. Entretanto, novos desafios se apresentam e no centro das dificuldades estão questões relativas à aferição e à garantia da qualidade estatística das informações produzidas. Com o entendimento de que a estrutura dos arcabouços para avaliação da qualidade estatística depende do contexto de sua aplicação, optou-se pelo foco nos indicadores de turismo. A escolha referente à delimitação setorial se deu pela percepção de que o turismo possui muita relevância na atividade econômica de Florianópolis e está na linha de frente do aproveitamento dos benefícios trazidos pelo Big Data. Nesse contexto, a presente tese tem por objetivo propor um *framework* para a avaliação da qualidade de indicadores de turismo produzidos a partir de Big Data. No âmbito da *Design Science* enquanto paradigma filosófico, utilizou-se do método *Design Science Research* para condução da pesquisa, com auxílio de técnicas do tipo pesquisa documental e bibliográfica – de forma sistemática e integrativa – durante a fase do desenvolvimento do artefato e de questionários aplicados a especialistas no tema – de acordo com a técnica Delphi – para fins de verificação de estrutura. Em paralelo, testou-se a aplicabilidade do *framework* em um contexto prático, utilizando-se de um indicador que se propõe a refletir a satisfação do turista de Florianópolis e de Balneário Camboriú, construído a partir de dados extraídos da plataforma TripAdvisor. Ao final, o *framework* resultante se diferencia dos já existentes por ser mais abrangente em termos de aspectos de qualidade que considera e ser o único até então percebido em nível nacional e também para o contexto dos indicadores de turismo. Sua estrutura considera três domínios, cinco dimensões e vinte e sete requisitos de qualidade, tendo cunho primordialmente conceitual, porém acompanhado de uma primeira proposta de protocolo para operacionalização prática, capaz de emitir pareceres de qualidade gerais ou parciais, a servir de referência para trabalhos futuros. Acredita-se que os resultados permitiram avançar frente às lacunas de pesquisa e se espera que contribuam para elucidar e disseminar ainda mais as possibilidades de inovação e complementariedade aos processos tradicionais de produção de estatísticas, assim como para orientar a aferição da qualidade dos produtos inovadores.

Palavras-chave: Indicadores socioeconômicos. Estatísticas públicas. Big Data. Qualidade estatística. Turismo.

ABSTRACT

Explicit knowledge representations as socioeconomic indicators are usually produced by public statistics institutions around public sector, academy and third sector, mainly through censuses, sample surveys and administrative records. Despite recognized by its scientific soundness, these information production processes are sometimes expensive, time-consuming, and complex to operate, since they consider predominantly human-led field surveys. However, the advent of the Data Revolution and Big Data are coming up with opportunities to address some of the socioeconomic indicators traditional production problems through the utilization of alternative data sources. Nevertheless, new challenges emerge and the measurement and guarantee of statistical quality of the produced indicators seems to be a core difficulty. Understanding that the structure of frameworks for statistics quality assessment depends on the context of its application, it was decided to focus here on tourism indicators. The sectoral delimitation occurs due to the relevance of tourism activity in Florianópolis's economy and to the fact of being in the forefront of taking advantage from Big Data benefits. In this context, the present thesis aims to propose a framework for assessing the statistical quality of tourism indicators produced from Big Data. Utilizing Design Science as the philosophical paradigm, the research was conducted through Design Science Research method, supported by documentary and bibliographic research techniques during the artifact development phase – in a systematic and integrative way – and questionnaires applied to subject specialists to verify its structure – according to the Delphi technique. In parallel, the applicability of the framework was tested in a practical context, making use of an indicator that aims to reflect tourist satisfaction in Florianópolis and Balneário Camboriú, built from TripAdvisor's data. The resulting framework differs from the existing ones by being more comprehensive in terms of quality aspects that considers and being the only one so far perceived at the national level and for the context of tourism indicators. Its structure has three domains, five dimensions and twenty-seven quality requirements, having primarily conceptual nature, but accompanied by a first practical operationalization protocol proposal, capable of issuing general or partial quality opinions, serving as a reference for future work. It is believed that the results have made it possible to advance through research gaps and are expected to contribute to further disseminate and encourage the possibilities of innovation in traditional statistics production processes, as well as to guide the quality measurement of innovative products.

Keywords: Socioeconomic indicators. Public statistics. Big data. Statistical quality. Tourism.

LISTA DE FIGURAS

Figura 1 – Fatores que compõem a relevância da pesquisa.	35
Figura 2 – Resumo de algumas dificuldades recentes dos produtores de estatísticas públicas com levantamentos do tipo de campo.	66
Figura 3 – Resumo de desafios para produtores de estatísticas públicas no contexto da utilização de Big Data.	90
Figura 4 – Etapas da <i>Design Science Research</i> (DSR).	103
Figura 5 – Resumo da sequência lógica percorrida até a definição do problema da pesquisa.	119
Figura 6 – Relações entre dimensões de qualidade de insumos, processo de produção e produtos estatísticos gerados a partir de Big Data.	122
Figura 7 – Estrutura e funcionamento do <i>framework</i> desenvolvido.	136
Figura 8 – Síntese do processo e resultados esperados da aplicação do <i>Framework</i>	137
Figura 9 – Recorte da estrutura original dos dados capturados no portal TripAdvisor.	149
Figura 10 – Distribuição das avaliações válidas coletadas de jan/2015 a dez/2020, por variáveis registradas e municípios selecionados.	150
Figura 11 – Nível de satisfação mensal do turista com as principais atrações locais de jan/2015 a dez/2020, por municípios selecionados.	150
Figura 12 – Quantidade de avaliações mensais consideradas de jan/2015 a dez/2020, por municípios selecionados.	152

LISTA DE QUADROS

Quadro 1 – Necessidades relativas à <i>frameworks</i> para a qualidade de produtos estatísticos gerados a partir de Big Data.....	28
Quadro 2 – Pesquisas do PPGE GC sobre construção, descoberta e aquisição de conhecimento a partir de fontes alternativas de dados.	34
Quadro 3 – Pesquisas do PPGE GC sobre engenharia e gestão do conhecimento no turismo.	34
Quadro 4 – Comparativo entre <i>frameworks</i> tradicionais de qualidade para estatísticas oficiais e públicas selecionados.	55
Quadro 5 – Conteúdos e possíveis fontes de informações sugeridas para o SITUR Fortalecido pela PNETUR.	70
Quadro 6 – Possíveis oportunidades trazidas pelo Big Data para a produção de estatísticas públicas e sua relação com as dificuldades da Figura 2.	81
Quadro 7 – Fontes de Big Data por tipo e áreas de aplicação já trabalhadas nas publicações investigadas.....	83
Quadro 8 – Fontes e aplicações de Big Data para indicadores e estatísticas de turismo.	86
Quadro 9 – Percentual de publicações classificadas nos tipos de finalidade da aplicação.	87
Quadro 10 – Revisões de literatura de temas correlatos ao proposto por esta seção.	88
Quadro 11 – Trechos associados a possíveis lacunas de pesquisa no âmbito da qualidade de produtos estatísticos gerados a partir de Big Data.	98
Quadro 12 – Bases filosóficas para a pesquisa científica.	102
Quadro 13 – Enquadramento metodológico da pesquisa.	103
Quadro 14 – Método de trabalho da pesquisa.	104
Quadro 15 – Condicionantes de busca das revisões sistemáticas de literatura.....	107
Quadro 16 – Etapas da técnica Delphi e critérios de referência considerados.	110
Quadro 17 – Especialistas participantes da etapa de avaliação do artefato.	111
Quadro 18 – Características do indicador de turismo utilizado na análise da aplicabilidade.....	116

Quadro 19 – Resumo dos objetivos e conclusões das revisões sistemáticas realizadas.....	127
Quadro 20 – Comparativo entre <i>frameworks</i> selecionados de qualidade para estatísticas oficiais e públicas produzidas a partir de Big Data.	130
Quadro 21 – Definições das dimensões de qualidade consideradas no <i>framework</i> proposto.	138
Quadro 22 – Relação entre domínios, requisitos e dimensões da qualidade estatística do <i>framework</i> proposto.....	141
Quadro 23 – Definições dos requisitos considerados no <i>framework</i> proposto.....	143
Quadro 24 – Aplicação do <i>framework</i> sobre o indicador do nível de satisfação do turista.	153
Quadro 25 – Solicitações de modificações e reflexões para o <i>framework</i>	166
Quadro 26 – Manifestações dos especialistas após a segunda rodada de questionários.....	178

LISTA DE TABELAS

Tabela 1 – Qualidade estatística do indicador de turismo, por domínios e dimensões.	163
Tabela 2 – Escore médio e distribuição dos especialistas por nível de concordância com os pontos avaliados do <i>framework</i> nas diferentes rodadas do Delphi.....	178

LISTA DE ABREVIATURAS E SIGLAS

AAPOR - *American Association for Public Opinion Research*
ABNT - Associação Brasileira de Normas Técnicas
ABS - Bureau de Estatística da Austrália
ANAC - Agência Nacional de Aviação Civil
ANTAQ - Agência Nacional de Transportes Aquaviários
ANTT - Agência Nacional de Transportes Terrestres
APNOS - *Accreditation Procedure for statistical data from Non-Official Sources*
BACEN - Banco Central do Brasil
BDQAF - *Big Data Quality Assessment Framework*
BDQF - *Big Data Quality Framework*
BDTEF - *Big Data Total Error Framework*
CADASTUR - Cadastro dos Prestadores de Serviços Turísticos
CAGED - Cadastro Geral de Empregados e Desempregados
CDC - Código de Defesa do Consumidor
CEBD - *UN Committee of Experts on Big Data and Data Science for Official Statistics*
CEMPRE - Cadastro Central de Empresas
CFQBD - *Conceptual Framework for Quality in Big Data*
CNPJ - Cadastro Nacional da Pessoa Jurídica
CNT - Conselho Nacional de Turismo
CONRE 4 - Conselho Regional de Estatística da 4ª Região
CoP - *European Statistics Code of Practice*
CPED - Centro de Pesquisa de Emprego e Desemprego
DQAF - *Data Quality Assessment Framework*
DQF - *Data Quality Framework*
DS - *Design Science*
DSR - *Design Science Research*
ECB - Banco Central Europeu
Embratur - Instituto Brasileiro de Turismo
ENCE - Escola Nacional de Ciências Estatísticas
ESAG - Centro de Ciências da Administração e Socioeconômicas

ESS - Sistema Estatístico Europeu

ETL - *Extract, Transform, Load*

EU - *European Union*

Eurostat - Gabinete de Estatísticas da União Europeia

FAO - Organização das Nações Unidas para a Alimentação e a Agricultura

FEE - Fundação de Economia e Estatística

FGV - Fundação Getúlio Vargas

Fiocruz - Fundação Oswaldo Cruz

FIPE - Fundação Instituto de Pesquisas Econômicas

FMI - Fundo Monetário Internacional

GDPR - Regulamento Geral sobre Proteção de Dados

GPS - Sistema de Posicionamento Global

GSBPM - *Generic Statistical Business Process Model*

IBGE - Instituto Brasileiro de Geografia e Estatística

IGTI - Núcleo de Estudos em Inteligência, Gestão e Tecnologias para Inovação

Inep - Instituto Nacional de Estudos e Pesquisas Educacionais

INFOPLAN - Conferência Nacional de Produtores e Usuários de Informações Estatísticas, Geográficas e Ambientais

INFRAERO - Empresa Brasileira de Infraestrutura Aeroportuária

Inpe - Instituto Nacional de Pesquisas Espaciais

IP - *Internet Protocol*

IPEA - Instituto de Pesquisa Econômica Aplicada

ISO - *International Organization for Standardization*

Komuso - *ESSnet on quality of multisource statistics*

LGPD - Lei Geral de Proteção aos Dados Pessoais

NASEM - *National Academies of Sciences, Engineering, and Medicine*

OCDE - Organização para a Cooperação e Desenvolvimento Econômico

ODS - Objetivos de Desenvolvimento Sustentável

OMT - Organização Mundial do Turismo

ONU - Organização das Nações Unidas

PED - Pesquisa de Emprego e Desemprego

PNAD - Pesquisa Nacional por Amostras de Domicílios

PNETUR - Plano Nacional Estratégico de Estatísticas Turísticas

PNT - Plano Nacional de Turismo

PRISMA - *Preferred Reporting Items for Systematic Reviews and Meta-Analyses*

PSH - Pesquisa de Serviços de Hospedagem

PPGEGC - Programa e Pós-Graduação em Engenharia e Gestão do Conhecimento

QAF - *Quality Assurance Framework*

RAIS - Relação Anual de Informações Sociais

RMPA - Região Metropolitana de Porto Alegre

RS - Rio Grande do Sul

MTE - Ministério do Trabalho e Emprego

MTur - Ministério do Turismo

NQAF - *National Quality Assurance Framework*

SCQG - *Statistics Canada Quality Guidelines*

SEBRAE - Serviço Brasileiro de Apoio às Micro e Pequenas Empresas

SES - Secretaria da Saúde do Estado

SITUR - Sistema de Nacional de Informação Turística

SNIPC - Sistema Nacional de Índices de Preços ao Consumidor

SNRHos - Sistema Nacional de Registro de Hóspedes

SQAF - *Statistics Quality Assurance Framework*

SQF - *Statistics Quality Framework*

TICs - Tecnologias da Informação e Comunicação

TQM - *Total Quality Management*

TDQM - *Total Data Quality Management*

TSE - *Total Survey Error*

TSQ - *Total Survey Quality*

Udesc - Universidade do Estado de Santa Catarina

UFRGS - Universidade Federal do Rio Grande do Sul

UFSC - Universidade Federal de Santa Catarina

UKSA - Autoridade de Estatísticas do Reino Unido

UN - Organização das Nações Unidas

UNECE - Comissão Econômica das Nações Unidas para a Europa

UNSC - Comissão de Estatística das Nações Unidas

UNWTO - Organização Mundial do Turismo das Nações Unidas

WTTC - *World Travel & Tourism Council*

WWW - *World Wide Web*

SUMÁRIO

1	INTRODUÇÃO	19
1.1	CONTEXTUALIZAÇÃO	19
1.2	OBJETIVOS	31
1.2.1	Objetivo Geral	31
1.2.2	Objetivos Específicos	31
1.3	ENQUADRAMENTO METODOLÓGICO	31
1.4	ADERÊNCIA AO EGC	32
1.5	RELEVÂNCIA	35
1.6	ORIGINALIDADE	38
1.7	DELIMITAÇÕES	39
1.8	EXPERIÊNCIA DO PESQUISADOR	40
1.9	ESTRUTURA DO DOCUMENTO	40
2	REFERENCIAL TEÓRICO	42
2.1	CONHECIMENTO	42
2.2	INOVAÇÃO	44
2.3	INDICADORES SOCIOECONÔMICOS E ESTATÍSTICAS PÚBLICAS	45
2.3.1	A produção tradicional	48
2.3.2	Qualidade estatística: dimensões e <i>frameworks</i> tradicionais	50
2.3.3	Dificuldades da produção tradicional	60
2.4	INDICADORES DE TURISMO	66
2.4.1	A produção tradicional no Brasil	68
2.5	BIG DATA E REVOLUÇÃO DOS DADOS	72
2.5.1	Valor para a construção de conhecimento	76
2.6	BIG DATA NA PRODUÇÃO DE INDICADORES SOCIOECONÔMICOS	77
2.6.1	Oportunidades	80

2.6.2	Possibilidades para o turismo	85
2.6.3	Desafios	89
2.6.4	A questão da qualidade	96
3	METODOLOGIA	100
3.1	PROCEDIMENTOS E TÉCNICAS DE INVESTIGAÇÃO	105
3.1.1	Pesquisa documental e bibliográfica	105
3.1.2	Questionário aplicado a especialistas	109
3.1.2.1	Mineração <i>web</i> para a geração de indicador de turismo	114
4	RESULTADOS	118
4.1	IDENTIFICAÇÃO DO PROBLEMA.....	118
4.2	CONSCIENTIZAÇÃO DO PROBLEMA	120
4.2.1	Dimensões e aspectos de qualidade	121
4.2.2	Operacionalização da mensuração	124
4.2.3	Frameworks específicos ou adaptáveis	125
4.3	REVISÃO SISTEMÁTICA DA LITERATURA.....	127
4.4	IDENTIFICAÇÃO DOS ARTEFATOS E CLASSES DE PROBLEMAS	128
4.5	PROPOSIÇÃO DE ARTEFATO.....	133
4.6	PROJETO DO ARTEFATO SELECIONADO	135
4.7	DESENVOLVIMENTO DO ARTEFATO	137
4.8	AVALIAÇÃO DO ARTEFATO.....	147
4.8.1	Aplicabilidade do <i>framework</i>	147
4.8.2	Verificação da estrutura <i>framework</i>	165
4.9	ESPECIFICAÇÃO DAS APRENDIZAGENS.....	179
5	CONSIDERAÇÕES FINAIS	182
5.1	CONCLUSÕES.....	182
5.1.1	Reflexão final	188
5.2	GENERALIZAÇÃO PARA UMA CLASSE DE PROBLEMAS	189
5.3	COMUNICAÇÃO DOS RESULTADOS	189

REFERÊNCIAS.....	191
APÊNDICE A – Carta convite enviada aos especialistas participantes.....	223
APÊNDICE B – Apresentação do <i>framework</i> para os especialistas.....	224
APÊNDICE C – Primeiro questionário aplicado aos especialistas.....	226
APÊNDICE D – Apresentação do <i>framework</i> reformulado para os especialistas	229
APÊNDICE E – Segundo questionário aplicado aos especialistas	231
APÊNDICE F – <i>Script</i> (Python) do <i>crawler</i> utilizado na mineração <i>web</i>.....	234
ANEXO A – Relação de fontes de dados sobre o turismo no Brasil....	235
ANEXO B – <i>UNECE Big Data Quality Framework</i>	236
ANEXO C – <i>Eurostat Accreditation Procedure for statistical data from Non-Official Sources</i>	237
ANEXO D – <i>AAPOR Big Data Total Error</i>.....	238
ANEXO E – <i>Conceptual Framework for Quality in Big Data</i>	239
ANEXO F – <i>Big Data Quality Assessment Framework</i>	240

1 INTRODUÇÃO

Este capítulo apresenta os elementos norteadores do trabalho, de forma resumida, com intuito de transmitir um entendimento inicial sobre o contexto, os objetivos e a organização do documento.

1.1 CONTEXTUALIZAÇÃO

O conhecimento, entendido como conteúdo ou processo efetivado por agentes humanos ou artificiais em atividades de geração de valor científico, tecnológico, econômico, social ou cultural (PACHECO, 2014), vem sendo percebido como um ativo intangível organizacional. Considerado um fator de produção, insumo para a cadeia de geração de valor (DRUCKER, 1993), tornou-se elemento central no contexto da economia pós-industrial da sociedade do conhecimento (BELL, 1999; UNESCO, 2005; CASTELFRANCHI, 2007), sendo reconhecido o mais poderoso produtor de riqueza, superando a importância da terra, do trabalho e do capital.

Sob a ótica da epistemologia dos cognitivistas (VENZIN et al., 1998; PACHECO, 2016), o conhecimento pode também ser explicitado para além das mentes humanas. Nesse contexto, dados e informações podem ser considerados representações de conhecimento explícito quando colocados em prática durante processos de tomadas de decisão, orientando a realização de tarefas e ações em busca por resultados concretos (SCHREIBER et al., 2000).

Assim, indicadores socioeconômicos podem ser vistos como uma relevante fonte de conhecimento explícito relativa ao ambiente externo às organizações. Caracterizam-se por medidas quantitativas capazes de ilustrar e comunicar um conjunto de fenômenos complexos de uma forma simples (OCDE, 2002; BOSSEL, 1999), geralmente relativos à aspectos econômicos, demográficos, sociais e ambientais de um território. Percebidos como bens públicos essenciais para o desenvolvimento de uma nação (UNECE, 1992), consistem em insumos fundamentais para o planejamento e a formulação de políticas e estratégias no

mundo contemporâneo (JANNUZZI, 2020; JANNUZZI; GRACIOSO, 2002; SCHNORR-BACKER, 2016).

Na administração pública, indicadores socioeconômicos podem fomentar melhores e mais eficazes e eficientes decisões no âmbito da operacionalização das políticas e estratégias, subsidiando processos de planejamento, implementação, monitoramento e avaliação (JANNUZZI, 2020; JANNUZZI, 2001), alcançando caráter institucional, com necessidade de serem periódicos e abrangentes, subsidiando e permitindo uma atuação autônoma, fundamentada e propositiva ao Estado (LINDBLOM, 1991). No setor privado, essas mesmas representações de conhecimento explícito podem ser utilizadas para orientar decisões de ações e investimentos com maior precisão e menor risco durante análises de mercado e identificação de oportunidades de negócio (LOURENÇO; ROMERO, 2002).

Em virtude da sua importância no suporte a processos de tomada de decisão, se faz necessário que tais indicadores sejam gerados com imparcialidade, confiabilidade, qualidade, respeitando padrões que permitam comparabilidade ao longo do tempo e entre diferentes localidades, entre outros princípios fundamentais (UNECE, 1992; IBGE, 2021).

Sob esses condicionantes, agentes que constroem indicadores socioeconômicos correspondem predominantemente aos agentes produtores de estatísticas públicas e oficiais, concentrados tradicionalmente no setor público, nos institutos de pesquisa, na academia e no terceiro setor. Produtores que: servem a todos os setores da sociedade, baseando-se nas melhores práticas e em critérios de qualidade orientados por estatísticos com independência profissional e objetividade (EUROSTAT, 2014b); encarregam-se de produzir e disponibilizar estatísticas sobre uma série de domínios e em diversos níveis de escala e granularidade (KITCHIN, 2015); e existem para subsidiar processos de tomadas de decisão com informações (HARWOOD; MAYER, 2016). Assim sendo, tais produtores utilizam-se das estatísticas públicas sob a forma de insumos para a construção de indicadores socioeconômicos – haja visto que estatísticas públicas são dados parcialmente preparados para uso na interpretação empírica de uma realidade (JANNUZZI, 2001).

Seus processos de produção têm sido tradicionalmente conduzidos à luz do rigor científico – com métodos de pesquisa consolidados. Em geral, utilizando-se de

pesquisas amostrais e censos – e, especialmente no contexto mais recente, valendo-se de registros administrativos e cadastros públicos (HACKL, 2016).

As pesquisas amostrais e os censos, práticas até então predominantes, são levantamentos que possuem utilidade e qualidade reconhecida por conta do sustento metodológico estatístico cientificamente consolidado, rigorosamente desenvolvido e padronizado ao longo de décadas (KITCHIN, 2015). São métodos sustentados pela teoria da probabilidade (COCHRAN, 1977) que permitem alto potencial analítico – ao viabilizarem que covariáveis sejam investigadas em conjunto com as variáveis principais de interesse – ao mesmo tempo que podem ser controlados pelo pesquisador (CITRO, 2014). Possibilitam que boa parte das fontes de erro associadas aos produtos sejam conhecidas – já estudadas, conforme sintetizam Brackstone (1999) e Biemer et al. (2014) –, facilitando mensurações relativas à qualidade e viabilizando processos inferenciais adequados para populações de interesse via tratamentos corretivos e validações estatísticas.

Entretanto, também são percebidas algumas dificuldades associadas ao modelo tradicional de produção de estatísticas públicas (CAUMO, 2020), fortemente baseados em levantamentos de campo com coleta conduzida por pessoas – processos por vezes caros, demorados e de operacionalização complexa (CITRO, 2014; HACKL, 2016). Publicações apontam para desafios relacionados a aspectos de: recursos – tempo e custo financeiro (HACKL, 2016; DAAS; PUTS, 2014; KITCHIN, 2015; TAM; CLARKE, 2014; EUROSTAT, 2014b); atendimento à demanda (HARWOOD; MAYER, 2016; FLORESCU et al., 2014); operacionalização e logística da coleta (CITRO, 2014; STRUIJS; BRAA, 2014; BRAAKSMA; ZEELENBERG, 2015); metodologia (MANSKI, 2014; HAND, 2015); e de posicionamento estratégico e sobrevivência para as instituições produtoras (DEMUNTER, 2017; STRUIJS; DE BROE, 2018; MACFEELY, 2016).

Algumas dessas dificuldades têm sido trabalhadas com estratégias que envolvem: uma maior utilização de dados disponíveis em registros administrativos e cadastro públicos; o emprego de técnicas de correção estatística; a informatização e a automatização dos processos tradicionais; a tentativa de realização de pesquisas postais, via telefone e internet; entre outras.

Em paralelo, o contexto recente de popularização da internet, de avanço das tecnologias da informação e de consolidação da era digital (SCHMIDT; COHEN, 2013) têm levantado outras oportunidades de inovação que podem viabilizar aos produtores de estatísticas públicas o contorno, pelo menos em alguma medida, às dificuldades do modelo tradicional. Por inovação, de acordo com o Manual de Oslo (OCDE, 2005), entende-se aqui como a implementação de um produto (bem ou serviço) novo ou significativamente melhorado, ou um processo, ou um novo método de marketing, ou um novo método organizacional nas práticas de negócios, na organização do local de trabalho ou nas relações externas. Nesta visão, o produto ou processo não precisa ser inédito no mundo para ser considerado inovador, porém deve o ser no mercado ou na empresa onde está sendo aplicado.

Tais oportunidades de inovação estão relacionadas com a utilização de novos tipos de dados como insumos no processo de produção de indicadores socioeconômicos. Dados oriundos de fontes diferentes das tradicionalmente utilizadas – pesquisas amostrais e censos em levantamentos de campo – na produção de estatísticas públicas e oficiais.

Trata-se aqui de dados que correspondem ao produto do recente contexto de avalanche de dados (MILLER, 2010; HELBING et al., 2016) pelo qual as sociedades vêm passando. Dados derivados dos registros dos rastros da interação humana e não humana – tanto inconsciente quanto deliberada – com dispositivos digitais inseridos no universo dos sensores, *scanners* e da internet das coisas, dos smartphones e aplicativos, das páginas da internet, dos mecanismos de buscas, das redes sociais virtuais, dos blogs e *web* fóruns, entre outros.

Esse tipo de dado digital é usualmente denominado Big Data (UNECE, 2013) que, por sua vez, se caracterizam por serem: velozes, gerados e disponibilizados em tempo real; vastos, produzidos em larga escala; e variados, menos estruturados e em diferentes tipos de formatos (EINAV; LEVIN, 2013; KITCHIN, 2015). Seu surgimento desencadeou o fenômeno que tem sido chamado de “Revolução dos Dados” (KITCHIN, 2014), se referindo à mudança da tipologia predominante dos dados produzidos pelas sociedades: dos antigos *small data* – analógicos, escassos e de acesso limitado – para os digitais – usualmente produzidos em vasta escala, com velocidade e em variedade de domínios e formatos (KITCHIN, 2015).

Antes mesmo de ter seu conceito formalizado, este fenômeno já era percebido como tendo características de inovação de ruptura ao desafiar o *status quo* de como os dados são produzidos, geridos, analisados, armazenados e utilizados, impactando na maneira como o conhecimento é produzido, os negócios são conduzidos e a governança é promulgada (CHRISTENSEN, 1997). Há alguns anos, Mayer-Schönberger e Cukier (2013) já diziam que o Big Data se tratava de um fenômeno que estaria pronto para “chacoalhar” tudo, dos negócios e das ciências até os sistemas de saúde, os governos, a educação, a economia, as humanidades e todos os outros aspectos da sociedade.

De fato, tais dados vêm sendo percebidos como uma nova classe de ativos, ou como “o novo petróleo” (WEF, 2011; BOSSOI, 2014), tamanha sua interessante relação de custo-benefício para processo de construção de conhecimento, principalmente no âmbito empresarial e da iniciativa privada. Como consequência, desempenham papel central em um dito período de *data economy* (OPHER et al., 2016) e de quarta revolução industrial (SCHWAB, 2016) – impulsionados pela transformação digital e pela inteligência artificial. Da mesma forma, sua relevância científica e de suporte socioeconômico também tem sido aos poucos percebida fora do ambiente dos negócios (ASQUER, 2013).

Produtores de estatísticas públicas e indicadores socioeconômicos também já analisam os possíveis benefícios que o Big Data e a Revolução dos Dados podem trazer, atentando também para sua real utilidade e viabilidade. No Brasil, por exemplo, o Instituto Brasileiro de Geografia e Estatística (IBGE) demonstrou estar atento ao contexto de oportunidades ao tratar o tema como central na terceira edição da Conferência Nacional de Produtores e Usuários de Informações Estatísticas, Geográficas e Ambientais (INFOPLAN), realizada em 2016, e, mais recentemente, ao adentrar como hub físico na *Global Platform* da Organização das Nações Unidas (ONU), proposta para fins de colaboração e construção de conhecimento no âmbito da ampliação da produção de estatísticas oficiais e públicas para o contexto de considerar também fontes alternativas de dados e métodos inovadores, sob gestão do *UN Committee of Experts on Big Data and Data Science for Official Statistics* (UN-CEBD). Investigam-se possibilidades de contorno, pelo

menos em alguma medida, de todos os tipos de problemáticas associadas ao modelo tradicional de produção mencionadas anteriormente (KITCHIN, 2015; TAM; CLARKE, 2014; FLORESCU et al., 2014; EUROSTAT, 2014; UNGP, 2012; HAND, 2015; BRAAKSMA; ZEELENBERG, 2015; HACKL, 2016).

Kitchin (2015), nesse contexto, é categórico ao reforçar a necessidade de uma significativa inovação disruptiva na produção de estatísticas públicas com o surgimento dessa série de oportunidades. Para Citro (2014), é obrigação que os produtores se movam do paradigma de produzir estimativas a partir de levantamentos tradicionais de campo – que cada vez mais se apresentam inadequadas para atender às expectativas do novo contexto de demandas (PINK et al., 2009) e são estranguladas por questões de custos e diminuição nas taxas de respostas (BRAAKSMA et al., 2014) – para o da produção que se aproveita de múltiplas fontes, beneficiando-se do contexto da “Revolução dos Dados”, objetivando a geração de estimativas que atendam de forma mais assertiva às demandas dos usuários.

Essa ideia de possível quebra de paradigma é tema de crucial importância para produtores de estatísticas públicas, uma vez que as oportunidades trazidas pela Revolução dos Dados poderiam estender a capacidade de produção de informação estatística, em alguma medida, para outras esferas, reduzindo a participação dos atuais principais agentes (STRUIJS; DE BROE, 2018; MACFEELY, 2016). Na mesma linha, Demunter (2017) é ainda mais radical ao afirmar que a utilização de Big Data é questão de sobrevivência para os produtores tradicionais.

Entretanto, esse debate não se dá por encerrado, especialmente no âmbito da gestão pública uma vez que neste domínio as estatísticas públicas assumem caráter institucional (JANNUZZI, 2020), tornando pouco provável a obsolência da produção estatística tradicional – além da percepção das estatísticas objetivas e independentes enquanto bens públicos essenciais em sociedades democráticas.

Em adição, também porque a utilização de Big Data na produção de estatísticas públicas e indicadores socioeconômicos traz diversas novas incertezas e dificuldades (CAUMO; SOUZA, 2021), despertando a necessidade por esforços e investimentos em pesquisa e articulação política. Dentre alguns desafios já percebidos, pode-se citar aspectos de:

- a) operacionalização da produção e metodologia (HASSANI; SAPORTA; SIRIMAL, 2014; TENNEKES; DE JONGE; DAAS, 2011; ZIKOPOULOS et al., 2012; FRY, 2008; DAAS et al., 2012; LIU; JIANG; HEER, 2013; COUPER, 2013; DAAS et al., 2013; UN, 2014a; REIMSBACH-KOUNATZE, 2015; BRAAKSMA; ZEELLENBERG, 2015; CITRO, 2014; TUFECKI, 2013; ASA, 2014; BUELENS et al., 2012; FLEKOVA; GUREVYCH, 2013; HASTIE; TIBSHIRANI; FRIEDMAN, 2009; DAAS, 2012; DAAS; PUTS, 2014; O'CONNOR et al., 2010);
- b) capacidade organizacional – financeira, tecnológica e humana (MILLS et al., 2012; NAS, 2013; SCHUTT; O'NEIL, 2013; PARISE; IYER; VESSET, 2012; SCANNAPIECO; VIRGILLITO; ZARDETTO, 2013; STRUIJS; BRAAKSMA; DAAS, 2014; TAM; CLARKE, 2014; CERVERA et al., 2014; DUNNE, 2013; DAVENPORT; PATIL, 2012; LAVALLE, 2011; UNGP, 2012; EUROSTAT, 2014b);
- c) regulação, legislação, ética e privacidade (WIRTHMANN, REIS, 2018; TAM; KIM, 2018; NASEM, 2017a; VACCARI, 2014; HACKL, 2016; VALE, 2015; STRUIJS; DAAS, 2013; DAAS et al., 2015; DE JONGE; VAN PELT; ROOS, 2012);
- d) redes de relacionamento externo (ESSNET, 2013; TAM; CLARKE, 2015a; KRATKE; BYIERS, 2014);
- e) cultura institucional (KITCHIN, 2015).

Destes, a aferição e a garantia da qualidade de estatísticas e indicadores gerados a partir de fontes de dados do tipo Big Data, presente no aspecto metodológico-operacional, é provavelmente o desafio central. Se não puderem ser de qualidade, indicadores socioeconômicos produzidos a partir de dados do tipo Big Data serão possivelmente desacreditados e descartados, uma vez que se destinam a suportar importantes processos de tomada de decisão tanto no setor público quanto no privado. Ou seja, a questão da qualidade estatística dos produtos é crucial para que as possibilidades de inovação trazidas pela Revolução dos Dados consolidem-se no campo dos indicadores socioeconômicos.

Dessa forma, pode-se interpretar que, no âmbito do contexto colocado, um arcabouço capaz de avaliar a qualidade de estatísticas públicas e indicadores socioeconômicos gerados a partir de Big Data será determinante para a efetivação de processos de transformação de dados e informações em conhecimento. Ou, em outras palavras, será decisivo para que os produtos estatísticos gerados sejam de fato utilizados na prática das tomadas de decisão, orientando a realização de tarefas e ações em busca por resultados concretos.

O conceito de qualidade de um produto estatístico – como uma estatística pública ou um indicador socioeconômico – aqui considerado entende-a como o grau de adequação do produto ao seu propósito de utilização, ou sua capacidade de atender as necessidades dos usuários (ONS, 2013) no que se refere a algumas dimensões ou princípios esperados (FAO, 2014; EUROSTAT, 2017). Na mesma linha, para UN (2018), o conceito de qualidade estatística é operacionalizado pela observação de todos os aspectos associados ao quão bem os processos e produtos estatísticos satisfazem as expectativas de seus usuários e outros stakeholders.

Representadas operacionalmente por atributos ou características observáveis ou mensuráveis, tais dimensões da qualidade são conceitos abstratos correspondentes a princípios desejados de um produto estatístico, que: variam de produto para produto conforme a importância percebida pela demanda, isto é, de acordo com o interesse – perspectivas, necessidades e prioridades – do usuário final (FAO, 2014); variam conforme o produto estatístico em questão, uma vez que algumas dimensões podem se aplicar menos ou mais de acordo com as características das fontes de dados e do resultado final desejado; e são frequentemente subdivididas perante domínios, como fazem, por exemplo, Eurostat (2019), ao organizar tais dimensões em ditas de ambiente institucional, de produção estatística e de produtos estatísticos, e (2) Agafitei et al. (2015), ao organizá-las em de entrada/insumos (*input*), de processo (*process/throughput*) e de saída (*output*).

Tal qualidade estatística tem sido aferida por meio de *frameworks* e manuais de boas práticas que se diferenciam pelas dimensões e domínios que consideram, assim como pelas métricas ou formas de operacionalização das dimensões de qualidade. Por *frameworks*, mais especificamente os ditos *frameworks* conceituais, entende-se aqui como estruturas ou artefatos construídos com o propósito de identificar elementos componentes de um sistema e suas relações a fim de nortear

análises, explicando os processos e prevendo os resultados (CARVALHO, 2013). Ou seja, correspondem, no presente contexto, a estruturas conceituais que identificam e formalizam dimensões e domínios de qualidade – bem como seus relacionamentos – de produtos estatísticos para fins de nortear o processo de aferição da qualidade.

Para estatísticas públicas geradas a partir de processos tradicionais, *frameworks* propostos por UN (2019), IBGE (2021), ABS (2009), *Statistics Canada* (2019), ECB (2008), FMI (2003) e (FAO, 2014), entre outros, já estão consolidados. Suas estruturas foram predominantemente constituídas a partir da combinação dos fundamentos do *Total Quality Management* (CROSBY, 1979) com os Princípios Fundamentais das Estatísticas Oficiais (UNECE, 1992), conforme Saebo (2016).

Para que atendam à sua finalidade, espera-se que tais *frameworks* de qualidade sejam suficientemente genéricos para abarcar diversos contextos, com importância dada a cada domínio e dimensão de qualidade decidida caso a caso (ECB, 2008) – com base nas características das fontes de dados e dos resultados finais desejados, necessidades da demanda.

Entretanto, é justamente pela necessidade de serem aplicáveis a diferentes contextos que *frameworks* para a qualidade estatística de produtos gerados a partir de Big Data são ainda incipientes (HAJNOVIC, 2018). Diferentemente dos métodos tradicionais de produção de estatísticas públicas, a produção a partir de Big Data envolve uma heterogeneidade muito grande de fontes de dados, geradas com diferentes procedimentos e em estruturas que dependem da aplicação às quais servem – e não necessariamente para fins estatísticos (ARACRI et al., 2018). Ou seja, parece que diretrizes únicas e regras gerais de decisão não serão suficientes para se trabalhar com a aferição da qualidade estatística em diferentes contextos de Big Data e de múltiplas fontes de dados (MÁSLANKOWSKI; NOWICKA, 2018; REINERT; STOLTZE, 2016).

Em termos de *frameworks* para a qualidade de produtos estatísticos gerados a partir de Big Data, foram identificadas as propostas de UNECE (2014), Eurostat (2014a), Kreuter et al. (2015)¹, Batini et al. (2015) e Cai e Zhu (2015). Entretanto, recentes publicações estão apontando para a necessidade de alterações ou

¹ Atualizado em Amaya, Biemer e Kinyon (2020).

expansões nos *frameworks* de qualidade até então propostos. Em geral, citam a necessidade de incorporação de outras dimensões e atributos de qualidade, de operacionalização de métricas mais viáveis do ponto de vista prático e de estruturas flexíveis para bom funcionamento dos *frameworks* em contextos com diferentes características, conforme sintetiza o Quadro 1.

Quadro 1 – Necessidades relativas à *frameworks* para a qualidade de produtos estatísticos gerados a partir de Big Data.

Publicação	Trecho	Pg.
De Waal, Delden e Scholtus (2019)	“Um tópico importante de trabalho futuro é o desenvolvimento de um <i>framework</i> sistemático para situações, métodos e métricas que podem surgir em contextos de múltiplas fontes.”	191
Barcaroli, Golini e Righi (2018)	“É crucial a definição de um <i>framework</i> metodológico que permita o uso eficiente de diferentes fontes, e também a avaliação da acurácia das estimativas obtidas na abordagem <i>model-based</i> .”	2
Brancato e Ascari (2018)	“Alguns tópicos relativos à confiabilidade, coerência e comparabilidade necessitam de análises e experimentações mais aprofundadas”.	8
Brancato e Di Consiglio (2018)	“ <i>Frameworks</i> teóricos para <i>multisource</i> e Big Data devem ser traduzidos em termos operacionais, contendo medidas e indicadores relevantes e informativos.”	8
Brancato et al. (2018)	“Necessidade de se repensar uma estratégia de qualidade eficiente e direcionada.”	4
	“Planejar métricas de mensuração da qualidade no novo ambiente de produção estatística.”	9
Hajnovic (2018)	“Por ser um campo emergente, há pouca orientação para a mensuração da qualidade das aplicações de Big Data e <i>data science</i> em estatísticas oficiais.”	1
Hand (2018)	“Construção de métricas de qualidade e <i>scorecards</i> de qualidade para bases de dados”.	10
Maslankowski e Nowicka (2018)	“Não há um <i>framework</i> unificado para qualidade de Big Data que pode ser aplicado para diferentes tipos de bases de dados.”	1
	“A variedade de <i>frameworks</i> de qualidade para Big Data a disposição permite criar um conjunto de indicadores capaz de aferir diferentes aspectos da usabilidade de uma base de dados.”	1
	“A solução está na criação de diferentes <i>frameworks</i> dependentes dos dados em utilização.”	1
Salgado et al. (2018)	“O QAF deve ser revisado em termos de potenciais novos indicadores.”	10
	“O paradigma do TSE precisa ser adaptado para o contexto de dados de registros telefônicos.”	10
De Waal, Delden e Scholtus (2017)	“Construir métricas e métodos para computá-las mais adequados para aplicações práticas reais.”	44
	“Aprofundar testes de adequação de indicadores propostos para a dimensão coerência no contexto de múltiplas fontes.”	44
	“Ampliar a quantidade de configurações de dados examinadas no projeto Komuso da ESSNET.”	45
	O autor apresenta também uma tabela com outros gaps pontuais e específicos.	46
NASEM (2017b)	“Novas fontes de dados requerem expansão e desenvolvimento adicional dos <i>frameworks</i> de qualidade existentes para que novos componentes e aspectos possam ser incluídos e enfatizados”	126
Hackl (2016)	“Um <i>framework</i> e critérios de qualidade necessitam ser desenhados ou adaptados para os vários tipos de Big Data e de produtos estatísticos que são baseados nesses dados.”	51
Reinert e Stoltze (2016)	“Necessidade de métodos padronizados que meçam a qualidade do <i>input</i> e que contenham indicadores quantitativos claros, especialmente para a dimensão acurácia.”	36
	“Está ficando claro que uma única lista de indicadores para todos os diferentes contextos de aplicação não se faz viável. Temos nos concentrados em conjuntos simples de indicadores bem adequados para o monitoramento contínuo da qualidade no contexto de uma fonte específica.”	36
	“Necessidades de trabalhos futuros que definam possíveis erros de cobertura.”	36
	“Trabalhos futuros podem focar no desenvolvimento de uma lista de indicadores que auxilie na avaliação da qualidade geral de uma fonte de dados antes mesmo da obtenção dos dados.”	37
Reis et al. (2016)	“Necessidade de complementar as diferentes propostas (BDQF, APNOS e BDTEF) e de estruturar os <i>links</i> entre a qualidade do <i>input</i> e do <i>process</i> com a qualidade do <i>output</i> .”	16
	“Elementos de cada um dos <i>frameworks</i> analisados podem ser combinados para gerar um <i>framework</i> de maior qualidade.”	16
	“Novas dimensões de qualidade, como complexidade, são também importantes, assim como a proposição de novos indicadores de qualidade.”	16
Agafitei et al. (2015)	“O ESS QAF ainda não é suficiente para cobrir a complexidade do processo de produção que se utiliza de múltiplas fontes de dados e métodos.”	207

Fonte: atualizado de Caumo e Souza (2021).

Face à complexidade de se construir uma estrutura única que se aplique a diferentes situações, parece razoável que *frameworks* para a qualidade de produtos estatísticos gerados a partir de Big Data sejam inicialmente desenhados para casos específicos, um a um (MÁSLANKOWSKI; NOWICKA, 2018), compartilhando ou partindo de uma base com pilares comuns.

Nesse contexto, um possível domínio a servir de ponto de partida é o dos indicadores sobre o turismo², importante setor para a atividade econômica de Florianópolis. Para Demunter (2017), o setor do turismo está na linha de frente das inovações associadas ao Big Data no que concerne a novos dados e métodos. Em adição, pesquisas tem apontado para uma certa insatisfação por parte dos usuários das estatísticas de turismo até então produzidas no Brasil, sugerindo a necessidade de melhoramentos, especialmente em termos de oportunidade temporal, variedade de temas/domínios, profundidade/granularidade e qualidade (BRASIL, 2016).

Sobre a utilização de fontes variadas de Big Data para geração de indicadores de turismo (CAUMO; BIZ; SOUZA, 2020), finalidades percebidas na literatura trataram sobre:

- a) quantidade e/ou distribuição espacial dos turistas (ALIS et al., 2015; DONOVAN; FLAHERTY; HEALY, 2017; PRAMANA et al., 2017; SETIADI; ULUWIYAH, 2017; GARCÍA et al., 2018; LIU et al., 2019; KOIDE et al., 2016; ÖNDER; KOERBITZ; HUBMANN-HAIDVOGEL, 2014; LIU; SHI, 2019; YANG et al., 2017);
- b) satisfação, opiniões, sentimentos e interesses manifestos (RAI, 2018; CLASTER; GHOTBI; SHANMUGANATHAN, 2015; MENCHAVEZ; ESPINOSA, 2015; SHIMADA et al., 2011; NOZAWA et al., 2017; GUO et al., 2017b; GUARDA et al., 2018; KITAYAMA, 2016; DE OLIVEIRA; BARACHO, 2018; LEAL et al., 2016);

² Cabendo adicionar que não foram encontrados frameworks para a qualidade de produtos estatísticos gerados a partir de Big Data voltados especificamente para o contexto do setor turístico e, assim sendo, as demandas trazidas pelo Quadro 1 se referem a produção estatística de forma genérica.

- c) fluxos turísticos, movimentações, atrações e locais visitados, tempo de estadia e gastos (SETIADI; ULUWIYAH, 2017; GARCÍA et al., 2018; HU et al., 2018; PROVENZANO; HAWELKA; BAGGIO, 2018; ICHIFUJI et al., 2016; DE OLIVEIRA; BARACHO, 2018);
- d) perfil, atributos e origem dos turistas (LIU; SHI, 2019; PARK et al., 2018; DE OLIVEIRA; BARACHO, 2018);
- e) impacto econômico do turismo (BERZINA; LAUBERTE, 2018);
- f) e outros aspectos (FENG; LI; ZHANG, 2013; SHI; ZHAO; CHEN, 2017; MA et al., 2018).

No Brasil, percebe-se algumas experiências já sendo conduzidas, como as do Observatório de Turismo do Espírito Santo e do Observatório do Turismo de Minas Gerais. O primeiro, utilizando dados de empresas telefônicas para gerar indicadores sobre a origem, o destino, o perfil sociodemográfico, o tipo/propósito do turismo, o tempo de estadia, os fluxos e locais mais visitados, entre outros (BRASIL, 2017). O segundo, capturando dados na *World Wide Web* para construir estatísticas relativas, dentre outros aspectos, ao quanto e ao quão bem os turistas e visitantes avaliam as cidades e os pontos turísticos (OLIVEIRA, 2017).

Cabe destacar que, no turismo, a visão sobre a utilidade de indicadores é a mesma dos indicadores socioeconômicos em geral. Segundo UNWTO (2007), trata-se de instrumentos essenciais para os processos de planejamento, gestão e monitoramento do turismo, fornecendo informação acurada para os processos de tomada de decisão. Além disso, são necessários para a formulação de estratégias de marketing, o fortalecimento das relações entre instituições, a avaliação da eficiência e da efetividade das decisões gerenciais e a mensuração do turismo no âmbito da economia nacional (UN, 2010).

Importância essa que foi recentemente corroborada em pesquisa realizada com reais e potenciais usuários de informações turísticas (BRASIL, 2016) durante a construção do Plano Nacional Estratégico de Estatísticas Turísticas (PNETUR) 2016-2020, do Ministério do Turismo (MTur), que percebeu que boa parcela dos entrevistados utiliza informações turísticas para fins profissionais por razão de ser muito necessário e que o aspecto relevância das informações fornecidas pelo MTur foi o mais bem avaliado pelos respondentes quando da investigação sobre a qualidade.

Nesse contexto, a presente tese se propõe a atuar na compreensão e solução do seguinte problema de pesquisa: como avaliar a qualidade estatística de indicadores de turismo gerados a partir de Big Data?

1.2 OBJETIVOS

Em consonância com o problema de pesquisa contextualizado, ficam estabelecidos os objetivos, geral e específicos, descritos abaixo.

1.2.1 Objetivo Geral

Propor um *framework* conceitual para a avaliação da qualidade estatística de indicadores de turismo produzidos a partir de Big Data.

1.2.2 Objetivos Específicos

- a) Identificar potenciais fontes de dados do tipo Big Data e correspondentes indicadores de turismo que podem auxiliar a gerar;
- b) Analisar a estrutura de *frameworks* para a qualidade de estatísticas públicas produzidas a partir de fontes tradicionais e de Big Data;
- c) Definir um arcabouço de requisitos, dimensões e domínios de qualidade aplicável aos indicadores de turismo produzidos a partir de Big Data;
- d) Analisar a aplicabilidade do *framework* em uma situação real;
- e) Verificar o *framework* proposto junto a produtores e usuários especialistas em estatísticas públicas e indicadores socioeconômicos.

1.3 ENQUADRAMENTO METODOLÓGICO

Para alcançar os objetivos, a presente tese se alinha metodologicamente com a *Design Science* (DS) enquanto base filosófica (HEVNER, 2007; VAISHNAVI; KUECHLER, 2009; DRESCH; LACERDA; ANTUNES JR., 2015), posicionando-se

alternativamente ao entendimento ontológico e epistemológico da ciência tradicional, amplamente utilizada quando do tratamento de questões naturais e sociais, em favor de uma ciência que se propõe ser mais adequada para lidar com problemas que tratem do artificial, criado pelo homem (SIMON, 1969), buscando gerar conhecimento para a concepção, o desenvolvimento e a avaliação de artefatos (VAN AKEN, 2004).

Partindo desse arcabouço e em linha com a visão de Dresch, Lacerda e Antunes Jr. (2015), este trabalho possui como razão para a realização da pesquisa a busca por uma solução para um problema prático, com objetivo de propor um artefato e prescrever soluções, utilizando-se de método científico de cunho dedutivo, indutivo e abdução, orientado pelo método de pesquisa *Design Science Research* (DSR) e executando, para tanto, técnicas de investigação do tipo pesquisa documental e bibliográfica e de questionários aplicados a especialistas.

A pesquisa bibliográfica foi realizada sob a forma de três revisões sistemáticas de literatura (MOHER et al., 2009), sendo a segunda delas integrativa (ERCOLE et al., 2014). A aplicação dos questionários a especialistas, produtores e usuários de estatísticas públicas e indicadores socioeconômicos, por sua vez, foi conduzida de acordo com a técnica Delphi (MARQUES; FREITAS, 2018; WRIGHT; GIOVINAZZO, 2000).

Os detalhamentos da abordagem e dos procedimentos metodológicos estão apresentados no Capítulo 3.

1.4 ADERÊNCIA AO EGC

A pesquisa aqui proposta entende o conhecimento enquanto ativo organizacional que pode também assumir forma explícita e ser efetivado por agentes não humanos, transferível e modelável, conforme visão de mundo dos cognitivistas – paradigma de referência, aprofundado na Seção 2.1. Possui como tema central os indicadores socioeconômicos, considerados, portanto, representações de conhecimento explícito ao passo que são utilizados para direcionar processos de tomadas de decisão em ações que buscam por resultados concretos. Ou seja, representações efetivadas para a geração de valor.

O objetivo proposto envolve o desenvolvimento de uma estrutura de conhecimento – *framework* conceitual – que formaliza a relação entre domínios e elementos de qualidade de indicadores socioeconômicos gerados a partir de Big Data. Ou seja, se propõe à construção de uma representação conceitual de um determinado conhecimento de domínio com finalidade de orientar o desenvolvimento e a aplicação de métodos que suportem a avaliação da qualidade de representações de conhecimento – indicadores socioeconômicos – utilizadas em atividades intensivas em conhecimento, em linha com a área de Engenharia do Conhecimento do Programa e Pós-Graduação em Engenharia e Gestão do Conhecimento (PPGEGC).

Assim, o artefato proposto corresponde não só a uma estrutura de conhecimento como também a um provável “efetivador de conhecimento”. Isso porque sua utilização permitirá que, quando avaliados positivamente em relação à sua qualidade, estatísticas públicas e indicadores socioeconômicos sejam colocados em prática durante processos de tomadas de decisão.

Em paralelo, o presente estudo traz contribuições ao difundir e avaliar a viabilidade – em termos de qualidade estatística – de novas possibilidades de construção, descoberta e aquisição de conhecimento – no caso, a partir de novas fontes, Big Data, tão abundantes e ainda pouco exploradas por diversas organizações. Além disso, o conhecimento que se espera gerar com a pesquisa poderá servir de base para que produtores de indicadores econômicos, sociais, demográficos e ambientais, viabilizem produtos e processos de produção inovadores.

Dessa forma, pelo paradigma de referência e pelo objeto que interessa a agentes produtores e usuários de indicadores socioeconômicos, este estudo se enquadra na linha de pesquisa “Engenharia do Conhecimento Aplicada às Organizações” do PPGEGC.

Ainda, esta tese dá continuidade a uma série de pesquisas do PPGEGC que também exploraram temas associados à construção, descoberta e aquisição de conhecimento a partir de novas – alternativas – fontes de dados. Dentre algumas, destacam-se as trazidas pelo Quadro 2.

Quadro 2 – Pesquisas do PPGE GC sobre construção, descoberta e aquisição de conhecimento a partir de fontes alternativas de dados.

Autor/Ano	Temática	Tipo
Pizzol (2014) e Schmitt (2013)	Tratam sobre a captura de dados da <i>web</i> para a Inteligência Competitiva de empresas.	Dissertação
Fernandes (2012)	Propõe um modelo de aquisição de conhecimento nas redes sociais para a identificação de oportunidades de negócio.	Dissertação
Speroni (2016)	Captura dados abertos para a construção de indicadores de inovação.	Tese
Dziekaniak (2012)	Propõe que governos utilizem as informações presentes em redes sociais para aprimoramento das plataformas de governo eletrônico.	Tese
Oliveira (2009)	Trabalha com comunicações realizadas pela população para o governo através de ferramentas colaborativas para construir conhecimento que auxilie o planejamento público.	Dissertação
Klein (2015)	Utiliza bases de Big Data e dados abertos para geração de conhecimento no âmbito das <i>smart cities</i> .	Dissertação
Woszezenk (2016)	Explora a explosão no volume de dados digitais para analisar a associação semântica e temporal entre elementos textuais.	Tese
Alvarez (2018) e Ribeiro (2018)	Capturam dados da <i>web</i> para gestão de ideias usando mineração de textos.	Dissertação

Fonte: elaborado pelo autor (2021).

Em termos de pesquisas do PPGE GC que abordam o tema da engenharia e da gestão do conhecimento no âmbito do turismo, destacam-se as apresentadas pelo Quadro 3.

Quadro 3 – Pesquisas do PPGE GC sobre engenharia e gestão do conhecimento no turismo.

Autor/Ano	Tema	Tipo
Muniz (2020)	Propõe um <i>framework</i> para a gestão inteligente da experiência turística.	Tese
Biz (2009)	Avalia os portais turísticos gerenciados pelas Organizações Públicas de Turismo quanto a sua capacidade de suporte à gestão do conhecimento.	Tese
Teixeira (2010)	Constrói um modelo tecnológico que dá suporte a gestão do conhecimento do turismo sustentável em Parques Nacionais.	Tese
Oliveira (2007)	Desenvolve um sistema de recomendação de hospedagem personalizado a usuários da internet.	Dissertação
Quevedo (2009)	Identifica estratégias desenvolvidas no gerenciamento de portais virtuais de turismo.	Dissertação

Fonte: elaborado pelo autor (2021).

1.5 RELEVÂNCIA

O presente trabalho se mostra relevante, conforme resume a Figura 1 por: (1) abordar um tema importante local e nacionalmente, (2) demarcar oportunidades e possibilidades de inovação e complementariedade nos processos tradicionais de produção de estatísticas públicas e indicadores socioeconômicos; e (3) aprofundar soluções para o provável principal desafio – aferição e garantia da qualidade – da geração de produtos estatísticos inovadores – construídos a partir de fontes de dados do tipo Big Data.

Figura 1 – Fatores que compõem a relevância da pesquisa.



Fonte: elaborado pelo autor (2021).

Inicialmente (1), considera-se esta tese relevante pelo simples fato de abordar a temática das estatísticas públicas e dos indicadores socioeconômicos, representações explícitas de conhecimento sobre aspectos econômicos, demográficos, sociais e ambientais de uma nação, importantes no suporte a tomadas de decisão, para formulação de políticas e estratégias e para geração de valor por parte de agentes públicos e privados. Além disso, foca no setor turístico, proeminente em Florianópolis, percebido como um dos possíveis grandes beneficiados pelo conhecimento aqui gerado (DEMUNTER, 2017) e carente de melhorias no âmbito da produção e disponibilização de estatísticas (BRASIL, 2016).

Em relação ao segundo aspecto (2), a relevância se inicia pela apresentação de oportunidades de geração de conhecimento a partir de novas fontes de dados – Big Data – e pela demarcação de possibilidades de inovação que permitiriam aos atuais produtores remodelar ou complementar seus processos tradicionais de produção – explicitando também um componente de contribuição científica.

Por vezes, a utilização de Big Data como fonte de dados durante a geração de produtos estatísticos é vista como obrigatória (CITRO, 2014). Mover-se do paradigma de produzir estimativas a partir de levantamentos tradicionais de campo – que, na visão de alguns autores, cada vez mais se apresentam inadequadas para atender às expectativas do novo contexto de demandas (PINK; BOROWIK; LEE, 2009) e são estranguladas por questões de custos e diminuição nas taxas de respostas (BRAAKSMA et al., 2014) – para o da produção que se aproveita de múltiplas fontes, beneficiando-se do contexto da “Revolução dos Dados”, poderia viabilizar o contorno de uma série de dificuldades enfrentadas pelos processos tradicionais (CAUMO, 2020). Kitchin (2015), nesse contexto, é categórico ao reforçar a necessidade de uma significativa inovação disruptiva na produção de estatísticas públicas a partir do surgimento de uma série de oportunidades.

Entretanto, nem todos corroboram tais ideias e alguns entendem que o produto da Revolução dos Dados não substituirá integralmente as fontes de dados atuais dos produtores de estatísticas públicas – até porque também está sujeito a desafios e complicadores (CAUMO, 2020) –, mas que deverá ter papel central enquanto ferramental complementar, como suplemento para os dados já existentes (HACKL, 2016) a partir de uma integração que poderá maximizar valor (HARWOOD; MAYER, 2016; DE WAAL; DELDEN; SCHOLTUS, 2020) em busca dos melhores

produtos possíveis que atendam às demandas dos usuários – resultados relevantes, oportunos, acurados e economicamente viáveis.

Além disso, a ideia de possível quebra de paradigma nos modelos tradicionais de produção é tema de crucial importância para produtores de estatísticas públicas uma vez que o Big Data poderia estender a capacidade de produção estatística para outras esferas, reduzindo a participação dos agentes até então predominantes (DEMUNTER, 2017; STRUIJS; DE BROE, 2018; MACFEELY, 2016). Para Demunter (2017), trata-se de uma questão de sobrevivência institucional. Caso organizações produtoras não atentarem para tais possibilidades de inovação, deverão perder espaço no setor das estatísticas públicas – até então monopolizado – uma vez que novos players estarão em condições de gerar indicadores socioeconômicos com baixo custo, em alto nível de detalhamento e oportunismo temporal (DEMUNTER, 2017) – cabendo adicionar a esse debate que, especialmente no âmbito da gestão pública, as estatísticas não podem ser vistas como bens mercadológicos em virtude da sua finalidade e do seu caráter institucional.

Por fim (3), a tese se faz relevante pela percepção de que a capacidade de aferir e garantir a qualidade dos produtos gerados pelos métodos inovadores será definidora do sucesso no aproveitamento das oportunidades que se apresentam aos produtores. Se não puderem ter sua qualidade garantida, indicadores socioeconômicos produzidos a partir de dados do tipo Big Data serão possivelmente descreditados, descartados ou pouco utilizados, uma vez que se destinam a suportar importantes processos de tomada de decisão tanto no setor público quanto no privado – como demonstra Brasil (2016) no âmbito das estatísticas e indicadores de turismo.

Em outras palavras, entende-se que um arcabouço capaz de auxiliar na avaliação da qualidade de estatísticas públicas e indicadores socioeconômicos gerados a partir de Big Data será determinante para processos de transformação de dados e informações em conhecimento. Ou seja, será condicionante da real utilização dos produtos estatísticos gerados na prática cotidiana das tomadas de

decisões, orientando a realização de tarefas e ações em busca por resultados concretos.

1.6 ORIGINALIDADE

Por ser um campo emergente, ainda há pouca orientação para a mensuração da qualidade das aplicações de Big Data em estatísticas públicas e oficiais (HAJNOVIC, 2018). Mesmo assim, alguns *frameworks* já foram propostos, como os sugeridos por UNECE (2014), Eurostat (2014a), Kreuter et al. (2015)³, Batini et al. (2015) e Cai e Zhu (2015).

Entretanto, diversas questões permanecem em aberto, se fazendo necessário incorporar outras dimensões de qualidade, desenvolver métricas mais viáveis na prática e flexibilizar os *frameworks* para bom funcionamento em contextos com diferentes características, conforme percepções encontradas na literatura – apresentadas no Quadro 1.

Para Másłankowski e Nowicka (2018), por exemplo, ainda não há um *framework* unificado para qualidade de Big Data que pode ser aplicado para diferentes tipos de bases de dados. Na mesma linha, Barcaroli, Golini e Righi (2018), apontam que se faz crucial a definição de um *framework* que permita o uso eficiente de diferentes fontes. Da mesma forma, De Waal, Delden e Scholtus (2019) destacam que o desenvolvimento de um *framework* sistemático para situações, métodos e métricas que podem surgir em contextos de múltiplas fontes é um tópico importante de trabalho futuro. Hackl (2016) já havia percebido o mesmo, afirmando que um *framework* e critérios de qualidade necessitam ser desenhados ou adaptados para os vários tipos de Big Data e de produtos estatísticos que são baseados nesses dados.

Porém, a complexidade do tema tem prejudicado a construção de soluções únicas aplicáveis a diferentes contextos. A principal razão é o fato de que dados do tipo Big Data possuem variadas estruturas e configurações e foram gerados a partir de diferentes processos de produção, com características que dependem da finalidade para o qual foram pensados – normalmente de interesses particulares e não de construção de estatísticas públicas (ARACRI et al., 2018).

³ Atualizado em Amaya, Biemer e Kinyon (2020).

Assim, a solução pode passar pela criação de diferentes *frameworks* para a qualidade de estatísticas públicas dependentes dos dados (Big Data) e do contexto em utilização, inicialmente desenhados para casos específicos, um a um (MÁSLANKOWSKI; NOWICKA, 2018). Reinert e Stoltze (2016) já haviam apontado para este caminho, afirmando que está ficando claro que uma única lista de métricas de qualidade para todos os diferentes contextos de aplicação não se faz viável. Dessa forma, seus estudos têm se concentrado em conjuntos de indicadores bem adequados para o monitoramento contínuo da qualidade no contexto de uma fonte específica.

Nesses termos, entende-se que a presente proposta de pesquisa está desenhada em consonância com uma demanda identificada na literatura especializada sobre o tema em análise.

1.7 DELIMITAÇÕES

Para fins de clarificação das fronteiras da investigação conduzida e dos elementos considerados na presente tese, inicia-se demarcando que a proposta central é a de proposição de um artefato, sob a forma de *framework* conceitual, com finalidade direta de auxiliar a aferição da qualidade de produtos estatísticos, mas que indiretamente pode ser utilizado para guiar a estruturação de processos de produção estatística.

Nesse contexto, cabe destacar que, do ponto de vista da utilização da solução, o artefato se restringe a:

- Orientar a avaliação da qualidade de indicadores socioeconômicos e estatísticas públicas sob a ótica dos princípios estatísticos;
- Aplicar-se a indicadores e estatísticas produzidas a partir de dados tipo Big Data; e
- Adequar-se primordialmente a indicadores e estatísticas de turismo.

Adicionalmente, no âmbito do método de construção da solução:

- O desenvolvimento é fruto de processo científico com componente criativo, baseado na análise de artefatos semelhantes, nas conclusões de

pesquisas já realizadas e no conhecimento do pesquisador e seus orientadores; e

- A avaliação, com verificação de estrutura a análise da aplicabilidade, é realizada por meio de consulta a especialistas selecionados de forma não probabilística, intencional, orientada por critérios norteadores, partindo de indicações dos orientadores e suas redes de contatos.

1.8 EXPERIÊNCIA DO PESQUISADOR

O autor desta tese de doutorado é estatístico, bacharel pela Universidade Federal do Rio Grande do Sul (UFRGS), mestre em Estudos Populacionais e Pesquisas Sociais pela Escola Nacional de Ciências Estatísticas (ENCE) do Instituto Brasileiro de Geografia e Estatística (IBGE).

É consultor estatístico autônomo e professor no Centro de Ciências da Administração e Socioeconômicas (ESAG) da Universidade do Estado de Santa Catarina (Udesc) desde o ano de 2017, ministrando as disciplinas de Métodos Estatísticos e Análise Estatística.

Possui experiência profissional na área de produção de estatísticas públicas e indicadores socioeconômicos, tendo trabalhado por 9 anos na Fundação de Economia e Estatística (FEE) do Rio Grande do Sul. Neste período, alcançou as posições de estatístico responsável e coordenador geral da Pesquisa de Emprego e Desemprego da Região Metropolitana de Porto Alegre (PED-RMPA), além de chefe do Centro de Pesquisa de Emprego e Desemprego (CPED) da FEE.

Além disso, foi conselheiro do Conselho Regional de Estatística da 4ª Região (CONRE 4) por 9 anos, chegando à presidência do órgão por dois mandatos, e analista – especialidade estatística – na Secretaria da Saúde do Estado (SES) do Rio Grande do Sul (RS).

1.9 ESTRUTURA DO DOCUMENTO

O presente documento está estruturado em cinco capítulos. Este primeiro contextualiza a problemática da pesquisa, apresentando as lacunas e os objetivos. No segundo, são definidos os conceitos que constituem o referencial teórico de

partida, construídos a partir dos resultados de três revisões sistemáticas da literatura, necessários para compreensão do contexto até o ponto em que o problema, a lacuna e os objetivos da tese ficam estabelecidos. No Capítulo 3, a metodologia da pesquisa é formalizada. O Capítulo 4, por sua vez, descreve, verifica e analisa a solução proposta para o atendimento dos objetivos geral e específicos. Ao final, no Capítulo 5, são apresentadas as conclusões e as considerações finais.

2 REFERENCIAL TEÓRICO

Este capítulo apresenta as definições referentes aos principais conceitos e construtos considerados no trabalho.

2.1 CONHECIMENTO

A definição de conhecimento aqui considerada é trazida de Pacheco (2014). Para o autor, trata-se de conteúdo ou processo efetivado por agentes humanos ou artificiais em atividades de geração de valor científico, tecnológico, econômico, social ou cultural. Nesses termos, pode-se dizer que o conhecimento é insumo para cadeias de geração de valor.

De acordo com essa visão, alinhada epistemologicamente com a corrente de pensadores ditos cognitivistas tais quais Herbert Simon, Noam Chomski e Marvin Minsky, o conhecimento não existe apenas sob a forma tácita, podendo também ser explicitado para além das mentes humanas (VENZIN et al., 1998; PACHECO, 2016) como conteúdo incorporado em agentes artificiais. Ou seja, existe conhecimento passível de formalização e codificação, de identificação, armazenamento e recuperação (WELLMAN, 2009).

Nesse contexto, dados e informações passam a ser considerados representações de conhecimento explícito quando colocados em prática durante processos de tomadas de decisão, orientando a realização de tarefas e ações em busca por resultados concretos (SCHREIBER et al., 2000). Ou seja, quando efetivadas no ciclo completo de conhecimento – identificação, criação, armazenamento, aplicação e compartilhamento – com propósito de geração de valor (PACHECO, 2014).

Frequentemente, este conhecimento passível de explicitação é o referenciado por “*know-what*”, ou o “saber o quê” (BROWN; DUGUID, 1998). Apesar disso, o elemento definidor da possibilidade de representação explícita de um conhecimento em maior ou menor grau é o seu tipo – ou a sua natureza.

Independente da sua forma, há muito já se percebe o conhecimento enquanto um importante fator de produção (DRUCKER, 1993) que assumiu o posto

de ativo – intangível – organizacional. Tornou-se, também, elemento central no contexto da economia pós-industrial das sociedades da informação e do conhecimento (BELL, 1999; UNESCO, 2005; CASTELFRANCHI, 2007), passando a ser reconhecido como o mais poderoso produtor de riqueza – superando a importância da terra, do trabalho e do capital.

Assim sendo, se faz importante que seja gerido de forma estratégica (SANTOS, 2001) para que impulse da melhor maneira possível o sucesso dos negócios de uma organização (DAVENPORT; PRUSAK, 1998). Nesses termos, há algumas décadas são propostos estudos e avanços no campo da Gestão do Conhecimento. Um campo multi e interdisciplinar relacionado à gestão estratégica da informação por meio da governança dos processos de conhecimento e com fins a melhorar a performance organizacional. Nas palavras de Davenport e Prusak (1998), a Gestão do Conhecimento é a coleção de processos que objetivam governar a criação, disseminação e uso do conhecimento para atingir os objetivos organizacionais.

Sob a ótica dos cognitivistas, que acreditam na existência de conhecimentos sob a forma explícita, a Gestão do Conhecimento corresponde a um arcabouço e um conjunto de ferramentas para melhorar a infraestrutura de conhecimento da organização que visa colocar o conhecimento correto para a pessoa correta na forma e no tempo corretos (SCHREIBER et al., 2000). Utilizando-se, para tanto, de sistemas técnico-sociais projetados para apoiar o fluxo de explicitação, disseminação e gestão do conhecimento organizacional – os ditos Sistemas de Gestão do Conhecimento. Ou seja, sistemas que fornecem meios de promoção da criação, explicitação, codificação, comunicação, disseminação, retenção e transferência do conhecimento estratégico, seja ele interno ou externo à organização.

Assim, a partir do momento que se vislumbra a possibilidade de transferir o conhecimento humano para sistemas ou de modelar o conhecimento e as técnicas de resolução de problemas de um especialista de domínio para um sistema (SCHREIBER et al., 2000), abre-se espaço para um processo de gestão que contempla elementos metodológicos da engenharia – viabilizando o surgimento da

Engenharia do Conhecimento. Uma disciplina que tem como aplicação mais importante a criação de sistemas especialistas sociotécnicos que representam conhecimento e apoiam atividades intensivas em conhecimento nas organizações (SCHREIBER et al., 2000), empregando, para tanto, metodologias, técnicas, linguagens e ferramentas específicas que viabilizem a identificação, aquisição, codificação, representação e uso do conhecimento. No cerne da sua atividade: a modelagem conceitual do conhecimento de diferentes níveis de detalhes.

2.2 INOVAÇÃO

Inovação é aqui considerada de forma ampla, conforme definição do Manual de Oslo (OCDE, 2005), entendida como a implementação de um produto (bem ou serviço) novo ou significativamente melhorado, ou um processo, ou um novo método de marketing, ou um novo método organizacional nas práticas de negócios, na organização do local de trabalho ou nas relações externas. Nesta visão, o produto ou processo não precisa necessariamente ser inédito no mundo para ser considerado inovador, porém deve o ser no mercado ou na empresa onde está sendo aplicado. No Brasil, em âmbito oficial, de acordo com a Lei 10.973 de 2 de dezembro de 2004, a definição diz que inovação é a introdução de novidade ou aperfeiçoamento no ambiente produtivo ou social que resulte em novos produtos, processos ou serviços.

Sendo mais específicos, Tidd, Bessant e Pavitt (2008) dizem que inovar consiste em transformar oportunidades que surgirem em novas ideias e então colocá-las em prática. Nesse entendimento, inovações se diferenciam de ideias e invenções ao serem implementadas e, conforme adicionam Schumpeter (1988) e Porter (1991), exploradas comercialmente. Ainda nessa linha, Barbieri (1990) vai além e fala na necessidade de sucesso comercial – além da simples exploração comercial.

Tal compreensão do valor econômico é frequentemente presente nas definições de inovação. Drucker (2002), por exemplo, afirma que a inovação é a capacidade de criar riqueza por meio de recursos. Um processo que, para Pinchot e Pellman (2004), depende da combinação de uma ideia, um intra-empendedor e um patrocinador.

Apesar desse entendimento, outros autores defendem que a inovação pode estar relacionada ao aspecto social, sem necessariamente ter por objetivo o sucesso comercial. André e Abreu (2006), por exemplo, diferenciam a inovação com fins econômicos daquela que tem por finalidade melhorar a qualidade de vida e promover alterações sociais em uma população, dita inovação social (CAJAIBA-SANTANA, 2014), em geral trabalhando novas e duradouras demandas e carências sociais não atendidas (CRISES, 2004; POL; VILLE, 2009).

Independente do entendimento, o consenso está na importância da inovação. Do ponto de vista das organizações, a prática de inovar pode representar fator crucial de sucesso, crescimento e sobrevivência em meio a um cenário globalizado e competitivo (PLENTZ; BERNARDES; FRAGA, 2015; TIDD; BESSANT; PAVITT, 2008; BROWN; EISENHARDT; 1995), garantindo uma série de vantagens aos inovadores – conforme elenca Porter (1989). Para regiões, estados e países, instigar que a inovação aconteça dentro de seus domínios é estimular o crescimento econômico e a elevação dos padrões de vida da população, promovendo o desenvolvimento socioeconômico (CRESPI et al., 2014).

Apesar dos reconhecidos benefícios, o ato de inovar é sempre um desafio, uma vez que seu grau de complexidade vem sendo aumentado pela crescente quantidade de variáveis e fontes de incerteza, internas e externas à uma organização, elevando o risco associado ao processo. Nesse sentido, o sucesso da inovação é extremamente dependente da disponibilidade e da apropriada utilização do conhecimento, tornando a Gestão do Conhecimento um fator determinante para tal desfecho (URGAL; QUINTÁS; ARÉVALO-TOMÉ, 2013; NELSON; WINTER, 1982; KOGUT; ZANDER, 1992).

2.3 INDICADORES SOCIOECONÔMICOS E ESTATÍSTICAS PÚBLICAS

De acordo com o dicionário, a palavra “indicador” pode ser um adjetivo atribuído a algo que indica, que dá a conhecer, assim como um substantivo associado a um aparelho que serve para indicar (INDICADOR, 2020). Ou seja, pode ser visto como um instrumento projetado para fornecer informação, com objetivo de

apontar ou mostrar algo. Assim, carregam informações que sumarizam as características ou destacam o que está acontecendo em um sistema (SAISANA; TARANTOLA, 2002).

Um indicador estatístico, mais especificamente, está representado sob a forma de uma medida estatística, geralmente quantitativa, utilizada para ilustrar e comunicar um conjunto de fenômenos complexos de uma forma simples (OCDE, 2002; BOSSEL, 1999), capaz de condensar a enorme complexidade do mundo numa quantidade manejável de informação significativa (BOSSEL, 1999), construído a partir de uma única ou por diversas variáveis ou dimensões.

Dentre esses, os indicadores ditos compostos, ou sintéticos, diferenciam-se dos indicadores ditos analíticos ao considerar várias dimensões ou variáveis, baseando-se em uma série de subindicadores que não necessariamente possuem uma unidade de medida comum (SAISANA; TARANTOLA, 2002). Seu estabelecimento tem como intuito possibilitar uma visão melhor da realidade complexa (NARDO et al., 2005). Em outras palavras, um indicador composto é formado quando uma série de índices que são compilados em uma única métrica com base em uma determinada equação matemática, comumente utilizado para se trabalhar com conceitos multidimensionais que não podem ser captados por um único parâmetro (NARDO et al., 2005). Além disso, geralmente é mais fácil para o público interpretar tais compilações do que identificar tendências em vários índices analisados separadamente (SALTELLI, 2007).

Quando tratam de aspectos econômicos, demográficos, sociais e ambientais de uma nação, tais indicadores estatísticos são ditos indicadores socioeconômicos. Neste domínio, indicadores comumente confundem-se com estatísticas públicas e/ou oficiais em virtude da capacidade de ambos em descrever numericamente algum fenômeno da realidade. Entretanto, estatísticas públicas e/ou oficiais são, na verdade, apenas dados parcialmente preparados para uso na interpretação empírica da realidade, consistindo em insumos para a construção de indicadores socioeconômicos (JANNUZZI, 2001).

Em outras palavras, uma estatística poderá ser considerada ou utilizada para a construção de um indicador quando for ampla e reconhecidamente eficaz para descrever determinado aspecto da realidade. Ou seja, é no valor contextual baseado em uma teoria social ou finalidade programática que os indicadores

socioeconômicos se diferem de estatísticas públicas. Afinal, indicadores são dotados de significado social substantivo, usados para substituir, quantificar ou operacionalizar um conceito social abstrato (DE VAUS, 2002; JANNUZZI, 2004).

De qualquer maneira, estatísticas são formas institucionalizadas de promover conhecimento para propiciar intervenções em uma realidade (FEIJÓ; VALENTE, 2005), correspondendo a complexas objetivações que revelam realidades previamente construídas e idealizadas, apoiadas por métodos de pesquisas avançados aceitos em várias ciências (SENRA, 2009).

Estatísticas oficiais, derivação da tradução do termo “*official statistics*”, por sua vez, correspondem àquelas produzidas por agências governamentais (SCHWARTZMAN, 1997) ou organizações internacionais tidas como oficiais. No Brasil, o termo mais utilizado, e muitas vezes empregado como sinônimo, é o de “estatísticas públicas”. Este, na verdade, representa um universo mais amplo de estatísticas pois inclui aquelas produzidas por órgãos não governamentais e/ou não percebidas como oficiais. Ou seja, estatísticas oficiais são chanceladas em nome do Estado ou de organizações internacionais tidas como oficiais enquanto as estatísticas públicas apenas carregam os mesmos princípios e objetivos, sem um carimbo oficial (IBGE, 2021).

Oficiais ou não, tais estatísticas são bens públicos essenciais para o desenvolvimento econômico, demográfico, social e ambiental de uma nação (UNECE, 1992). Frequentemente utilizadas para geração de indicadores socioeconômicos, convertem-se em insumos fundamentais para o planejamento e a formulação de políticas e estratégias no mundo contemporâneo (JANNUZZI; GRACIOSO, 2002; SCHNORR-BACKER, 2016; JANNUZZI, 2002).

No âmbito do poder público, podem subsidiar políticas e suas correspondentes ações, desde a formulação até a etapa de avaliação. No Brasil, por exemplo, conforme contextualiza e exemplifica Jannuzzi (2020), as estatísticas públicas, quando construídas em linha com demandas originárias de problemas socioeconômicos da população, têm possibilitado uma série de avanços no desenho e na gestão de políticas públicas nos mais diversos domínios. Por esse motivo, tais informações estatísticas alcançam um diferenciado caráter institucional, com

necessidade de que sejam periódicas e abrangentes, tendo em vista sua propriedade em subsidiar e permitir uma atuação autônoma e fundamentada ao Estado no âmbito da gestão pública (LINDBLUM, 1991) – assumindo papel propositivo, menos restrito do que o trazido pela ótica pluralista de análise de políticas públicas (ROCHA, 2005) e o associado à perspectiva analítica marxista convencional (JANNUZZI, 2020).

Além disso, tais estatísticas ainda contribuem para a consolidação da democracia e do exercício da cidadania à medida que permitem que a sociedade civil em geral – incluindo a academia – atue com controladoria e desenvolvimento paralelo de conhecimento.

No aspecto privado, não são menos importantes. O planejamento e as ações conduzidas por empresários e empreendedores são mais precisos e envolvem menores riscos quando realizados em posse de análises conjunturais bem fundamentadas. Ainda, estatísticas públicas podem agregar conhecimento relevante quando da avaliação das possibilidades de crescimento de mercados e da identificação de oportunidades locais.

Desta forma, ao subsidiarem ações em busca por resultados concretos em meio a processos de tomada de decisão, indicadores socioeconômicos podem ser considerados representações de conhecimento explícito (SCHREIBER et al., 2000).

2.3.1 A produção tradicional

No contexto do Brasil republicano, a produção de estatísticas públicas tem contribuído para documentar de forma detalhada e abrangente, precisa e sistemática, a história nacional – social e econômica – até então apenas retratada por viajantes estrangeiros do século XVI, pelos registros subnotificados de nascimentos, casamentos e óbitos nas paróquias e pela contabilidade imprecisa de produtos de exportação do Brasil colônia e do Brasil império (JANNUZZI, 2020).

A partir de tais estatísticas, hoje produzidas por instituições como o IBGE, o Instituto Nacional de Estudos e Pesquisas Educacionais (Inep), o Instituto Nacional de Pesquisas Espaciais (Inpe), a Fundação Oswaldo Cruz (Fiocruz), os departamentos de estatísticas e pesquisas dos Ministérios, os órgãos subnacionais de planejamento e estatística, entre outros, tem sido possível ao Brasil avançar no

âmbito do desenho e da gestão de políticas públicas nos mais diversos domínios (JANNUZZI, 2020) – reforçando aqui o caráter institucional da informação estatística, além da necessidade de que seja periódica e abrangente.

Assim, de modo geral, organizações produtoras de estatísticas públicas são encarregadas de produzir e disponibilizar estatísticas sobre uma série de domínios e em diversos níveis de escala e granularidade (KITCHIN, 2015). Mais ainda, servem a todos os setores da sociedade, públicos e privados, subsidiando processos de tomadas de decisão com informações (HARWOOD; MAYER, 2016). Assim, pelo caráter da atividade e pelo fato de que estatísticas públicas são insumos para a construção de indicadores, a produção de indicadores socioeconômicos acaba por estar fortemente associada – confundindo-se, por vezes – à produção de estatísticas oficiais e/ou públicas.

Pela importância da sua missão, é necessário que produtores se baseiem nas melhores práticas e em critérios de qualidade, sob orientação de estatísticos com independência profissional e objetividade (EUROSTAT, 2014). Necessita-se que os produtos, por sua vez, sejam imparciais, confiáveis, de qualidade, respeitem padrões que permitam comparabilidade ao longo do tempo e entre diferentes localidades, entre outros princípios fundamentais (UNECE, 1992; IBGE, 2021).

Como fonte de conhecimento adicional, os artigos de Schwartzman (1997) e Feijó e Valente (2005) fazem uma importante reflexão sobre aspectos associados à legitimidade, à credibilidade e à confiabilidade das estatísticas públicas, contextualizando com o papel das instituições produtoras e, no âmbito do primeiro, apresentando um estudo de caso sobre o IBGE. Senra (1999), em tema correlato, discorre sobre a necessidade da existência de um órgão coordenador da produção nacional de estatísticas públicas. Nesses termos, espera-se que organizações produtoras de estatísticas públicas atuem à luz do rigor científico internacionalmente consolidado e com autonomia dentro da estrutura governamental local.

Dado o caráter da atividade e dos produtos gerados, agentes produtores acabam por concentrar-se predominantemente no setor público, nos institutos de pesquisa, na academia e no terceiro setor. Em geral, utilizando-se de práticas que envolvem pesquisas amostrais e censos – e, mais recentemente, valendo-se de

registros administrativos e cadastros públicos (HACKL, 2016). O IBGE, maior produtor de estatísticas públicas do Brasil, responsável por diversas estatísticas oficiais do país, concentra grande parte da sua produção de indicadores sociais, econômicos, demográficos e ambientais em processos que envolvem censos ou pesquisas amostrais.

As pesquisas amostrais e os censos são levantamentos que possuem utilidade e qualidade reconhecida por conta do sustento metodológico estatístico cientificamente consolidado (IBGE, 2021), rigorosamente desenvolvido e padronizado ao longo de décadas (KITCHIN, 2015). São métodos sustentados pela teoria da probabilidade (COCHRAN, 1977) que permitem alto potencial analítico ao viabilizarem que covariáveis sejam investigadas em conjunto com as variáveis principais de interesse ao mesmo tempo que podem ser controlados pelo pesquisador (CITRO, 2014). Além disso, possibilitam que boa parte das fontes de erro associadas aos produtos sejam conhecidas, conforme sintetizam Brackstone (1999) e Biemer et al. (2014), facilitando mensurações relativas à qualidade e viabilizando processos inferenciais adequados para populações de interesse por meio de tratamentos corretivos e ajustes estatísticos.

2.3.2 Qualidade estatística: dimensões e *frameworks* tradicionais

Reconhecida a importância da qualidade dos produtos e processos conduzidos por produtores de estatísticas públicas e indicadores socioeconômicos, esforços em pesquisa sobre o tema têm sido empregados desde muito tempo. Como ponto de partida cabe, entretanto, explicitar a definição de qualidade de um produto estatístico.

O conceito de qualidade considerado aqui parte da definição da *International Organization for Standardization* (ISO), o qual qualidade é a totalidade de atributos e características de um produto ou serviço que culmina na sua capacidade de satisfazer necessidades explícitas ou implícitas (ABNT, 2005). Órgãos estatísticos utilizaram-se desta para definir o que viria a ser a qualidade de produtos estatísticos, ou, simplesmente, a qualidade estatística. FAO (2014) e Eurostat (2017) compartilham que se trata do grau em que produtos estatísticos atendem a requisitos em algumas dimensões ou princípios esperados. Em linha, para ONS

(2013) a qualidade de um produto estatístico pode ser definida como sua adequação ao seu propósito de utilização, ou sua capacidade de atender as necessidades dos usuários. Para UN (2018), a operacionalização da qualidade se faz pela observação da combinação de todos os aspectos associados a quão bem processos e produtos estatísticos satisfazem as expectativas de usuários e outros *stakeholders*, isto é, das características associadas a produtos adequados à demanda dos usuários e produzidos por meio de processos bem estruturados.

Dessas definições pode-se extrair que a qualidade estatística é fruto da combinação de conceitos abstratos correspondentes a princípios desejados de um produto estatístico, aqui chamados de “dimensões de qualidade”, representados operacionalmente por uma série de atributos ou características observáveis ou mensuráveis, que:

- a) variam conforme a importância percebida pela demanda, isto é, de acordo com o interesse – perspectivas, necessidades e prioridades – do usuário final, comprometendo produtores com a maior quantidade possível de dimensões de qualidade – para que os produtos tenham maior alcance e disseminação perante diferentes perfis de usuários (FAO, 2014);
- b) variam conforme o produto estatístico em questão, uma vez que algumas dimensões podem se aplicar menos ou mais de acordo com as características das fontes de dados e do resultado final desejado;
- c) são frequentemente subdivididas perante domínios, como fazem, por exemplo, (1) Eurostat (2019), ao organizar tais dimensões em ditas de ambiente institucional, processos de produção estatística e produtos estatísticos, e (2) Agafitei et al. (2015), ao organizá-las em de entrada/insumos (*input*), processo (*process/throughput*) e saída (*output*).

Nesse contexto, instrumentos já desenvolvidos para aferição da qualidade de produtos estatísticos diferenciam-se pelas dimensões e domínios que consideram, assim como pelas métricas ou formas de operacionalização das dimensões de qualidade.

Por vezes, tais instrumentos são denominados *frameworks* conceituais, definidos genericamente por estruturas ou artefatos construídos com o propósito de

identificar elementos componentes de um sistema e suas relações a fim de nortear análises, explicando os processos e prevendo os resultados (CARVALHO, 2013). Ou seja, correspondem no presente contexto a estruturas conceituais que identificam e formalizam dimensões e domínios de qualidade – bem como seus relacionamentos – de produtos estatísticos para fins de nortear o processo de aferição da qualidade.

Em âmbito geral, a maioria dos sistemas de gestão ou *frameworks* para aferição da qualidade de produtos e serviços partem dos princípios (SAEBO, 2016) do *Total Quality Management* (TQM) – Crosby (1979). Com foco específico sobre dados e informações, sob a ótica da ciência da informação, Zhu et al. (2012) propuseram o *Total Data Quality Management* (TDQM). Outros modelos e *frameworks* também propostos e utilizados no âmbito da ciência da computação e da tecnologia da informação podem ser encontrados em Jaya et al. (2017) e Batini e Scannapieco (2006).

Para o caso de indicadores socioeconômicos e estatísticas públicas, entretanto, em virtude das características dos produtos e da demanda, tais princípios não são suficientes pois, dentre outras razões, por exemplo, não possuem abertura para se considerar a avaliação da qualidade quanto à proteção dos dados pessoais, à independência dos produtores e à imparcialidade dos resultados (SAEBO, 2016), imprescindíveis de serem consideradas no âmbito da produção estatística.

Dessa forma, o comparativo que pode ser feito entre os conceitos de qualidade estatística e qualidade de dados é que o primeiro é mais amplo, abrangendo o segundo em seu escopo, uma vez que se refere a produtos estatísticos que se utilizam de dados para serem gerados. Na qualidade de dados, aspectos relacionados ao ambiente institucional, ao processo de produção e ao quanto bem a necessidade do usuário é satisfeita, dentre outros, por exemplo, não são usualmente considerados.

Em indicadores socioeconômicos, características e valores desejados associam-se aos ditos Princípios Fundamentais das Estatísticas Oficiais (UNECE, 1992). Primeiramente pelo fato de que indicadores socioeconômicos são produtos estatísticos majoritariamente construídos a partir de estatísticas oficiais e públicas. Em segundo lugar, devido ao fato de que produtores de estatísticas oficiais são, por princípio, organizações altamente engajadas do ponto de vista do rigor científico

considerado. Assim, a aferição da qualidade dos indicadores produzidos passou a ser conduzida por *frameworks* propostos à luz do trabalho de UNECE (1992).

Os Princípios Fundamentais das Estatísticas Oficiais (UNECE, 1992) são balizadores para o processo de produção e disseminação de estatísticas públicas, adotados pela Comissão de Estatística das Nações Unidas (UNSC) em 1994 e mantidos até os dias atuais, compostos por dez diretrizes: relevância, imparcialidade e igualdade de acesso; padrões profissionais e ética; responsabilidade e transparência; prevenção do mau uso dos dados; eficiência; confidencialidade; legislação; coordenação nacional; uso de padrões internacionais; cooperação internacional.

À luz dos Princípios Fundamentais das Estatísticas Oficiais e do TQM, o Gabinete de Estatísticas da União Europeia (Eurostat) lançou o *European Statistics Code of Practice* (CoP) em 2005 – revisado em 2017 (EUROSTAT, 2017) após atualização da legislação estatística que reforçou o sistema estatístico europeu ao fortalecer o papel dos institutos nacionais de estatística (Regulação 2015/759 EU). Seu lançamento original estimulou a construção de outros códigos de recomendação para boas práticas estatísticas, como o *Recommendation of the OECD Council on good statistical practice*, da Organização para a Cooperação e Desenvolvimento Econômico (OCDE, 2015), o *Code of Practice for Official Statistics*, da Autoridade de Estatísticas do Reino Unido – revisado recentemente (UKSA, 2018) – e, no Brasil, o Código de Boas Práticas das Estatísticas do Instituto Brasileiro de Geografia e Estatística (IBGE, 2021), atualmente em sua segunda edição.

A versão revisada do CoP contempla uma declaração de qualidade endereçada ao Sistema Estatístico Europeu (ESS) e um *framework* de orientação para as práticas estatísticas, o *Quality Assurance Framework* (QAF), documentado em versão mais recente por Eurostat (2019). Conforme Saebo (2016), praticamente todos os países europeus conduzem sua produção de estatísticas oficiais por meio de uma abordagem sistemática de gestão da qualidade baseada nos princípios do CoP.

Além da Eurostat, com o ESS QAF (EUROSTAT, 2019), diversos agentes envolvidos ou preocupados com produção de estatísticas oficiais e públicas também já desenvolveram *frameworks* de qualidade. Dentre alguns, pode-se citar:

- a) *National Quality Assurance Framework* (NQAF), da Organização das Nações Unidas (UN, 2019);
- b) *Data Quality Framework* (DQF), do Bureau de estatística da Austrália (ABS, 2009);
- c) *Statistics Canada Quality Guidelines* (SCQG), do instituto de estatística do Canadá (STATISTICS CANADA, 2019);
- d) *Statistics Quality Framework* (SQF), do Banco Central Europeu (ECB 2008);
- e) *Data Quality Assessment Framework* (DQAF), do Fundo Monetário Internacional (FMI, 2003);
- f) *Statistics Quality Assurance Framework* (SQAF), da Organização das Nações Unidas para a Alimentação e a Agricultura (FAO, 2014);
- g) Código de Boas Práticas das Estatísticas, do Instituto Brasileiro de Geografia e Estatística (IBGE, 2021).

A título de breve comparação, o Quadro 4 apresenta domínios, dimensões e a quantidade de indicadores ou elementos/atributos considerados por cada um dos *frameworks* mencionados. Como ponto de partida para aprofundamento sugere-se Eurostat (2019) e HIQA (2018), que exploram as dimensões trazidas pelo ESS QAF e analisam em profundidade uma série de propostas de *frameworks*. No Brasil, o Código de Boas Práticas das Estatísticas do Instituto Brasileiro de Geografia e Estatística (IBGE, 2021), em sua segunda edição, revisada, apesar de soar como um documento orientador com diretrizes e recomendações para práticas estatísticas, possui também um caráter de *framework* para aferição da qualidade estatística. Com 17 dimensões (8, 4 e 5 em cada domínio) e 73 boas práticas (39, 17 e 17 em cada domínio), o código do IBGE atua internamente ao instituto como uma diretriz para o trabalho de produção estatística e, externamente, como um instrumento para prover confiança aos informantes e aos diversos grupos de usuários quanto ao cumprimento de rígidos padrões de excelência e de requisitos internacionais que pautam as melhores práticas de reconhecidos Institutos Nacionais de Estatística em todo o mundo (IBGE, 2021).

Quadro 4 – Comparativo entre *frameworks* tradicionais de qualidade para estatísticas oficiais e públicas selecionados.

Framework	Domínio	Dimensão	Indicadores/requisitos/elementos/práticas
ESS QAF v2.0 (Eurostat, 2019)	Ambiente institucional	Independência; Coordenação e cooperação; Permissão de coleta e acesso aos dados; Recursos adequados; Comprometimento com a qualidade; Confidencialidade; Imparcialidade e objetividade	37 indicadores (8, 3, 4, 4, 4, 6 e 8 em cada dimensão)
	Processos de produção estatística	Metodologia sólida; Procedimentos estatísticos apropriados; Cargas de pesquisa não excessiva sobre respondentes; Custo benefício	24 indicadores (7, 7, 6 e 4 em cada dimensão)
	Produtos estatísticos	Relevância; Acurácia e confiabilidade; Pontualidade e oportunismo temporal; Coerência e comparabilidade; Acessibilidade	23 indicadores (3, 3, 5, 5 e 7 em cada dimensão)
UN NQAF (UN, 2019)	Sistema estatístico nacional	Coordenação do sistema nacional de estatística; Relacionamento entre usuários, produtores e outros; Padrões estatísticos	14 requisitos (4, 7 e 3 em cada dimensão)
	Ambiente institucional	Independência profissional; Imparcialidade e objetividade; Transparência; Confidencialidade estatística e segurança; Comprometimento com a qualidade; Adequação dos recursos	29 requisitos (3, 7, 2, 6, 8 e 3 em cada dimensão)
	Processos de produção estatística	Metodologia sólida; Custo benefício; Procedimentos estatísticos apropriados; Cargas de pesquisa não excessiva sobre respondentes	20 requisitos (5, 6, 5 e 4 em cada dimensão)
	Produtos estatísticos	Relevância; Acurácia e confiabilidade; Pontualidade e oportunismo temporal; Acessibilidade e clareza; Coerência e comparabilidade; Metadados	24 requisitos (4, 3, 4, 7, 3 e 3 em cada dimensão)
ABS DQF (ABS, 2009)	Não explícitos	Ambiente institucional; Relevância; Oportunismo temporal; Acurácia; Coerência; Interpretabilidade; Acessibilidade	29 indicadores (6, 7, 2, 6, 4, 2 e 2 em cada dimensão)
SCQG (<i>Statistics Canada</i> , 2019)	Dados e Processo	Qualidade dos dados (Relevância; Oportunismo temporal e pontualidade; Acurácia e confiabilidade; Acessibilidade e clareza; Interpretabilidade; Coerência e comparabilidade) e do processo.	31 indicadores para a qualidade dos dados (5, 7, 7, 6, 3, 3 em cada dimensão)
ECB SQF (ECB 2008)	Ambiente institucional	Independência e prestação de contas; Autoridade para coletas; Imparcialidade e objetividade; Confidencialidade estatística; Coordenação e cooperação; Recursos e eficiência	26 elementos (5, 1, 10, 3, 4 e 3 em cada dimensão)
	Processos de produção estatística	Metodologia sólida e procedimentos estatísticos adequados; Custo benefício e carga não excessiva sobre respondentes	9 elementos (6 e 3 em cada dimensão)
	Produtos estatísticos	Relevância e completude; Acurácia e confiabilidade; Consistência/coerência; Pontualidade e oportunismo temporal; Acessibilidade e clareza	28 elementos (5, 5, 5, 8 e 3 em cada dimensão)
FMI DQAF (FMI, 2003)	Não explícitos	Pré-requisitos de qualidade; Garantia de integridade; Metodologia sólida; Acurácia e confiabilidade; Facilidade de manutenção; Acessibilidade	51 indicadores (10, 8, 6, 10, 8 e 9 em cada dimensão)
FAO SQAF (FAO, 2014)	Ambiente institucional	Independência profissional e imparcialidade; Confidencialidade estatística; Comprometimento com a qualidade; Padrões internacionais; Cooperação; Coordenação com outros produtores	31 elementos (8, 7, 6, 4, 3 e 3 em cada dimensão)
	Processos de produção estatística	Metodologia sólida e procedimentos estatísticos adequados; Custo benefício; Carga não excessiva sobre respondentes	21 elementos (11, 4 e 6 em cada dimensão)
	Produtos estatísticos	Relevância; Acurácia e confiabilidade; Pontualidade e oportunismo temporal; Coerência e comparabilidade; Acessibilidade e clareza	21 elementos (4, 5, 4, 4 e 4 em cada dimensão)
Código de Boas Práticas das Estatísticas (IBGE, 2021)	Ambiente institucional e coordenação	Independência; Coordenação do SNE; Mandato estatístico; Sigilo; Adequação dos recursos; Compromisso c/ a qualidade; Imparcialidade e objetividade; Cooperação e participação internacional	39 boas práticas (7, 3, 3, 7, 5, 4, 8, 2 em cada dimensão)
	Processos estatísticos	Metodologia sólida; Processos estatísticos adequados; Solicitação não excessiva de dados; Relação entre custo e eficácia	17 boas práticas (6, 5, 3, 3 em cada dimensão)
	Produtos estatísticos	Relevância; Precisão e exatidão; Oportunidade e pontualidade; Coerência e comparabilidade; Acessibilidade e clareza	17 boas práticas (4, 3, 3, 3, 4 em cada dimensão)

Fonte: atualizado de Caumo (2020).

Cabe destacar que, em relação à mensuração da acurácia, dimensão presente em todos os *frameworks* de qualidade apresentados e que diz respeito ao quão corretamente um resultado estatístico estima ou descreve o parâmetro populacional de interesse que se propõe a medir, a abordagem de fundo predominantemente utilizada é a do *Total Survey Error* (TSE), com raízes descritas em Groves e Lyberg (2010), Biemer (2010) e Biemer et al. (2017). A abordagem do TSE se baseia no *total survey error paradigm* (PLATEK; SARNDAL, 2001), no qual se orienta a partir da identificação das grandes fontes de erros de uma pesquisa para que então esforços e recursos para minimizá-las possam ser direcionados. Tal paradigma é parte do conceito maior de *Total Survey Quality* (TSQ), que aborda a qualidade no contexto da adequação da estimativa produzida ao seu fim (JURAN; GRZYNA, 1980).

O TSE considera a acurácia como o aspecto principal da qualidade associada a um resultado estatístico gerado por uma pesquisa do tipo *survey*, usualmente representada pelo erro quadrático médio – uma medida composta que observa o viés (parcialidade) e a variância (precisão) de tal resultado. Quando da condução de um levantamento amostral, tais componentes, (i) viés e (ii) variância, podem ser entendidos como:

- i. o tamanho do erro sistemático das estimativas geradas em um determinado contexto. Isto é, o quanto a média das estimativas geradas por um processo de estimação repetido diversas vezes com diferentes amostras se distancia do verdadeiro valor do parâmetro populacional que as estimativas se propõem a estimar. Quando um processo de estimação – ou simplesmente um estimador – consegue acertar, na média, seu alvo de pesquisa, ele não terá viés e será dito imparcial (ou não enviesado);
- ii. o tamanho do erro aleatório das estimativas geradas em um determinado contexto. Isto é, o tamanho da variabilidade das estimativas geradas por um processo de estimação repetido diversas vezes com diferentes amostras. Quando um processo de estimação – ou simplesmente um estimador – gerar resultados que variam pouco, ele terá boa precisão (baixa variância ou erro aleatório). Se um estimador sem viés tiver a

menor variância dentre todos os demais possíveis estimadores imparciais, ele será dito o estimador mais eficiente.

No âmbito do TSE, o erro quadrático médio é impactado por fontes de erros dos tipos: amostrais, tais quais de desenho amostral, tamanho amostral e de escolha do estimador; e não amostrais, como os de especificação, não respostas, cobertura, mensuração e processamento dos dados (BIEMER, 2010; BIEMER et al., 2014). Para fins de breve exemplificação:

- i. se uma pesquisa possui taxas de não respostas não uniformemente distribuídas pelos diferentes perfis de respondentes, isto é, na qual alguns perfis específicos de respondentes se negaram a responder a pesquisa em níveis diferentes de outros perfis, os resultados estarão sujeitos à erros ditos não amostrais e podem se apresentar enviesados, com erros sistemáticos – caso não passem por tratamentos corretivos. Da mesma forma, se uma pesquisa foi incapaz de garantir uma cobertura adequada a uma população alvo de interesse, deixando de fora algumas parcelas populacionais, seu processo de estimação poderá também estar enviesado uma vez que a amostra poderá não ser representativa da população;
- ii. se uma pesquisa considerou um tamanho de amostra muito pequeno, os resultados poderão estar sujeitos a altos erros aleatórios, com um processo de estimação de alta variância. Isto é, as estimativas poderão variar muito pois dependem de quem foram os integrantes da pequena amostra investigada. Da mesma forma, processos de estimação poderão apresentar variabilidades – ou erros aleatórios – diferentes, sendo mais ou menos precisos, caso tenham considerado distintos métodos de seleção – desenhos de amostragem – dos participantes da amostra.

Para quando da realização de pesquisas mais complexas ou da utilização de dados que derivam de fontes diferentes de *surveys*, como os de registros administrativos (DAAS et al., 2009; NEDERPELT, 2010; NEDERPELT; DAAS, 2012; IWIG et al., 2013; WALLGREN; WALLGREN, 2014), importantes discussões são feitas por Biemer et al. (2017) e Reinert e Stoltze (2016).

É possível perceber que, dos códigos de recomendação e *frameworks* aqui apresentados, alguns têm por objetivo servir apenas de sugestão – ou ponto de

partida genérico – para países quando da estruturação de seu sistema de produção estatística, identificando atividades, métodos e ferramentas que podem ser utilizadas. Entretanto, cabe ressaltar que em geral este tipo de documentação vem evoluindo: (1) passando de “*assessment*” para “*assurance*”, isto é, de suporte para adoção de práticas internas para ferramenta de certificação; (2) ampliando sua aplicabilidade daquela restrita a produtos estatísticos gerados a partir de pesquisas amostrais do tipo *survey* para uma utilizável em produtos gerados a partir de outras fontes – como registros administrativos (ONS, 2013); (3) e expandindo a aferição da qualidade dos produtos para a qualidade dos processos de produção como um todo, com sistemas e metodologias padrões, impulsionados pela introdução do *Generic Statistical Business Process Model* (GSBPM) (UNECE, 2013a) – que passou a servir de base para a grande maioria dos países europeus na estruturação de seus modelos de processo de negócio estatístico (EUROSTAT, 2016).

Um exemplo representativo dessa evolução é o *Guidelines for Measuring Statistical Output Quality – Version 4.1* (ONS, 2013), do Departamento Nacional de Estatísticas do Reino Unido. O documento fornece um checklist de medidas e indicadores para mensuração da qualidade estatística que engloba diversos aspectos evolutivos mencionados. O manual sugere métricas para todas as etapas do GSBPM no âmbito da mensuração de cada uma das cinco dimensões de qualidade de produtos estatísticos consideradas pela ESS (relevância, acurácia e confiabilidade, pontualidade e oportunismo temporal, acessibilidade e clareza, coerência e comparabilidade), tanto para contextos de dados oriundos de pesquisas amostrais quanto de registros administrativos ou de múltiplas fontes.

Em suma, conforme observa Agafitei et al. (2015), a aferição da qualidade da produção tradicional de estatísticas públicas alcançou um estágio de maturidade que considera três domínios: de insumos (*input*), associada às fontes produtoras e aos dados brutos; de processos de produção (*process/throughput*), associada aos métodos e procedimentos empregados durante a geração dos produtos estatísticos; e de produtos finais (*output*), associada a características que atendem aos interesses dos usuários. Os mesmos autores destacam que, apesar da mensuração da qualidade dos domínios de *input* e *output* ser tratada de forma consensual

perante os diversos mecanismos, a dimensão da qualidade do processo é algumas vezes avaliada sob a ótica do GSBPM e do TQM, com foco nas etapas do processo em si e objetivando melhorias de interesse do produtor, e outras vezes avaliada sob a ótica de como cada etapa do processo de produção influencia na qualidade do produto final, com foco na demanda – no que deseja o usuário final.

De um modo geral, percebe-se que *frameworks* de qualidade devem ser suficientemente genéricos para abarcar diversos contextos, com importância dada a cada domínio e dimensão de qualidade decidida caso a caso (ECB, 2008) – com base nas características das fontes de dados e dos resultados finais desejados, necessidades da demanda.

Essa percepção também existe fora do escopo das estatísticas públicas e oficiais, como, por exemplo, quando do estudo da qualidade de dados e informações, conforme a revisão de literatura no âmbito da ciência da informação de Fagundes, Macedo e Freund (2017). Para Valente e Fujino (2016), mesmo que não exista consenso sobre o conceito de qualidade e até mesmo sobre o significado exato de cada dimensão (BATINI et al., 2009), diversas são as proposições de dimensões de qualidade de dados e informações, dependentes da abordagem e da vertente de cada aplicação. Percepção carregada até os dias atuais, conforme corroboram Gudivada, Apon e Ding (2017) e Jaya et al. (2017) ao dizerem que a qualidade dos dados está altamente associada com o contexto de utilização final e sua sinergia com as necessidades dos clientes, habilidades de utilização e de acesso aos dados. Em linha, Wang e Strong (1996) propõem que a qualidade de dados seja considerada de maneira genérica para que possa ser aplicada a diferentes domínios e contextos, classificando 15 dimensões de qualidade em 4 categorias: intrínsecas (associada à acurácia), contextuais (associada à relevância), representacionais (associada à forma de representação) e de acessibilidade (associada à forma de acesso).

2.3.3 Dificuldades da produção tradicional

Apesar de reconhecidas pelo rigor científico e pelos avançados estudos de mensuração da qualidade dos produtos gerados, algumas práticas do processo

tradicional de produção de estatísticas públicas não estão isentas de algumas adversidades (CAUMO, 2020).

Estudos têm apontado para o crescimento de problemas de cunho operacional e metodológico até então não bem explorados e que podem impactar no conhecimento sobre a qualidade dos resultados gerados (MANSKI, 2014). Em geral, dificuldades associadas aos levantamentos de campo do tipo *survey*, sejam eles censos ou pesquisas amostrais, especialmente quando domiciliares, de grande porte e conduzidos por pessoas. Estes possuem operacionalização complicada – por vezes complexa – e sujeita a uma série de contratempos, tendendo a serem demorados e podendo serem caros (CITRO, 2014; HACKL, 2016).

Em certa medida, a maioria dessas problemáticas operacionais e logísticas (BRAAKSMA; ZEELENBERG, 2015) acabam por impactar na taxa de aproveitamento – ou taxa de respostas – dos levantamentos (CITRO, 2014; DE LEEUW; DE HEER, 2002). Ou seja, alteram negativamente a razão entre os indivíduos efetivamente investigados e a quantidade inicialmente planejada. Aspecto problemático, pois, além de significar desperdício de recursos e fonte para uma série de diferentes erros, a tentativa de corrigi-lo pode ampliar o viés dos resultados (GROVES; PEYTCHEVA, 2008; FRICKER; TOURANGEAU, 2010). Em relação aos fatores que podem impactar na taxa de aproveitamento, pode-se citar:

- a) aumento das recusas em fornecer respostas – totais ou em itens específicos – por descrença na pesquisa – no governo ou no entrevistador – ou simplesmente por fadiga – especialmente em investigações longas que acabam por desencadear o que é conhecido por *response burden* (HACKL, 2016; STRUJIS; BRAAKSMA; DAAS, 2014);
- b) informações coletadas nem sempre são precisas pois suas metodologias e constructos teóricos por vezes se apresentam desatualizados frente a um contexto de constantes e cada vez mais frequentes transformações sociológicas ou também pelo fato de que os indivíduos não entenderam o questionamento, mentem ou simplesmente acreditam e comunicam ações que, na prática, não correspondem à realidade (CITRO, 2014; HAND, 2015);

- c) no caso de pesquisas domiciliares, está cada vez mais difícil encontrar indivíduos nas residências, seja pelo fato de que as famílias estão menores ou porque passam mais tempo fora de casa⁴;
- d) o trabalho de entrevistador de campo exige esforços de treinamento, atualização e motivação constantes por parte do empregador e necessita de grande dedicação e múltiplas habilidades por parte do profissional. Em paralelo, geralmente é mal remunerado e acaba por apresentar alta rotatividade. Tais questões são problemáticas pois um satisfatório desempenho por parte do entrevistador é crucial para o alcance de maior aproveitamento e para a qualidade dos resultados gerados;
- e) também para pesquisas domiciliares, a violência urbana pode intimidar e enclausurar moradores, fazendo com que não aceitem receber o entrevistador, da mesma forma que pode influenciar no acesso e no trânsito do entrevistador à determinadas áreas da cidade;
- f) ainda para pesquisas domiciliares, o crescimento na quantidade de condomínios fechados – verticais e horizontais – com regulamento de acesso rigidamente restrito tem dificultado cada vez mais o acesso dos entrevistadores à domicílios específicos.

Além dessas questões, perturbações na ordem social podem, por vezes, desencadear problemas de operacionalização ainda maiores em uma pesquisa de campo, culminando na reestruturação, na interrupção, no adiamento ou até mesmo na inviabilização completa do levantamento. Como exemplo: a recente pandemia causada pelo Covid-19, popularmente conhecido como “Novo Coronavírus”, que inviabilizou temporariamente todas as pesquisas conduzidas em campo em diversos países.

No Brasil, seus efeitos foram devastadores a ponto de interferir na realização do Censo Demográfico do Instituto Brasileiro de Geografia e Estatística (IBGE), previsto para 2020. Face as complicações trazidas pela doença, a coleta dos dados populacionais foi transferida para 2021⁵. Além disso, a Pesquisa Nacional por Amostras de Domicílios Contínua (PNAD Contínua), também do IBGE, foi,

⁴ Com a epidemia de Covid 19, este aspecto perdeu validade ao colocar as famílias em isolamento domiciliar e impulsionar o remodelamento das formas de trabalho (mais remoto, à distância). Porém, há a expectativa de que volte a ser válido, pelo menos em parte, quando da resolução das atuais adversidades.

⁵ Posteriormente adiado para 2022 também por questões políticas e financeiras, somadas a de saúde pública.

excepcionalmente durante o período da pandemia, realizada por telefone – alteração que pode mexer com a qualidade dos resultados ou até prejudicar as séries históricas em caso de mudança na taxa de aproveitamento ou do surgimento do dito *mode effect* (GROVES, 1989). Da mesma forma, a pesquisa de preços em locais de compras para fins de cálculo da inflação, base do Sistema Nacional de Índices de Preços ao Consumidor (SNIPC), até então parcialmente conduzida através de coleta presencial de dados, passou a ser integralmente realizada pela internet ou pelo telefone.

No âmbito metodológico, parte dos problemas (MANSKI, 2014; HAND, 2015) estão associados à dificuldade de se conseguir sistemas/marcos de referência – cadastros/listas com a relação de indivíduos/entidades – completos e atualizados da população alvo – podendo culminar em problemas de cobertura e representatividade (CITRO, 2014). Ou seja, sem um cadastro completo e atualizado da população alvo de interesse, pesquisas podem acabar por não cobrir – abranger – uma população em sua totalidade, gerando resultados que poderão não ser representativos do todo. Do ponto de vista técnico, sorteios probabilísticos de indivíduos (ou entidades) que comporão a amostra a ser investigada serão prejudicados e os cálculos dos indicadores de qualidade dos resultados da pesquisa poderão gerar valores distorcidos. Ou seja, as garantias trazidas pela teoria da amostragem poderão ser invalidadas e os reais viés e variância das estatísticas poderão permanecer desconhecidos nos casos em que problemas nos cadastros não forem explícitos e tratados.

Além das problemáticas metodológico-operacionais, muitas vezes censos e pesquisas amostrais sob a forma de levantamentos de campo podem demandar alto custo financeiro (HACKL, 2016; DAAS; PUTS, 2014; KITCHIN, 2015; TAM; CLARKE, 2014; EUROSTAT, 2014a) de planejamento e execução, assim como tomar tempo considerável até a definitiva disponibilização dos resultados – especialmente se as pesquisas forem do tipo domiciliares. Nesse contexto, portanto, os produtos podem demandar elevados recursos e correm o risco de serem lançados com defasagem considerável em relação ao período que se referem.

No Brasil, um caso de elevada repercussão foi o do adiamento do Censo Demográfico do IBGE de 1990 para 1991. Sob justificativa de crise econômica, a contratação dos funcionários temporários necessários para condução do censo foi somente autorizada pelo governo federal em julho de 1990. Dessa forma, não houve tempo hábil para que se pudesse concluir as etapas planejamento e treinamento e então partir para a execução do campo ainda naquele ano (IBGE, 2003). Ou seja, a pesquisa esteve à mercê de condicionantes políticos e orçamentários. Caso que, de forma semelhante, parece estar acontecendo atualmente, em 2021.

Em adição às questões metodológico-operacionais, de custo e de tempo, outros desafios têm se apresentado aos produtores pelo lado da demanda – e para que se mantenham relevantes, é fundamental que produtores continuem gerando produtos e serviços que atendam aos interesses de seus usuários (HARWOOD; MAYER, 2016; FLORESCU et al., 2014), incluídas as instituições de Estado para fins de planejamento e formulação de políticas públicas (JANNUZZI, 2020).

Usuários, financiadores e provedores de estatísticas públicas cada vez mais desejam: estimativas que tratem sobre muitos aspectos e domínios; tenham profundidade no sentido de permitirem aberturas para unidades geográficas pequenas e específicas; sejam oportunas temporalmente; de mais qualidade; mais baratas; de relevância ao retratar exatamente aquilo que se deseja; e menos onerosas (HOLT, 2007; CITRO, 2014) – características para as quais ainda se poderia adicionar a necessidade de comparabilidade entre países.

Tal qualificação da demanda pode ser vista como um reflexo da percepção dos indivíduos acerca do aumento do poder tecnológico disponível e de uma consequente maior elucidação sobre a utilidade e o manuseio dos dados, criando a noção de indivíduos *data savvy* – que cada vez mais demandam esclarecimentos mais sofisticados e exigem respostas mais rápidas (HARWOOD; MAYER, 2016).

Por fim, alguns desafios, aqui chamados de estratégicos, estão relacionados ao impacto da pressão das transformações externas e do surgimento de aspirantes a produtores de estatísticas públicas na sobrevivência das instituições produtoras tradicionais (DEMUNTER, 2017; STRUIJS; DE BROE, 2018; MACFEELY, 2016). Sobre esse aspecto, praticamente todas as problemáticas derivam da imagem de que novas tecnologias – e, conseqüentemente, processos mais rápidos e econômicos – são absorvidas de maneira mais ágil pelo setor privado, convertendo-

se em ganho competitivo. Dessa forma, novas tecnologias e fontes alternativas de dados facilitarão a quebra do “monopólio” da produção de estatísticas públicas dos atuais e tradicionais produtores, reduzindo sua participação e estendendo a capacidade de produção estatística para outras esferas.

Sobre esse aspecto, cabe ressaltar que estatísticas públicas objetivas e independentes são bens públicos essenciais em sociedades democráticas e, especialmente no âmbito da gestão pública, não podem ser vistas como bens mercadológicos em virtude do seu caráter institucional, com necessidade de que sejam periódicas e abrangentes, de construção planejada a partir das demandas por problemas socioeconômicos da população, para fins de subsídio à formulação e à gestão de políticas (JANNUZZI, 2020). Nesse contexto, a obsolência da produção estatística vigente parece pouco provável.

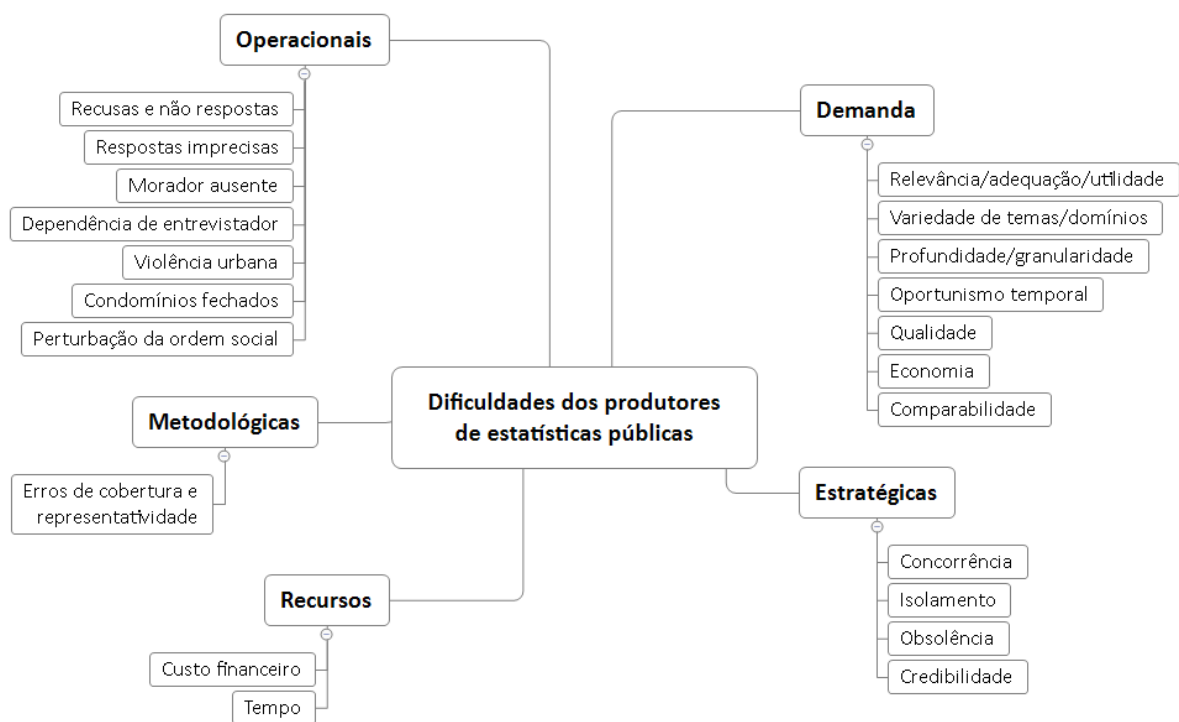
Em suma, o contexto geral de dificuldades dos produtores de estatísticas públicas elencado anteriormente foi percebido como tendo cinco grandes grupos de problemáticas, conforme resume a Figura 2: (1) operacionais; (2) metodológicas; de (3) recursos; de atendimento à (4) demanda; e de sobrevivência e posicionamento (5) estratégico para as instituições produtoras.

Alternativas de contorno às problemáticas tradicionais têm evoluído desde sempre, girando em torno: do emprego de métodos de correção estatística (CITRO, 2014); da modernização dos processos tradicionais – especialmente com a informatização e a automatização de tarefas antes manuais; da substituição de algumas pesquisas e questões por uma maior utilização de dados disponíveis em registros administrativos e cadastro públicos; de tentativas de realização de pesquisas postais, via telefone e internet; de combinações de dados de distintas fontes; entre outras.

Tal temática desperta atenção e está inserida na agenda de diversas instituições pelo mundo. No Brasil, por exemplo, o IBGE vem demonstrando estar atento a tais dificuldades ao tratar assunto como central na terceira edição da Conferência Nacional de Produtores e Usuários de Informações Estatísticas, Geográficas e Ambientais (INFOPLAN), realizada em 2016. Além disso, o instituto se tornou, recentemente, um dos quatro *hubs* físicos da *Global Platform* pelo mundo,

uma plataforma da divisão de estatísticas do departamento de economia e temas sociais da Organização das Nações Unidas (ONU), gerenciada pelo *UN Committee of Experts on Big Data and Data Science for Official Statistics (UN-CEBD)*, que se propõe a ser um ecossistema de serviço de nuvem para sustentar a colaboração internacional de conhecimentos sobre o desenvolvimento de uma produção de estatísticas oficiais e públicas que incorpore também, de forma complementar, fontes alternativas de dados e métodos inovadores no âmbito da mensuração dos Objetivos de Desenvolvimento Sustentável (ODS) e da Agenda 2030 para o desenvolvimento sustentável – tema também explorado em Jannuzzi e Carlo (2018).

Figura 2 – Resumo de algumas dificuldades recentes dos produtores de estatísticas públicas com levantamentos do tipo de campo.



Fonte: atualizada de Caumo (2020).

2.4 INDICADORES DE TURISMO

Para tratar de indicadores de turismo, um grupo específico de indicadores socioeconômicos, cabe inicialmente definir o conceito de turismo. Nesta tese, a definição considerada é a trazida pela Organização Mundial do Turismo das Nações Unidas (UNWTO), criada em 1946 – passando a integrar as Nações Unidas somente

em 2003 – e comprometida desde então, dentre outras tarefas, com a mensuração do turismo para fins de gerar e aprofundar conhecimento sobre o setor no mundo. Para a UNWTO (2008b), do ponto de vista da demanda, o turismo compõe-se das atividades de visitantes que viajam para e permanecem em locais fora de seu ambiente usual, adquirindo bens e serviços, por não mais de um ano consecutivo para lazer, negócios e outros fins não relacionados com o exercício de uma atividade remunerada dentro do local visitado. Por outro lado, do ponto de vista da oferta, turismo é entendido como o conjunto de atividades produtivas que atendem principalmente tais visitantes (UNWTO, 2008b).

No que diz respeito aos indicadores, a visão sobre a sua utilidade, no âmbito do turismo, é a mesma atribuída aos indicadores socioeconômicos em geral. Segundo a UNWTO (2007), trata-se de instrumentos essenciais para os processos de planejamento, gestão e monitoramento do turismo, fornecendo informação acurada para os processos de tomada de decisão tanto por parte da demanda quanto por parte da oferta.

Mais especificamente, pelo lado da oferta, indicadores são necessários para a formulação de estratégias de marketing, o fortalecimento das relações entre instituições, a avaliação da eficiência e da efetividade das decisões gerenciais e a mensuração do turismo no âmbito da economia nacional (UN, 2010). Ainda, a importância do monitoramento e da criação de ações estratégicas baseada em informação é maior especialmente em nível local e regional, pois é ali que o impacto gerado pelo turismo se manifesta e que os recursos, a infraestrutura e os equipamentos devem ser geridos (SANTOS; INÁCIO, 2018). Pelo lado da demanda, por sua vez, indicadores de turismo podem auxiliar em processos de tomadas de decisão quando do planejamento e da execução de um roteiro ou atividade turística.

Por vezes, indicadores de turismo são agrupados justamente como sendo relativos às perspectivas da demanda ou da oferta turística (EUROSTAT, 2014c; UN, 2010; BAGGIO, 2019), seguindo estrutura também considerada no *International Recommendations for Tourism Statistics* (UNWTO, 2008a) e no *Tourism Satellite Account: Recommended Methodological Framework* (UNWTO, 2008b), ambos publicados pela Organização Mundial do Turismo há mais de 10 anos e que até hoje

orientam a produção estatística relativa ao turismo em grandes organizações como o *World Travel & Tourism Council* (WTTC), representante global do setor privado de viagens e turismo, a Organização para a Cooperação e Desenvolvimento Econômico (OCDE), a divisão de estatística da união europeia (Eurostat) e, naturalmente, a Organização das Nações Unidas (UN).

Mais recentemente, a classificação da própria Organização Mundial do Turismo tem subdividido os indicadores de turismo (UNWTO, 2019) como sendo referentes ao: (i) turismo de entrada; (ii) turismo doméstico; (iii) turismo de saída; (iv) indústria do turismo; (v) empregos; (vi) complementares (macroeconômicos). Tal subdivisão tem sido utilizada em publicações como o *Compendium of Tourism Statistics* e o *Yearbook of Tourism Statistics*, que constituem as principais bases de dados e informações da UNWTO sobre turismo no mundo – completos guias de referência para a mensuração e a análise comparativa do setor turístico de mais de 200 países utilizando-se de um conjunto de mais de 145 indicadores.

2.4.1 A produção tradicional no Brasil

No Brasil, o Ministério do Turismo (MTur) coordenou, em convênio pactuado junto à Fundação Getúlio Vargas (FGV) e à Fundação Instituto de Pesquisas Econômicas (FIPE), a elaboração do Plano Nacional Estratégico de Estatísticas Turísticas (PNETUR), no âmbito do Plano Nacional de Turismo (BRASIL, 2018a), construído de forma coletiva com as demais instituições pertencentes ao Sistema Nacional de Turismo – Instituto Brasileiro de Turismo (Embratur), Conselho Nacional de Turismo (CNT) e Fórum Nacional de Secretários e Dirigentes Estaduais de Turismo.

A construção do PNETUR (BRASIL, 2016) visou avançar no alinhamento das estatísticas nacionais de turismo às recomendações do *International Recommendations for Tourism Statistics* (UNWTO, 2008a) e na operacionalização dos compromissos de harmonização de pesquisas aprovadas pelos países participantes do Projeto CONESUL em termos de: contribuir para a melhoria da qualidade das estatísticas turísticas; avançar na comparabilidade intrarregional da atividade turística; impulsionar a padronização e harmonização dos principais instrumentos de observação e mensuração da atividade turística; e avançar no

processo de renovação das operações estatísticas sob responsabilidade do Ministério do Turismo, componentes do núcleo básico do Sistema Nacional de Informações Turísticas, referente ao turismo receptivo, interno e ocupação hoteleira, entre outros (BRASIL, 2018a).

Em outras palavras, o PNETUR tratou de diagnosticar e traçar um plano estratégico de estatísticas turísticas nacionais para auxiliar a produção, o armazenamento e disseminação de estatísticas, padronizadas e comparáveis, sobre a atividade turística no Brasil (BRASIL, 2016). Nesses termos, o Ministério do Turismo é, no país, ao lado do Sistema Nacional de Turismo, o grande catalizador da produção de estatísticas sobre o turismo ao apoiar – com suporte técnico – a identificação, geração e análise de informações e estatísticas subnacionais necessárias para a gestão da atividade turística. Para tanto, orienta sobre a reunião, sistematização e disponibilização de dados estatísticos e informações relativas aos mercados emissores consolidados e potenciais, além de realizar estudos, análises e levantamentos que propiciam o conhecimento do comportamento e das tendências do turismo nacional e internacional (BRASIL, 2018a).

Durante a construção do PNETUR (BRASIL, 2016) foi realizado um levantamento de todas as operações estatísticas e fontes de dados sobre turismo que, naquele momento, compunham o Sistema Nacional de Informação Turística (SITUR). Como resultado, o relatório aponta 46 diferentes pesquisas amostrais, censos, registros administrativos ou operações estatísticas de variadas organizações que podem ser utilizadas como fontes de dados sobre o turismo. Ao perceber uma quantidade considerável de informação relevante sendo produzida de maneira isolada por algumas instituições, o relatório sugere, ao final, para fins de evolução do SITUR para o SITUR Fortalecido, que sejam estabelecidos convênios de troca de informações com o Instituto Brasileiro de Geografia e Estatística (IBGE), a Polícia Federal e o Banco Central (BRASIL, 2016), além de citar também outros órgãos como responsáveis pela produção e divulgação de dados relevantes para o setor do turismo – IPEA, Polícia Rodoviária Federal, Secretaria de Aviação Civil, ANAC, ANTAQ, Ministério da Fazenda, Ministério do Desenvolvimento, Indústria e Comércio

Exterior, Estados, Municípios, representantes do setor privado, academia e demais entidades relacionadas ao setor de turismo.

Após também sugerir reformulações nas pesquisas de turismo internacional e de turismo interno realizadas pelo MTur, o PNETUR propõe uma estrutura final para a base de dados do SITUR Fortalecido subdividida em quatro grandes grupos de conteúdo: demanda turística; oferta turística; meios de hospedagem; e outros indicadores síntese (BRASIL, 2016). As principais fontes de dados – acompanhadas de suas características – sugeridas pelo documento estão apresentadas no Quadro 5.

Quadro 5 – Conteúdos e possíveis fontes de informações sugeridas para o SITUR Fortalecido pela PNETUR.

Conteúdo	Principal fonte de dados	Síntese metodológica da fonte
Demanda turística	Pesquisa de Turismo Internacional	Pesquisa em campo com amostra de 45.000 turistas espalhados por 25 pontos de entrada e saída do Brasil (edição de 2018).
	Pesquisa de Turismo Interno/Doméstico	Pesquisa em campo domiciliar com amostra de pouco mais de 37.000 domicílios particulares permanentes espalhados por 137 municípios do país (edição de 2012).
	Pesquisas de Ocupação Hoteleira	Variadas pesquisas realizadas por consultorias ou agentes locais ou regionais.
	Diretório de Meios de Hospedagem	Variados registros administrativos (CADASTUR e SNRHos do MTur, CEMPRE e PSH do IBGE, e CAGED/RAIS do MTE).
Oferta turística	Diretório de Estabelecimentos Turísticos	Variados registros administrativos (CADASTUR e SNRHos do MTur).
	Pesquisa Anual de Serviços	Pesquisa por Questionário eletrônico auto preenchido em amostra de empresas espalhadas pelo país (edição de 2017).
	Relação Anual de Informações Sociais	Registro administrativo com informações de pessoas jurídicas inscritas no CNPJ.
Meios de hospedagem	Diretório de Meios de Hospedagem	Variados registros administrativos (CADASTUR e SNRHos do MTur, CEMPRE e PSH do IBGE, e CAGED/RAIS do MTE).
	Pesquisas de Ocupação da Hotelaria	Variadas pesquisas realizadas por consultorias ou agentes locais ou regionais.
Outros indicadores síntese	Não mencionado	Não se aplica.

Fonte: elaborado pelo autor (2021) com base em Brasil (2016).

Em adição, uma relação de fontes de dados turísticos do Brasil é disponibilizada como anexo do Anuário Estatístico de Turismo, do Ministério do Turismo, compilado e divulgado anualmente. A relação, para consulta, está aqui disponibilizada sob a forma de anexo – ANEXO A – e aponta para as seguintes

instituições detentoras das informações: MTur, FIPE, Polícia Federal, BACEN, FGV, INFRAERO, ANAC, ANTT, IPEA, MTE e IBGE.

Pode-se perceber que, a partir de todas as fontes percorridas e que subsidiam o SITUR, dados de turismo no Brasil possuem origem fortemente concentrada em pesquisas amostrais conduzidas em campo ou de modo eletrônico e em registros administrativos ou cadastros mantidos por organizações governamentais ou públicas. Ou seja, estão sujeitos às mesmas problemáticas que rondam a produção tradicional de estatísticas públicas, apresentadas na Subseção 2.3.3.

Além disso, durante a elaboração do Plano Nacional Estratégico de Estatísticas Turísticas 2016-2020 foi realizada uma pesquisa com objetivo de perceber a opinião de reais e potenciais usuários do SITUR. Para tanto, investigou-se 294 reais e potenciais usuários de todas as unidades da federação, distribuídos por instituições do tipo universidades, secretarias de turismo e órgãos distritais, municipais ou estaduais, administração federal, associações, fóruns e conselhos, organismos financeiros, empresas privadas e sistema S. Dos resultados, chama a atenção que (BRASIL, 2016):

- i. as fontes de dados turísticos mais citadas foram: OMT, IBGE, MTur, FGV, SEBRAE, Observatório de Turismo de São Paulo;
- ii. aproximadamente 77% dos respondentes fazem uso de informações do SITUR para atividade profissional, sendo que cerca de 40% desses o faz porque é muito necessário;
- iii. informações sobre demanda turística são buscadas por aproximadamente 91% dos respondentes que fazem uso de informações do SITUR;
- iv. dentre os aproximadamente 23% que mencionaram não utilizar os dados do SITUR, cerca de 57% justificaram dizendo não encontrar o que precisa ou não encontrar dados relevantes. Dentre esses últimos, as justificativas foram, principalmente, a falta de conteúdo e indicadores de interesse (78%), a desatualização da informação (60%) e a falta de informação na sua granularidade (nível geográfico de região, estado ou município) de interesse (88%);

- v. dentre as respostas válidas (90%), 78% dos entrevistados disseram que a qualidade da informação do SITUR pode melhorar ou que os dados não possuem credibilidade; e
- vi. em relação às dimensões de qualidade, o percentual de respondentes que atribuiu nível regular ou baixo foi de aproximadamente: 74% para a pontualidade; 58% para precisão; 49% para facilidade de interpretação e para acessibilidade; 44% para coerência; 34% para credibilidade; e 28% para relevância.

Percebe-se, portanto, que uma parcela considerável utiliza informações turísticas para fins profissionais por razão de ser muito necessário. Além disso, outra parcela, também relevante, não utiliza as informações principalmente por não ter o que precisa, pela desatualização ou pelo fato de o dado não estar no nível geográfico de interesse. Em termos de qualidade, a avaliação geral não é boa e, quando desagregada, aponta para um maior descontentamento com a pontualidade e a precisão – sendo o aspecto relevância o mais bem avaliado.

Ou seja, os dados da pesquisa realizada corroboram a importância e relevância das informações turísticas e fornecem indícios sobre o tamanho do impacto das problemáticas que cercam a produção tradicional de estatísticas públicas e indicadores socioeconômicos, principalmente em aspectos de: oportunismo temporal, variedade de temas/domínios, profundidade/granularidade e qualidade.

2.5 BIG DATA E REVOLUÇÃO DOS DADOS

O contexto recente de popularização da internet, avanços tecnológicos e consolidação da era digital (SCHMIDT; COHEN, 2013), elementos do fenômeno chamado de nova onda dos TICs⁶, vem possibilitando a criação de uma avalanche de dados diariamente (MILLER, 2010; HELBING et al., 2016).

⁶ Caracterizada pela massificação: dos dispositivos digitais nas casas, ambientes de trabalho e espaços públicos; da computação distribuída e em nuvem; das redes sociais; e da internet das coisas (Kitchin, 2014). Em Jug (2014), o fenômeno associado é referido como “a terceira plataforma”.

Tais dados, ditos dados digitais, derivam dos rastros de eventos da atividade humana e não humana – tanto inconscientes, ao terem suas atividades capturadas de forma implícita ou despercebida, quanto conscientes, ao deliberadamente executarem uma tarefa – capturados por dispositivos digitais inseridos no universo dos sensores, GPSs, câmeras, *scanners* e da internet das coisas, dos telefones móveis, aplicativos e transações eletrônicas, das páginas da internet, dos mecanismos de buscas, das redes sociais virtuais, dos *blogs* e *web fóruns*, entre outros (WEF, 2010; KITCHIN, 2014).

Pesquisas apontam para um crescimento dessa “nova” classe de dados – os digitais – na ordem de 100% ao ano (HELBING et al., 2016) à medida que o “universo digital” cresceu em média 30% ao ano pelos últimos trinta anos – cerca de cinco vezes mais rápido que o crescimento da economia mundial (SAMMAN, 2015). Em 2015, por exemplo, conforme Helbing et al. (2016), o mundo produziu mais dados digitais do que havia produzido em toda história da humanidade até 2014. Em adição, uma pesquisa conduzida pela empresa de tecnologia Cisco, realizada em 2018, concluiu que o volume de tráfego de IPs em redes mundiais de 2022 deverá exceder aquele até então percebido em toda a era da internet até 2016.

A tendência é de aceleração nas taxas de crescimento. Segundo Schmidt e Cohen (2013), se o ritmo atual da inovação tecnológica for mantido, a maioria da população da Terra, estimada em oito bilhões de pessoas, estará online em 2025 – saindo, em uma geração, da quase total falta de acesso a informações para o domínio de toda a informação do mundo através de um aparelho que cabe na palma da mão. Com a continuidade da expansão dos adventos da internet das coisas e da transformação digital, logo a quantidade de dados digitais no mundo dobrará a cada doze horas (HELBING et al., 2016).

Com a recente explosão na produção de dados digitais, a disponibilidade por este tipo de dado atualmente supera a de dados analógicos ou mecânicos – por vezes chamados de “*small data*”, predominantemente escassos e de acesso limitado, produzidos por levantamentos do tipo *survey* e por outros métodos tradicionais de pesquisa (KITCHIN, 2015). A este processo de transformação da

predominância de disponibilidade por “*small data*” do passado para a de dados digitais vem sendo atribuído o nome de Revolução dos Dados (KITCHIN, 2014).

Protagonistas da Revolução dos Dados, produzidos, armazenados e disponibilizados em grande volume, com velocidade e em variedade de domínios e formatos (KITCHIN, 2015), os dados digitais são frequentemente tratados simplesmente por Big Data – como faz UNECE (2013b) ao apresentar e classificar⁷ as possíveis fontes desse tipo de dado.

Esses três “V’s” – volume, velocidade e variedade – formam o núcleo base dos conceitos mais genéricos, e aqui adotados, de Big Data (KITCHIN, 2013). Nesse contexto, Big Data são dados que estão disponibilizados em tempo real, em larga escala, menos estruturados e em diferentes tipos de variáveis (EINAV; LEVIN, 2013). Para Kitchin (2015), dados com características de: vastos; velozes; variados; resolutos; relacionais; de baixo custo; e cada vez mais abertos e acessíveis. Além disso, trata-se de conjuntos de dados cujo tamanho ultrapassa a capacidade das ferramentas tradicionais de bancos de dados para capturar, armazenar, gerenciar e analisar (MANYIKA et al, 2011). Ou seja:

Vastas quantidades de dados digitais dinâmicos e variados, que são facilmente combinados, compartilhados e distribuídos por redes de tecnologia da informação e comunicação, analisados por uma nova geração de métodos e ferramentas de análise de dados desenvolvida para lidar com dados em abundância – em oposição aos métodos e ferramentas antigamente desenvolvidos para lidar com dados em escassez. (KITCHIN, 2014, p. XV)

Assim, esta definição de Big Data não só engloba os ditos dados digitais como pode também contemplar dados de registros administrativos de Estados e organizações – cadastros de indivíduos de uma população, de beneficiários de programas sociais, de usuários de um serviço, de clientes de um produto, etc. (KITCHIN, 2014) – à medida que tais registros estão cada vez mais abundantes, com variedade de conteúdo e sendo gerados de forma dinâmica.

Dentro do universo do Big Data, outro conceito em utilização é o de “*crowdsourcing data*”. Introduzido por Howe (2006) e definido por Schenk e Guittard

⁷ Em taxonomia proposta pela UNECE (2013), as fontes de Big Data podem ser classificadas em: redes sociais (informações de base humana); sistemas tradicionais de organizações (dados mediados por processos); e internet das coisas (dados gerados por máquinas).

(2009), se referem a dados de origem semelhante à do Big Data, extraídos da rede, porém restritos àqueles produzidos intencionalmente pelos próprios indivíduos em um ambiente colaborativo de compartilhamento – e sem imposição relacionada ao grande volume. O boom desse tipo de dado se deu com o advento da segunda geração da *World Wide Web* (Web 2.0), de conceito proposto por O'Reilly (2005). Conforme Costa et al. (2009), nesse contexto uma grande massa de indivíduos passou a ser não apenas consumidora de informações como também produtora, à medida que deixam “rastros” deliberados em dispositivos conectados em rede.

Além desses, o conceito de *open data* – dados abertos – também vem sendo bastante utilizado em paralelo ao advento do Big Data, sendo assim chamados os dados que são tornados técnica e legalmente abertos, isto é, em formatos prontos para leitura computacional e com permissão de livre acesso para uso comercial e não comercial (WB, 2014). Tais tipos de dados têm sido alvo de políticas de estímulo e incentivos a partir de esforços em regulação por parte de governos (JUG, 2014). Porém, essa já não tem sido uma diretriz restrita ao setor público, uma vez que a abertura de dados tem sido colocada também na agenda do setor privado. Wong (2012) traz alguns exemplos de abertura de dados no setor privado e alega que tal movimentação ocorre a partir da compreensão da importância dos negócios em redes, dos modelos de *open source* e da *open innovation*.

De maneira geral, tais os movimentos acontecem pelo entendimento de que os benefícios da disponibilização da maior quantidade possível de dados podem auxiliar tanto o setor público – na formulação e na avaliação de políticas públicas – quanto o setor privado – na gestão das atividades das empresas em busca de produtividade, competitividade e conhecimento de mercados – e a sociedade em geral (englobando o terceiro setor e a academia) – proporcionando aumento na produção de conhecimento, na transparência e nas atividades de controladoria social, solidificando o processo democrático e o entendimento de informação pública enquanto bem público.

2.5.1 Valor para a construção de conhecimento

Antes mesmo de fazer parte do contexto formalizado como Revolução dos Dados, o crescimento na geração de dados digitais já era percebido como tendo características de inovação de ruptura ao desafiar o *status quo* de como os dados são produzidos, geridos, analisados, armazenados e utilizados, impactando na maneira como o conhecimento é produzido, os negócios são conduzidos e a governança é promulgada (CHRISTENSEN, 1997).

Para Mayer-Schönberger e Cukier (2013), este fenômeno tem modificado a maneira como muitas organizações norteiam seus processos de tomadas de decisão. Especialmente para atores, indivíduos e organizações que produzem conhecimento e orientam ações e processos de tomadas de decisão a partir de dados, assumindo um comportamento *data-driven*, a exploração do ambiente de Big Data é possibilidade de usufruir de uma vasta fonte de matéria-prima (ASKITAS; ZIMMERMANN, 2015). O que se acredita é que através dessa exploração seja possível adquirir e descobrir conhecimento relevante, transformando dados em conhecimento através da identificação de padrões compreensíveis que sejam válidos, novos e potencialmente úteis do ponto de vista prático (FAYYAD, 1996).

O potencial do Big Data vem sendo percebido tanto para fins econômicos quanto sociais e científicos uma vez que oferecem a perspectiva de registros sobre os mais diversos aspectos da vida dos indivíduos de uma população, viabilizando a investigação e a construção de conhecimento a respeito de desde aspectos sociais até análises econômicas e de mercado. Há alguns anos, Mayer-Schönberger e Cukier (2013) já diziam que o surgimento do Big Data se tratava de um fenômeno que estaria pronto para “chacoalhar” tudo, dos negócios e das ciências até os sistemas de saúde, os governos, a educação, a economia, as humanidades e todos os outros aspectos da sociedade. Ou seja, abrem-se, portanto, oportunidades de inovação que podem alcançar os mais diversos setores da sociedade, gerando, por vezes, soluções radicais e até disruptivas para processos, produtos, serviços e modelos organizacionais tradicionais.

Em termos econômicos, organizações do ambiente empresarial já perceberam o valor do Big Data, passando a fazer deste uma ferramenta efetiva para superar a concorrência (UNGP, 2012). Entretanto, ao passo que as empresas

estão tendo sucesso demonstrando os benefícios que podem acumular a partir do Big Data, efeitos mais substantivos ou qualitativos – principalmente em termos de práticas sociais e de causas públicas – ainda estão apenas começando a serem percebidos pelo setor público, pela academia e pelo terceiro setor (ASQUER, 2013). O relatório de UN (2014b), por exemplo, coloca a necessidade de os países construírem agendas que mobilizem esforços para a compreensão sobre o real potencial da utilização do produto da Revolução dos Dados em práticas voltadas para o desenvolvimento sustentável.

Em virtude de tais percepções, o Big Data vem sendo tratado como um valioso recurso do século XXI, uma nova classe de ativos, ou como “o novo petróleo” (WEF, 2011; BOSSOI, 2014; ROTELLA, 2012), tamanha sua interessante relação de custo-benefício para processo de construção de conhecimento, principalmente no âmbito empresarial e da iniciativa privada. Em termos de exemplos concretos, Manyika et al. (2011) e CEBR (2012) estimaram o valor adicional potencial do Big Data em diversos setores das economias dos Estados Unidos e do Reino Unido, respectivamente, encontrando cifras que estão na casa de trilhões de dólares ao ano.

Como consequência, o Big Data vem desempenhando papel central em um dito período de *data economy* (OPHER et al., 2016) e de quarta revolução industrial (SCHWAB, 2016) – impulsionados pela transformação digital e pela inteligência artificial.

2.6 BIG DATA NA PRODUÇÃO DE INDICADORES SOCIOECONÔMICOS

Os produtores de estatísticas públicas e/ou oficiais também estão atentos ao fenômeno da Revolução dos Dados e do Big Data, avaliando a dimensão da oportunidade, do ônus e do bônus, da utilização destas novas fontes de dados como insumos para a produção de estatísticas públicas e indicadores socioeconômicos (UN, 2014a; EUROSTAT, 2013; ASQUER, 2013). No Brasil, por exemplo, a inclusão recente do IBGE como *hub* físico da *Global Platform* da ONU, sob gestão do *UN Committee of Experts on Big Data and Data Science for Official Statistics* (UN-

CEBD), é indicativo de que a produção de estatísticas oficiais e públicas que incorpora fontes alternativas de dados e métodos inovadores está sendo avaliada.

Isso não se dá por acaso, mas sim porque muitos dos métodos tradicionais de produção de estatísticas são fortemente concentrados em pesquisas amostrais ou censos via levantamentos em campo com coleta conduzida por pessoas (BRAAKSMA et al., 2014) que, apesar de apresentarem qualidade e validade científica reconhecida, estão sujeitos uma série de contratempos e dificuldades práticas, conforme apresentado na Subseção 2.3.3 – principalmente de ordem financeira, temporal e operacional (KITCHIN, 2015; HACKL, 2016; DAAS; PUTS, 2014; CITRO, 2014).

Destacando que, conforme também mencionado na Subseção 2.3.3, diversas alternativas de tratamento a tais dificuldades têm sido utilizadas pelos produtores tradicionais – especialmente envolvendo a modernização dos processos, a utilização de dados de registros administrativos e cadastros públicos e a implementação de novos modos de coleta das informações. Entretanto, as alternativas em geral buscam viabilizar os métodos tradicionais de produção através de melhorias e inovações normalmente incrementais nos mesmos.

A utilização de Big Data, por sua vez, poderia atuar para contribuir, pelo menos em alguma medida, com o contorno a diversas dificuldades tradicionais ao mesmo tempo, através da modificação ou de uma complementariedade na maneira como os processos são conduzidos e os produtos estatísticos são construídos, utilizando-se de insumos que derivam de fontes completamente novas.

Surge o questionamento, porém, sobre a exequibilidade e a prosperidade de uma quebra de paradigma no *status quo* da produção de estatísticas por este meio. Será mesmo que a incorporação das novas fontes de dados nos modelos de produção dos agentes produtores pode representar uma inovação que traria economia de tempo e dinheiro, viabilizando produtos mais ágeis e responsivos ao contexto social evolutivo vigente e que possuam, ao mesmo tempo, garantias do ponto de vista da qualidade e confiabilidade?

Para alguns autores, sim. Kitchin (2015) é categórico ao reforçar a necessidade de uma significativa inovação disruptiva na produção de estatísticas públicas a partir do surgimento de uma série de oportunidades. Para Citro (2014), trata-se de obrigação mover-se do paradigma de produzir as melhores estimativas

possíveis a partir de levantamentos tradicionais de campo – que cada vez mais se apresentam inadequadas para atender às expectativas do novo contexto de demandas (PINK; BOROWIK; LEE, 2009) e são estranguladas por questões de custos e diminuição nas taxas de respostas (BRAAKSMA et al., 2014) – para o da produção que se aproveita de múltiplas fontes, beneficiando-se do contexto da Revolução dos Dados, visando a geração das melhores estimativas possíveis e que atendam às demandas dos usuários – resultados relevantes, oportunos, acurados e economicamente viáveis.

Essa ideia de possível quebra de paradigma é um tema de crucial importância para produtores de estatísticas públicas, uma vez que poderiam estender a capacidade de produção de informação estatística para outras esferas, reduzindo a participação dos atuais tradicionais e principais agentes: os institutos oficiais de estatística (DEMUNTER, 2017; STRUIJS; DE BROE, 2018; MACFEELY, 2016). Para Demunter (2017), trata-se de uma questão de sobrevivência para os produtores tradicionais. Com o Big Data, novos *players* estariam em condições de gerar indicadores socioeconômicos com baixo custo, em alto nível de detalhamento e oportunismo temporal, modificando a estrutura desse setor até então monopolizado.

Sobre esse aspecto, cabe ressaltar que estatísticas públicas objetivas e independentes são bens públicos essenciais em sociedades democráticas e, especialmente no âmbito da gestão pública, não podem ser vistas como bens mercadológicos em virtude do seu caráter institucional.

Dessa forma, é pouco provável que os produtos da Revolução dos Dados venham a substituir integralmente as fontes de dados atuais dos produtores de estatísticas públicas – até porque também estão sujeitos a desafios e complicadores, elencados na Subseção 2.6.3 –, mas é possível que exerçam papel central como suplemento para os dados já existentes (HACKL, 2016) no âmbito de uma integração que poderá maximizar valor (HARWOOD; MAYER, 2016; DE WAAL; DELDEN; SCHOLTUS, 2020).

2.6.1 Oportunidades

O aproveitamento das oportunidades trazidas pela utilização de Big Data como matéria-prima para a produção estatística, na tentativa de contornar diversas das dificuldades dos processos tradicionais (Subseção 2.3.3), pode se dar, segundo Florescu et al. (2014), em diferentes medidas, que dependem se inovação a ser conduzida no processo tradicional tem por objetivo:

- a) substituir total ou parcialmente as práticas tradicionais com interesse de geração de estatísticas equivalentes, com mesma perspectiva e utilidade (contexto de manutenção do propósito original);
- b) trazer uma nova perspectiva de análise para os domínios já trabalhados (contexto de adicionar ou complementar aos processos tradicionais);
- c) contribuir como informações auxiliares aos processos tradicionais (contexto de melhoramento das práticas existentes); ou
- d) alterar tanto a perspectiva de abordagem e análise quanto as fontes e os métodos de produção (contexto de propor uma alternativa ao *status quo* de análise de certo domínio).

Cabe destacar que o objetivo do processo de inovação não se define apenas por interesse do produtor, mas também por uma análise da qualidade dos insumos de Big Data – que podem apresentar adversidades, conforme apresentado na Subseção 2.6.3. Ou seja, se define também por uma análise que permitirá dizer o quão satisfatoriamente os indicadores inovadores serão capazes de substituir os tradicionais.

Independente do objetivo da inovação e dos possíveis entraves da utilização de Big Data para a produção de estatísticas públicas e indicadores socioeconômicos, uma série de oportunidades já foram registradas na literatura. Em uma perspectiva genérica, partindo dos possíveis benefícios mencionados por Kitchin (2015) e considerando também as contribuições de Tam e Clarke (2014), Florescu et al. (2014), Eurostat (2014a), UNGP (2012), Hand (2015), Braaksma e Zeelenberg (2015) e Hackl (2016), foi possível construir o resumo apresentado no Quadro 6. As oportunidades estão relacionadas com os tipos de dificuldades dos produtores – organizados na Figura 2 da Subseção 2.3.3 – que podem auxiliar a resolver.

Quadro 6 – Possíveis oportunidades trazidas pelo Big Data para a produção de estatísticas públicas e sua relação com as dificuldades da Figura 2.

Dificuldades	Tipo	Oportunidade
Operacionais	Todos	Reduzir ou eliminar a necessidade de conduzir levantamentos de campo.
Recursos	Custo financeiro	Reduzir custos dos processos de produção, hoje fortemente baseados em coletas em campo, contornando problemáticas de orçamento.
	Tempo	Reduzir tempo dos processos de produção tradicionais.
Metodológicas	Erros de cobertura	Alcançar populações completas em algumas situações específicas.
Demanda	Relevância	Captar informações mais aderentes ao contexto social evolutivo uma vez que podem se referir a reflexos de transações, interações e comportamentos que os indivíduos de fato executam – e não simplesmente dizem, como em uma pesquisa tradicional.
	Variedade	Gerar resultados para novas áreas/temas de investigação, atualmente não trabalhadas.
	Profundidade	Gerar resultados mais detalhados e com aberturas para pequenas localidades e subgrupos populacionais específicos com o aumento da quantidade de informações disponíveis.
	Oportunismo	Produzir estatísticas com maior frequência ou de forma contínua em tempo real, agregando valor pelo oportunismo e utilidade prática decisória (ações preventivas e de monitoramento rápido).
	Qualidade	Gerar estimativas mais precisas (menor variância) ao contornar a queda da taxa de aproveitamento das pesquisas tradicionais e alcançar uma quantidade maior de informações.
	Economia	Otimizar processos de trabalho, gerando ganhos de eficiência percebidos pelo usuário.
	Comparabilidade	Facilitar comparações entre regiões e países, permitindo que sejam construídas para quaisquer países, independente da sua estrutura.
Estratégicas	Concorrência	Desenvolver produtos de dados mais ricos a partir do <i>link</i> entre diversas bases de dados e/ou de perspectivas longitudinais.
		Liberar o quadro funcional para tarefas de maior valor agregado.
		Promover maior visibilidade e utilização de estatísticas produzidas.
		Explorar novos mercados como novas estatísticas e domínios.
		Complementar, substituir, melhorar e adicionar às bases de dados já existentes.
	Isolamento	Promover a aproximação, a colaboração e a co-criação com outros <i>players</i> da indústria de dados, produtores de dados digitais e especialistas das áreas das ciências da computação e de dados.
	Obsolência	Explorar novas técnicas de análises de dados para obtenção de <i>insights</i> diferenciados.
	Credibilidade	Servir de dados auxiliares para aumentar a qualidade de processos tradicionais de produção (para variáveis de estratificação, imputação de valores faltantes, tratamento de anomalias).
Reforçar a qualidade (validade e consistência) e a credibilidade de estatísticas tradicionais, funcionando como um contracheque.		

Fonte: atualizado de Caumo (2020).

Nesses termos, a literatura investigada sugere, conforme apresentado no Quadro 6, que o Big Data poderia contribuir, em alguma medida e nem sempre concomitantemente, para o tratamento de todas as dificuldades que produtores de estatísticas públicas enfrentam.

Tais oportunidades foram identificadas a partir das possibilidades de utilização de Big Data advindos das fontes apresentadas no Quadro 7, construído a partir Kitchin (2015) e ampliado com contribuições dos autores relacionados no corpo do quadro. Pelas aplicações encontradas, percebe-se que as fontes alternativas de dados oferecem extensa oportunidade de produção de conhecimento sobre os mais diversos domínios, com replicação imaginável para outros aqui não relacionados.

Cabe adicionar aqui o exemplo recente da utilização de Big Data para a construção de conhecimento sob a forma de indicadores socioeconômicos durante a pandemia de Covid-19 no Brasil. Em um contexto de extrema adversidade, com impossibilidade de condução dos levantamentos de campo, bases dos processos tradicionais de produção estatística, dados de telefonia móvel foram utilizados para que se pudesse gerar indicadores relativos ao nível de isolamento da população, subsidiando processos de tomadas de decisão de gestores públicos para fins de estruturação de políticas. Em outra aplicação prática, indicadores de inflação construídos pelo IBGE no âmbito do Sistema Nacional de Índices de Preços ao Consumidor (SNIPC) deixaram de ter uma parcela da coleta de preços realizada de forma presencial – substituída por uma coleta na *web*.

Quadro 7 – Fontes de Big Data por tipo e áreas de aplicação já trabalhadas nas publicações investigadas.

Tipo	Fonte de dados	Publicações relacionadas	Áreas de aplicação
Comunicação móvel	Dados de telefones móveis (registro de ligações, mensagens e posicionamento)	Ahas (2008), Ahas et al. (2010), Berzina e Lauberte (2018), Blumenstock et al. (2018), Coudin e Poulhes (2019), Dattilo et al. (2016), de Jonge, Van Pelt e Roos (2012), Eurostat (2014a), Heerschap et al. (2014), Karlberg e Skaliotis (2013), Kroon (2019), Lestari (2019), Nurmi e Piela (2019), Pramana et al. (2017), Setiadi e Uluwiyah (2017), Tennekes e Offermans (2014), Tiru (2014), SNZ (2012) e UN (2014a).	Turismo, população, migração, mobilidade, tráfego, redes de relacionamento, pobreza, saúde, segregação social.
Internet e crowdsourcing	Redes sociais virtuais (Facebook, Twitter, Instagram, LinkedIn, Tumblr, etc.)	Chew e Eysenbach (2010), Claster, Ghotbi e Shanmuganathan (2015), Daas e Puts (2014), Daas et al. (2012), Hu et al. (2018), Java (2007), Leyva e Coronado (2019), Nozawa et al. (2017), Menchavez e Espinosa (2015), Provenzano, Baggio e Hawelka (2018) e UN (2014a).	Comportamento, confiança do consumidor, consumo, economia, emprego, ensino superior, força de trabalho, humor, inflação, informações sobre empresas, imóveis, migração, mobilidade, PIB, população, potencial de mercado, preços, qualidade dos serviços urbanos, renda, sociedade da informação, opinião, redes de relacionamento, saúde, sentimento, turismo, urbanismo, uso da terra.
	Comunidades de compartilhamentos de imagens e vídeos (YouTube, Flickr, Picasa, Panoramio, etc.)	ESSC (2014), Kitayama (2016) e Önder, Koerbitz e Hubmann (2014).	
	Blogs, <i>web forums</i> e <i>wikis</i>	Alis et al. (2015), Donovan, Flaherty e Healy (2017), Liu e Shi (2019), Shi, Zhao e Chen (2017) e Yang et al. (2017).	
	Buscas na <i>web</i>	Ginsberg et al. (2008), Nooraeni, Purba e Yudho (2019), Rai (2018), Samanta (2019) e Valdivia et al. (2010).	
	Logs de uso da <i>web</i> e de sistemas computacionais	EC (2012) e Ruslan, Tetiana e Kateryna (2019).	
	<i>Websites</i> de acomodação, hotelaria e passagens aéreas, <i>e-commerce</i> , shoppings, mercados, empregos, imóveis, organizações públicas e privadas, etc.	Barcaroli et al. (2016), Cavallo e Rigobon (2016), Cavallo (2013), deHaan e Hendriks (2013), Dufour et al. (2019), Feng et al. (2013), Groves (2011), Guo et al. (2016b), Heerschap et al. (2014), Hoekstra et al. (2012), Ismail et al. (2019), Khee, Tan e Hoong (2019), Oliveira e Baracho (2018), Polidoro et al. (2015), Silva e Miranda (2019), Summa (2014) e Virgillito (2014).	
	Websites de Informação Geográfica Voluntária	Dufour et al. (2019), ESSC (2014), Stubbings e Peskett (2018).	
Aplicativos e sistemas proprietários online	Dufour et al. (2019) e Park et al. (2018).		
Sensores e Internet das coisas	Meteorologia e poluição (científicos)	Broe, Snijkers e Schouten (2019), López-Gonzales et al. (2019), Mokhtar et al (2019) e UN (2014a).	Agricultura, consumo, cultivo de drogas (crime), economia, estilo de vida, emissão de poluentes, habitação, meteorologia, migração, mobilidade, pesca, silvicultura, tráfego, transportes, turismo, uso da terra.
	Dispositivos digitais pessoais móveis e <i>wearables</i>	Broe, Snijkers e Schouten (2019).	
	Beacons	Guarda et al. (2018).	
	Rastreamento veicular	UN (2014a).	
	GPS e Wi-Fi	Ichifuji et al. (2016) e Koide et al. (2016).	
	Trânsito	Daas et al. (2012), Fornaro e Luomaranta (2019) e van Ruth (2014).	
	Medidores inteligentes (energia, água, etc.)	AlQadhi et al. (2019), Dummee e Trongthamakit (2019) e ESSC (2014).	
	Imagens de satélite	Brezzi e Sanchez-Serra (2014), van Donkelaar et al. (2016), Dufour et al. (2019), Figueroa e Mayorga (2015), Schnorr-Bäcker (2016), Seid (2019), Sharnoby (2019), UN (2017) e UN (2014a).	
	Identificação automática de embarcações	ESSC (2014) e Nolan et al. (2019).	
Movimentação de voos	ESSC (2014).		
Registros e scanners	Transações comerciais, do varejo, etc.	Dufour et al. (2019), Huang (2019), Kasperuniene e Limpach (2019), Silva e Miranda (2019) e UN (2014a).	Agricultura, comportamento, consumo, demografia, educação, estilo de vida, economia, finanças, inflação, mercado de trabalho, turismo, preços, saúde, vendas.
	Cartões de crédito	Camachoa e Doménech (2019), ESSC (2014), Garcia et al. (2018) e Kroon (2019).	
	Transações e estoques financeiros e bancários	Buckmann, et al. (2019), Gravesteijn et al. (2019), Mydin et al. (2019) e Salemink et al. (2019).	
	Administrativos (médicos, impostos, de programas sociais, empresariais, etc.)	Abidin, Khairi e Ling (2019), Ahmad et al. (2019), Beahan (2019), Frolova (2019), Humbelin e Farys (2019), Saint-Pierre (2019) e Salemink et al. (2019).	

Fonte: atualizado de Caumo (2020).

2.6.2 Possibilidades para o turismo

No âmbito do turismo, importante setor para a atividade econômica de Florianópolis, a utilização de Big Data vem se consolidando ao redor do mundo a partir da percepção de que tais dados podem se tratar de importante matéria prima para suportar processos de geração de conhecimento e de tomadas de decisão em destinos turísticos. Demunter (2017), por exemplo, corrobora que o turismo está na linha de frente das inovações associadas ao Big Data no que concerne a utilização de novas fontes e métodos. No Brasil, o contexto traz oportunidades para que o setor promova inovações e contorne, pelo menos em alguma medida, a insatisfação manifesta por usuários em relação aos atuais produtos estatísticos – Subseção 2.4.1.

À título de ilustração referente a algumas experiências de utilização em andamento, destaca-se o caso do Observatório de Turismo do Estado do Espírito Santo que, à exemplo de outras experiências antes realizadas na Espanha (ROCASALVATELLA, 2014), na Argentina (BUENOS AIRES, 2020) e em outros países pelo mundo, fechou parceria com empresas telefônicas para fins de obtenção e análise de dados relativos aos sinais de telefonia móvel com intuito de gerar conhecimento turístico (BRASIL, 2017) – acerca de características, ações e comportamentos de turistas e excursionistas. Com tais dados, o Observatório foi capaz de perceber a origem, o destino, o perfil sociodemográfico, o tipo/propósito do turismo, o tempo de estadia, os fluxos e locais mais visitados, entre outros. Em Minas Gerais, percebe-se a experiência do Observatório do turismo de Minas Gerais. A organização captura dados na *World Wide Web*, mais especificamente no portal TripAdvisor, para construção de estatísticas relativas, dentre outros aspectos, ao quanto e ao quão bem os turistas e visitantes avaliam as cidades e os pontos turísticos do estado (OLIVEIRA, 2017).

Em termos globais, percebe-se a utilização de uma série de fontes de dados do tipo Big Data com fins de geração de conhecimento turístico sob a forma de indicadores em diversos contextos de aplicação, conforme apresenta o Quadro 8.

Quadro 8 – Fontes e aplicações de Big Data para indicadores e estatísticas de turismo.

Fonte de dados	Aplicação	Publicação
Múltiplas fontes	Dados de posicionamento de telefones móveis, registros de pagamentos eletrônicos e preços e produtos turísticos ofertados em páginas da <i>web</i> para quantidade, tempo de estadia, gastos e locais visitados por turistas.	García et al. (2018)
	Expedia e TripAdvisor para avaliação e classificação de hotéis.	Leal et al. (2016)
	Baidu e plataforma de turismo inteligente baseada em dados de telefonia para quantidade de turistas que entram.	Liu et al. (2019)
	TripAdvisor e Yelp para probabilidade de um <i>review</i> sobre hotel ser útil para outras pessoas.	Ma et al. (2018)
	Google e Facebook para opinião sobre os centros de saúde utilizados pelo turismo médico.	Rai (2018)
Beacons	Tendências e interesses dos turistas.	Guarda et al. (2018)
ChinaZ.com	Estrutura da rede das 120 principais páginas Chinesas de informação turística da <i>web</i> .	Feng, Li e Zhang (2013)
Dados de posicionamento de telefones móveis	Impacto econômico do turismo.	Berzina e Lauberte (2018)
	Quantidade de visitas ao país.	Pramana et al. (2017)
	Quantidade de turistas estrangeiros e suas movimentações.	Setiadi e Uluwiyah (2017)
Flickr	Avaliação, reconhecimento e classificação de pontos turísticos (foco na identificação de “Anabas”).	Kitayama (2016)
	Quantidade e distribuição espacial de turistas (demanda).	Önder, Koerbitz e Hubmann (2014)
TripAdvisor	Satisfação, perfil dos turistas e períodos relevantes de visitação de municípios e atrações em Minas Gerais.	Oliveira e Baracho (2018)
	Satisfação e opinião dos turistas.	Guo et al. (2016b)
Twitter	Motivações, percepções, preocupações e sentimentos daqueles que já realizaram e dos que consideram realizar turismo médico.	Claster, Ghotbi e Shanmuganathan (2015)
	Padrões de movimentações (rotas entre atrações) dos turistas.	Hu et al. (2018)
	Sentimento/polaridade de manifestações de turistas.	Menchavez e Espinosa (2015) e Shimada et al. (2011)
	Comportamento e interesses dos turistas.	Nozawa et al. (2017)
	Redes de fluxos turísticos entre países.	Provenzano, Baggio e Hawelka (2018)
Sina Weibo (China)	Quantidade de chegadas e origem dos turistas.	Liu e Shi (2019)
	Relação entre a popularidade de atrações turísticas e as multidões que de fato atraem.	Shi, Zhao e Chen (2017)
	Quantidade de turistas estrangeiros.	Yang et al. (2017)
Sistema proprietário de reservas online	Comportamento e atributos do turista que retorna ao hotel.	Park et al. (2018)
Wi-fi logs	Movimentações dos turistas.	Ichifuji et al. (2016)
	Quantidade de turistas em tempo real.	Koide et al. (2016)
Wikipedia	Quantidade de turistas que saem do Reino Unido por mês.	Alis et al. (2015)
	Quantidade de turistas.	Donovan, Flaherty e Healy (2017)

Fonte: atualizado de Caumo, Biz e Souza (2020).

Dentre todas as fontes percebidas, a maioria está presente na *World Wide Web* (páginas, redes sociais, *blogs*, mecanismos de busca, entre outros). As demais fontes de Big Data percebidas, compreendidas por sensores e internet das coisas (*beacons* e pontos de *wi-fi*), *scanners* (registros de pagamentos eletrônicos), *smartphones* (dados de posicionamento e registros de ligações) e outros dispositivos digitais não classificados como *WWW*, se fizeram presentes em 29% das aplicações práticas observadas. Desta parcela, metade envolve a utilização de dados oriundos de serviços de telefonia.

No âmbito do universo de fontes presentes na *WWW*, chama atenção o não aproveitamento da rede social Instagram nas aplicações práticas. A hipótese aqui especulada é de que a política de utilização de dados fortemente restritiva do Instagram impeça uma série de coletas e acabe por afastar interessados. A rede Facebook, por sua vez, apareceu como fonte em apenas uma publicação, sugerindo que fenômeno semelhante possa estar no cerne da justificativa. Em sentido oposto, a rede Twitter foi a fonte presente na *WWW* com maior quantidade de aparições nas pesquisas investigadas.

Em termos de finalidade das aplicações práticas, ou seja, dos possíveis domínios de cobertura dos indicadores de turismo subsequentes, as mais frequentes são aquelas que envolvem a estimação ou a projeção da quantidade ou demanda de turistas de entrada ou saída e aquelas que trabalham com satisfação, opinião, sentimentos e interesses manifestos. Ambas aparecem em aproximadamente 36% das publicações. A totalidade das demais finalidades de aplicação encontradas no levantamento estão apresentadas no Quadro 9.

Quadro 9 – Percentual de publicações classificadas nos tipos de finalidade da aplicação.

Finalidade	Publicações
Quantidade e/ou distribuição espacial dos turistas	35,7 %
Satisfação, opiniões, sentimentos e interesses manifestos	35,7 %
Fluxos turísticos, movimentações, atrações e locais visitados, tempo de estadia e gastos	21,4 %
Perfil, atributos e origem dos turistas	10,7 %
Impacto econômico	3,6 %
Outros	10,7 %

Fonte: retirado de Caumo, Biz e Souza (2020).

Outras publicações analisadas não propuseram ou relataram experiências próprias de utilização de Big Data para a construção de indicadores e estatísticas de turismo, mas citaram finalidades e aplicações realizadas por terceiros. É o caso das revisões de literatura cujos objetivos estão se encontram apresentados no Quadro 10.

Quadro 10 – Revisões de literatura de temas correlatos ao proposto por esta seção.

Publicação	Objetivo
Centobelli e Ndou (2019)	Análise de 109 artigos sobre Big Data no turismo, com identificação dos possíveis impactos gerados no setor, funcionalidades, tecnologias e metodologias.
Li et al. (2018)	Análise de 165 artigos para identificação de fontes e aplicações de Big Data no turismo, destacando tipos de dados e técnicas analíticas.
Liang et al. (2017)	Análise de 57 artigos com foco na percepção de Tecnologias da Informação e comunicação que vêm sendo adotadas no setor turístico Chinês.
Mariani et al. (2018)	Análise de 96 artigos sobre Big Data no turismo para identificação aplicações práticas com fontes e tipos de dados e métodos de coleta, análise e visualização utilizados.
Shoval e Ahas (2016)	Análise de 45 artigos com foco na percepção de tecnologias utilizadas para rastreamento de turistas.
Theilwall (2019)	Revisão das principais abordagens de análise de sentimentos com foco em descrições práticas de como os métodos funcionam e podem ser aplicados no turismo.

Fonte: retirado de Caumo, Biz e Souza (2020).

Além das aplicações e finalidades apresentadas nos Quadros 8 e 9 e nas revisões de literatura apresentadas no Quadro 10, cabe destacar que outras duas utilizações de Big Data para o setor turístico foram amplamente percebidas durante a análise aqui realizada: as construções de sistemas customizados de informação e de recomendação turística (BIN et al., 2018; BOULAALAM et al., 2018; CHEN; CHEN, 2017; DEZFOULI; SHAHRAKI; ZAMANI, 2018; FAN; DU; ZHOU; 2015; GUO et al., 2017a; GUO et al., 2016a; GUO et al., 2017b; HAN et al., 2018; HAWELKA et al., 2013; HUSAIN; DIH, 2012; LEAL et al., 2018; LUZ et al., 2013; RONG, 2016; SHEN et al., 2016; XIONG et al., 2017; YIFAN et al., 2016; YOUNG; VERHULST, 2017; ZHANG; PAN; XU, 2016; ZHANG et al., 2016). Em geral, os autores utilizam-se das fontes de Big Data para o desenvolvimento ou para a melhoria nos resultados e no desempenho de plataformas que: compilam, qualificam e disponibilizam informações e ofertas turísticas; ou descrevem e sugerem atrações, rotas e pontos de interesse.

2.6.3 Desafios

Apesar do atraente contexto de oportunidades, percebe-se, através da literatura analisada, que muitos desafios ainda estão colocados quando da utilização de Big Data para a geração de estatísticas públicas e indicadores socioeconômicos.

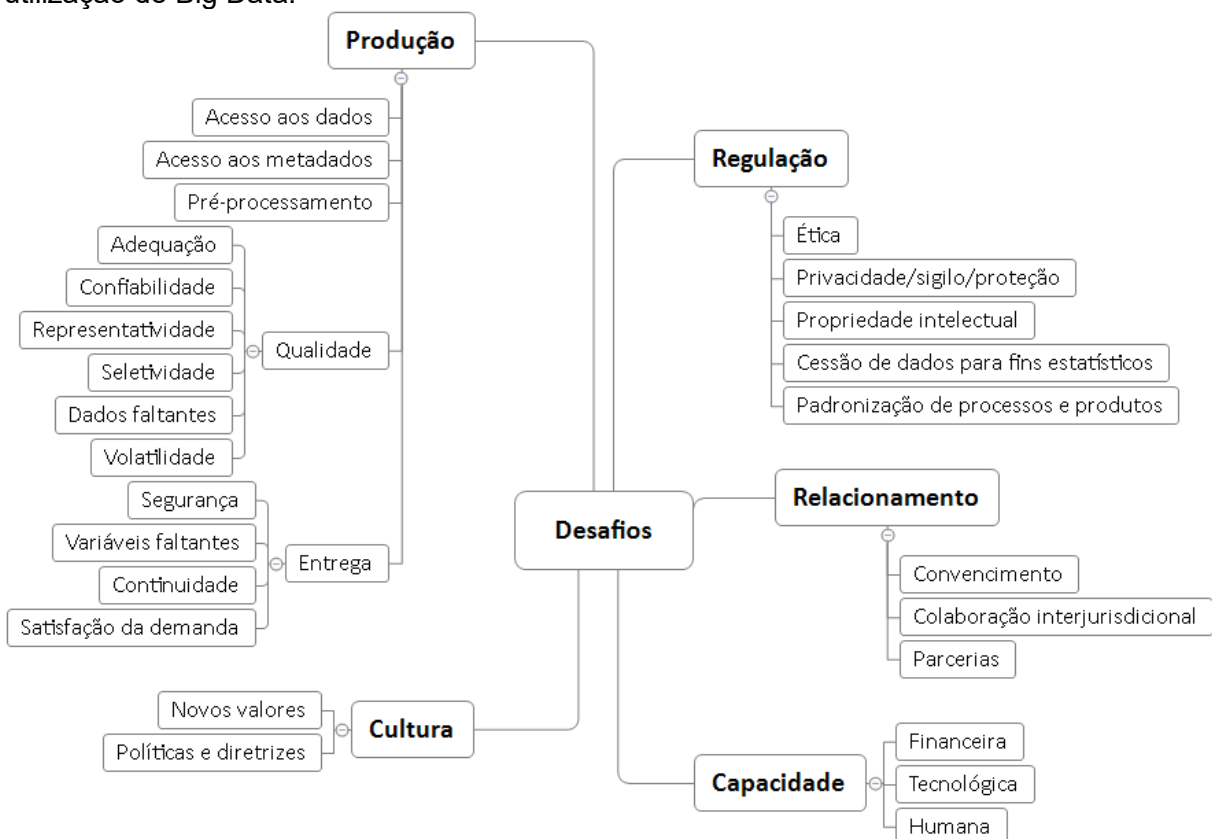
Tais desafios derivam do fato de que o Big Data é geralmente produto de atividades conduzidas por instituições privadas para seus próprios interesses, sem a intenção de que sigam padrões que permitam sua validade para a utilização como insumos durante a construção de estatísticas públicas (KITCHIN, 2015). Assim, grande parte das problemáticas passa pela mensuração da qualidade dos dados nos moldes tradicionalmente conhecidos (UN, 2014a; REIMSBACH-KOUNATZE, 2015), conforme descrito na Subseção 2.3.2, uma vez que os dados digitais podem se apresentar cheios de ruídos, com dados estimulados, falsos e de cobertura parcial e seletiva da população (DAAS et al., 2013; KITCHIN, 2014) – fatos que reforçam, mais uma vez, o argumento que levanta a possível incapacidade dos dados digitais substituírem por completo os insumos utilizados na produção estatística tradicional.

Autores como Struijs, Braaksma e Daas (2014), Tennekes e Offermans (2014) e Vaccari (2014) estabelecem uma classificação que distingue os desafios em seis grandes tipologias: de legislação; de privacidade; financeiros; gerenciais; tecnológicos; e metodológicos. Conforme Vaccari (2014), em termos de legislação, a problemática se refere à necessidade de abertura da possibilidade de acesso às bases que são produto da Revolução dos Dados – muitas vezes privadas. Sobre privacidade, as questões são relacionadas às garantias de que dados disponibilizados serão protegidos em termos de confidencialidade. Em termos financeiros, o custo pelo acesso aos dados – que por vezes serão buscados no setor privado – pode surpreender em certas ocasiões. Do ponto de vista da gestão, a utilização de fontes derivadas da Revolução dos Dados em instituições com ambientes de produção tradicional deverá demandar esforços para novas práticas e modelos. Tecnicamente, os desafios estão na avaliação da capacidade computacional e da mudança necessária de paradigma requerido em termos de tecnologia da informação. E metodologicamente, por fim, as questões se referem ao impacto das novas fontes sobre os métodos tradicionais de coleta, processamento e

disseminação dos dados, com uma provável e consequente adaptação e desenvolvimento de novos métodos de aferição da qualidade e de validação estatística.

Aqui, entretanto, em virtude da percepção de questões adicionais, optou-se por agrupar os desafios de outra maneira, conforme apresenta a Figura 3. Foram considerados como grandes grupos os aspectos: (1) regulação e legislação; (2) cultura organizacional; (3) relacionamento externo; (4) processo de produção de estatísticas; e (5) capacidades organizacionais.

Figura 3 – Resumo de desafios para produtores de estatísticas públicas no contexto da utilização de Big Data.



Fonte: atualizada de Caumo (2020).

Avanços no âmbito de (1) regulação e legislação são importantes pilares que podem consolidar e escalonar o processo de inovação na produção de estatísticas públicas. Nesse contexto, se faz necessária a formalização de normas que tratem sobre:

- a) questões éticas no que diz respeito à definição do limite entre bens públicos e privados, confidencialidade, transparência, igualdade de

acesso e obrigatoriedade de divulgação de informações relevantes sobre o uso e a qualidade dos produtos (WIRTHMANN; REIS, 2018; TAM; KIM, 2018; NASEM, 2017a);

- b) direito à privacidade, ao sigilo, à proteção contra intrusão, à necessidade de concordância e à clareza sobre a finalidade de uso de dados autorizados e não autorizados pelos indivíduos fontes, mesmo quando não estão conscientes dos rastros que deixam e dados que geram, muitas vezes integrados, *linkados* ou cruzados com outras bases para fins diferentes do preconizado originalmente (VALE, 2015; STRUIJS; DAAS, 2013, DAAS et al., 2015; DE JONGE; VAN PELT; ROOS, 2012; HACKL, 2016; TAM; CLARKE, 2015a);
- c) respeito à propriedade intelectual (KITCHIN, 2015);
- d) obrigatoriedade de certas organizações em disponibilizar bases de dados digitais para que produtores licenciados possam gerar produtos com fins de estatísticas públicas (VACCARI, 2014; HACKL, 2016);
- e) requisitos de padronização para os produtos gerados assim como de métodos e técnicas para os processos de produção, garantindo rigor e qualidade metodológica, viabilizando comparações temporais e entre diferentes países/regiões (KITCHIN, 2015).

Sobre o item b), destaca-se, no Brasil, o lançamento da Lei Geral de Proteção aos Dados Pessoais (LGPD), em vigência desde 2020, descendente direta do Regulamento Geral sobre Proteção de Dados (GDPR) da União Europeia, construída à luz do Marco Civil da Internet e do Código de Defesa do Consumidor (CDC). A LGPD se propõe a regulamentar o tratamento que é dado às informações de pessoas colhidas por parte de empresas, especialmente na internet, via formulários, abordando aspectos desde a coleta até a classificação, o processamento, o armazenamento, e principalmente a utilização e a transferência de tais dados. Entre outras disposições, proíbe qualquer empresa de transmitir esses dados sem consentimento expresso dos titulares.

Estruturado o aspecto normativo, produtores de estatísticas públicas vão precisar remodelar seus processos internos de gestão. Em termos de (2) cultura organizacional, necessitarão direcionar esforços para:

- a) trabalhar a mudança de visão estratégica, focando na disseminação contínua dos novos valores (KITCHIN, 2015);
- b) estabelecer políticas e diretrizes alinhadas com o novo contexto através de um processo de gestão que prepare a organização para assumir novas tarefas e responsabilidades (KITCHIN, 2015).

Ainda sobre a gestão institucional, porém agora considerando o (3) relacionamento com o ambiente externo, surge a necessidade de os produtores atuarem no sentido de:

- a) convencer detentores sobre a importância das estatísticas públicas (VACCARI, 2014) e a consequente necessidade de liberação das bases de dados, nem sempre respeitada mesmo quando prevista legalmente (ESSNET, 2013), estabelecendo uma relação de confiança a partir da garantia de que os dados serão mantidos em sigilo e em segurança (TAM; CLARKE, 2015a);
- b) estabelecer colaborações interjurisdicionais com outros produtores, no âmbito da *open innovation*, que permitam o enfrentamento aos desafios que se colocam, o avanço conjunto e a homogeneização de processos de produção e produtos (VALE, 2015; KITCHIN, 2015; TAM; CLARKE, 2015a);
- c) formalizar parcerias com detentores de Big Data, especialmente com o oferecimento de produtos de interesse mútuo (KRATKE; BYIERS, 2014; HACKL, 2016; VALE, 2015).

Para que a produção de estatísticas públicas alcance produtos que justifiquem a inovação, produtores vão precisar contornar dificuldades metodológico-operacionais que poderão surgir. Estas correspondem, talvez, ao mais preocupante entrave relacionado à utilização de Big Data, principalmente no quesito qualidade dos insumos – que podem apresentar anomalias inerentes e não passíveis de correção. Caso partam de dados questionáveis e não consigam ser gerados com qualidade, indicadores socioeconômicos produzidos a partir de Big Data serão possivelmente descartados. Dessa forma, os produtos estatísticos inovadores não serão utilizados na prática das tomadas de decisão, orientando a realização de tarefas e ações em busca por resultados concretos. Ou seja, processos de transformação de dados e informações em conhecimento deixarão de acontecer.

Nesse contexto, já se pode perceber algumas problemáticas que podem vir a afetar os (4) processos de produção e, conseqüentemente, os produtos estatísticos gerados:

- a) restrição de acesso às bases de dados, matéria prima para os produtos finais, de posse predominantemente do setor privado (TAM; CLARKE, 2015a);
- b) restrição de acesso a aspectos metodológicos e metadados que auxiliam no entendimento da estrutura e do conteúdo das bases de Big Data (KITCHIN, 2015);
- c) necessidade de mais tempo de pré-processamento das bases – compreensão e organização da estrutura, conteúdo, padrões e da qualidade dos dados (detecção de erros, anomalias e valores faltantes, etc.) – até que possam ser utilizadas (HASSANI; SAPORTA; SIRIMAL, 2014; TENNEKES; DE JONGE; ZIKOPOULOS et al., 2012; FRY, 2008; DAAS, 2012; LIU; JIANG; HEER, 2013);
- d) dificuldade no atendimento à qualidade de seus produtos (VALE, 2015; HACKL, 2016). Avaliando a viabilidade metodológica do novo processo, certificando veracidade (acurácia, fidelidade), considerando incerteza, erro (cobertura, seleção, mensuração e resposta), viés, confiança e calibração (KITCHIN, 2015; COUPER, 2013), uma vez que as bases de dados digitais podem não ser confiáveis por conta da presença de ruídos, dados estimulados, falsos ou incompletos (DAAS; PUTS, 2014), aspectos na maioria das vezes desconhecidos (UN, 2014a; REIMSBACH-KOUNATZE, 2015).
- e) dificuldade em garantir produtos que transmitam confiança (BRAAKSMA; ZEELENBERG, 2015), com relevância para o usuário através da garantia de que estejam mensurando e refletindo de maneira adequada os fenômenos aderentes às questões de pesquisa (KITCHIN, 2015; CITRO, 2014);
- f) atenção com a representatividade dos resultados para a população alvo de interesse, uma vez que dados digitais não tiveram sua produção

- planejada e controlada como nos métodos tradicionais (HACKL, 2016; CITRO, 2014; TUFECKI, 2013; DAAS et al., 2013; TAM; CLARKE, 2014);
- g) possível seletividade nas fontes de dados – que acaba enviesar resultados ou por fazer com que representem apenas uma parcela da população alvo de interesse (BRAAKSMA; ZEELENBERG, 2015; DAAS et al., 2015; ASA, 2014; BUELENS et al., 2012; FLEKOVA; GUREVYCH, 2013; HASTIE; TIBSHIRANI; FRIEDMAN, 2009);
- h) dados faltantes, especialmente em virtude de problemas com a rede ou com o servidor (DAAS, 2012);
- i) volatilidade devido à flutuação presente em alguns registros, especialmente de sensores, podendo fazer com que estimativas se apresentem distorcidas quando realizadas em alta resolução (BRAAKSMA; ZEELENBERG, 2015; DAAS et al., 2015; DAAS; PUTS, 2014; O'CONNOR et al., 2010);
- j) cuidados com a segurança nos produtos e processos de modo que o sigilo das informações seja preservado (TAM; CLARKE, 2015a);
- k) falta de variáveis chaves para integração com bases de dados tradicionais (BRAAKSMA; ZEELENBERG, 2015) ou covariáveis para análises completas;
- l) incertezas quanto a garantia de que as fontes de dados digitais não modificarão a estrutura dos dados ou descontinuarão o registro e a disponibilização dos dados (TAM; CLARKE, 2014; DAAS et al., 2015; KITCHIN, 2015);
- m) dificuldades de atendimento às expectativas do novo perfil de demandas, com coleta, processamento e disponibilização em tempo real de informações com valor agregado – riqueza e utilidade de conteúdo (PARISE; IYER; VESSET, 2012).

E, por fim, os objetivos serão alcançados apenas se produtores tiverem (5) capacidade no que diz respeito a:

- a) recursos financeiros, principalmente para o caso de necessidade de aquisição de bases de Big Data de empresas privadas (MILLS et al., 2012; HACKL, 2016);

- b) recursos tecnológicos – infraestrutura de TI e ferramentas – com alta capacidade e eficiência para coletar e processar dados em tempo real – incluindo dados de streaming (NAS, 2013; SCHUTT; O’NEIL, 2013; PARISE; IYER; VESSET, 2012; SCANNAPIECO; VIRGILLITO; ZARDETTO, 2013; STRUIJS; BRAAKSMA; DAAS; 2014; TAM; CLARKE, 2014; CERVERA et al., 2014; HACKL, 2016; TAM; CLARKE, 2015a; KITCHIN, 2015). Ao mesmo tempo, se faz necessário que os sistemas garantam a proteção e a segurança das informações armazenadas (DUNNE, 2013);
- c) recursos humanos, com profissionais qualificados (DAVENPORT; PATIL, 2012; DAAS et al., 2015; SCHUTT; O’NEIL, 2013; HACKL, 2016) que consigam avaliar e prover soluções – novos métodos e técnicas (LAVALLE, 2011) – tanto no âmbito tecnológico, para questões de TI (DUNNE, 2013) – com conhecimentos de engenharia de dados, computação de alta performance, data warehousing (HACKL, 2016), Hadoop e bases de dados SQL e noSQL (VALE, 2015) –, quanto metodológico e analítico, com habilidades estatísticas (VALE, 2015; HACKL, 2016) – para (1) soluções de desenvolvimento de métodos para *Big Data analytics* (KITCHIN, 2015; UNGP, 2012; EUROSTAT 2014a; TAM; CLARKE, 2014), (2) avaliação da qualidade dos resultados (LAVALLE, 2011), (3) extração de conhecimento com visualização (KITCHIN, 2015) e mineração (DUNNE, 2013), (4) exploração inicial dos dados baseada em visualização (TENNEKES; DE JONGE; DAAS, 2011; ZIKOPOULOS et al., 2012; FRY, 2008; DAAS et al., 2012; LIU; JIANG; HEER, 2013), (5) correção de dados faltantes (DAAS, 2012) e da seletividade (DAAS et al., 2015) com métodos de predição (BUELENS et al., 2012; FLEKOVA; GUREVYCH, 2013; HASTIE; TIBSHIRANI; FRIEDMAN, 2009) e mineração de dados (VALE, 2015), (6) tratamento da volatilidade com médias móveis e técnicas de filtragem avançada (DAAS et al., 2015), e (7) garantia de sigilo das informações divulgadas (TAM; CLARKE, 2015a). Além desses, é de fundamental importância a disposição de profissionais que entendam de ética e governança de

dados – adaptada para o contexto de Big Data (KIM; CHO, 2017) – e sejam criativos (VALE, 2015) para que possam propor soluções que ampliem a aplicabilidade, a utilidade e, conseqüentemente, o valor dos produtos (LAVALLE, 2011).

Nesse caminho de possível mudança de paradigma, estão colocados, ainda, inúmeros riscos a serem considerados, que, na visão de Kitchin (2015), contemplam: a possibilidade de desistência em meio à transição; a perda de reputação e de confiabilidade; a violação de privacidade e de segurança da informação; o acesso inconsistente às fontes alternativas e possíveis problemas de continuidade dos produtos gerados; uma possível resistência da população e de produtores das fontes de dados alternativos; a fragmentação das abordagens entre diferentes países e instituições; cortes e restrições de recursos disponíveis; e a competitividade com a possibilidade de transferência da produção para diferentes organizações do mercado.

2.6.4 A questão da qualidade

Dentre os desafios que permeiam a produção de estatísticas públicas e indicadores socioeconômicos a partir de Big Data, identificados na Subseção 2.6.3, a aferição e a garantia da qualidade dos produtos estatísticos gerados é possivelmente a problemática central. Se não puderem ser de qualidade, indicadores socioeconômicos produzidos a partir de dados do tipo Big Data serão possivelmente descartados, uma vez que se destinam a suportar importantes processos de tomada de decisão tanto no setor público quanto no privado. Ou seja, a questão da qualidade dos produtos é crucial para que as possibilidades de inovação trazidas pela Revolução dos Dados consolidem-se no campo dos indicadores socioeconômicos.

Dessa forma, pode-se interpretar que, no âmbito do desafio destacado, um arcabouço capaz de avaliar a qualidade de estatísticas públicas e indicadores socioeconômicos gerados a partir de Big Data será determinante para o processo de transformação de dados e informações em conhecimento. Ou, em outras palavras, será decisivo para que os produtos estatísticos gerados sejam de fato utilizados na

prática das tomadas de decisão, orientando a realização de tarefas e ações em busca por resultados concretos.

Em termos do que já se avançou no desenvolvimento de pesquisas, Brancato e Di Consiglio (2018) perceberam quatro correntes que abordam a questão da qualidade de produtos estatísticos gerados a partir de Big Data sob as óticas:

- a) da extensão dos conceitos e métodos tradicionais de aferição da qualidade estatística para o contexto de uma produção baseada em Big Data (STRUIJS; DAAS, 2014; UNECE, 2014; KREUTER et al., 2015; AMAYA; BIEMER; KINYON, 2020; BAKER, 2017; CAI; ZHU, 2015);
- b) da percepção de possíveis fontes de erros quando da tentativa de execução de aplicações práticas – reais – de produção de estatísticas públicas a partir de Big Data (LUGOMER et al., 2017; HSIEH; MURPHY, 2017; ESSNET, 2018; WANG et al., 2012; BARCAROLI; GOLINI; RIGHI, 2018);
- c) do grau de confiabilidade das fontes de Big Data e de usabilidade delas (EUROSTAT, 2014a); e
- d) do processo de produção estatística (TAO; GAO, 2016; CAI; ZHU, 2015).

Além de perceber tais correntes, Brancato e Di Consiglio (2018) também listam e relacionam algumas das principais dimensões de qualidade no contexto do Big Data encontradas na literatura com os domínios considerados nos *frameworks* tradicionais: insumos, processo e produto.

Em termos de ferramentas para aferição da qualidade de estatísticas públicas – e, conseqüentemente, indicadores socioeconômicos – produzidas a partir de Big Data, percebe-se⁸: o *UNECE Big Data Quality Framework* (UNECE, 2014); o *Eurostat Accreditation Procedure for statistical data from Non-Official Sources* (EUROSTAT, 2014a); o *AAPOR Big Data Total Error Framework* (KREUTER et al., 2015)⁹; o *Conceptual Framework for Quality in Big Data* (BATINI et al., 2015); e o *Big Data Quality Assessment Framework* (CAI; ZHU, 2015).

⁸ Cabendo destacar aqui que não foram encontrados *frameworks* para a qualidade de produtos estatísticos gerados a partir de Big Data voltados especificamente para o contexto do setor turístico e, assim sendo, as demandas trazidas pelo Quadro 11 se referem à produção estatística a partir de Big Data de forma genérica, como um todo.

⁹ Atualizado e operacionalizado em Amaya, Biemer e Kinyon (2020).

Porém, mesmo com as proposições existentes, a literatura ainda parece apontar para a necessidade de avanços e aperfeiçoamentos, sinalizando a existência de lacunas de pesquisa, conforme sintetiza o Quadro 11.

Quadro 11 – Trechos associados a possíveis lacunas de pesquisa no âmbito da qualidade de produtos estatísticos gerados a partir de Big Data.

Publicação	Trecho	Pg.
De Waal, Delden e Scholtus (2019)	“Um tópico importante de trabalho futuro é o desenvolvimento de um <i>framework</i> sistemático para situações, métodos e métricas que podem surgir em contextos de múltiplas fontes.”	191
Barcaroli, Golini e Righi (2018)	“É crucial a definição de um <i>framework</i> metodológico que permita o uso eficiente de diferentes fontes, e também a avaliação da acurácia das estimativas obtidas na abordagem <i>model-based</i> .”	2
Brancato e Ascari (2018)	“Alguns tópicos relativos à confiabilidade, coerência e comparabilidade necessitam de análises e experimentações mais aprofundadas”.	8
Brancato e Di Consiglio (2018)	“ <i>Frameworks</i> teóricos para <i>multisource</i> e Big Data devem ser traduzidos em termos operacionais, contendo medidas e indicadores relevantes e informativos.”	8
Brancato et al. (2018)	“Necessidade de se repensar uma estratégia de qualidade eficiente e direcionada.”	4
	“Planejar métricas de mensuração da qualidade no novo ambiente de produção estatística.”	9
Hajnovic (2018)	“Por ser um campo emergente, há pouca orientação para a mensuração da qualidade das aplicações de Big Data e <i>data science</i> em estatísticas oficiais.”	1
Hand (2018)	“Construção de métricas de qualidade e <i>scorecards</i> de qualidade para bases de dados”.	10
Maslankowski e Nowicka (2018)	“Não há um <i>framework</i> unificado para qualidade de Big Data que pode ser aplicado para diferentes tipos de bases de dados.”	1
	“A variedade de <i>frameworks</i> de qualidade para Big Data a disposição permite criar um conjunto de indicadores capaz de aferir diferentes aspectos da usabilidade de uma base de dados.”	1
	“A solução está na criação de diferentes <i>frameworks</i> dependentes dos dados em utilização.”	1
Salgado et al. (2018)	“O QAF deve ser revisado em termos de potenciais novos indicadores.”	10
	“O paradigma do TSE precisa ser adaptado para o contexto de dados de registros telefônicos.”	10
De Waal, Delden e Scholtus (2017)	“Construir métricas e métodos para computá-las mais adequados para aplicações práticas reais.”	44
	“Aprofundar testes de adequação de indicadores propostos para a dimensão coerência no contexto de múltiplas fontes.”	44
	“Ampliar a quantidade de configurações de dados examinadas no projeto Komuso da ESSNET.”	45
	O autor apresenta também uma tabela com outros gaps pontuais e específicos.	46
NASEM (2017b)	“Novas fontes de dados requerem expansão e desenvolvimento adicional dos <i>frameworks</i> de qualidade existentes para que novos componentes e aspectos possam ser incluídos e enfatizados”	126
Hackl (2016)	“Um <i>framework</i> e critérios de qualidade necessitam ser desenhados ou adaptados para os vários tipos de Big Data e de produtos estatísticos que são baseados nesses dados.”	51
Reinert e Stoltze (2016)	“Necessidade de métodos padronizados que meçam a qualidade do <i>input</i> e que contenham indicadores quantitativos claros, especialmente para a dimensão acurácia.”	36
	“Está ficando claro que uma única lista de indicadores para todos os diferentes contextos de aplicação não se faz viável. Temos nos concentrados em conjuntos simples de indicadores bem adequados para o monitoramento contínuo da qualidade no contexto de uma fonte específica.”	36
	“Necessidades de trabalhos futuros que definam possíveis erros de cobertura.”	36
	“Trabalhos futuros podem focar no desenvolvimento de uma lista de indicadores que auxilie na avaliação da qualidade geral de uma fonte de dados antes mesmo da obtenção dos dados.”	37
Reis et al. (2016)	“Necessidade de complementar as diferentes propostas (BDQF, APNOS e BDTEF) e de estruturar os <i>links</i> entre a qualidade do <i>input</i> e do <i>process</i> com a qualidade do <i>output</i> .”	16
	“Elementos de cada um dos <i>frameworks</i> analisados podem ser combinados para gerar um <i>framework</i> de maior qualidade.”	16
	“Novas dimensões de qualidade, como complexidade, são também importantes, assim como a proposição de novos indicadores de qualidade.”	16
Agafitei et al. (2015)	“O ESS QAF ainda não é suficiente para cobrir a complexidade do processo de produção que se utiliza de múltiplas fontes de dados e métodos.”	207

Fonte: atualizado de Caumo e Souza (2021).

A análise aprofundada da estrutura dos *frameworks* já existentes e dos trechos que sinalizam eventuais lacunas atende aos objetivos desta pesquisa e está, portanto, apresentada no Capítulo 4 – referente aos resultados da tese.

3 METODOLOGIA

Quando da realização de uma pesquisa científica, admite-se uma série de pressupostos acerca do que se compreende e reconhece por conhecimento científico. Trata-se da concepção ou paradigma filosófico da ciência de base – ou simplesmente visão de mundo. Um entendimento que, segundo Saunders, Lewis e Thornhill (2012), está associado: ao ponto de vista do pesquisador sobre a natureza das coisas ou do funcionamento da realidade ou do mundo, descrito por “ontologia”; e ao que é considerado para uma comunidade científica ou determinado campo de estudo como um conhecimento aceitável, descrito por “epistemologia”. Partindo de, e alinhadas a tais pressupostos, decisões sobre métodos de pesquisa e técnicas ou procedimentos ou estratégias de investigação a serem empregadas na pesquisa podem então ser tomadas (CRESWELL, 2010).

Na visão de Saunders, Lewis e Thornhill (2012) e de Newman e Benz (1998), paradigmas filosóficos não devem ser julgados comparativamente. São adequados, cada um, para atender determinados objetivos, distintos entre as pesquisas. Ou seja, o entendimento é de que a delimitação de um paradigma de pesquisa tanto orienta os métodos e procedimentos quanto depende do objetivo da investigação.

No caso do presente estudo, a *Design Science* (DS), originária de pesquisas seminais realizadas por Simon (1969) e Gibbons et al. (1994), é utilizada como base filosófica (HEVNER, 2007; VAISHNAVI; KUECHLER, 2009; DRESCH; LACERDA; ANTUNES JR., 2015). A escolha pela *Design Science* se dá pelo fato de que se propõe a ser uma abordagem mais adequada para pesquisas que prescrevem artefatos como modelos e sistemas de informação e conhecimento, principalmente dentro da linha de pesquisa de gestão do conhecimento, na qual as metodologias mais clássicas têm alcance limitado (BARBOSA; BAX; 2017). Afinal, a *Design Science*, que pode ser compreendida como a ciência do projeto – que lida com o artificial, criado pelo homem (SIMON, 1969) –, é uma abordagem que busca gerar conhecimento para a concepção, o desenvolvimento e a avaliação de artefatos (VAN AKEN, 2004).

Nessa linha, a ciência tradicional, que lida com o natural e o social e que tem por objetivos centrais explorar, descrever, explicar e, quando possível, prever (VAN AKEN, 2004), em geral, não possuiria paradigmas de base adequados para quando o objetivo da pesquisa está em projetar, prescrever ou criar algo novo (DRESCH, 2018). Para contornar a problemática, a *Design Science* (DS) se oferece com a função de servir de base para o desenvolvimento de conhecimentos que os profissionais, nas organizações, possam utilizar para projetar soluções para seus problemas, ou para gerar melhorias e tecnologias de maneira geral (VAN AKEN, 2004). Dessa forma, pesquisas com essa abordagem estão, em geral, relacionadas à problemas complexos, relevantes, que consideram o contexto em que seus resultados serão aplicados (BURGOYNE; JAMES, 2005) e culminam no desenvolvimento de conhecimento essencialmente multidisciplinar e prescritivo (LACERDA et al., 2013).

Assim, o propósito da DS se alinha com o objetivo central dessa pesquisa, pois um *framework* conceitual é também reconhecido como um artefato (VAISHNAVI; KUECHLER, 2009) uma vez que corresponde a uma representação simbólica ou uma instanciação física, assim como modelos, construtos, métodos, instancicações e sistemas de informações e de conhecimento (MARCH; SMITH, 1995; HEVNER et al., 2004). Ou seja, por *framework* conceitual está se considerando aqui uma estrutura ou artefato construído com o propósito de identificar elementos componentes de um sistema e suas relações a fim de nortear análises, explicando os processos e prevendo os resultados (CARVALHO, 2013), ou, em outras palavras, fornecendo o suporte necessário à compreensão e à comunicação de estrutura e relacionamento dentro de um sistema com um propósito definido (SHEHABUDENN; PROBERT; PHAAL, 2000).

Cabe destacar que o entendimento da *Design Science* enquanto paradigma ou base filosófica – assim como são na ciência tradicional do contexto natural e social, por exemplo, o positivismo nas pesquisas ditas quantitativas e o interpretativismo nas pesquisas ditas qualitativas – é ainda difuso e incipiente (PURAO, 2013; DRESCH, 2018). Aqui, entretanto, partiremos da visão de Vaishnavi e Kuechler (2009), Dresch, Lacerda e Antunes Jr. (2015) e Hevner (2007) sobre a base epistemológica e ontológica da DS, conforme apresenta o Quadro 12.

Quadro 12 – Bases filosóficas para a pesquisa científica.

Publicação	Positivismo	Interpretativismo	Design Science
Ontologia	Uma única realidade: conhecida e probabilística	Múltiplas realidades; socialmente construídas	Múltipla, contextualmente situada de acordo com o estado atual do que está sendo estudado; habilitada para estudos sócio técnicos
Epistemologia	Objetiva: imparcial, observador em busca da verdade	Subjetiva. Os valores e o conhecimento emergem da interação entre pesquisador e participantes	Conhecendo por meio do fazer: restrita a construção dentro de um contexto, a circunscrição iterativa que revela o significado do conhecimento

Fonte: adaptado de Vaishnavi e Kuechler (2009, p. 9).

No âmbito da DS como base científica, o método de pesquisa aqui considerado foi o *Design Science Research* (DSR). Trata-se de um dos métodos de pesquisa que podem ser utilizados para operacionalizar a construção de conhecimento no contexto do paradigma DS (CHAKRABARTI, 2010), permitindo a condução de pesquisas que se ocupam do projeto e da avaliação de artefatos (MARCH; SMITH, 1995; VAISHNAVI; KUECHLER, 2009). Ressalta-se, entretanto, que a *Design Science* não se preocupa com a ação em si mesma, mas com o conhecimento que pode ser utilizado para projetar as soluções (VAN AKEN, 2004). Nesse sentido, o conhecimento na *Design Science* é generalizável quando válido para uma dada classe de casos (VAN AKEN, 2004), aqui entendidos como uma classe de problemas.

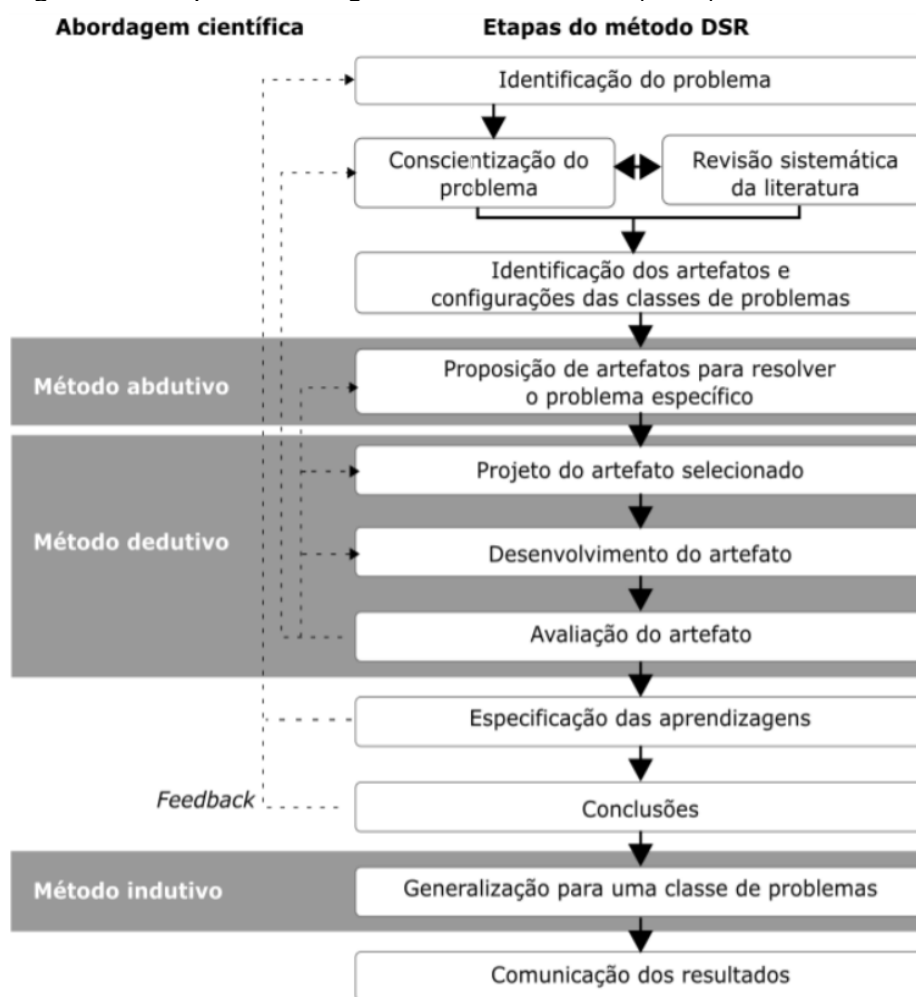
A maneira de condução do método DSR aqui considerada foi aquela elaborada por Dresch, Lacerda e Antunes Jr. (2015) pelo fato de que a concepção dos autores é produto de uma extensa revisão, análise e combinação de outras 10 propostas de estrutura e orientação para aplicação do método, publicadas entre 1980 e 2011. Com isso, alcançaram o roteiro mais amplo – com maior quantidade de etapas – e detalhado dentre todas as propostas observadas. Assim, tendo como referencial o trabalho dos três autores, a presente pesquisa se caracteriza pelo enquadramento metodológico descrito no Quadro 13.

De acordo com a proposta de Dresch, Lacerda e Antunes Jr. (2015), a condução do método *Design Science Research* (DSR) obedece a doze etapas, conforme apresenta a Figura 4.

Quadro 13 – Enquadramento metodológico da pesquisa.

Aspecto	Enquadramento
Base filosófica	<i>Design Science (DS)</i>
Razão para realização da pesquisa	Solução para um problema prático
Objetivo da pesquisa	Projetar artefato e prescrever soluções
Abordagem/método científico	Abdução, Dedução e Indução
Método de pesquisa	<i>Design Science Research (DSR)</i>
Procedimentos e técnicas de investigação	Documental, Bibliográfica e Questionário

Fonte: elaborado pelo autor (2021) com base em Dresch, Lacerda e Antunes Jr. (2015).

Figura 4 – Etapas da *Design Science Research (DSR)*.

Fonte: retirada de Dresch, Lacerda e Antunes Jr. (2015, p.125)

Nesse contexto, o método de trabalho desta pesquisa, em termos da articulação entre as etapas da DSR, as práticas de investigação de apoio e os objetivos específicos da presente tese de doutorado, é apresentado no Quadro 14.

Quadro 14 – Método de trabalho da pesquisa.

Etapa da DSR	Finalidade	Prática de apoio	Objetivo a que atende
i. Identificação do problema	Definição clara e objetiva de um problema importante de relevância justificada para formalização da questão de pesquisa.	1. Pesquisa bibliográfica e documental	Objetivos específicos a) e b).
ii. Conscientização do problema	Entendimento em profundidade dos aspectos relacionados ao contexto e às causas do problema para que se consiga, posteriormente, especificar as funcionalidades, performance esperada e os requisitos do artefato.		
iii. Revisão sistemática da literatura	Busca por conhecimento para auxiliar a etapa de conscientização do problema.		
iv. Identificação dos artefatos e configurações das classes de problemas	Percepção de outros artefatos já propostos no âmbito da classe de problemas considerada para fins de desenvolvimento de solução mais assertiva e satisfatória.		
v. Proposição de artefato para resolver o problema específico	Formalização descritiva da solução a ser desenvolvida para que se perceba o quanto se diferencia dos artefatos já propostos no contexto de interesse.		
vi. Projeto do artefato selecionado	Desenho do projeto de artefato para que se perceba tanto as características internas quanto o contexto externo no qual irá operar, com procedimentos, métodos de avaliação e resultados esperados.	2. Questionário aplicado a especialistas	Objetivo específico c).
vii. Desenvolvimento do artefato	Desenvolvimento do artefato em si para que se alcance sua forma funcional e se registre a heurística da construção a servir de base para novas pesquisas.		
viii. Avaliação do artefato	Observar e medir o comportamento do artefato na solução do problema em contexto real ou experimental para se alcançar uma versão final e se registrar a heurística contingencial.		
ix. Especificação das aprendizagens	Formalização do aprendizado alcançado para que a pesquisa seja útil para subsidiar futuras construções de conhecimento.		
x. Conclusões	Apresentação dos resultados e limitações da pesquisa, direcionando estudos futuros.		
xi. Generalização para uma classe de problemas	Apontar para possibilidades de generalização da solução do problema específico para uma classe de problemas.		
xii. Comunicação dos resultados	Publicações para disseminação do conhecimento gerado.		

Fonte: elaborado pelo autor (2021) com base em Dresch, Lacerda e Antunes Jr. (2015).

Tendo como base a estrutura metodológica apresentada, os capítulos seguintes – Resultados e Considerações Finais – estão organizados de acordo com a sequência de etapas do método DSR. Os conteúdos referentes à etapa de identificação do problema até a etapa de especificação das aprendizagens apresentados no Capítulo 4, enquanto os da etapa de conclusões até a de comunicação dos resultados no Capítulo 5.

3.1 PROCEDIMENTOS E TÉCNICAS DE INVESTIGAÇÃO

No âmbito da *Design Science Research* (DSR), foram adotados como técnicas de investigação de apoio: a pesquisa documental e bibliográfica, utilizada como auxílio para a condução das etapas de i – identificação do problema – até vii – desenvolvimento do artefato –; e o questionário aplicado a especialistas, utilizado como base para a condução das etapas vii – desenvolvimento do artefato – e viii – avaliação do artefato. As descrições dos procedimentos utilizados em cada uma das técnicas estão nas Subseções 3.1.1 e 3.1.2, respectivamente.

3.1.1 Pesquisa documental e bibliográfica

A técnica documental, com objetivo de fornecer as primeiras informações sobre o assunto a ser estudado, e a bibliográfica, com objetivo de informar ao pesquisador o que já tem sido desenvolvido em determinada área (SAUNDERS et al., 2012), foram realizadas sob a forma de revisões sistemáticas (FORBES, 1998) ou sistemáticas e integrativas (ERCOLE et al., 2014), em três diferentes momentos, de acordo com três diferentes propósitos:

- i. apresentar o estado da arte da produção científica sobre a utilização de Big Data na produção de estatísticas públicas no que diz respeito à identificação de oportunidades – com possíveis fontes de dados e aplicações práticas já propostas – e de desafios – com problemáticas e riscos – já percebidos ou previstos;
- ii. resumir a produção técnico científica sobre a aferição da qualidade de indicadores socioeconômicos produzidos a partir de Big Data,

identificando dimensões e *frameworks* que têm sido considerados e propostos para tal finalidade;

- iii. sintetizar a produção científica sobre o uso de dados do tipo Big Data na geração de indicadores de turismo no que concerne a possíveis fontes e aplicações práticas.

Partiu-se, em cada um dos três momentos, de uma pergunta claramente formulada, utilizando-se de métodos sistemáticos e explícitos, para: identificar, selecionar e avaliar criticamente pesquisas relevantes; e coletar e analisar dados dos estudos (GREEN; HIGGINS, 2005), com critérios para a busca que seguiram o método PRISMA (MOHER et al., 2009). Ao final, os documentos coletados foram analisados em busca de conteúdos que permitiram responder a cada propósito. Os condicionantes de busca e seleção dos documentos em cada um dos momentos estão apresentados no Quadro 15.

Os resultados das três revisões sistemáticas subsidiaram a compreensão do contexto e a estruturação do problema, da lacuna, dos objetivos, do referencial teórico da tese e contribuíram com o desenvolvimento do artefato, sendo publicados, respectivamente, no:

- i. livro “População, Território e Estatísticas Públicas”, comemorativo aos 20 anos da pós-graduação da Escola Nacional de Ciências Estatísticas, do Instituto Brasileiro de Geografia e Estatística (CAUMO, 2020); no
- ii. periódico Planejamento e Políticas Públicas do Instituto de Pesquisa Econômica Aplicada (CAUMO; SOUZA, 2021); e na
- iii. revista E-Tech, do Serviço Nacional de Aprendizagem Industrial (CAUMO; BIZ; SOUZA, 2020), assim como nos anais da 15^o Conferência Internacional de Gestão do Conhecimento e do 2^o Seminário Internacional de Universidades Corporativas e Escolas de Governo (CAUMO; BIZ; SOUZA, 2019).

Quadro 15 – Condicionantes de busca das revisões sistemáticas de literatura.

Condicionante	Revisão para o propósito i	Revisão para o propósito ii	Revisão para propósito o iii
Base de dados	Scopus (Elsevier); Web of Science (Clarivate Analytics); Scielo.org	Scopus (Elsevier); Web of Science (Clarivate Analytics); Scielo.org	Scopus (Elsevier); Web of Science (Clarivate Analytics).
Data da busca	19 de setembro de 2017	23 de junho de 2018	23 de abril de 2019
Tipo de busca	Título, resumo e palavras-chave (Scopus); Tópico (Web of Science); Padrão (Scielo)	Título, resumo e palavras-chave (Scopus); Tópico (Web of Science); Padrão (Scielo)	Título, resumo e palavras-chave (Scopus); Tópico (Web of Science).
Critérios de busca	Conter pelo menos uma expressão de cada um dos dois grupos: (1) <i>official statistics, public statistics</i> ; (2) <i>big data, data revolution, data avalanche, digital data, digital era, internet of things, open data, crowdsourcing data, web 2.0</i> .	Conter pelo menos uma expressão de cada um dos três grupos: (1) <i>official statistics, public statistics, official indicators, public indicators, socioeconomic statistics, socioeconomic indicators, public index, official index, socioeconomic index</i> ; (2) <i>quality framework, statistical quality, quality of statistics, statistical thinking, data quality, quality of data, statistical literacy, quality assessment, assessment of quality, information quality, quality of information</i> ; (3) <i>big data, multisource, data science, digital data, personal data, crowdsourcing data, social data, crowdsensing data, social media, social network, sensor data</i> .	Conter pelo menos uma expressão de cada um dos dois grupos: (1) <i>Tourism statistics, tourism data, tourism indicators, tourism information, touristic information, tourism research</i> ; (2) <i>Big data, multisource, data science, digital data, social media, facebook, Instagram, social data, personal data, crowdsourcing, crowdsensing, social network, twitter, smart destination, smart tourism destination</i> .
Retornos	178 (106; 72; e 0)	21 (13; 8; e 0)	256 (157; e 99)
Excluídos	122 (37 repetidos e 85 fora de escopo)	13 (6 repetidos e 7 fora de escopo)	81 (75 repetidos e 6 fora de escopo)
Busca adicional	Não realizada	Google ¹	Não realizada
Critério de busca adicional	Não realizada	Conter pelo menos uma expressão de cada um dos três grupos: (1) <i>official statistics, public statistics</i> ou <i>official indicators</i> ; (2) <i>quality framework, statistical quality, data quality, quality assessment</i> ou <i>information quality</i> ; e (3) <i>big data</i> ou <i>multisource</i> .	Não realizada
Retornos	Não realizada	46 (selecionados após análise dos 100 primeiros resultados) ¹	Não realizada
Total de documentos	56	54	175

Fonte: elaborado pelo autor (2021).

Chamada: ¹uma vez que o algoritmo de classificação e ordenamento dos resultados utilizados pelo Google considera a experiência do usuário no que se refere a mais de 200 fatores (GOOGLE, 2019), é de conhecimento deste autor que os 100 primeiros resultados não necessariamente serão os mesmos para diferentes usuários.

3.1.2 Questionário aplicado a especialistas

A aplicação de questionários em especialistas, usuários e produtores de estatísticas de públicas e indicadores socioeconômicos, foi realizada com intuito de apoiar o processo de avaliação – verificação de estrutura e aplicabilidade – do *framework* aqui proposto, almejando o alcance de uma versão final para o artefato.

Para tanto, apoiou-se na técnica Delphi, originária do trabalho de Dalkey e Helmer (1963), sob justificativa de ser adequada para quando se deseja encontrar um consenso com relação a um determinado assunto ou problema, fundamentando-se em um grupo de especialistas (FRACIONE, 1990). Seu funcionamento consiste na aplicação em um conjunto de questionários, respondidos de maneira sequencial, individualmente pelos participantes anonimizados e separados geograficamente, intercalados por *feedbacks*. Ou seja, com base nas opiniões manifestas em cada rodada de aplicação do questionário, o artefato é então modificado e colocado para análise em nova rodada de investigação. Para fins de convergência para uma resposta final coletiva, a interação entre os especialistas é construída através da transmissão de informações resumidas sobre as respostas do grupo à rodada anterior de aplicação do questionário, permitindo que justificativas dadas sejam observadas e posicionamentos possam ser repensados (OSBORNE et al., 2003).

Nesses termos, há concordância de que os quatro pilares fundamentais da técnica sejam: o anonimato, o uso de especialistas, a aplicação de rodadas interativas e com *feedback* e a busca por consenso (SÁFADI, 2001).

Dessa forma, a técnica Delphi pode ser vista como uma forma de estruturar o processo de comunicação grupal, de modo a permitir que um conjunto de indivíduos, tratados como um todo, possa trabalhar com problemas complexos (LINSTONE; TUROFF, 2002). Corresponde, portanto, a uma abordagem ampla, completa e enriquecedora, muito eficaz para a promoção da interdisciplinaridade a partir do aproveitamento do conhecimento, da experiência e da criatividade de um painel de especialistas.

Para a implementação da técnica, foram considerados os processos e diretrizes propostos por Marques e Freitas (2018) e Wright e Giovinazzo (2000). O primeiro por ser fruto de análise e combinação de outras 6 propostas de condução

da técnica e o segundo por ser frequentemente utilizado no PPGEGC/UFSC. De forma compilada, o Quadro 16 apresenta as etapas e os correspondentes critérios de referência considerados nesta tese.

Quadro 16 – Etapas da técnica Delphi e critérios de referência considerados.

Etapa Delphi	Crítérios de referência
i. Escolha do grupo de especialistas	Busca por pelo menos 10 especialistas, produtores ou usuários de estatísticas públicas ou de turismo, de perfis heterogêneos em termos de experiência, áreas de especialidade e perspectivas em relação ao problema, selecionados por meio de amostragem não probabilística de forma intencional a partir das redes de cooperação deste autor e seus orientadores.
ii. Construção do questionário 1	Opção por questões menos estruturadas, mais abertas, para que fatores e temas importantes emergjam com maior liberdade, detalhamento e menos vieses.
iii. Pré-teste	Simulação em pesquisadores da rede de colaboração deste autor.
iv. Primeiro contato e convite para participação	Atenção para causar engajamento perante os convidados com contatos personalizados e apontamento da utilidade benefícios dos resultados (por e-mail).
v. Aplicação do questionário	Opção pela realização via internet, utilizando-se de ferramentas apropriadas para <i>design</i> e aplicação de questionários.
vi. Análise dos dados	Análise predominantemente qualitativa (de conteúdo).
vii. Construção de novo(s) questionário(s), com feedback	Apresentação de <i>feedback</i> da análise dos resultados do questionário anterior e utilização de questões semiestruturadas, registrando escalas de importância e permitindo manifestações livres adicionais.
viii. Aplicação do novo questionário	Opção pela realização via internet, utilizando-se de ferramentas apropriadas para <i>design</i> e aplicação de questionários.
ix. Análise dos novos dados	Análise predominantemente quantitativa (estatística descritiva).
x. Análise da convergência	Busca por mínimo duas e no máximo quatro rodadas, sendo interrompidas quando forem atingidos níveis satisfatórios de estabilidade e consenso nas respostas – pelo menos dois terços classificando um item com quatro ou cinco na escala Likert de cinco pontos.
xi. Conclusões e relatório final	Redação das conclusões e do relatório final da pesquisa, com envio do material para os respondentes.

Fonte: elaborado pelo autor com base em Marques e Freitas (2018) e Wright e Giovinazzo (2000).

Na primeira etapa, treze (13) especialistas foram intencionalmente selecionados e convidados a participarem da pesquisa. Desses, dois rejeitaram o convite por questão de indisponibilidade de tempo e outros dois desistiram do processo no decorrer da aplicação dos questionários. Cabe destacar que os nomes foram selecionados em virtude de possuírem alta expertise acadêmica e profissional associada ao tema desta tese, a partir das redes de contatos deste autor e seus orientadores/coorientadores, formando um grupo heterogêneo em termos de áreas de conhecimento e organizações de vínculo atual e passado, conforme perfis descritos no Quadro 17.

Quadro 17 – Especialistas participantes da etapa de avaliação do artefato.

Nome	Vínculo	Mini currículo acadêmico	Mini currículo profissional
Alexandre Fernandes Barbosa	Gerente do Núcleo de Informação e Coordenação do Ponto BR (NIC.br).	Pós-Doutorado pela <i>Ecole des Hautes Etudes Commerciales</i> (Canadá). Doutor em administração de empresas pela FGV-SP com tese sobre o tema de Governo Eletrônico. Mestrado em <i>Business Administration</i> pela <i>University of Bradford</i> (Inglaterra) e em Ciências da Computação pela UFMG, com dissertação sobre redes de comunicação de dados. Especialização em Marketing (FGV), Finanças Empresariais (FGV) e graduação em Engenharia Elétrica (PUC-MG).	Gerente do Núcleo de Informação e Coordenação do Ponto BR (NIC.br) no CETIC.br desde 2009, vice-presidente de tecnologia da Certisign em 2007 e 2008, diretor de consultoria da Brisa entre de 2003 a 2007, diretor de parcerias e interconexão da Pegasus Telecom de 2001 a 2003, diretor de desenvolvimentos de negócios da Vésper em 2000 e 2001, e gerente sênior Telemig (Telemar) de 1985 a 2000.
Arnaldo Lyrio Barreto	Analista do IBGE na área de TI desde 1986, Diretor Adjunto da Diretoria de Informática do IBGE desde 2016, Professor da FGV-RJ e FGV-SP desde 2001 nas áreas de gerenciamento de projetos, gestão estratégica de TI e segurança da informação.	Professor e coordenador em várias IES, mestrado na área de Educação e Novas Tecnologias, doutorado em Epistemologia com tese sobre as classificações de raça/cor e ocupação nos Censos Demográficos brasileiros e pós-doutorado em Sociologia e Inovação na Universidade de Essex – Inglaterra.	Atuação direta em toda a parte de gestão das tecnologias de informação e comunicação, sendo o responsável direto pelo desenvolvimento do Plano Diretor de TI e do documento de Estratégia Geral de TI, que verificam o estado atual de TI e o planejamento estratégico do IBGE, indicando ações futuras e prováveis para a condução dos assuntos de informática. Atuação direta também nos assuntos de Gestão e Curadoria de Dados e dos que tangenciam a Lei Geral de Proteção de Dados Pessoais, com seus impactos na gestão de acervo de informações já existentes, quanto ao uso dos mesmos em pesquisas e cadastros, disseminação dos mesmos e acordos de uso de bases de dados de outros órgãos para melhoria das informações e cadastros utilizados pelo IBGE.
Heimar de Fatima Marin	Profa. Titular aposentada da UNIFESP, Editora - Chefe do <i>International Journal of Medical Informatics</i> (Elsevier) e Coordenadora Curso de Especialização em Informática em Saúde (PROADI DigiSUS & Terminologias).	Graduação em Enfermagem pela Universidade Federal de São Paulo (1983), Mestrado em Enfermagem (UNIFESP, 1991) e Doutorado em Ciências Biológicas com foco em Informática em Saúde (Biologia Molecular) também pela UNIFESP (1994). Fez o pós-doutoramento em Informática Clínica no <i>Center for Clinical Computing</i> na <i>Harvard Medical School</i> (<i>Harvard University</i>) de 1996 a 1998.	Em 2001 fez o concurso de Livre Docência na Faculdade de Medicina da Universidade de São Paulo. Em 2004, recebeu o título de <i>Fellow</i> do <i>American College of Medical Informatics</i> , tendo sido a primeira brasileira nomeada neste Colégio. Em 2006, tornou-se Professora Titular da Escola Paulista de Enfermagem, Universidade Federal de São Paulo, defendendo a tecnologia e comunicação na enfermagem e na saúde. Foi Presidente da Sociedade Brasileira de Informática em Saúde de 2004 a 2008 e Vice-Presidente de 2010 a 2013. Em 2009, assumiu a Coordenação do Programa de Pós-graduação em Informática em saúde, reformulado em 2010 para Gestão e Informática em Saúde da UNIFESP. Em 2009 fundou o <i>Journal of Health Informatics</i> da Sociedade Brasileira de Informática em Saúde e foi editora Chefe até 2015. Em 2012 foi convidada pelo CETIC (Centro de Pesquisas e Estatísticas), do NIC.BR (Núcleo de Informação e Coordenação do Ponto BR), para coordenar cientificamente o projeto de Tecnologias de Informação e Comunicação em Saúde (TIC Saúde BRASIL), função exercida até o presente. Atua também como Coordenadora do Curso de Especialização em Informática em Saúde no Hospital Sirio Libanês e é Editora-Chefe do <i>International Journal of Medical Informatics</i> , Elsevier, Londres.

Leo Cremonesi	Estatístico jornalista, <i>Head do FT Rankings</i> , na instituição <i>Financial Times</i> e <i>Chartered Statistician (Cstat)</i> na <i>Royal Statistitcal Society</i> .	Bacharel em Estatística na Escola Nacional de Ciências Estatísticas (ENCE/IBGE) com Pós-graduação em educação na Fundação Getúlio Vargas, MBA na <i>London School of Economics</i> e professor convidado na <i>London School of Economics</i> .	Atua há 20 anos como estatístico em diferentes institutos de pesquisa de mercado, com a maior parte do tempo na Ipsos MORI. É estatístico jornalista no <i>Financial Times</i> (Londres) há 2 anos, além de professor convidado da <i>London School of Economics</i> há 5 anos. Possui como característica principal: escrever sobre estatística e "traduzir" <i>outputs</i> estatísticos de forma que todos entendam.
Marcelo Trindade Pitta	Coordenador de Métodos Quantitativos no Cetic.br, departamento do NIC.br.	Mestrado em Estudos Populacionais e Pesquisas Sociais (ENCE/IBGE), com dissertação intitulada "Uso de Modelos de Espaços de Estado para a Estimção do Efeito de Vício de grupos de Rotação na PME/IBGE". Graduação em Estatística pela ENCE/IBGE. Pós-Graduação em Análise de Sistemas pela UVA-RJ, com trabalho de conclusão intitulado "Análise de Correspondências dos dados da Pesquisa sobre Comunidades Eclesiais Católicas".	Atualmente é coordenador de Métodos Quantitativos com função de planejamento, seleção e análise de amostras de pesquisas anuais sobre Tecnologia da Informação e Comunicação no Cetic.br (Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação). Na Fundação SEADE, foi chefe de divisão de estatística com responsabilidade por planejar e analisar dados para produção de estatísticas públicas do Estado de São Paulo. No Centro de Estatística Religiosa e Investigações Sociais (CERIS) foi gerente da área de estatística com função de planejamento, seleção e análise de amostras de pesquisas anuais sobre presença da Igreja Católica no Brasil. No Instituto Superior de Estudos da Religião (ISER), foi estatístico responsável pelo planejamento e análise de pesquisas sociais. Também já prestou serviços de planejamento de pesquisa e análise de dados para diversas instituições: CEERT, <i>Bankble Frontiers</i> , <i>Worleyparsons</i> , UNESCO, Unitrabalho, Observatório Social Cut (OS-CUT), Centro de Informações e Dados do Rio de Janeiro (CIDE) e Empresa MULTICANAL.
Maria Luiza Barcellos Zacharias	Aposentada do IBGE. Sócia-Fundadora e Diretora de Novos Negócios da HORUS Inteligência de mercado.	Graduação em Estatística (ENCE, 1984), doutorado em Administração (UFRJ/COPPEAD, 2009). Atuou lecionando como professora colaboradora do Programa de Pós-Graduação (Mestrado e Doutorado) da ENCE, ministrando as disciplinas de Métodos de Pesquisas e Levantamentos e de Qualidade de Dados. Professora de Metodologia Científica do MBA em Gerenciamento de Projetos da Fundação Getúlio Vargas (FGV), fazendo também orientação de trabalhos de fim de curso. Professora de Pesquisa de Mercado na ESPM.	Foi pesquisadora de carreira do IBGE, tendo coordenado equipes e projetos de grande porte da instituição, como o Cadastro Central de Empresas, e atuado na área de métodos de pesquisas, gerenciando a área de qualidade estatística do IBGE. Professora de graduação e de pós-graduação em diversas IES. Especialista e membro da Comissão de Assessoramento Técnico-Pedagógico, especializada em questionários para aferir a qualidade da educação básica do Inep. Atualmente, é responsável pelo desenvolvimento de novos negócios e parcerias da HORUS, empresa de inteligência competitiva de mercado que utiliza técnicas estatísticas e de Big Data para fornecer informações e indicadores de desempenho sobre o mercado varejista de bens de consumo.
Maria Luiza Guerra de Toledo	Professora e Pesquisadora na Escola Nacional de Ciências Estatísticas (ENCE) - IBGE.	Doutora em Engenharia de Produção pela Universidade Federal de Minas Gerais (2014), Bacharel (2005) e Mestre (2001) em Estatística pela Universidade Federal de Minas Gerais.	De 2005 a 2011, atuou como Instrutora e Consultora certificada pela Minitab, LCC, uma empresa privada situada em <i>State College</i> , Pensilvânia, EUA, fornecedora de softwares e serviços para melhoria da qualidade e educação estatística. Atuou em treinamentos e no desenvolvimento de projetos de gestão da qualidade em dezenas de organizações, entre elas Votorantim, Unilever, Gerdaul, LG Electronics, Nycomed, GM, Faber Castell, entre outras. Entre 2011 e 2014, foi Professora Adjunta no Departamento de Engenharia de Produção da Universidade Federal de Ouro Preto (UFOP). Desde Jun/2014, atua como Professora e Pesquisadora em Informações Geográficas e Estatísticas na Escola Nacional de Ciências Estatísticas do Instituto Brasileiro de Geografia e Estatística. Em 2017, esteve como Professora Visitante na <i>Université Grenoble Alpes</i> (França). Tem experiência nas áreas de Probabilidade e Estatística, atuando principalmente nos seguintes temas: controle estatístico de qualidade, análise de sobrevivência, confiabilidade e manutenção.

Tiago Mendes Dantas	Data Science Manager na Wildlife Studios.	Graduado em Estatística (ENCE, 2009), Mestre em Engenharia Elétrica (PUC-RIO, 2011) e Doutor em Engenharia de Produção (PUC-RIO, 2018).	Atualmente responsável pelo trabalho de Ciência de Dados associado à aquisição de usuários (estratégias de leilão de anúncios/Real-Time Bidding, modelagem preditiva, etc). Foi Head de Data Science da Blu e Gerente de Desenvolvimento e pesquisa do IBGE.
Zélia Magalhães Bianchini	Aposentada do Instituto Brasileiro de Geografia e Estatística (IBGE), tendo prestado serviços à Instituição por 38 anos.	Licenciada em Matemática pela Faculdade de Filosofia Ciências e Letras de São José do Rio Preto - SP (1975). Mestre em Estatística pelo Instituto de Matemática Pura e Aplicada (IMPA) no Rio de Janeiro, em 1978, com tese intitulada "Eficiências de Pitman e Bahadur para o problema de escala no caso de Weibull".	Servidora do IBGE (agosto de 1978 a agosto de 2016), tendo atuado grande parte do tempo na unidade de metodologia da Diretoria de Pesquisas, que chefiou de 1992 a janeiro de 1999. Atuou como orientadora técnica na área de metodologia, especialmente, em amostragem, processo de produção de informações, confidencialidade, princípios fundamentais das estatísticas oficiais e qualidade em estatísticas oficiais. Assessora e Diretora Adjunta da Diretoria de Pesquisas do IBGE de 1999 até sua aposentadoria em março de 2016, acumulando vasta experiência como representante do IBGE em diversos fóruns nacionais e internacionais. Assessora da Presidência do IBGE (março a agosto de 2016), atuando na coordenação do conteúdo programático da Reunião Preparatória para a INFOPLAN (3a Conferência Nacional de Produtores e Usuários de Informações Estatísticas, Geográficas e Ambientais), realizada no Rio de Janeiro, em dezembro de 2016. Professora colaboradora da Escola Nacional de Ciências Estatísticas do IBGE (ENCE): na Graduação em Estatística (área de amostragem - de 1990 a 1993 e de 1999 a 2003); no Curso de Desenvolvimento de Habilidades em Pesquisa – CDHP (módulo de Amostragem - cinco edições); no Curso "Princípios Fundamentais das Estatísticas Oficiais" do Plano Anual de Treinamento (2011 a 2015); e conteudista em EaD.

Fonte: elaborado pelo autor (2021).

Após serem convidados formalmente, por e-mail, para participação na pesquisa, conforme carta apresentada no APÊNDICE A, os especialistas receberam uma apresentação sobre a proposta trazida nesta tese, operacionalizada através da plataforma Prezi, conforme demonstra o APÊNDICE B, e em seguida responderam, à distância, via internet, por meio da ferramenta JotForms, do dia 05 de março ao dia 01 de abril de 2021, o questionário trazido pelo APÊNDICE C, semi estruturado e com diversos espaços para manifestações livres, com muitas questões em formato qualitativo, testado preliminarmente junto ao Núcleo de Estudos em Inteligência, Gestão e Tecnologias para Inovação (IGTI), o grupo de pesquisa deste autor no PPGECC.

Na sequência, após compiladas as manifestações, apresentadas de forma anônima ao grupo (Subseção 4.8.2), modificado e reapresentado o *framework* (APÊNDICE D), os especialistas responderam então ao segundo questionário, dessa vez predominantemente estruturado, com questões formatadas em escalas de Likert de cinco pontos, conforme demonstra o APÊNDICE E – operacionalizado entre os dias 05 e 30 de maio de 2021.

Por fim, o processo se encerrou, conforme apresentado na Subseção 4.8.2, quando do atingimento de níveis satisfatórios de concordância dos especialistas com a estrutura do *framework* – com todos os respondentes avaliando cada um dos itens em nível 4 ou 5 da escala –, com resultados compilados e novamente transmitidos ao grupo de especialistas.

3.1.2.1 Mineração *web* para a geração de indicador de turismo

A aplicação de questionários a especialistas, procedimento empregado durante a etapa de avaliação do *framework* proposto, serviu não só para a verificação da estrutura do artefato, como também para análise da aplicabilidade do instrumento em um contexto prático de avaliação da qualidade estatística de um indicador de turismo gerado a partir de Big Data coletados na internet, obtidos em tarefa conhecida por mineração da *web*, descrita por Cooley, Mobasher e Srivastava (1997) como o processo de descoberta e análise de dados e informações úteis na *web* – que pode ser subdividida em três tipos: mineração de conteúdo; mineração de estrutura; e mineração de uso.

Neste caso, conduziu-se uma mineração de conteúdo na *web*, que foca em informações disponíveis dentro de documentos e páginas da *web*, realizada por um software computacional do tipo rastreador da rede, também conhecido como *web crawler*, que navega pela *web* de uma forma metódica e automatizada (CHEONG, 1996).

Em termos concretos, o *web crawler* foi desenvolvido em linguagem de programação Python (ROSSUM, 1996), com script apresentado no APÊNDICE F, e serviu para a coleta de informações do portal TripAdvisor¹⁰, utilizadas como insumos para a construção de um indicador que se propõe a refletir o nível de satisfação do turista com as atrações locais dos municípios de Florianópolis e Balneário Camboriú¹¹, ambos pertencentes ao estado de Santa Catarina, com especificações apresentadas no Quadro 18 – indicador que poderia ser considerado uma aproximação para o nível de satisfação geral do turista com as atrações dos municípios.

Cabe destacar que apenas 11 atrações turísticas foram consideradas em ambos os municípios de Florianópolis e Balneário Camboriú, escolhidas por estarem mais bem classificadas pelo critério "favoritos dos viajantes", do TripAdvisor, desde que contivessem um total de pelo menos 500 avaliações já recebidas – considerando a situação no período da coleta dos dados, entre 20 e 24 de janeiro de 2021.

É reconhecido aqui que o indicador considerado pode apresentar limitações, que podem inserir vieses nos resultados, tais quais: um desalinhamento entre a população alvo do indicador, turistas que visitam as atrações locais, e os efetivos participantes da amostra, com possibilidade de inclusão de "não turistas"; uma seletividade na amostra, especialmente se o perfil dos visitantes das atrações selecionadas que deixam *feedback* em língua portuguesa no TripAdvisor for diferente do perfil dos demais visitantes, que não registram suas manifestações na plataforma ou que registraram em outra língua; má especificação das quantidades e

¹⁰ Tal prática só é permitida, conforme os termos de uso do portal, mediante solicitação por parte do pesquisador, sucedida de aceite e concordância manifesta formalmente por escrito pelo TripAdvisor.

¹¹ Escolhidos por serem dois dos principais destinos turísticos indutores de Santa Catarina, assim considerados nos trabalhos do Índice de Competitividade dos Destinos Turísticos, da parceria entre Ministério do Turismo (MTur), Serviço Brasileiro de Apoio às Micro e Pequenas Empresas (SEBRAE) e Fundação Getúlio Vargas (FGV).

dos critérios de seleção de atrações locais consideradas em cada município que, se fossem outros, poderiam alterar os resultados do indicador; a utilização de registros de avaliações falsas, no sentido de terem sido propositalmente direcionadas pelos usuários respondentes para beneficiar ou prejudicar a avaliação geral de uma determinada atração ou localidade; a não validade da suposição de que os registros dos insumos – avaliações coletadas – estão em nível de mensuração intervalar, apresentando equidistância numérica (diferenças constantes) entre os cinco níveis da escala (possíveis respostas), arbitrariamente pontuados com valores de 1 a 5; entre outras.

Quadro 18 – Características do indicador de turismo utilizado na análise da aplicabilidade.

Característica	Descrição
Nome	Nível de satisfação do turista com as atrações locais.
Formato	Valor no universo dos números reais pertencente ao intervalo de 1 até 5, onde 5 representa o nível mais alto de satisfação.
Insumos	Avaliações em escala de <i>likert</i> de 5 pontos (1: Horrível, 2: Ruim, 3: Razoável, 4: Muito bom, 5: Excelente), variável originalmente qualitativa em escala ordinal, registradas por visitantes de atrações locais que deixaram <i>feedback</i> em língua portuguesa no portal TripAdvisor.
Fórmula de cálculo	Média das avaliações registradas pelos visitantes em todas as atrações locais consideradas (pressupondo insumos em nível de mensuração intervalar). Esta fórmula de cálculo equivale a uma média ponderada das avaliações médias de todas as atrações locais consideradas. A ponderação é realizada de modo que atrações mais avaliadas recebam maior peso e exerçam maior influência sobre o resultado do indicador.
Período	Experiência turística de janeiro de 2015 a dezembro de 2020.
Localidade	Municípios de Florianópolis (SC) e Balneário Camboriú (SC).
Atrações consideradas	Em Florianópolis (11): Santo Antônio de Lisboa, Lagoinha do Leste, Ilha do Campeche, Praia da Joaquina, Praia de Canasvieiras, Praia da Daniela, Jurerê, Praia de Barra da Lagoa, Praia Mole, Dunas da Joaquina, Costa da Lagoa. Em Balneário Camboriú (11): Parque Unipraias Camboriú, Morro do Careca, Passeio San Miguel, Praia das Laranjeiras, Cristo Luz, Molhe – Barra Sul, Avenida Atlântica, Praia do Estaleirinho, Praia do Estaleiro, Praia Central, Praia de Taquaras.
Principais possíveis limitações	Alinhamento entre população alvo e população de pesquisa, representatividade da amostra, critério de seleção e quantidade de atrações locais, manifestações falsas, violação de pressuposto de escala intervalar nas avaliações.

Fonte: elaborado pelo autor (2021).

Ressalta-se que o foco neste momento não está na construção do melhor indicador possível para o nível de satisfação do turista nas localidades consideradas,

mas sim na simulação da aplicação do *framework* proposto nesta tese. Para fins de melhoramento do indicador em si, estudos específicos deverão ser conduzidos.

4 RESULTADOS

Este capítulo está estruturado de modo a apresentar os resultados da pesquisa de acordo com as etapas de i – identificação do problema – até ix – especificação das aprendizagens – preconizadas pelo método *Design Science Research* (DRESCH; LACERDA; ANTUNES JR., 2015). As demais etapas – de x até xii – estão contempladas no último capítulo deste documento.

4.1 IDENTIFICAÇÃO DO PROBLEMA

O caminho percorrido até a identificação do problema, representado pelo resumo articulado da Figura 5, foi auxiliado por três revisões sistemáticas de literatura, conforme metodologia já apresentada no Capítulo 3.

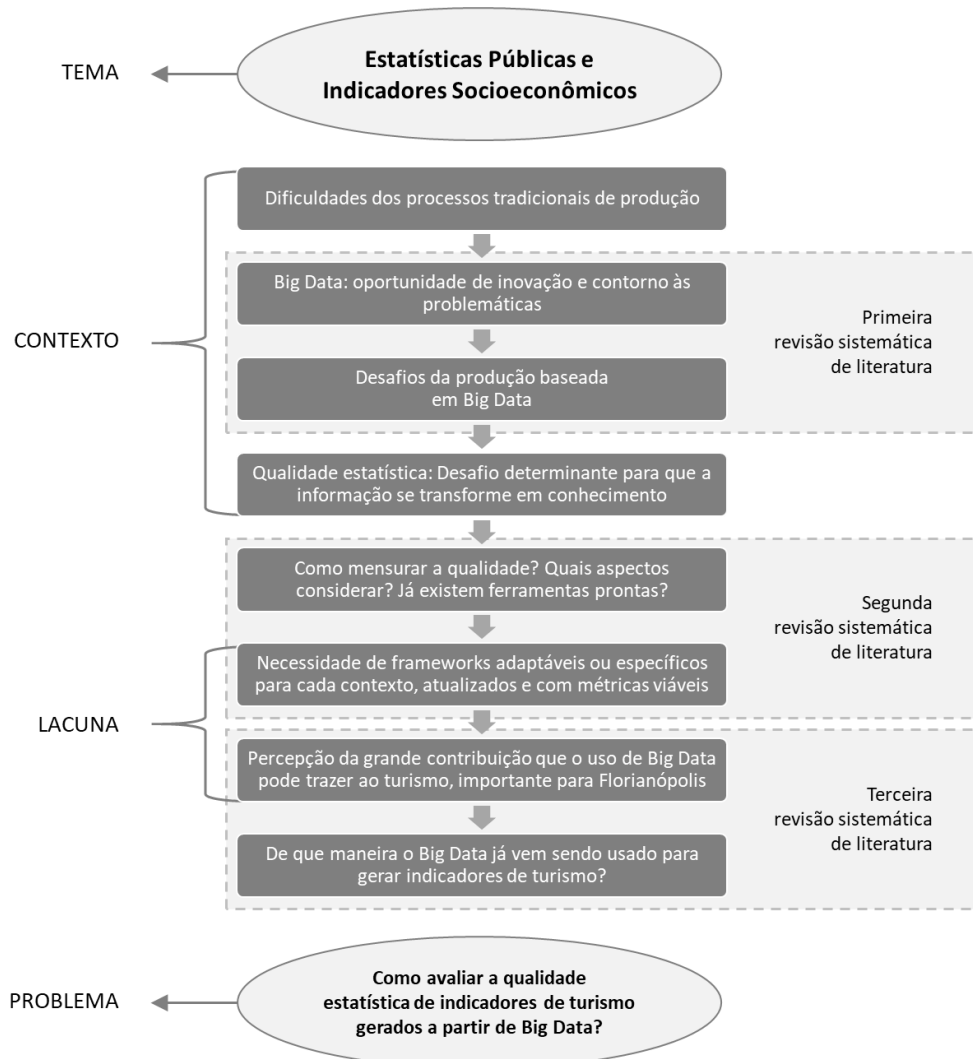
Partindo do tema estatísticas públicas e indicadores socioeconômicos, escolhido por sua relação com os históricos profissional e acadêmico deste autor, sintetizados na Seção 1.8, elencou-se uma série de dificuldades que afetam os atuais produtores dessas tipologias de produtos estatísticos em seus processos tradicionais (Subseção 2.3.3). Em seguida, percebeu-se que o contexto de Big Data pode trazer oportunidades de inovação na geração dos produtos, abrindo espaço para novas possibilidades de contorno às problemáticas tradicionais (Subseção 2.6.1).

Entretanto, a produção de estatísticas públicas e indicadores socioeconômicos baseada em Big Data também está sujeita a desafios (Subseção 2.6.3). Dentre eles, o central é o da garantia da qualidade, aspecto determinante da continuidade no uso dos produtos inovadores (Subseção 2.6.4). Durante a investigação sobre como qualidade de tais produtos vem sendo tratada por pesquisadores e especialistas, detectou-se que o tema ainda carece de aprofundamento de pesquisa e que a mensuração da qualidade depende das características do contexto de aplicação.

Nesse sentido, optou-se por iniciar delimitando uma área de interesse dentro dos indicadores socioeconômicos e estatísticas públicas. A escolha foi por focar nos indicadores de turismo (Seção 2.4). Como justificativa, a importância do turismo para

Florianópolis e a percepção de que o setor pode muito se beneficiar da utilização de Big Data em um contexto em que os usuários, pelo menos no Brasil, têm manifestado insatisfação perante as estatísticas tradicionalmente produzidas (Subseção 2.6.2).

Figura 5 – Resumo da sequência lógica percorrida até a definição do problema da pesquisa.



Fonte: elaborado pelo autor (2021).

Dessa forma, foi possível chegar ao seguinte problema de pesquisa: como avaliar a qualidade estatística de indicadores de turismo gerados a partir de Big Data?

O problema é aqui entendido como relevante por uma série de aspectos:

- a) aborda estatísticas e indicadores socioeconômicos, importantes no suporte a tomadas de decisão de agentes públicos e privados;
- b) foca no setor turístico, proeminente para Florianópolis, percebido como um dos possíveis grandes beneficiados com o uso de Big Data e carente de melhores estatísticas;
- c) apresenta oportunidades de geração de conhecimento a partir de novas fontes de dados;
- d) apresenta oportunidades de contorno às dificuldades enfrentadas nos processos tradicionais de produção de estatísticas públicas;
- e) evidencia a necessidade e subsidia alterações de posicionamento estratégico para os atuais produtores de estatísticas públicas;
- f) explora e apresenta soluções para um tema crucial para a efetivação dos indicadores em conhecimento, determinante do aproveitamento dos produtos estatísticos gerados a partir de Big Data: a qualidade.

4.2 CONSCIENTIZAÇÃO DO PROBLEMA

Para fins de aprofundamento na problemática e de especificação dos atributos que foram contemplados pelo artefato proposto, diferenciando-o do que já existe e preenchendo lacunas na demanda, esta seção apresenta os resultados da análise da literatura no que diz respeito às dificuldades que pesquisadores tem encontrado quando da avaliação da qualidade de indicadores socioeconômicos e estatísticas públicas produzidas a partir de Big Data.

A literatura destaca, conforme os trechos que sugerem a existência de lacunas de pesquisa, sintetizados no Quadro 11 da Subseção 2.6.4, que diversas questões atuais de pesquisa permanecem em aberto. Em suma, identifica-se a necessidade de alterações ou expansões nas ferramentas para análise da qualidade de produtos estatísticos gerados a partir de Big Data até então propostas, abordando tópicos relativos a:

- i. incorporação de novas dimensões e aspectos de qualidade;
- ii. operacionalização factível da mensuração da qualidade, com proposição de métricas focadas no domínio do produto e mais viáveis do ponto de

vista prático, especialmente para as dimensões acurácia e confiabilidade, coerência e comparabilidade;

- iii. estruturação dos *frameworks* em arcabouços flexíveis, adaptáveis para contextos de diferentes características, ou específicos, desenhados para as características de uma situação particular.

As Subseções 4.2.1, 4.2.2 e 4.2.3, trazem o detalhamento de cada um dos itens percebidos. Tais questões foram adereçadas, como tentativa de atendê-las, quando da construção do artefato proposto para solucionar o problema de pesquisa.

4.2.1 Dimensões e aspectos de qualidade

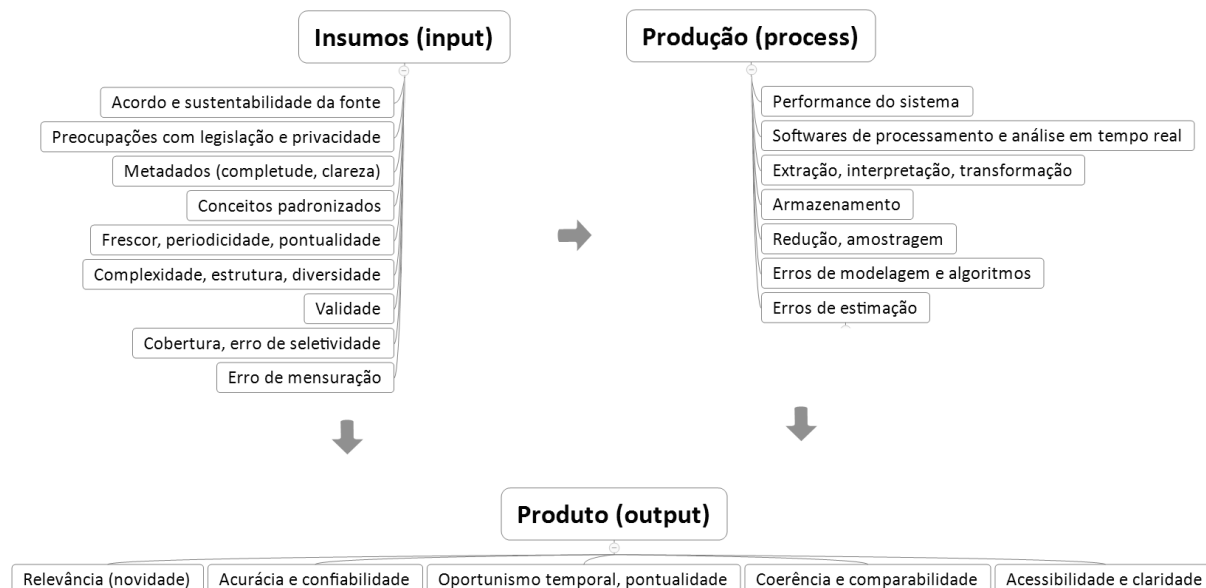
Conforme apresentou o Quadro 11, a literatura investigada sugere a necessidade de se considerar novas dimensões e aspectos quando da avaliação da qualidade de produtos estatísticos gerados a partir de Big Data (NASEM, 2017b; REIS et al., 2016).

Para a compreensão em profundidade deste item, o trabalho de Brancato e Di Consiglio (2018) pode ser utilizado como ponto de partida. Os autores listam as principais dimensões de qualidade identificadas na literatura e percorrem o trabalho de classificá-las, uma a uma, conforme apresenta a Figura 6, dentre domínios equivalentes aos considerados em diversos *frameworks* tradicionais de aferição da qualidade estatística: de (1) insumos e condições institucionais (*input*); de (2) processo de produção de estatísticas (*process*); e de (3) produtos estatísticos gerados (*output*). Além disso, as dimensões pertencentes aos domínios (1) e (2) são brevemente relacionadas com as dimensões do domínio dos produtos estatísticos – consideradas as mesmas reconhecidas pela ESS (EUROSTAT, 2019) e amplamente utilizadas em diversos *frameworks* tradicionais.

Percebeu-se que diversos autores corroboram as dimensões reunidas por Brancato e Di Consiglio (2018). Nem todos consideram todas as dimensões ao mesmo tempo, concentrando-se em esferas específicas. Alguns, por sua vez, mencionaram dimensões não explicitamente contempladas no resumo da Figura 6. A grande maioria, entretanto, possui alguma relação e/ou poderia ser facilmente

incorporada – tratada como elemento ou atributo operacional (sub-dimensão) – às dimensões de qualidade elencadas na figura apresentada.

Figura 6 – Relações entre dimensões de qualidade de insumos, processo de produção e produtos estatísticos gerados a partir de Big Data.



Fonte: traduzida e adaptada de Brancato e Di Consiglio (2018).

A tradicional dimensão relevância poderia abarcar os aspectos mínima granularidade, mencionado por NASEM (2017b) e Gurwitz (2012), e usabilidade, citado por UNECE (2014). Extração, interpretação e transformação, mencionadas na Figura 6, poderiam contemplar os elementos processos de gestão e governança de dados, com integração, pareamento, limpeza, processamento, ponderação e imputação de dados, citadas por uma série de autores investigados (VAJU; MESZAROS, 2018; THOMPSON, 2017; MASLANKOWSKI; NOWICKA, 2018; LAURO; TRAVERSO, 2018; HAND, 2018; STRUIJS; DAAS, 2014; NASEM, 2017b; BIFFIGNANDI; SIGNORELLI, 2015; UNECE, 2014; ULUWIYAH, 2016; KIM; CHU, 2017). Variabilidade, volatilidade, instabilidade e interrupção das fontes (SILVA, 2016; BIFFIGNANDI; SIGNORELLI, 2015; HAJNOVIC; 2018) entrariam em sustentabilidade da fonte, em linha com mudanças metodológicas que podem vir a acontecer (HAND, 2018) e culminar na perda da comparabilidade temporal. Comunicação de incertezas (HAND, 2018) e interpretabilidade (ABS, 2009; STATISTICS CANADA, 2019; HACKL, 2016; TAM; CLARKE, 2015b; NASEM, 2017b) entrariam em acessibilidade e clareza – possuindo relação também com os

metadados. Veracidade (BERGAMASCHI et al., 2016; SILVA, 2016), consistência (LAURO; TRAVERSO, 2018; TAM; CLARKE, 2015b), reprodutibilidade (STRUIJS; DE BROE, 2018) e auditabilidade (CAI; ZHU, 2015) seriam incorporadas tanto pela acurácia e confiabilidade quanto pela coerência e comparabilidade. Alta dimensionalidade, mencionada na publicação de Biffignandi e Signorelli (2015), poderia ser contemplada pelas dimensões redução, modelagem e/ou estimação. Eficiência, performance e infraestrutura computacional (AL-HAJJAR et al., 2015; TAM; CLARKE, 2015b; ULUWIYAH, 2016), assim como escalabilidade e interoperabilidade (MACFEELY, 2016; HILBERT, 2016), relacionam-se com as dimensões performance do sistema e softwares de processamento, tendo as duas últimas uma relação também com a coerência e a comparabilidade.

Por outro lado, a revisão de literatura realizada apontou para outros aspectos que parecem não diretamente se relacionar com as dimensões apresentadas na Figura 6. Autores como Hackl (2016) e Tam e Clarke (2015b) entendem o componente custo como um elemento de qualidade. Também na linha dos recursos utilizados, Batini et al. (2015) defende que para serem de qualidade, os produtos devem atentar ao aspecto da redundância – com minimalidade, compactude e concisão. Os aspectos transparência, integridade, independência e ambiente da instituição produtora (NASEM, 2017b; UNECE, 2014) também não estão explicitamente contemplados. E, por fim, a possibilidade de manipulação – no sentido pejorativo, de ação motivada por má fé – dos dados de algumas fontes também deve ser levada em consideração, conforme apontam Wirthmann e Reis (2018) e MacFeely (2016). A incorporação de tais aspectos merece ser pelo menos ponderada quando da proposição de ferramentas de aferição da qualidade de produtos estatísticos.

A título de conhecimento adicional, externo ao escopo das estatísticas públicas e oficiais, Fagundes, Macedo e Freund (2017) e Jaya et al. (2017) realizam revisões de literatura sobre o processo de transição da qualidade de dados para a de Big Data sob a ótica da ciência da informação. No âmbito da computação, Gudivada, Apon e Ding (2017) fazem uma importante discussão sobre as dimensões de qualidade de dados para Big Data e *Machine Learning*. Na tecnologia da

informação, Aracri et al. (2018) defendem uma abordagem baseada em ontologia para aferição da qualidade de dados.

4.2.2 Operacionalização da mensuração

No que diz respeito à operacionalização da aferição das dimensões da qualidade estatística, foram percebidos, conforme sintetizou o Quadro 11, diversos autores que apontam para a necessidade de se avançar na proposição de métricas ou indicadores mais viáveis do ponto de vista prático (BARCAROLI; GOLINI; RIGHI, 2018; BRANCATO; ASCARI, 2018; BRANCATO; DI CONSIGLIO, 2018; BRANCATO et al., 2018; HAJNOVIC, 2018; HAND, 2018; SALGADO et al., 2018; DE WAAL; DELDEN; SCHOLTUS, 2017; REINERT; STOLTZE, 2016; REIS et al., 2016).

Para Agafitei et al. (2015), a mensuração deve ocorrer exclusivamente com base nas dimensões do domínio *output*, apesar de reconhecerem que as demais, associadas aos insumos e aos processos de produção, também influenciam na qualidade final de produtos estatísticos. Os autores argumentam que a mensuração de todas as dimensões nos três domínios – *input*, *process*, *output* – frequentemente não se faz viável na prática do contexto de Big Data, especialmente quando diversas fontes de dados estão sendo integradas (*multisource*) durante a geração de um produto. E como bases de Big Data não são originalmente estruturadas para responder aos questionamentos inerentes ao contexto das estatísticas públicas, provavelmente necessitarão ser combinadas com dados de outras fontes (pesquisas do tipo *survey*, registros administrativos ou até mesmo outras bases de Big Data), fazendo com que a complexidade dos contextos de múltiplas fontes (*multisource*) seja corriqueira (REIS et al., 2016).

Ao defenderem sua proposta, os autores consideram as dimensões de qualidade de produto da ESS (EUROSTAT, 2019): (a) relevância, associada ao atendimento das necessidades e expectativas dos usuários por parte do produto; (b) acurácia e confiabilidade, associadas à capacidade do produto de retratar a realidade de forma precisa, não enviesada e confiável; (c) temporalidade e pontualidade, associadas a uma entrega rápida e no momento que se fizer útil; (d) coerência e comparabilidade, associadas a utilização de conceitos padronizados amplamente aceitos, permitindo combinação e comparação com estatísticas de

outras localidades e temporalidades; (e) e acessibilidade e clareza, associadas a produtos entregues de maneira conveniente, em formatos claros e com informações relevantes completas e compreensíveis.

Dentre essas, algumas dimensões podem ser operacionalmente aferidas em estatísticas produzidas a partir de Big Data da mesma forma que são em estatísticas tradicionais. As dimensões (a) relevância e (e) acessibilidade e clareza, por exemplo, podem ser igualmente percebidas e mensuradas em ambos os contextos – cabendo destacar que produtos gerados de Big Data provavelmente terão clareza prejudicada uma vez que podem ser originários de bases privadas. As dimensões (c) temporalidade e pontualidade são também igualmente percebidas e mensuradas em ambos os cenários, com expectativas de mais satisfatórias quando se trabalhando com Big Data. Assim, as dificuldades de mensuração da qualidade no âmbito dos produtos concentrar-se-ão nas dimensões (b) acurácia e confiabilidade e (d) coerência e comparabilidade, uma vez que são altamente impactadas pelas características dos institutos produtores e dos dados brutos (insumos), pelos processos de produção (métodos e procedimentos) e pelas práticas de integração de dados quando os produtos são agregados de fontes diferentes (*multisource*).

Apesar do cenário desafiador, De Waal, Delden e Scholtus (2017) propõem maneiras de mensuração da qualidade em termos de (b) acurácia e (d) coerência em seu *framework* para avaliação da qualidade de produtos estatísticos derivados de múltiplas fontes – considerando exclusivamente pesquisas do tipo *survey* e registros administrativos. Para lidar com a heterogeneidade das fontes, consideram seis diferentes configurações de bases de dados a partir dos aspectos: nível de agregação; unidades de pesquisa; variáveis mensuradas; cobertura; aspectos temporais; população alvo; e tipo de dado.

4.2.3 Frameworks específicos ou adaptáveis

Também com base nos autores elencados no Quadro 11, percebeu-se a necessidade de que artefatos construídos para orientação da avaliação da qualidade de produtos estatísticos gerados a partir de Big Data possuam estruturas ou flexíveis, adaptáveis para contextos de diferentes características, ou específicas,

desenhadas para as especificidades de uma situação particular (DE WALL; DELDEN; SCHOLTUS, 2019; BARCAROLI; GOLINI; RIGHI, 2018; HAJNOVIC, 2018; MASLANKOWSKI; NOWICKA, 2018; SALGADO et al., 2018; HACKL, 2016; REINERT; STOLTZE, 2016; AGAFITEI et al., 2015).

Como justificativa, coloca-se o argumento de que a mensuração da qualidade de estatísticas públicas e de indicadores socioeconômicos gerados a partir de Big Data depende das características do contexto em aplicação, não se fazendo possível estabelecer e observar um mesmo conjunto fixo de dimensões e aspectos de qualidade em todos os possíveis produtos estatísticos. Dependendo do produto, algumas dimensões e aspectos de qualidade não farão sentido, não sendo aplicáveis ou adequados.

Para compreensão deste item, há de se ter presente que a qualidade de um produto estatístico é resultado da combinação de diversos fatores: das características dos insumos (fontes e dados brutos de entrada), da forma como os processos de integração de múltiplas fontes de dados são conduzidos, da forma como o processo de produção estatística é conduzido, assim como do seu propósito final de utilização e de como é percebido por diferentes usuários – alinhamento às expectativas da demanda.

Por esses fatores, a problemática do estabelecimento de aspectos e dimensões definitivas de qualidade em um artefato orientador da análise da qualidade acaba por ser ainda maior no caso de produtos estatísticos gerados de maneira inovadora, utilizando-se de Big Data, comparativamente aos processos tradicionais.

Primeiramente, porque dados do tipo Big Data, utilizados como insumos, possuem uma variedade muito maior de formatos, estruturas e peculiaridades, conforme destacam Aracri et al. (2018). A isso, os autores acrescentam o frequente obscuro entendimento sobre qual semântica os dados do tipo Big Data trazem ou devem trazer – uma vez que são normalmente gerados para fins de atendimento à sua aplicação privada, por meio de processos não pensados ou planejados para fins de estatísticas públicas. Sobre esse aspecto, Loshin (2014) vai além e adiciona que um mesmo conjunto de dados pode possuir diferentes significados em distintos contextos. Sinalizações de que, além das características dos insumos e dos processos utilizados, o propósito de utilização dos produtos e os objetivos de a

quem interessa, diferentes usuários ou produtores, devam também ser considerados durante a concepção do instrumento de aferição da qualidade estatística.

4.3 REVISÃO SISTEMÁTICA DA LITERATURA

As três revisões sistemáticas que subsidiaram as etapas de identificação (Seção 4.1) e conscientização (Seção 4.2) do problema, conduzidas conforme metodologia descrita na Subseção 3.1.1, estão com objetivos individuais e conclusões resumidos no Quadro 19.

Quadro 19 – Resumo dos objetivos e conclusões das revisões sistemáticas realizadas.

Revisão	Objetivo	Conclusões
1	Apresentar o estado da arte da produção científica sobre a utilização de Big Data na produção de estatísticas públicas no que diz respeito à identificação de oportunidades – com possíveis fontes de dados e aplicações práticas já propostas – e de desafios – com problemáticas e riscos – já percebidos ou previstos.	Considerar produzir estatísticas públicas a partir de processos inovadores e fontes alternativas se faz crucial e estratégico aos agentes atuais. Entretanto, desafios – como o da mensuração da qualidade dos produtos – se colocam como elementos de aprofundamentos de pesquisa.
2	Resumir a produção técnico científica sobre a aferição da qualidade de indicadores socioeconômicos produzidos a partir de Big Data, identificando dimensões e <i>frameworks</i> que têm sido considerados e propostos para tal finalidade.	<i>Frameworks</i> únicos e estáticos não são suficientes para todas as situações uma vez que as dimensões de qualidade dependem das características do contexto em aplicação. Além disso, por vezes se faz inviável mensurar a qualidade nas dimensões dos domínios <i>input</i> e <i>process</i> . Os artefatos já propostos ainda precisam evoluir em termos de contemplar a possibilidade de adaptabilidade/flexibilização, considerar novas dimensões e sugerir métricas mais viáveis.
3	Sintetizar a produção científica sobre o uso de dados do tipo Big Data na geração de indicadores de turismo no que concerne a possíveis fontes e aplicações práticas.	As principais fontes de dados percebidas foram aquelas ligadas à <i>World Wide Web</i> , sendo o Twitter a plataforma mais utilizada. Das fontes não ligadas à <i>WWW</i> , os dados de empresas de telefonia apareceram como os mais utilizados. Em termos de aplicações práticas, ou finalidades, as principais encontradas se referem: à estimação ou à projeção da quantidade ou demanda de turistas de entrada ou saída; ou à compreensão da satisfação, opinião, sentimentos e interesses manifestos.

Fonte: elaborado pelo autor (2021).

Através da primeira revisão, publicada no (i) livro “População, Território e Estatísticas Públicas”, comemorativo aos 20 anos da pós-graduação da Escola Nacional de Ciências Estatísticas, do IBGE (CAUMO, 2020), foi possível compreender a dimensão da aplicabilidade e da relevância do tema. Com a segunda, publicada no (ii) periódico Planejamento e Políticas Públicas do Instituto de Pesquisa Econômica Aplicada (CAUMO; SOUZA, 2021), conseguiu-se identificar lacunas de pesquisa atuais sobre o tema para subsidiar a construção do objetivo do presente trabalho. A terceira, veiculada na (iii) revista E-Tech, do Serviço Nacional de Aprendizagem Industrial (CAUMO; BIZ; SOUZA, 2020), assim como nos anais da 15ª Conferência Internacional de Gestão do Conhecimento e do 2º Seminário Internacional de Universidades Corporativas e Escolas de Governo (CAUMO; BIZ; SOUZA, 2019), viabilizou a compreensão de possibilidades de aplicação do tema no âmbito do setor turístico para que, posteriormente, pudesse-se avançar nos estudos de aferição da qualidade dos indicadores gerados.

4.4 IDENTIFICAÇÃO DOS ARTEFATOS E CLASSES DE PROBLEMAS

Mesmo que o artefato tenha sido proposto para atender à produção de indicadores inovadores de turismo, delimitação setorial em que não foram encontrados protocolos para a avaliação da qualidade estatística, espera-se que possa ser expandido ou adaptado para uma classe maior de problemas: a aferição da qualidade de quaisquer indicadores socioeconômicos ou estatísticas públicas gerados a partir de Big Data.

Para tal finalidade, os *frameworks* construídos para contextos de dados oriundos de registros administrativos correspondem àqueles que parecem mais se adequar frente aos tradicionais predominantemente desenhados para contextos de pesquisas amostrais do tipo *survey*, uma vez que fontes administrativas e de Big Data compartilham uma série de características (REIS et al., 2016).

Sobre dados derivados de registros administrativos, Reinert e Stoltze (2016) fazem uma extensa revisão – com testes práticos – em 13 proposições de métodos – orientações, *checklists*, *frameworks* – de avaliação da qualidade construídos entre 2009 e 2014 por instituições de diversos países e uniões continentais. Conforme os autores, todas as propostas possuem indicadores agrupados em dimensões e

hiperdimensões de qualidade semelhantes, com algumas apresentando um ou outro tipo diferente de categoria. Apesar disso, por mais que diversas propostas possuam dimensões e hiperdimensões comuns, a maioria possui indicadores próprios específicos, provavelmente pelo fato de que foram pensadas para contextos e estatísticas de interesse particular (BRANCATO et al., 2017).

Tal trabalho de Reinert e Stoltze (2016) faz parte de um conjunto de entregas do *ESSnet Quality of Multisource Statistics* (Komuso) que, conforme Brancato e Di Consiglio (2018), se trata de um projeto de pesquisa que está reunindo os principais modelos de aferição da qualidade de registros administrativos para compor um *framework* que relaciona as dimensões de qualidade dos outputs de processos estatísticos com as principais fontes de erro em contextos *multisource*, de dados advindos de múltiplas fontes (BRANCATO; ASCARI, 2018; DE WAAL; DELDEN; SCHOLTUS, 2020).

Entretanto, conforme aponta Reis et al. (2016), registros administrativos são constituídos em ambientes controlados, de finalidade preestabelecida, em contraposição ao processo de geração nem sempre controlado de Big Data, tornando necessária a constituição de *frameworks* específicos para indicadores gerados a partir de Big Data.

Em termos de artefatos – sob a forma de *frameworks* conceituais – já propostos especificamente para orientar a avaliação da qualidade no contexto de Big Data, foram percebidos os seguintes:

- i. *UNECE Big Data Quality Framework* (UNECE, 2014);
- ii. *Eurostat Accreditation Procedure for statistical data from Non-Official Sources* (EUROSTAT, 2014a);
- iii. *AAPOR Big Data Total Error Framework* (KREUTER et al., 2015)¹²;
- iv. *Conceptual Framework for Quality in Big Data* (BATINI et al., 2015);
- v. *Big Data Quality Assessment Framework* (CAI; ZHU, 2015).

O Quadro 20 apresenta um resumo comparativo das dimensões de qualidade contempladas em cada proposta, sendo as duas primeiras propostas diretamente associadas à produção de estatísticas públicas e as outras três mais genéricas. As cinco propostas aparecem sob a forma de anexos – de B até F.

¹² Atualizado e operacionalizado em Amaya, Biemer e Kinyon (2020).

Quadro 20 – Comparativo entre *frameworks* selecionados de qualidade para estatísticas oficiais e públicas produzidas a partir de Big Data.

Framework	Hiperdimensões	Domínios	Dimensões – Elementos
BDQF (UNECE, 2014)	a. Fonte b. Metadados c. Dados	i. Entrada (origem) ii. Meio (processo) iii. Saída (produto)	Ambiente institucional (i.a.; iii.a.)
			Privacidade e segurança (i.a.; iii.a.)
			Complexidade (i.b.; iii.b.)
			Completitude (i.b.)
			Usabilidade (i.b.)
			Coerência e consistência (i.b.; i.c.; iii.c.), e <i>Linkabilidade</i> (i.b.; ii.c.; iii.c.)
			Validade (i.b.; i.c.; iii.c.)
			Aspectos temporais (i.b.; iii.c.)
			Acurácia e seletividade (i.c.; iii.c.)
			Acessibilidade e clareza (iii.b.)
			Relevância (iii.b.)
APNOS (EUROSTAT, 2014a)	a. Fonte (conteúdo) b. Metadados c. Dados agregados d. Microdados e. Fonte (instituição)	Entrada (origem)	Ambiente institucional (i.a.; iii.a.)
			Uso esperado
			Usabilidade
		Saída (produto)	Cooperação
			Relevância
			Acurácia e confiabilidade
			Pontualidade e oportunismo temporal
			Coerência e comparabilidade
			Acessibilidade e clareza
			Novidade
BDTEF (KREUTER et al., 2015) ⁷	Não explícitas.	Geração (origem)	Sinal com ruído
			Sinal perdido
			Incompletude
			Seleção não aleatória
			Metadados faltantes
		ETL (extração, transformação, carregamento)	Especificação
			Combinação
			Codificação
			Edição
			Manipulação
		Análise	Integração de dados
			Acúmulo de ruídos
			Correlações espúrias
CFQBD (BATINI et al., 2015)	Não explícitas.	Não explícitas.	Endogeneidade incidental
			Acurácia – acurácia, correção, validade e precisão
			Completude – completude, pertinência e relevância
			Redundância – redundância, minimalidade, compactude e concisão
			Legibilidade – legibilidade, compreensibilidade, clareza e simplicidade
			Acessibilidade – acessibilidade e disponibilidade
			Consistência – consistência, coesão e coerência
			Veracidade – veracidade, credibilidade, confiabilidade e reputação
BDQAF (CAI; ZHU, 2015)	Não explícitas.	Não explícitas.	Disponibilidade – acessibilidade, temporalidade e autorização
			Usabilidade – definição/documentação, credibilidade e metadados
			Confiabilidade – acurácia, integridade, consistência, completude e auditabilidade
			Relevância – adequação ao uso
			Apresentação – legibilidade e estrutura

Fonte: atualizado de Caumo e Souza (2021).

Das propostas encontradas, o *framework* proposto por UNECE (2014) – *Big Data Quality Framework* (BDQF) – é reconhecido pelos próprios proponentes como preliminar e de cunho orientador às práticas de aferição da qualidade estatística. Foi idealizado a partir de uma adaptação dos *frameworks* tradicionais orientados para a qualidade de dados administrativos, considerando os domínios de *input*, *process* e *output*, com três hiperdimensões – fonte, metadados e dados. Sua estrutura sugere dimensões e indicadores que podem ser utilizados na avaliação da qualidade dos insumos mesmo antes dos dados disponíveis (nas hiperdimensões fonte e metadados), fornecendo a ideia de qualidade potencial dos indicadores e demais produtos estatísticos (REIS et al., 2016). No âmbito do processo, o *framework* não é objetivo, trazendo apenas princípios gerais, sem sugestão de indicadores.

A proposta de Eurostat (2014a), intitulada *Accreditation Procedure for statistical data from Non-Official Sources* (APNOS), também é uma extensão de *frameworks* para estatísticas oficiais geradas a partir de registros administrativos, passando a considerar qualquer outro tipo de dado secundário (REIS et al., 2016). Sua estrutura se utiliza de um procedimento subdividido em cinco estágios. O primeiro estágio pode ser conduzido sem a posse de dados e fornece uma percepção sobre a potencialidade em termos de cobertura, unidades e variáveis, temporalidade e frequência. O segundo estágio explora as possibilidades de obtenção dos dados, passando por níveis de agregação e formato. No terceiro, a qualidade dos dados é avaliada de forma concreta. O quarto estágio envolve uma análise de custo-benefício e de riscos, gerando orientações sobre a viabilidade da continuidade da utilização do Big Data. Por fim, o último corresponde ao firmamento de acordo de cooperação entre o provedor do Big Data e o produtor interessado.

O instrumento proposto por Kreuter et al. (2015), o *Big Data Total Error Framework* (BDTEF), é uma construção da *American Association for Public Opinion Research* (AAPOR), organização dedicada a estudos no campo da opinião pública – atitudes, normas, valores e comportamentos. Corresponde a uma extensão do *Total Survey Error* (TSE), passando a considerar fontes de erros específicas de Big Data (BRANCATO; DI CONSIGLIO, 2018). Esse considera especificamente a acurácia, em termos de viés e erro padrão, dos produtos gerados, avaliando as bases de dados em termos de possíveis erros presentes nas “linhas, colunas e células” em três

etapas: geração dos dados; extração, transformação e carregamento (*Extract, Transform, Load* – ETL); e análise. Recentemente foi atualizado em termos de componentes e elementos de qualidade que considera, além de trazer também uma aplicação de operacionalização prática – baseada em análises subjetivas e qualitativas de cada aspecto observado (AMAYA, BIEMER; KINYON, 2020).

Já Batini et al. (2015) propuseram o *Conceptual Framework for Quality in Big Data* (CFQBD), que mede a qualidade do ponto de vista do usuário e considera, além das dimensões, três coordenadas que orientam e determinam a avaliação da qualidade: o tipo dos dados, a fonte dos dados e o domínio de aplicação. Ou seja, os autores compreendem que os critérios para mensuração dependem do contexto para o qual estão sendo aplicados (FAGUNDES; MACEDO; FREUND, 2017). Destacam, por exemplo, que a acurácia depende do tipo do dado, realizando testes em mapas, dados relacionais, *linked open data* e textos não estruturados – percebendo impactos no que concerne à dimensão completude e outras, também. Citro (2014) corrobora esse entendimento ao afirmar que a mensuração de alguns componentes da acurácia baseados no TSE (não respostas, erro de mensuração, erro de processamento de dados, etc.), propostos por Biemer et al. (2014), não se aplicam a dados oriundos da interação humana com redes sociais virtuais e mecanismos de buscas na internet, por exemplo.

Assim como o CFQBD, a proposta de Cai e Zhu (2015), o *Big Data Quality Assessment Framework* (BDQAF), também foca na qualidade da perspectiva do usuário e da satisfação dos seus interesses. Os autores consideram cinco dimensões subdivididas em elementos e sugerem alguns indicadores para mensuração de cada aspecto. Além do *framework*, disponibilizaram também um *Quality Assessment Process for Big Data*.

A aplicabilidade e a suficiência dos *frameworks* identificados carecem ser mais profundamente testadas em situações reais diversas. O que se percebe de antemão é que nenhum cobre todas as dimensões e elementos levantados na Seção 4.2.1. Além disso, nenhum é voltado para o contexto específico das estatísticas de turismo e, por mais que apresentem sugestões de aspectos de qualidade a serem observados, nenhum operacionaliza uma proposta com quantificação do nível de qualidade do produto estatístico – questões essas

percebidas também pelos diversos estudos que sugerem a necessidade de ajustes ou expansões de tais *frameworks*, conforme mostra o Quadro 11 da Subseção 2.6.4.

4.5 PROPOSIÇÃO DE ARTEFATO

Em virtude das análises descritas nas seções anteriores, idealizou-se um artefato para atender ao objetivo geral desta tese com características de:

- a) ser um arcabouço primordialmente conceitual, capaz de orientar a construção de parecer sobre o nível de qualidade de um produto estatístico gerado a partir de Big Data;
- b) ser voltado à percepção do usuário, compreendendo a qualidade no âmbito da adequação do produto ao seu uso e focando, nesses termos, na sua avaliação sob a ótica das dimensões do domínio de produto – aqui consideradas as componentes da qualidade estatística;
- c) ser abrangente no que diz respeito aos elementos – aspectos de qualidade – que considera (*comprehensive*);
- d) ser específico no sentido de voltado para o contexto da produção de indicadores e estatísticas de turismo; e
- e) ser estruturado com base em (i) dimensão, (ii) requisito e (iii) domínio, definidos por:
 - i. um conceito abstrato associado a um princípio de qualidade desejado para um produto estatístico, operacionalizado por um grupo de requisitos que, quando observados conjuntamente, se propõem a retratar seu nível de qualidade;
 - ii. um atributo, característica, componente ou elemento, que se propõe a operacionalizar, em conjunto com outros atributos, uma dimensão de qualidade, tornando-a tangível, observável ou mensurável, e que, portanto, se espera estar compreendido pelo produto estatístico para fins de atendimento às expectativas dos usuários ou *stakeholders*;
 - iii. o ambiente, ou a etapa, de pertencimento do requisito no fluxo de produção estatística.

Destaca-se que, além da sua estrutura básica, conceitual, alinhada com a expectativa do objetivo geral desta tese, o *framework* foi desenvolvido acompanhado, adicionalmente, de um protótipo, um ponto de partida para estudos futuros, de protocolo de operacionalização prática, com intuito de se contribuir também com pesquisas que focarão em aprimorar uma forma mais prática e factível de mensuração da qualidade estatística em produtos gerados a partir de Big Data – um dos aspectos percebidos como lacuna dentro do tema durante a realização das revisões bibliográficas.

Nesses termos, parte-se do entendimento do conceito de qualidade estatística como a capacidade de atender as necessidades dos usuários no âmbito da combinação de conceitos abstratos correspondentes a princípios desejados (dimensões) de um produto estatístico. Ou seja, a qualidade estatística é um conceito multidimensional, sendo aqui considerada o produto da composição das dimensões preconizadas pelo *European Statistics Code of Practice* (EUROSTAT, 2017), seguindo padronização internacional.

A tentativa de materializar tais dimensões abstratas de qualidade se dá por meio da observação de uma série de aspectos associados à qualidade de um produto estatístico, aqui ditos requisitos. Ou seja, os requisitos de qualidade correspondem a uma tentativa de definição operacional, prática e mensurável, de uma dimensão de qualidade. A relação de requisitos considerada no *framework* proposto tem por finalidade a ser a mais abrangente possível (*comprehensive*), contemplando todos os fatores determinantes da qualidade estatística mencionados por pesquisadores percebidos ao longo da construção desta tese. Cabe destacar que os componentes chamados hiperdomínios, considerados em alguns *frameworks* analisados, estão implicitamente contemplados sob a forma de requisitos de qualidade, por razões de tentativa de redução de complexidade – sem perda de conteúdo.

Tais requisitos de qualidade podem ser observados em diferentes ambientes de pertencimento, ou etapas, no âmbito do fluxo de produção estatística, ditos domínios. Os domínios aqui considerados são insumos (entrada), produção (processo) e produto, seguindo uma estrutura já consolidada e aceita quando da aferição da qualidade estatística.

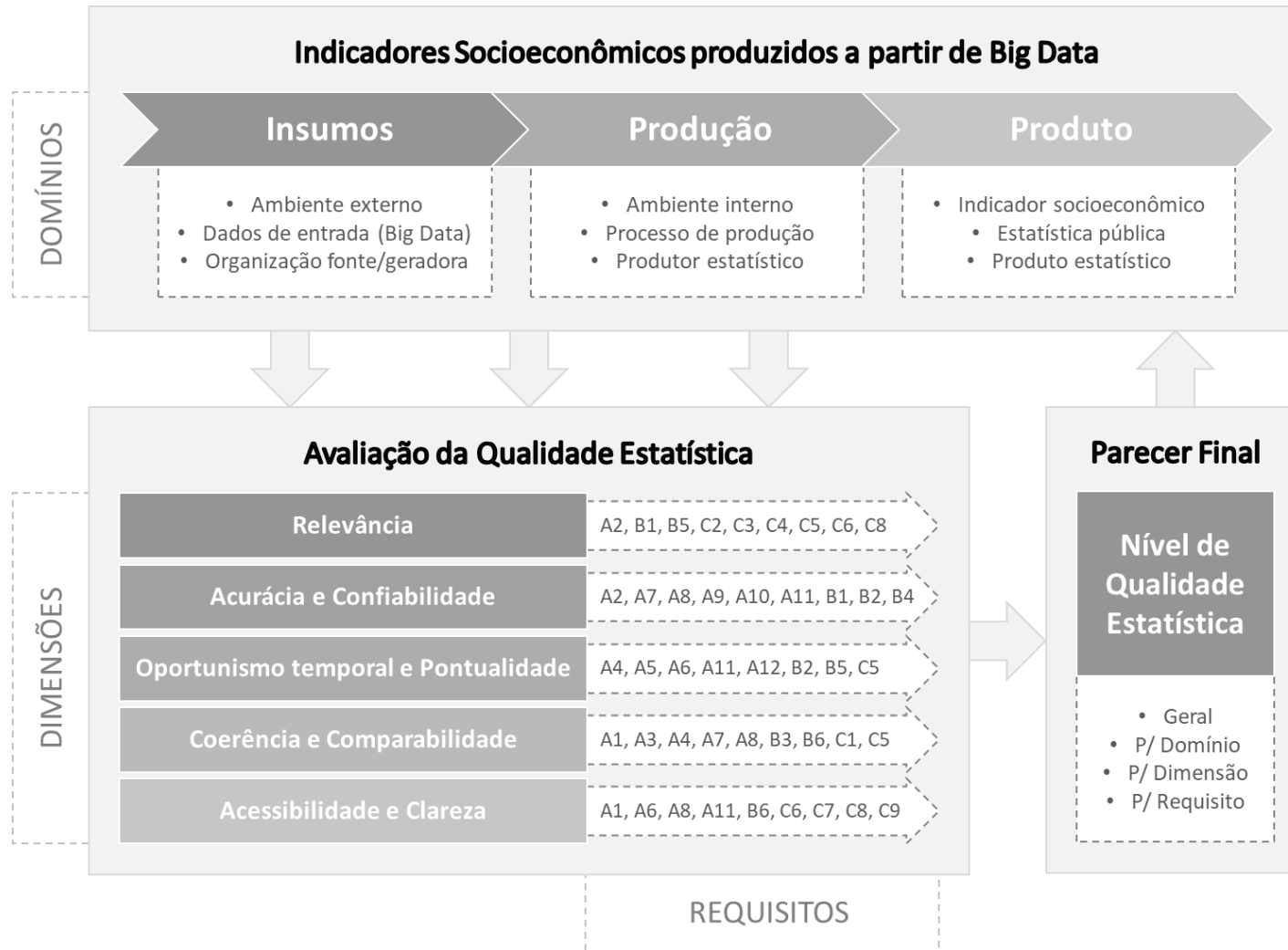
4.6 PROJETO DO ARTEFATO SELECIONADO

Para a construção de um parecer sobre o nível de qualidade de um indicador de turismo gerado a partir de Big Data, o artefato sugere um protocolo norteador da análise de uma série abrangente de requisitos de qualidade, considerando a observação ou mensuração quanto a presença ou intensidade de atributos ou características de qualidade no produto estatístico em aferição.

Alguns dos requisitos possuem relação com o domínio dos insumos (entrada), outros com o domínio da produção (processo) e outros com o do produto. Todos, entretanto, são avaliados somente no que diz respeito ao impacto que causam nas dimensões associadas ao produto final – ou seja, na qualidade estatística. Da mesma forma, cada uma das dimensões da qualidade estatística é representada por alguns requisitos somente – que podem variar dependendo do produto estatístico e da expectativa do usuário final.

Assim, conforme representa visualmente a Figura 7, a qualidade estatística – qualidade final do produto – é conceito resultante da composição das 5 dimensões/princípios de qualidade do produto, sendo a qualidade de cada dimensão operacionalizada no ambiente prático pela observação de uma série de requisitos de qualidade, que podem estar associados aos domínios de insumos (entrada), de produção (processo) ou de produto. Dessa forma, se faz possível uma representação aproximada acerca da qualidade de um produto estatístico em nível geral (qualidade estatística), subdividida por dimensões ou segmentada por domínios do fluxo de produção estatística – referente aos insumos, processos de produção e/ou produto final.

Figura 7 – Estrutura e funcionamento do *framework* desenvolvido.

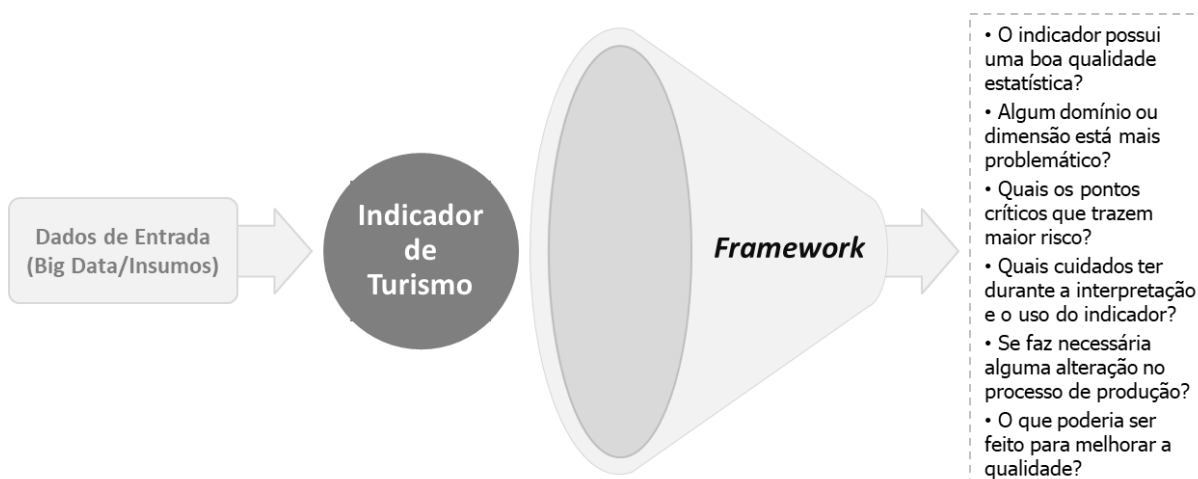


Fonte: elaborado pelo autor (2021).

Ou seja, a lista dos requisitos de qualidade é base de uma matriz de informações que os relaciona, indicando pertencimento ou influência, a um domínio e a uma ou mais possíveis dimensões. Após a observação dos requisitos na prática, a matriz direciona a construção de pareceres de qualidade completos e parciais – por dimensões e domínios.

Na posse dos pareceres de qualidade, esforços podem ser empregados para melhoramento e correção de falhas nos domínios e dimensões mais críticos, de menor qualidade, prioritários, conforme processo sintetizado pela Figura 8. Dessa forma, o *framework* não somente contribui com o apontamento de cuidados e potenciais riscos que permeiam o indicador e devem ser considerados durante sua utilização e interpretação por parte do usuário final, como também orienta produtores estatísticos no âmbito da condução dos seus processos.

Figura 8 – Síntese do processo e resultados esperados da aplicação do *Framework*.



Fonte: elaborado pelo autor (2021).

4.7 DESENVOLVIMENTO DO ARTEFATO

O artefato solução está estruturado sobre as cinco dimensões da qualidade estatística definidas no Quadro 21 – compreendidas da mesma forma que Eurostat (2017) e UN (2019), referências mundiais em estudos e práticas associadas à produção de estatísticas públicas.

Quadro 21 – Definições das dimensões de qualidade consideradas no *framework* proposto.

Dimensão	Definição
Relevância	O quão bem o indicador de turismo satisfaz a necessidade do usuário (EUROSTAT, 2017). Os indicadores devem satisfazer atuais e/ou emergentes necessidades ou requisitos dos usuários. Trata-se de um critério subjetivo que dependerá da necessidade do usuário e que coloca aos produtores o desafio de ponderar as necessidades mais importantes e de maior prioridade de atuais e potenciais usuários, considerando as restrições dos recursos disponíveis (UN, 2019). Ou seja, por relevância entende-se como o quanto o indicador é capaz de satisfazer a necessidade dos usuários (UN, 2019).
Acurácia e Confiabilidade	A quão acurada e confiável é a forma em que o indicador de turismo consegue representar a realidade (EUROSTAT, 2017). Os produtores devem desenvolver, produzir e disseminar indicadores que retratam a realidade de forma acurada e confiável, com a acurácia refletindo o grau em que o indicador corretamente descreve o fenômeno que se propõe a mensurar, ou, em outras palavras, a proximidade dos resultados aos verdadeiros valores (UN, 2019). Dessa forma, por acurácia entende-se como a proximidade do valor do indicador ao verdadeiro valor numérico da estatística que se propõe a mensurar, enquanto por confiabilidade se refere a proximidade do valor do indicador inicialmente calculado para o valor subsequente, se os valores preliminares forem divulgados (UN, 2019).
Oportunismo temporal e Pontualidade	O quão atual e pontual é o indicador de turismo divulgado (EUROSTAT, 2017). Os produtores devem minimizar a demora entre os períodos de referência e a divulgação dos indicadores, com oportunismo temporal se referindo ao quão rapidamente entre os dois momentos os dados são disponibilizados aos usuários e pontualidade se referindo a se os resultados são entregues nas datas prometidas ou anunciadas (UN, 2019). Dessa forma, oportunismo temporal está associado ao intervalo de tempo entre o final do período de referência do indicador e momento da divulgação do indicador, com a pontualidade se referindo ao intervalo de tempo entre o momento da disponibilização do indicador e a data em que deveria ter sido divulgado (UN, 2019).
Coerência e Comparabilidade	O quão consistente internamente e ao longo do tempo, e comparável entre regiões e países, é o indicador de turismo, possibilitando combinar e utilizar conjuntamente dados relacionados provenientes de diferentes fontes (EUROSTAT, 2017). Os produtores devem desenvolver, produzir e disseminar indicadores consistentes no sentido de permitirem a combinação e o uso conjunto com dados relacionados, inclusive derivados de outras fontes. Além disso, que possibilitem comparações ao longo do tempo e entre diferentes áreas, países e regiões (UN, 2019). Dessa forma, a coerência correspondem à habilidade de se confiavelmente combinar o indicador com outras bases de dados de diferentes maneiras e para diversos usos, tendo o termo consistência frequentemente utilizado como sinônimo, enquanto a comparabilidade se refere ao quanto as diferenças nos valores do indicador entre distintas áreas geográficas, domínios e momentos no tempo, refletem unicamente as disparidades entre os verdadeiros valores do indicador nos contextos em comparação (UN, 2019).
Acessibilidade e Clareza	O quão claro e compreensível é o indicador de turismo, divulgado de um modo adequado e conveniente, disponível e acessível de forma imparcial, acompanhado de metadados e outras informações de apoio e orientação (EUROSTAT, 2017). Os produtores devem garantir que os indicadores que desenvolvem, produzem e disseminam sejam encontrados e obtidos sem dificuldades, apresentados de forma clara e compreensível, disponíveis e acessíveis a todos os diferentes usuários de forma igualitária e imparcial em diferentes e convenientes formatos alinhados com os princípios dos dados abertos. Além disso, com providências tomadas para que microdados possam ser acessados para fins de pesquisa, respeitadas as políticas estabelecidas de confidencialidade estatística (UN, 2019). Ou seja, o conceito de acessibilidade se refere à facilidade e às condições em que o indicador pode ser obtido, com clareza correspondendo à disponibilidade de documentação apropriada sobre o indicador e à qualquer tipo de assistência adicional aos usuários (UN, 2019).

Fonte: elaborado pelo autor (2021) com base em Eurostat (2017) e UN (2019).

Os domínios considerados, (i) insumos (entrada), (ii) produção (processo) e (iii) produto, por sua vez, aqui se referem:

- i. ao ambiente externo ao da produção do indicador de turismo, às características e ao contexto das organizações geradoras do Big Data e dos dados brutos de entrada (Big Data/insumos);
- ii. ao ambiente interno ao da produção do indicador de turismo, às características e ao contexto da instituição produtora estatística e dos métodos, técnicas e procedimentos utilizados durante a produção do indicador quando da manipulação dos dados brutos de entrada;
- iii. ao ambiente do objeto resultante do processo de produção do indicador de turismo, às características e atributos do produto finalizado.

Já os requisitos, componentes principais do *framework*, foram definidos inicialmente mediante análise de outros artefatos já existentes e de resultados de pesquisas e publicações sobre o tema, produto de uma das revisões sistemáticas de literatura aqui realizadas, posteriormente melhorados, verificados e consolidados durante a aplicação de questionários em especialistas (Seção 4.8).

Ao final deste processo, alcançou-se uma quantidade de 27 requisitos – com definições descritas no Quadro 23 – entendidos como relacionados de forma direta às dimensões e pertencendo aos domínios considerados conforme apresenta o Quadro 22.

Destaca-se que um requisito pode-se relacionar direta e indiretamente com as dimensões de qualidade e que, por vezes, o não atendimento de um requisito por parte de um produto estatístico – como uma estatística pública – pode gerar impactos indiretos até mesmo em todas as dimensões. Por exemplo, se os insumos (Big Data) apresentarem alta complexidade estrutural, como pode acontecer em informações textuais livres, dados não estruturados, o requisito de baixa complexidade estrutural não será atendido e conseqüentemente a qualidade do indicador poderá ser afetada indiretamente em todas as dimensões, pois poderá haver maior chance de erros de modelagem e inferência, impactando na acurácia, maior necessidade por tempo de processamento, interferindo no aspecto do oportunismo temporal, além de maior espaço para interpretações errôneas e conceitos mal definidos e/ou ajustados, com influência sobre a coerência e a comparabilidade, a clareza e, conseqüentemente, a usabilidade e a relevância do produto estatístico.

Quadro 22 – Relação entre domínios, requisitos e dimensões da qualidade estatística do *framework* proposto.

Domínio	Requisito		Dimensão da Qualidade Estatística				Possíveis Sub Dimensões	
	Código	Descrição	Relevância	Acurácia e Confiabilidade	Oportunismo temporal e Pontualidade	Coerência e Comparabilidade		Acessibilidade e Clareza
Insumos	A1	Base legal				X	X	Continuidade; Legalidade.
	A2	Credibilidade dos insumos	X	X				Credibilidade; Veracidade; Verificabilidade; Reputação; Natureza.
	A3	Sustentabilidade dos insumos				X		Continuidade; Consistência.
	A4	Acordo formal de cooperação			X	X		Continuidade; Periodicidade; Legalidade.
	A5	Imediatismo dos insumos			X			Imediatismo; Atualidade; Frescor.
	A6	Facilidade de aquisição			X		X	Disponibilidade; Alcançabilidade; Custo; Aquisição.
	A7	Padronização dos conceitos		X		X		Consistência; Padronização.
	A8	Metadados dos insumos		X		X	X	Transparência; Documentação; Auditabilidade.
	A9	Representatividade da amostra		X				Representatividade; Seletividade; Completude; Esparsidade.
	A10	Aderência dos registros		X				Exatidão; Correção; Validade; Precisão; Adequação.
	A11	Complexidade estrutural		X	X		X	Complexidade; Linkabilidade; Legibilidade; Compreensibilidade; Simplicidade.
	A12	Otimização dos insumos			X			Redundância; Minimalidade; Compactação; Concisão.
Produção	B1	Credibilidade do produtor	X	X				Credibilidade; Reputação; Autoridade.
	B2	Infraestrutura e recursos humanos		X	X			Infraestrutura; Recursos.
	B3	Padronização dos processos				X		Consistência; Padronização.
	B4	Rigor dos processos		X				Robustez; Calibração; Ajuste; Solidez metodológica; Adequação.
	B5	Agilidade da produção	X		X			Agilidade; Automatização.
	B6	Documentação dos processos				X	X	Transparência; Documentação; Auditabilidade.
Produto	C1	Padronização do indicador				X		Consistência; Padronização.
	C2	Adequação conceitual	X					Adequação; Utilidade.
	C3	Grau de inovação	X					Inovatividade.
	C4	Percepção ética	X					Ética; Moral; Reputação.
	C5	Periodicidade praticada	X		X	X		Periodicidade; Regularidade.
	C6	Metadados do indicador	X				X	Transparência; Documentação.
	C7	Capacitação para uso					X	Instrução; Suporte; Capacitação; Orientação.
	C8	Disponibilidade do indicador	X				X	Disponibilidade; Alcançabilidade; Custo; Aquisição.
	C9	Disseminação e promoção					X	Facilidade; Disseminação; Divulgação; Promoção.
Total		27	9	9	8	9	9	

Fonte: elaborado pelo autor (2021).

Entretanto, para demarcar as relações diretas principais, elencadas no Quadro 22, utilizou-se como critério a análise de uma pergunta aplicada a cada uma das cinco dimensões: “Quais são os requisitos essenciais devem ser atendidos para que o indicador tenha boa qualidade em relação ao que propõe esta dimensão?”. Seguindo este referencial, avaliou-se que cada uma das cinco dimensões teria seu nível de qualidade aferido a partir da observação dos requisitos conforme a seguinte estrutura:

- i. Relevância (9) – Credibilidade dos insumos, Credibilidade do produtor, Agilidade da produção, Adequação conceitual, Grau de inovação, Percepção ética, Periodicidade praticada, Metadados do indicador, Disponibilidade do indicador;
- ii. Acurácia e Confiabilidade (9) – Credibilidade dos insumos, Padronização dos conceitos, Metadados dos insumos, Representatividade da amostra, Aderência dos registros, Complexidade estrutural, Credibilidade do produtor, Infraestrutura e recursos humanos, Rigor dos processos;
- iii. Oportunismo temporal e Pontualidade (8) – Acordo formal de cooperação, Imediatismo dos insumos, Facilidade de aquisição, Complexidade estrutural, Otimização dos insumos, Infraestrutura e recursos humanos, Agilidade da produção, Periodicidade praticada;
- iv. Coerência e Comparabilidade (9) – Base legal, Sustentabilidade dos insumos, Acordo formal de cooperação, Padronização dos conceitos, Metadados dos insumos, Padronização dos processos, Documentação dos processos, Padronização do indicador, Periodicidade praticada;
- v. Acessibilidade e Clareza (9) – Base legal, Facilidade de aquisição, Metadados dos insumos, Complexidade estrutural, Documentação dos processos, Metadados do indicador, Capacitação para uso, Disponibilidade do indicador, Disseminação e promoção.

As possíveis sub dimensões trazidas pelo Quadro 22 se referem a aspectos de qualidade desejáveis em produtos estatísticos – percebidos nas revisões bibliográficas realizadas – que são utilizados de forma sinônima aos requisitos aqui propostos ou se apresentam em nível de detalhamento mais aprofundado dos elementos aqui considerados, se fazendo contemplados, cobertos, ou influenciados pelo que preconiza o requisito aos quais estão associados. Dessa forma, percebe-se

que o atendimento ou não de um requisito pode impactar em um ou mais micro aspecto de qualidade desejável para um produto estatístico.

No geral, o que se pode perceber é uma maior influência dos requisitos associados aos domínios de insumos e de processo nas dimensões acurácia e confiabilidade, oportunismo temporal e pontualidade, e coerência e comparabilidade, enquanto os requisitos de produto impactam majoritariamente nas dimensões relevância e acessibilidade e clareza – conforme esperado. As definições de cada um dos 27 requisitos estão apresentadas no Quadro 23.

Quadro 23 – Definições dos requisitos considerados no *framework* proposto.

Código	Referência	Definição
A1	Base legal	Se refere à necessidade de que exista uma legislação local estabelecida assim como os termos de consentimento, políticas de privacidade e termos de uso do Big Data que respeitem possibilidades, restrições e obrigações quando da manipulação dos dados de entrada (Big Data/insumos) e da divulgação do indicador de turismo, passando por aspectos de produção/coleta, armazenamento, manutenção, acesso, técnicas de análise e processamentos, finalidade, segurança, confidencialidade, privacidade, propriedade e direitos autorais, etc., garantindo autorização e legalidade operacional tanto à fonte produtora do Big Data quanto à organização produtora do indicador de turismo. O não atendimento deste requisito diminui a qualidade do indicador ao propiciar uma eventual descontinuidade nos dados de entrada e na divulgação do indicador final.
A2	Credibilidade dos insumos	Se refere à necessidade de que os dados de entrada (Big Data/insumos) derivem de instituições com autoridade e boa reputação, transparentes, íntegras e independentes, e sejam verdadeiros, críveis e verificáveis - transmitindo confiança, livres de "fake news" e eventuais manipulações ou direcionamentos propositais. Para observação desse ponto pode ser interessante atentar para a natureza da fonte dos insumos e seu processo de geração de dados (se de redes sociais e gerados por seus utilizadores humanos, se de sensores e gerados por máquinas, se registros de organização oficial, etc.) e para a existência de algum órgão autônomo auditor, regulador ou atestador da veracidade da informação. O não atendimento deste requisito coloca em dúvida a qualidade do indicador, impactando principalmente em sua usabilidade e, conseqüentemente, sua relevância.
A3	Sustentabilidade dos insumos	Se refere à necessidade de que a fonte dos dados de entrada (Big Data/insumos) seja sustentável ao longo do tempo, no sentido de não interromper a geração ou a disponibilização dos dados (por encerramento das atividades, orientação estratégica, etc.), não ser volátil, instável ou com risco de disrupção, e não modificar métodos ou processos que possam gerar inconsistências ou descontinuidade incorrigível/incontornável de teor, conteúdo, formato ou semântica nos dados. O não atendimento deste item prejudica a qualidade do indicador de turismo ao permitir que séries temporais sejam rompidas e comparações e análises no tempo se restrinjam.
A4	Acordo formal de cooperação	Se refere à necessidade de existência de um acordo de cooperação formalizado entre a fonte do Big Data e a organização produtora do indicador de turismo, destacando objetivos finais, conteúdo das entregas, permissões e restrições, contrapartidas, frequência e cronogramas, etc. O não atendimento deste item diminui a qualidade do indicador ao deixá-lo suscetível a atrasos na sua divulgação, inconsistências (alterações de conteúdo), interrupções, descontinuidades.

A5	Imediatismo dos insumos	Se refere à necessidade de que os dados de entrada (Big Data/insumos) sejam registrados no momento de sua ocorrência (período a que se referem) e cheguem (sejam transferidos/repassados) à organização produtora do indicador de turismo imediatamente, em tempo real. A demora entre o período de referência do acontecimento e o seu registro pode aumentar a quantidade de erros de registro e, assim como a demora na transferência para a organização produtora do indicador, pode tornar o indicador final demasiadamente defasado - e conseqüentemente sem utilidade prática.
A6	Facilidade de aquisição	Se refere à necessidade de que os dados de entrada (Big Data/insumos) sejam acessados, pegos, capturados ou, quando for o caso, comprados, a um baixo custo - financeiro e operacional. O não atendimento deste item diminui a qualidade do indicador ao dificultar sua produção, podendo torná-lo inviável ou irrelevante por questões temporais ou de recursos - financeiros ou operacionais.
A7	Padronização dos conceitos	Se refere à necessidade de que definições (sobre domínios, conceitos, construtos, elementos, classificações, etc.) utilizadas pela fonte do Big Data sigam padrões reconhecidos e aceitos perante os stakeholders, nacional e internacionalmente, com consistência temporal. O não atendimento deste requisito reduz a qualidade do indicador à medida que pode restringir sua comparabilidade entre diferentes períodos de tempo, organizações produtoras, regiões e países.
A8	Metadados dos insumos	Se refere à necessidade de existência de metadados, seguindo padrões e classificações consistentes e uniformes de auditabilidade, que descrevam por completo e de forma clara tanto o conteúdo (terminologia, conceitos, significados, escopo e cobertura, período de referência, especificação geográfica, etc.) quanto o formato dos dados de entrada, isto é, do Big Data que está sendo utilizado como insumo para a produção dos indicadores de turismo. O não atendimento deste requisito prejudica a qualidade do produto estatístico à medida que torna obscura a detecção de erros nos insumos (impactando na acurácia e na confiabilidade) e propicia interpretações equivocadas ou distorcidas, dificultando por conseqüência a construção da documentação sobre o uso adequado do indicador.
A9	Representatividade da amostra	Se refere à necessidade de que os dados de entrada (Big Data/insumos) contemplem a população alvo por completa (censo) ou, quando de amostra, não apresentem problemas imensuráveis ou incorrigíveis de amostragem viesada, seletividade dos registros ou falhas de cobertura, tanto no âmbito de parcelas populacionais quanto em termos temporais e espaciais, estando completos, não esparsos, sem problemas de registros/linhas/elementos duplicados, omissos/faltantes ou não pertencentes à população alvo de interesse ou de variáveis/colunas/atributos com valores faltantes (<i>missing</i>). O não atendimento deste requisito diminui a qualidade do indicador de turismo ao dificultar o processo de estimação, podendo inserir vieses de tendenciosidade no resultado, prejudicando a acurácia.
A10	Aderência dos registros	Se refere à necessidade de que os dados de entrada (Big Data/insumos) sejam aderentes e reflitam a realidade de interesse de forma correta, exata, precisa e consistente, não possuindo erros de coleta/registro/tabulação derivado de processamento (integração, codificação, edição, etc.), de imputação ou de geração/mensuração - com instrumento de coleta/captura/mensuração garantido e confiável, com validade de conteúdo ou construto. O não atendimento deste requisito diminui a qualidade do indicador de turismo ao inserir vieses nem sempre conhecidos no resultado, prejudicando a acurácia e a confiabilidade, propiciando análises e conclusões distorcidas em relação ao objeto de análise.
A11	Complexidade estrutural	Se refere à necessidade de que os dados de entrada (Big Data/insumos) não sejam complexos, sejam bem estruturados, especificados e legíveis, com rótulos e conteúdos das variáveis/atributos claros e corretamente definidos, sem ambiguidades e com facilidade de compreensão, inclusive semântica, contendo todas as variáveis chaves necessárias para <i>linkagem</i> e integração de bases de dados. O não atendimento deste requisito diminui a qualidade do indicador de turismo ao criar dificuldades de processamento, modelagem e inferência, podendo culminar em erros de <i>linkagem</i> e mensuração, inserindo vieses nem sempre conhecidos no resultado, prejudicando a acurácia e a confiabilidade, interferindo na clareza e no tempo de produção do indicador de turismo.

A12	Otimização dos insumos	Se refere à necessidade de que os dados de entrada (Big Data/insumos) se restrinjam ao conjunto mínimo de informações necessárias para representar a realidade de interesse, compactos, concisos e sem redundâncias. O não atendimento deste requisito pode diminuir a qualidade do indicador de turismo quando criar dificuldades de processamento, requerendo o investimento de uma quantidade maior de recursos, impactando no custo e na temporalidade do indicador.
B1	Credibilidade do produtor	Se refere à necessidade de que o indicador de turismo seja produzido e disponibilizado por uma organização especializada nacional ou internacionalmente, do ramo, com autoridade e boa reputação, transparente, íntegra e independente, com produtos assinados por profissionais especialistas reconhecidos. O não atendimento deste requisito diminui a qualidade do indicador de turismo ao reduzir sua credibilidade, impactando na sua utilização, afetando sua relevância.
B2	Infraestrutura e recursos humanos	Se refere à necessidade de que a organização produtora do indicador de turismo disponha de recursos humanos e não humanos para propor, conduzir e manter os processos conforme o esperado. Em relação aos recursos humanos, se faz necessário que haja profissionais - especialistas de domínio, pesquisadores, estatísticos, cientistas de dados e da computação, administradores de sistema - com habilidades e recursos para manter os processos sem erros. Na parte de infraestrutura de tecnologia da informação, pode ser necessário que sistemas e componentes apresentem performance suficiente (disponibilidade, tempo de resposta, processamento, eficiência, escalabilidade, interoperabilidade) para lidar com dados complexos em processos de integração e transformação. O não atendimento deste requisito prejudica a qualidade do indicador ao torná-lo suscetível a erros de estimação e atrasos de entrega/divulgação.
B3	Padronização dos processos	Se refere à necessidade de que os métodos, técnicas, protocolos e procedimentos (coleta, pré processamento, análise da qualidade, codificação, edição, ETL, modelagem e de integração de dados, amostragem ou redução de dados, algoritmos de inferência, mineração, etc.) utilizados pelo produtor durante todas as etapas da concepção do indicador de turismo sigam padrões reconhecidos e aceitos perante os <i>stakeholders</i> , nacional e internacionalmente, claros e consistentes temporalmente, que não modifiquem os conceitos e domínios inerentes aos dados de entrada. O não atendimento deste requisito reduz a qualidade do indicador à medida que pode restringir sua comparabilidade entre diferentes períodos de tempo, organizações produtoras, regiões e países.
B4	Rigor dos processos	Se refere à necessidade de que os métodos, técnicas, protocolos e procedimentos sejam robustos e sólidos metodologicamente, contemplando procedimentos de gestão, controle, checagem e verificação de consistência, para garantir coleta, pré processamento e ETL, <i>parsing</i> , validação, análise da qualidade, curadoria, governança, tradução, limpeza, codificação, recodificação, agregação, edição, modelagem e integração de dados que não causem erros de registro ou mensuração, assim como com amostragem ou redução de dados ou dimensionalidade que não diminuam a representatividade e algoritmos e modelos de inferência, ponderação, imputação e mineração que não causem erros de estimação e de pressupostos. O não atendimento deste requisito diminui a qualidade do indicador de turismo ao inserir vieses de tendenciosidade no resultado, impactando na integridade dos dados, prejudicando a acurácia e a confiabilidade.
B5	Agilidade da produção	Se refere à necessidade de que o indicador de turismo seja processado (construído) e divulgado no menor tempo possível (se possível com todas as etapas automatizadas) imediatamente após o recebimento dos dados de entrada (Big Data/insumos) pela organização produtora. A demora no processamento do indicador deixará sua divulgação distante do período a que se refere, reduzindo sua qualidade, podendo torná-lo irrelevante, sem utilidade prática.

B6	Documentação dos processos	Se refere à necessidade de existência de documentação, seguindo padrões e classificações consistentes e uniformes de auditabilidade, que descreva por completo e de forma clara, inclusive com apresentação de indicadores de qualidade, o processo de produção do indicador de turismo produzido a partir de Big Data em todas as suas etapas - sejam elas de processamento computacional ou de métodos e modelos estatísticos utilizados, destacando como a eventual complexidade dos dados de entrada (formato, estrutura, conteúdo, hierarquias, etc.) foi tratada. O não atendimento deste requisito prejudica a qualidade do produto final ao inviabilizar processos de auditoria ou de replicação dos resultados, além de propiciar o uso inadequado do indicador.
C1	Padronização do indicador	Se refere à necessidade de que o indicador de turismo calculado reflita um construto e siga padrões de formato reconhecidos e aceitos perante os <i>stakeholders</i> , compatível e integrável com indicadores equivalentes em âmbito nacional e/ou internacional, claros e consistentes temporalmente. O não atendimento deste requisito reduz a qualidade do indicador à medida que pode restringir sua comparabilidade entre diferentes períodos de tempo, organizações produtoras, regiões e países.
C2	Adequação conceitual	Se refere à necessidade de que o indicador final reflita um contexto (domínios, conceitos, construtos, elementos, classificações, escopo e cobertura, período de referência, detalhamento geográfico, granularidade, tipo de indicador, etc.) alinhado com as expectativas do usuário, carregando em si o conteúdo necessário a permitir a geração de <i>insights</i> para fenômenos reais de interesse, sendo válido do ponto de vista da sua utilidade conceitual. O não atendimento deste requisito reduz a qualidade do indicador à medida que pode restringir sua utilidade, afetando sua relevância.
C3	Grau de inovação	Se refere à necessidade de que o indicador de turismo carregue um diferencial frente ao que já existe e é produzido no ambiente tradicional de produção estatística, seja trazendo um novo conteúdo, refletindo domínios e construtos até então não explorados, ou por reduzir a quantidade de recursos necessários no processo de produção. O não atendimento deste requisito reduz a relevância do indicador pois não promove e motiva esforços para sua construção e manutenção.
C4	Percepção ética	Se refere à necessidade de que o indicador de turismo seja positivamente percebido pelos <i>stakeholders</i> , com ações conduzidas pela organização produtora do indicador que transpareçam a não infração a questões éticas e morais de privacidade e segurança da informação tanto no que diz respeito à origem dos insumos quanto aos processos (práticas, métodos, técnicas, etc.) utilizados durante a produção do indicador final. O não atendimento deste item reduz a qualidade do indicador ao afetar a reputação dos envolvidos no processo, impactando no uso do produto final.
C5	Periodicidade praticada	Se refere à necessidade de que o indicador de turismo seja disponibilizado rigorosamente em uma periodicidade regular preestabelecida, necessitando, para tanto, que os dados de entrada também cheguem ao produtor estatístico seguindo um cronograma rígido. O não atendimento deste requisito diminui a qualidade do indicador ao reduzir sua utilidade à medida que não poderá ser considerado para alguns tipos de análises - como o de séries temporais, por exemplo.
C6	Metadados do indicador	Se refere à necessidade de existência de metadados, seguindo padrões e classificações consistentes e uniformes, que descrevam por completo e de forma clara tanto o conteúdo, definições, hierarquias agrupadas, intervalos de valores válidos, escopo e cobertura, período de referência, especificação geográfica, etc., quanto o formato do indicador de turismo produzido a partir de Big Data, permitindo ao usuário facilmente perceber se atende às suas expectativas ou não. O não atendimento deste item prejudica a qualidade do produto final por reduzir a capacidade de compreensão do conteúdo e consequentemente o potencial de utilização do usuário final, além de propiciar o uso inadequado do indicador.

C7	Capacitação para uso	Se refere à necessidade de existência de oferta de capacitação para uso do indicador, acompanhado de orientações, instruções e alertas objetivos e claros sobre restrições, incertezas, limitações e cuidados, operacionais e legais (ocasionados por questões de complexidade nos dados de entrada, especificidades do processo de produção do indicador de turismo ou por aspectos de legislação ou de acordos de cooperação), que o usuário deve ter ao interpretar e utilizar o indicador de turismo produzido a partir de Big Data. O não atendimento deste requisito prejudica a qualidade do produto final ao propiciar o uso inadequado ou ilegal do indicador.
C8	Disponibilidade do indicador	Se refere à necessidade de que o indicador e suas informações de apoio (metadados, documentações, etc.) estejam disponíveis ao grande público de forma gratuita e em formato que atende as especificações dos dados ditos abertos. O não atendimento deste requisito prejudica a qualidade do produto final ao restringir o acesso ao indicador.
C9	Disseminação e promoção	Se refere à necessidade de existência de uma política com estratégia de divulgação e disseminação do uso do indicador, contemplando um sistema de fácil acesso (plataforma, interface, aplicativo, <i>website</i> ou outra ferramenta) que contenha os dados, metadados, documentações, etc., para fins de facilitar e promover junto ao grande público o acesso, a interpretação e a utilização do indicador. O não atendimento deste requisito prejudica a qualidade do produto final ao dificultar o acesso e a promoção do indicador.

Fonte: elaborado pelo autor (2021).

Nesses termos, a aplicação prática do *framework* proposto consiste na análise e avaliação do atendimento aos requisitos preestabelecidos por parte do indicador, subsidiando a geração de um parecer sobre o nível da sua qualidade estatística. Para fins de simulação e ponto de partida para estudos futuros, uma proposta preliminar de operacionalização é apresentada na Subseção 4.8.1.

4.8 AVALIAÇÃO DO ARTEFATO

A etapa de avaliação do artefato foi conduzida com apoio da aplicação de questionários em especialistas, seguindo a técnica Delphi, conforme descrito na Subseção 3.1.2, tendo por finalidade tanto verificar a estrutura do *framework* quanto observar a aplicabilidade do mesmo em um contexto real, viabilizando que necessidades de ajustes fossem percebidos, subsidiando o consenso quanto à sua melhor versão final.

4.8.1 Aplicabilidade do *framework*

Anterior à verificação da estrutura do *framework* junto aos especialistas, a etapa de avaliação do artefato tratou de observar a aplicabilidade prática do

instrumento em um contexto real de produção de um indicador de turismo a partir de dados do tipo Big Data, com intuito de contribuir com a percepção das necessidades de alterações para fins de melhoramento de estrutura e viabilidade de operacionalização. Este procedimento foi também elucidado ao grupo de especialistas investigado para que pudessem formular seus pareceres durante o preenchimento dos questionários.

Nesse contexto, optou-se por trabalhar com um indicador que se propõe a refletir o nível de satisfação do turista com as principais atrações locais dos municípios de Florianópolis (SC) e Balneário Camboriú (SC), uma aproximação para o nível de satisfação geral do turista para os municípios, construído, conforme descrito na Subseção 3.1.2.1, por este pesquisador – aqui tido como a “instituição” produtora estatística, do indicador de turismo – a partir de dados extraídos da plataforma TripAdvisor – o portal aqui considerado a organização fonte do Big Data, dos insumos.

Cabe reforçar que o indicador considerado pode apresentar limitações, que podem inserir vieses nos resultados, tais quais: um desalinhamento entre a população alvo do indicador, turistas que visitam as atrações locais, e os efetivos participantes da amostra, com possibilidade de inclusão de “não turistas”; uma seletividade na amostra, especialmente se o perfil dos visitantes das atrações selecionadas que deixam *feedback* em língua portuguesa no TripAdvisor for diferente do perfil dos demais visitantes, que não registram suas manifestações na plataforma ou que registraram em outra língua; má especificação das quantidades e dos critérios de seleção de atrações locais consideradas em cada município que, se fossem outros, poderiam alterar os resultados do indicador; a utilização de registros de avaliações falsas, no sentido de terem sido propositalmente direcionadas pelos usuários respondentes para beneficiar ou prejudicar a avaliação geral de uma determinada atração ou localidade; a não validade da suposição de que os registros dos insumos – avaliações coletadas – estão em nível de mensuração intervalar, apresentando equidistância numérica (diferenças constantes) entre os cinco níveis da escala (possíveis respostas), arbitrariamente pontuados com valores de 1 a 5; entre outras.

Entretanto, o foco neste momento não está na construção do melhor indicador possível para o nível de satisfação do turista nas localidades consideradas,

mas sim na observação da possível aplicabilidade do *framework* proposto nesta tese. Para fins de melhoramento do indicador em si, estudos específicos deverão ser conduzidos.

Dito isso, os insumos foram capturados em 22 atrações turísticas – 11 de cada um dos municípios analisados –, considerando o período de mês da experiência de janeiro de 2015 a dezembro de 2020, tabulados conforme a estrutura exemplificada pela Figura 9, com registro, para cada avaliação, de 5 variáveis – município, atração turística, nota da avaliação, mês da experiência turística e mês de realização da avaliação.

Figura 9 – Recorte da estrutura original dos dados capturados no portal TripAdvisor.

Munic	Atracao	Nota	Data_Exp	Data_Ava
Florianópolis	Lagoinha do Leste	50	abril de 2019	jan de 2020
Florianópolis	Lagoinha do Leste	50	fevereiro de 2019	jan de 2020
Florianópolis	Lagoinha do Leste	40	janeiro de 2020	jan de 2020
Florianópolis	Lagoinha do Leste	50	dezembro de 2019	jan de 2020
Florianópolis	Lagoinha do Leste	50	dezembro de 2019	dez de 2019
Florianópolis	Lagoinha do Leste	50	março de 2019	dez de 2019
Florianópolis	Lagoinha do Leste	50		dez de 2019
Florianópolis	Lagoinha do Leste	50	novembro de 2019	dez de 2019

Fonte: elaborado pelo autor (2021).

Os registros, que apesar de originalmente já bem estruturados, foram submetidos a um trabalho de pré-processamento para transformação e padronização das variáveis nota da avaliação, mês da experiência e mês da avaliação, seguido de limpeza (exclusão) daqueles com valor faltante (*missing*) na variável referente ao mês da experiência – única variável a apresentar valores faltantes.

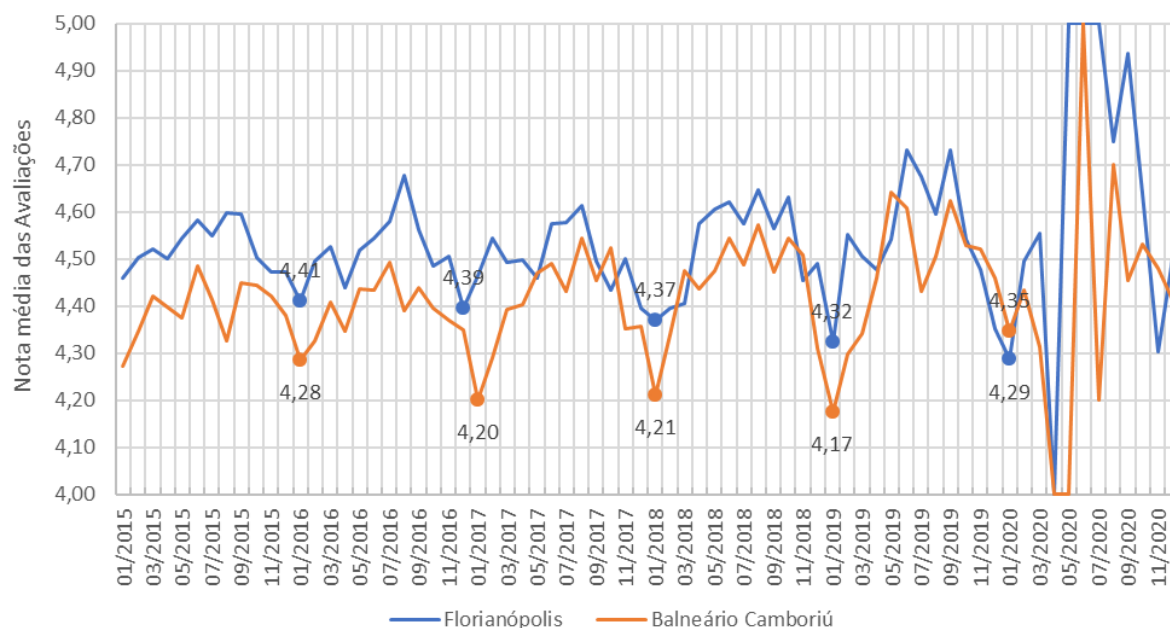
Ao final, os registros válidos totalizaram 50.445 casos, distribuídos pelas 5 variáveis conforme apresenta a Figura 10. Foram calculados os resultados mensais do indicador para as duas localidades no período de tempo considerado, disponibilizados sob a forma de séries temporais na Figura 11. Reforça-se que o cálculo dos indicadores pressupôs avaliações (insumos) em nível de mensuração intervalar.

Figura 10 – Distribuição das avaliações válidas coletadas de jan/2015 a dez/2020, por variáveis registradas e municípios selecionados.

Variável	Município		Município	Atração	Avaliações	
	Florianópolis	Balneário Camboriú				
Ano	2015	8839	5775	Florianópolis	Costa da Lagoa	1653
	2016	9922	6778		Dunas da Joaquina	1564
	2017	5097	3580		Ilha do Campeche	2004
	2018	2878	2484		Jurerê	5289
	2019	1908	1917		Lagoinha do Leste	1330
	2020	659	608		Praia da Daniela	1494
	Jan	6316	3289		Praia da Joaquina	4140
Mês	Fev	3646	1933	Praia de Barra da Lagoa	1683	
	Mar	3006	1856	Praia de Canasvieiras	1664	
	Abr	2296	1665	Praia Mole	2376	
	Mai	1412	1173	Santo Antônio de Lisboa	6106	
	Jun	1001	944	Balneário Camboriú	Avenida Atlântica	1145
	Jul	1447	1994		Cristo Luz	2074
	Ago	1042	1049		Molhe - Barra Sul	1527
	Set	1599	1370		Morro do Careca	1321
	Out	1782	1965		Parque Unipraias	7623
	Nov	2361	1760		Passeio San Miguel	528
	Dez	3395	2144		Praia Central	1941
	Nota	5	18142		11908	Praia das Laranjeiras
4		8173	6486		Praia de Taquaras	336
3		2376	2077		Praia do Estaleirinho	496
2		400	464		Praia do Estaleiro	451
1		212	207		Total	50445

Fonte: elaborado pelo autor (2021).

Figura 11 – Nível de satisfação mensal do turista com as principais atrações locais de jan/2015 a dez/2020, por municípios selecionados.



Fonte: elaborado pelo autor (2021).

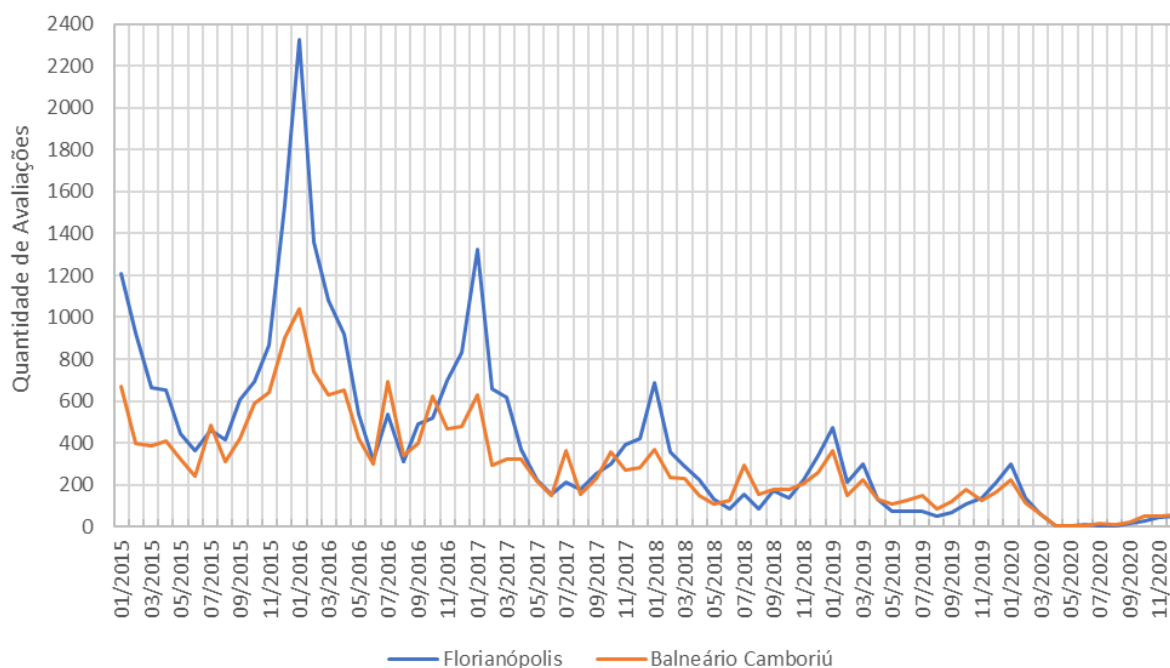
A partir do gráfico é possível observar um aparente nível de satisfação turística mais baixo em Balneário Camboriú, na comparação com o de Florianópolis, praticamente em todo o período analisado. Além disso, percebe-se a existência de um comportamento cíclico nas séries temporais de ambos os municípios, com momentos de níveis de satisfação mais baixos se repetindo em torno dos meses de dezembro e janeiro em todos os anos, sugerindo uma menor satisfação do turista nos períodos de alta temporada. Em termos de tendência, parece não existir para o período analisado, com uma aparente estabilidade das notas médias em torno dos níveis médios globais de 4,54 para Florianópolis e 4,42 para Balneário Camboriú.

Entretanto, percebe-se um aumento da volatilidade nos valores do indicador nos meses mais recentes. Inicialmente, imaginou-se que seria fruto exclusivamente da provável pouca quantidade de avaliações realizadas no ano de 2020, em virtude da baixa atividade turística, consequência da pandemia de Covid-19. Com poucas avaliações, uma amostra menor, há de se esperar maior variância nos estimadores e, naturalmente, maior oscilação nas estimativas – valores percebidos pelo indicador.

Quando da observação da quantidade de avaliações realizadas mês a mês percebeu-se que sim, o ano de 2020 apresentou quantidades de avaliações baixíssimas em alguns meses, principalmente de abril até outubro, conforme mostra o gráfico da Figura 12.

Porém, nota-se também que o aumento da volatilidade do indicador não é fruto exclusivo da queda da atividade turística ocasionada pela pandemia, mas também de uma tendência de queda na quantidade de avaliações sendo realizadas no portal TripAdvisor ao longo do período analisado. Especula-se que, dentre as possíveis razões para esse fenômeno, o que pode estar acontecendo é que a partir do momento que uma atração turística já apresenta uma quantidade alta de avaliações no portal, novos usuários se sentem menos motivados ou percebem menor necessidade de colocarem ali suas novas avaliações. Ou, em uma outra hipótese, o fenômeno pode ser simplesmente fruto de uma eventual redução na quantidade de pessoas que utiliza o portal TripAdvisor.

Figura 12 – Quantidade de avaliações mensais consideradas de jan/2015 a dez/2020, por municípios selecionados.



Fonte: elaborado pelo autor (2021).

De qualquer modo, a análise da quantidade de avaliações realizadas ao longo do tempo aponta para a necessidade de cuidados quando da construção, análise e disponibilização sob a forma de séries temporais de indicadores como esse aqui construído.

Após construído e analisado o indicador de turismo, o diagnóstico correspondente a aplicação do *framework* pôde então ser simulado, utilizando, para tanto, a estrutura preliminar de aplicação – diagnóstico e análise – aqui desenvolvida, um protótipo, conforme apresenta o Quadro 24. Na estrutura utilizada, o indicador é analisado em relação à cada um dos requisitos de qualidade, tomando como ponto de partida um roteiro de diagnóstico – sob a forma de uma pergunta – seguido de uma avaliação, em uma escala de 5 pontos, para o nível de atendimento do requisito, de mais insatisfatório para mais satisfatório. As escalas utilizadas durante a avaliação de cada um dos requisitos não foram validadas, servindo aqui apenas de referencial, de ponto de partida, de sugestão de caminho e, mais objetivamente, de suporte para o processo de observação da possível aplicabilidade do instrumento nos moldes da estrutura proposta.

Quadro 24 – Aplicação do *framework* sobre o indicador do nível de satisfação do turista.

Requisito	Proposta de Diagnóstico	Análise
A1. Base legal	<p>Pergunta: Existe uma base legal que contemple todos os aspectos de possibilidades, restrições e obrigações para o uso dos dados (produção/coleta, armazenamento, manutenção, acesso, técnicas de análise e processamentos, finalidade, segurança, confidencialidade, privacidade, direitos autorais, etc.), legitimando a atuação da fonte do Big Data e do produtor do indicador de turismo?</p> <p>1: Não existe base legal no âmbito do contexto de interesse. 2: Sim, mas com a grande maioria dos aspectos não contemplados. 3: Sim, mas com uma quantidade intermediária de aspectos não contemplados. 4: Sim, mas com alguns poucos aspectos não contemplados. 5: Sim, contemplando todos os aspectos.</p> <p>Com quem verificar: Fonte do Big Data e organização produtora do indicador</p>	<p>Em relação a esse requisito, três são os principais aspectos que devem ser observados: existência de legislação local que trata do uso de informações disponíveis e “deixadas” na internet (pessoais, privadas e públicas); existência e adequação dos termos de autorização de uso de dados e de consentimento com a política de privacidade do portal que o usuário gerador da informação assina quando faz uma avaliação; e existência e respeito aos termos de condições de uso das informações disponibilizadas pelo portal por parte do produtor estatístico.</p> <p>Sobre o primeiro aspecto, a Lei Geral de Proteção aos Dados pessoais (LGPD) é a principal hoje existente e seu propósito de proteger os dados pessoais e comportamentais dos usuários dos portais de internet – deixados deliberada ou despercebidamente durante a navegação – funciona como norteador da política de privacidade estabelecida pelo portal fonte do Big Data. Em termos de política de privacidade, que expõe ao usuário o que poderá ser feito com as informações deixadas durante o seu uso do portal, também é existente¹³, e sua não adequação à LGPD pode culminar em penalidades para a organização fonte dos dados. Ainda, no que diz respeito as avaliações realizadas por usuários, principal insumo aqui utilizado, estas são realizadas pelo próprio usuário em consonância com a assinatura de termos de conteúdo¹⁴ e com consentimento, através do termo de uso¹⁵, de quais as informações prestadas que serão publicitadas – para fins de contribuir com a experiência de outros usuários. E por fim, o portal possui e disponibiliza os termos de uso⁷ das informações presentes na plataforma, destacando atividades proibidas – que podem afetar o trabalho do produtor de estatísticas públicas. No caso, se faz necessário, principalmente, que todos os produtos gerados a partir dos dados extraídos do website não sejam construídos para fins comerciais não autorizados e sejam precedidos de permissão expressa pelo portal.</p> <p>Avaliação: 5 – Sim, contemplando todos os aspectos.</p>

¹³ <https://tripadvisor.mediaroom.com/br-privacy-policy> - Acessado em 06 de janeiro de 2021.

¹⁴ <https://www.tripadvisor.com/hc/pt-br/articles/360008133913-Pol%C3%ADtica-de-conte%C3%BAdo-do-Tripadvisor> - Acessado em 06 de janeiro de 2021.

¹⁵ <https://tripadvisor.mediaroom.com/BR-terms-of-use> - Acessado em 06 de janeiro de 2021.

A2. Credibilidade dos insumos	<p>Pergunta: Os dados de entrada (Big Data/insumos) são verdadeiros, críveis, verificáveis e regulados, originários de instituições autoridades e/ou de boa reputação?</p> <p>1: Não, a péssima credibilidade inviabiliza o indicador. 2: A credibilidade dos insumos é precária. 3: A credibilidade dos insumos é intermediária. 4: A credibilidade dos insumos é boa. 5: Sim, os insumos são completamente críveis e verificáveis.</p> <p>Com quem verificar: Fonte do Big Data ou organizações terceiras independentes</p>	<p>A organização fonte dos insumos é referência mundial em informações turísticas geradas em modelo de <i>crowdsourcing</i>, sendo assim bastante utilizado e conseqüentemente bem-visto. Por outro lado, não é uma instituição autoridade que possui por finalidade produzir informação turística para fins e seguindo padrões estatísticos, sendo tampouco regulada por alguma instituição auditora, movida essencialmente por seus interesses privados. Os insumos são gerados diretamente pelos usuários do portal e, portanto, há espaço para manipulações através de manifestações falsas. Entretanto, o portal exige que os usuários assinem um termo de conteúdo⁶, que trata da obrigatoriedade de manifestações idôneas, e possui políticas e estratégias para proteção, rastreamento¹⁶ e correção de avaliações injustas¹⁷, falsas¹⁸ e tendenciosas¹⁹. Assim sendo, acredita-se que o percentual de registros deturpados não exerça influência dominante sobre os resultados – também devido à natureza do conteúdo, relacionado principalmente a atrações turísticas geridas predominantemente para fins públicos, e a grande quantidade de manifestações. Em termos positivos, os registros dos usuários são integralmente disponibilizados pela plataforma em sua forma original, sem cortes ou transformações.</p> <p>Avaliação: 4 – A credibilidade dos insumos é boa.</p>
A3. Sustentabilidade dos insumos	<p>Pergunta: A fonte dos dados de entrada (Big Data/insumos) e seu conteúdo oferecido são sustentáveis temporalmente?</p> <p>1: Não, completamente insustentável. 2: Há uma grande chance de descontinuidade. 3: Há uma chance intermediária de descontinuidade. 4: Há uma pequena chance de descontinuidade. 5: Sim, completamente sustentável.</p> <p>Com quem verificar: Fonte do Big Data</p>	<p>A organização fonte dos insumos existe desde 2000, estando hoje sólida sob o rótulo de mais popular e maior comunidade de viagens do mundo, com conteúdo gerado pelos próprios usuários mediante assinatura de termos de conteúdo⁸ e de permissão de uso dos dados⁷. Dessa forma, avalia-se que as únicas possibilidades de descontinuidade do conteúdo hoje disponibilizado estariam associadas ao surgimento de uma inovação de ruptura no mercado das informações turísticas ou de uma modificação no formato de como as manifestações são registradas no portal junto aos usuários.</p> <p>Avaliação: 4 – Há uma pequena chance de descontinuidade.</p>

¹⁶ <https://www.tripadvisor.com.br/TripAdvisorInsights/w3690> - Acessado em 06 de janeiro de 2021.

¹⁷ <https://www.tripadvisor.com.br/TripAdvisorInsights/w3680> - Acessado em 06 de janeiro de 2021.

¹⁸ <https://www.tripadvisor.com.br/TripAdvisorInsights/w3688> - Acessado em 06 de janeiro de 2021.

¹⁹ <https://www.tripadvisor.com.br/TripAdvisorInsights/w3683> - Acessado em 06 de janeiro de 2021.

A4. Acordo formal de cooperação	<p>Pergunta: Existe um acordo formalizado entre a fonte do Big Data e a organização produtora do indicador que contemple aspectos de objetivo final, conteúdo das entregas, permissões e restrições, contrapartidas, frequência e cronogramas, etc.?</p> <p>1: Não existe acordo formal de cooperação. 2: Sim, mas contempla poucos aspectos. 3: Sim, mas contempla uma quantidade intermediária de aspectos. 4: Sim, contemplando quase todos os aspectos. 5: Sim, contemplando todos os aspectos.</p> <p>Com quem verificar: Fonte do Big Data e/ou organização produtora do indicador</p>	<p>Existe uma autorização formal para uso dos dados do portal da organização fonte dos insumos para os fins aqui propostos, mas que não garante a continuidade da disponibilização dos dados da maneira como atualmente acontecem. Assim sendo, o portal detentor das informações turísticas de base reconhece apenas parcialmente a importância do seu conteúdo no âmbito do trabalho realizado pelo produtor do indicador de turismo, sem assumir qualquer comprometimento ou cooperação com a saudável disponibilização do produto estatístico no médio e longo prazo.</p> <p>Avaliação: 3 – Sim, mas contempla uma quantidade intermediária de aspectos.</p>
A5. Imediatismo dos insumos	<p>Pergunta: Os dados de entrada (Big Data/insumos) chegam à organização produtora do indicador de turismo no momento imediato da sua ocorrência (no seu período de referência)?</p> <p>1: Não, chegam tão defasados que se tornam inúteis. 2: Não, chegam em tempo insatisfatório. 3: Não, mas chegam em tempo aceitável. 4: Não, mas chegam em tempo satisfatório. 5: Sim, em tempo real.</p> <p>Com quem verificar: Fonte do Big Data</p>	<p>Os dados de entrada são disponibilizados pela plataforma praticamente em tempo real, passado somente o intervalo de tempo necessário para que a plataforma verifique se a avaliação não viola algum requisito dos termos de conteúdo e uso. Entretanto, pode haver uma lacuna entre o momento da experiência turística e o da manifestação do turista no portal – o visitante pode registrar sua avaliação alguns meses (ou até anos) depois da visita realizada. Dessa forma, consultas referentes a datas passadas de experiência turística, quando realizadas em diferentes momentos, podem retornar uma quantidade diferente de avaliações, implicando em eventuais revisões no resultado do indicador calculado – conforme novas avaliações são inseridas no portal. De qualquer forma, acredita-se que o conteúdo dos insumos aqui considerados seja gerado em tempo próximo ao da ocorrência da experiência turística – até porque essa é a proposta do portal, maximizando o valor percebido do conteúdo junto a outros usuários – e que novas e posteriores avaliações não sejam destoantes daquelas rapidamente realizadas.</p> <p>Avaliação: 3 – Não, mas chegam em tempo aceitável.</p>
A6. Facilidade de aquisição	<p>Pergunta: Os dados de entrada (Big Data/insumos) são facilmente acessados ou comprados (quando for o caso), demandando poucos recursos?</p> <p>1: Não, a péssima alcançabilidade inviabiliza o indicador. 2: A alcançabilidade dos insumos é precária. 3: A alcançabilidade dos insumos é intermediária. 4: A alcançabilidade dos insumos é boa. 5: Sim, os insumos possuem excelente alcançabilidade.</p> <p>Com quem verificar: Fonte do Big Data</p>	<p>Os insumos são gratuitos e de livre acesso por parte do produtor do indicador de turismo, mediante acordo e autorização prévia do portal fonte do Big Data, sendo o custo de acessibilidade associado exclusivamente à construção do mecanismo e/ou ao trabalho de coleta dos dados.</p> <p>Avaliação: 4 – A alcançabilidade dos insumos é boa.</p>

A7. Padronização dos conceitos	<p>Pergunta: As definições conceituais dos insumos estão padronizadas e são consistentes conforme preconizam os stakeholders?</p> <p>1: Não, há completa desconfiança em relação a padronização e consistência dos conceitos. 2: Os conceitos são padronizados e/ou consistentes de forma precária. 3: Os conceitos são padronizados e/ou consistentes de forma intermediária. 4: Os conceitos são padronizados e/ou consistentes de forma boa. 5: Sim, todos os conceitos são padronizados e consistentes. Com quem verificar: Fonte do Big Data</p>	<p>Os insumos se referem a avaliações feitas sobre atrações bem especificadas, todas em escalas de cinco níveis explícitos, bem definidos e consistentes temporal e regionalmente (Horível, Ruim, Razoável, Muito bom, Excelente), sem margem para confusão. Avaliação: 5 – Sim, todos os conceitos são padronizados e consistentes.</p>
A8. Metadados dos insumos	<p>Pergunta: Existem metadados que descrevem os dados de entrada (Big Data/insumos)?</p> <p>1: Não, nenhum. 2: Sim, porém com uma grande quantidade de informações faltantes ou não compreensíveis. 3: Sim, porém com uma quantidade intermediária de informações faltantes ou não compreensíveis. 4: Sim, porém com uma pequena quantidade de informações faltantes ou não compreensíveis. 5: Sim, cobrindo todas as informações necessárias de forma compreensível. Com quem verificar: Fonte do Big Data</p>	<p>Mesmo que os dados se apresentem claros e bem estruturados, não existem metadados uma vez que os insumos são extraídos diretamente do <i>front end</i> da plataforma da organização detentora do Big Data. Avaliação: 1 – Não, nenhum.</p>
A9. Representatividade da amostra	<p>Pergunta: Os dados de entrada (Big Data/insumos) estão completos e bem distribuídos no tempo, no espaço e perante todas as parcelas da população que se propõem a representar?</p> <p>1: Não, há muitos vieses desconhecidos e/ou incorrigíveis. 2: Não, há alguns vieses desconhecidos e/ou incorrigíveis. 3: Não, porém todos os vieses são conhecidos, apesar de nem todos serem passíveis de medidas corretivas ou paliativas. 4: Não, porém todos os vieses são conhecidos e passíveis de medidas corretivas ou paliativas. 5: Sim, os dados derivam de um censo ou de uma amostra sem vieses. Com quem verificar: Fonte do Big Data</p>	<p>Os insumos são gerados por uma amostra de turistas – aquela que se dispôs a acessar o portal e fornecer sua avaliação – que não se sabe ser representativa ou não, com existência de viés de seletividade ou falha de cobertura ou não – tanto em termos de parcelas populacionais quanto em termos temporais ou espaciais. De qualquer forma, parte-se do pressuposto que a estrutura de seletividade da amostra de turismo seja semelhante perante as diferentes atrações turísticas, localidades investigadas e momentos no tempo, possibilitando, pelo menos em alguma medida, comparações temporais e entre atrações e localidades. Além disso, é reconhecido que a amostra de turistas associados às avaliações analisadas aumenta com o passar do tempo, à medida que novas avaliações sobre períodos passados são realizadas tardiamente. Nesse caso, pressupõe-se também que o padrão das manifestações tardias seja semelhante ao das manifestações mais próximas do momento da experiência. Em relação aos registros coletados, apresentaram valores faltantes na variável data da experiência em um percentual de 1,89% dos casos, podendo sua distribuição perante os diferentes tipos de avaliações ser analisada. Avaliação: 2 – Não, há alguns vieses desconhecidos e/ou incorrigíveis.</p>

A10. Aderência dos registros	<p>Pergunta: Os dados de entrada (Big Data) estão livres de erros de registro, coleta, processamento, tabulação ou mensuração?</p> <p>1: Não, há muitos erros desconhecidos e/ou incorrigíveis. 2: Não, há alguns erros desconhecidos e/ou incorrigíveis. 3: Não, porém todos os erros são conhecidos, apesar de nem todos serem passíveis de medidas corretivas ou paliativas. 4: Não, porém todos os erros são conhecidos e passíveis de medidas corretivas ou paliativas. 5: Sim, os dados não apresentam erros. Com quem verificar: Fonte do Big Data</p>	<p>Os insumos são gerados diretamente pelo turista, usuário do portal, em escalas de cinco níveis explícitos, bem definidos, sem inconsistências, com pouquíssima margem para erros de registro ou de mensuração. Além disso, não sofrem manipulações ou transformações por parte da instituição fonte do Big Data, praticamente eliminando os possíveis erros de processamento, sendo disponibilizados e podendo serem coletados na sua forma original. Além disso, antes de incluídas no portal, as avaliações são submetidas às práticas de validação da plataforma (identificação proteção, rastreamento⁸ e correção de avaliações injustas⁹, falsas¹⁰ e tendenciosas¹¹).</p> <p>Avaliação: 5 – Sim, os dados não apresentam erros.</p>
A11. Complexidade estrutural	<p>Pergunta: Os dados de entrada (Big Data/insumos) estão bem estruturados, especificados e compreensíveis, não apresentando complexidade?</p> <p>1: Não, há muitos complicadores desconhecidos e/ou incorrigíveis. 2: Não, há alguns complicadores desconhecidos e/ou incorrigíveis. 3: Não, porém todos os complicadores são conhecidos, apesar de nem todos serem passíveis de medidas corretivas ou paliativas. 4: Não, porém todos os complicadores são conhecidos e passíveis de medidas corretivas ou paliativas. 5: Sim, os dados não apresentam complicadores. Com quem verificar: Fonte do Big Data</p>	<p>Os insumos estão em escalas de cinco níveis, com todas as demais variáveis claras e sem complexidade alguma nos registros – compreensíveis, com formato estruturado, bem especificado e sem inconsistências.</p> <p>Avaliação: 5 – Sim, os dados não apresentam complicadores.</p>
A12. Otimização dos insumos	<p>Pergunta: Os dados de entrada (Big Data/insumos) correspondem ao conjunto mínimo necessário para representar e analisar o contexto de interesse?</p> <p>1: Não, a grande maioria dos dados não são utilizados durante a construção do indicador. 2: Não, boa parte dos dados não são utilizados durante a construção do indicador. 3: Não, alguma parte dos dados não são utilizados durante a construção do indicador. 4: Não, porém apenas uma pequena parte dos dados não são utilizados durante a construção do indicador. 5: Sim, todos os dados são utilizados durante a construção do indicador. Com quem verificar: Fonte do Big Data</p>	<p>Em virtude de serem coletados pelos próprios produtores do indicador de turismo, os insumos são concisos e mínimos, estando otimizados para o propósito final de interesse – contendo apenas as informações relevantes e necessárias para a construção do indicador.</p> <p>Avaliação: 5 – Sim, todos os dados são utilizados durante a construção do indicador.</p>

B1. Credibilidade do produtor	<p>Pergunta: O indicador de turismo é produzido e atestado por profissionais especializados de uma organização reconhecida?</p> <p>1: Não, há completa incerteza em relação à organização e aos profissionais envolvidos.</p> <p>2: A organização e os profissionais envolvidos possuem credibilidade precária.</p> <p>3: A organização e os profissionais envolvidos possuem credibilidade intermediária.</p> <p>4: A organização e os profissionais envolvidos possuem credibilidade boa.</p> <p>5: Sim, a organização e os profissionais envolvidos possuem credibilidade excelente.</p> <p>Com quem verificar: Organização produtora do indicador</p>	<p>No presente estudo, o indicador está sendo produzido por um profissional individual, estatístico, com longa expertise na construção de estatísticas públicas e indicadores socioeconômicos. Entretanto, não se trata de uma organização estruturada e reconhecida pelo trabalho no ramo.</p> <p>Avaliação: 4 – A organização e os profissionais envolvidos possuem credibilidade boa.</p>
B2. Infraestrutura e recursos humanos	<p>Pergunta: Existem na organização produtora do indicador de turismo recursos humanos e não humanos adequados para uma boa manutenção do processo de produção?</p> <p>1: Não, os recursos são completamente inadequados.</p> <p>2: Existe uma estrutura mínima, porém muito vulnerável.</p> <p>3: Existe uma estrutura intermediária, porém instável.</p> <p>4: Existe uma boa estrutura, porém com algumas falhas.</p> <p>5: Sim, os recursos são perfeitamente adequados.</p> <p>Com quem verificar: Organização produtora do indicador</p>	<p>Uma vez que os insumos são compostos de poucas variáveis ou atributos, com uma quantidade não tão grande de novos registros a cada mês, a infraestrutura e os recursos humanos necessários são poucos. Dessa forma, um profissional, um <i>crawler</i>, uma base de dados estruturados e um <i>script</i> de processamento e geração dos indicadores são suficientes para manter o processo de produção dentro do esperado.</p> <p>Avaliação: 5 – Sim, os recursos são perfeitamente adequados.</p>
B3. Padronização dos processos	<p>Pergunta: As etapas do processo de produção do indicador de turismo são claras, consistentes e padronizadas conforme preconizam os stakeholders?</p> <p>1: Não, há completa incerteza em relação às etapas do processo de produção.</p> <p>2: Não, muitas etapas do processo de produção não são claras, padronizadas e/ou consistentes.</p> <p>3: Não, uma quantidade intermediária de etapas do processo de produção não são claras, padronizadas e/ou consistentes.</p> <p>4: Não, algumas poucas etapas do processo de produção não são claras, padronizadas e/ou consistentes.</p> <p>5: Sim, todas as etapas do processo de produção são claras, padronizadas e consistentes.</p> <p>Com quem verificar: Organização produtora do indicador</p>	<p>O processo se inicia com a coleta dos dados via <i>crawler</i>, com posterior verificação de equivalência entre os registros coletados e os disponíveis no portal. As etapas seguintes envolvem apenas pequenas transformações e padronizações nos dados originais, sem modificação da essência do conteúdo, limpeza de valores faltantes, com exclusão de registros, e estimação, com a utilização de estimadores simples. É, portanto, um processo de baixa complexidade e sem maiores manipulações ou transformações nos insumos. É composto por poucas etapas, todas claras e consistentes perante diferentes localidades, atrações e intervalos de tempo.</p> <p>Avaliação: 5 – Sim, todas as etapas do processo de produção são claras, padronizadas e consistentes.</p>

B4. Rigor dos processos	<p>Pergunta: As etapas do processo de produção do indicador de turismo são robustas, ajustadas e controladas de modo a não gerar erros de mensuração, representatividade e inferência?</p> <p>1: Não, há completa incerteza em relação às etapas do processo de produção. 2: Não, muitas etapas do processo de produção não são adequadas. 3: Não, uma quantidade intermediária de etapas do processo de produção não são adequadas. 4: Não, algumas poucas etapas do processo de produção não são adequadas. 5: Sim, todas as etapas do processo de produção são adequadas.</p> <p>Com quem verificar: Organização produtora do indicador</p>	<p>O processo – que se inicia com a coleta dos dados via <i>crawler</i> e é seguido por apenas pequenas transformações e padronizações nos dados originais, sem modificação da essência do conteúdo, limpeza de valores faltantes, com exclusão de registros, e estimação, com a utilização de estimadores simples – é verificado em todas as etapas, não ocasionando erros e inconsistências nos resultados.</p> <p>Avaliação: 5 – Sim, todas as etapas do processo de produção são adequadas.</p>
B5. Agilidade da produção	<p>Pergunta: O indicador de turismo é processado (construído) e disponibilizado imediatamente após o recebimento dos dados de entrada (Big Data/insumos)?</p> <p>1: Não, chega tão defasado que se torna inútil. 2: Não, chega em tempo insatisfatório. 3: Não, mas chega em tempo aceitável. 4: Não, mas chega em tempo satisfatório. 5: Sim, em tempo real.</p> <p>Com quem verificar: Organização produtora do indicador</p>	<p>O processamento (construção) do indicador se inicia imediatamente após o término da coleta dos dados no portal. Entretanto, as etapas referentes ao pré-processamento dos dados e à geração das estimativas não estão automatizadas, dependendo ainda de um profissional que manualmente as realiza.</p> <p>Avaliação: 4 – Não, mas chega em tempo satisfatório.</p>
B6. Documentação dos processos	<p>Pergunta: Existe documentação que descreva por completo, apresentando inclusive indicadores de qualidade, o processo de produção do indicador de turismo (desde o processamento computacional até os métodos estatísticos empregados)?</p> <p>1: Não existe. 2: Sim, porém com uma grande quantidade de informações faltantes ou não compreensíveis. 3: Sim, porém com uma quantidade intermediária de informações faltantes ou não compreensíveis. 4: Sim, porém com uma pequena quantidade de informações faltantes ou não compreensíveis. 5: Sim, cobrindo todas as informações necessárias de forma compreensível.</p> <p>Com quem verificar: Organização produtora do indicador</p>	<p>Não existe ainda uma documentação que explicita por completo e de forma clara todo o processo de produção do indicador de turismo.</p> <p>Avaliação: 1 – Não existe.</p>

C1. Padronização do indicador	<p>Pergunta: O conteúdo do indicador é claro, consistente e padronizado conforme preconizam os stakeholders?</p> <p>1: Não, há completa incerteza em relação ao conteúdo do indicador de turismo. 2: O conteúdo do indicador de turismo é claro, padronizado e consistente de forma precária. 3: O conteúdo do indicador de turismo é claro, padronizado e consistente de forma intermediária. 4: O conteúdo do indicador de turismo é claro, padronizado e consistente de forma boa. 5: Sim, o conteúdo do indicador de turismo é completamente claro, padronizado e consistente.</p> <p>Com quem verificar: Organização produtora do indicador</p>	<p>O indicador que se propõe a refletir o nível de satisfação do turista com as principais atrações locais dos municípios analisados, de baixa complexidade, é claro e consistente ao destacar quais atrações estão sendo consideradas e ao utilizar avaliações, conceitos e fórmulas de cálculo de mesma estrutura perante diferentes localidades, atrações e intervalos de tempo.</p> <p>Avaliação: 5 – Sim, o conteúdo do indicador de turismo é completamente claro, padronizado e consistente.</p>
C2. Adequação conceitual	<p>Pergunta: O indicador de turismo reflete o contexto esperado pelo usuário, subsidiando insights para fenômenos reais de interesse?</p> <p>1: Não, o contexto em nada consegue atender às expectativas dos usuários. 2: O contexto atende às expectativas dos usuários de forma precária. 3: O contexto atende às expectativas dos usuários de forma intermediária. 4: O contexto atende às expectativas dos usuários de forma boa. 5: Sim, o contexto atende perfeitamente às expectativas dos usuários.</p> <p>Com quem verificar: Organização produtora do indicador</p>	<p>Partindo do pressuposto de que o interesse do usuário deste indicador está na compreensão do nível de satisfação do turista com as atrações locais dos municípios de Florianópolis e de Balneário Camboriú, avalia-se que o principal aspecto que pode não estar sendo atendido se refere à cobertura do indicador em termos da quantidade limitada de atrações que considera em seu cálculo.</p> <p>Além disso, outra questão que pode causar desalinhamento com as expectativas dos usuários é a presença de uma eventual seletividade seguida de falhas de cobertura na amostra.</p> <p>Avaliação: 3 – O contexto atende às expectativas dos usuários de forma intermediária.</p>
C3. Grau de inovação	<p>Pergunta: O indicador de turismo é inovador, no sentido de carregar algum diferencial (domínio temático ou geográfico, economia de recursos financeiros, tecnológicos ou humanos, tempo de produção, etc.) frente ao que é tradicionalmente produzido?</p> <p>1: Não, o indicador não apresenta diferencial. 2: Sim, o indicador apresenta diferencial em relação a um aspecto. 3: Sim, o indicador apresenta diferencial em relação a dois aspectos. 4: Sim, o indicador apresenta diferencial em relação a três aspectos. 5: Sim, o indicador apresenta diferencial em relação a quatro ou mais aspectos.</p> <p>Com quem verificar: Organização produtora do indicador</p>	<p>Não se tem conhecimento de outro indicador que reflita a satisfação dos turistas de Florianópolis e Balneário Camboriú com possibilidade de divulgação em periodicidade regular, gerado de forma praticamente gratuita, com mínima utilização de recursos humanos e tecnológicos.</p> <p>Avaliação: 5 – Sim, o indicador apresenta diferencial em relação a quatro ou mais aspectos.</p>

C4. Percepção ética	<p>Pergunta: Os dados de entrada (Big Data/insumos) e os processos de produção do indicador final são gerados e conduzidos de forma ética, respeitando privacidade e segurança, conforme preconizam os stakeholders?</p> <p>1: Não, insumos e processos são percebidos como completamente inadequados. 2: Não, insumos e processos são percebidos como muito inadequados. 3: Não, insumos e processos são percebidos como razoavelmente inadequados. 4: Não, insumos e processos são percebidos como um pouco inadequados. 5: Sim, insumos e processos são percebidos como adequados.</p> <p>Com quem verificar: Organização produtora do indicador</p>	<p>Uma vez que os insumos são gerados pelos próprios usuários, que assinam termo de consentimento e que reconhecem que suas manifestações serão acessíveis para o grande público, abertamente e com identificação do autor, o indicador não deverá sofrer com percepção negativa dos stakeholders.</p> <p>Avaliação: 5 – Sim, insumos e processos são percebidos como adequados.</p>
C5. Periodicidade praticada	<p>Pergunta: O indicador de turismo é disponibilizado com periodicidade regular, seguindo cronograma preestabelecido?</p> <p>1: Não, sua divulgação é esporádica, aleatória. 2: Sim, mas com cronograma raramente respeitado. 3: Sim, porém com cronograma instável. 4: Sim, com cronograma normalmente respeitado. 5: Sim, com cronograma seguido rigorosamente.</p> <p>Com quem verificar: Organização produtora do indicador</p>	<p>Não, ainda não há perspectiva de construção de um cronograma de divulgação com periodicidade regular preestabelecida.</p> <p>Avaliação: 1 – Não, sua divulgação é esporádica, aleatória.</p>
C6. Metadados do indicador	<p>Pergunta: Existem metadados (independente do quão acessíveis estão) que descrevem o indicador de turismo?</p> <p>1: Não, nenhum. 2: Sim, porém com uma grande quantidade de informações faltantes ou não compreensíveis. 3: Sim, porém com uma quantidade intermediária de informações faltantes ou não compreensíveis. 4: Sim, porém com uma pequena quantidade de informações faltantes ou não compreensíveis. 5: Sim, cobrindo todas as informações necessárias de forma compreensível.</p> <p>Com quem verificar: Organização produtora do indicador</p>	<p>Ainda não existem metadados formalizados que descrevam por completo e de forma clara o indicador de turismo.</p> <p>Avaliação: 1 – Não, nenhum.</p>
C7. Capacitação para uso	<p>Pergunta: Existe material de capacitação e orientação (independente do quão acessível está) sobre a utilização apropriada do indicador de turismo, destacando cuidados e restrições?</p> <p>1: Não existe. 2: Sim, porém com uma grande quantidade de informações faltantes ou não compreensíveis. 3: Sim, porém com uma quantidade intermediária de informações faltantes ou não compreensíveis. 4: Sim, porém com uma pequena quantidade de informações faltantes ou não compreensíveis. 5: Sim, cobrindo todas as informações necessárias de forma compreensível.</p> <p>Com quem verificar: Organização produtora do indicador</p>	<p>Ainda não existe nenhum tipo de documentação ou material que capacite e oriente o usuário final sobre a correta interpretação, análise e utilização do indicador de turismo.</p> <p>Avaliação: 1 – Não existe.</p>

C8. Disponibilidade do indicador	<p>Pergunta: O acesso ao indicador de turismo e às informações de apoio (metadados, documentações, etc.) é gratuito, em formato aberto?</p> <p>1: Não, para nenhum item. 2: Sim, porém com uma grande quantidade de itens não gratuitos ou livres. 3: Sim, porém com uma quantidade intermediária de itens não gratuitos ou livres. 4: Sim, porém com uma pequena quantidade de itens não gratuitos ou livres. 5: Sim, para todos os itens.</p> <p>Com quem verificar: Organização produtora do indicador</p>	<p>A disponibilização de qualquer conteúdo referente ao indicador ainda não foi pensada e organizada, sendo hoje inexistente.</p> <p>Avaliação: 1 – Não, para nenhum item.</p>
C9. Disseminação e promoção	<p>Pergunta: Existe uma estratégia de disseminação, com plataforma, interface, aplicativo, website ou outra ferramenta que facilite e promova junto ao público o acesso, a interpretação e a utilização do indicador (contendo os dados, metadados, documentações, etc.)?</p> <p>1: Não existe. 2: Sim, porém com uma grande quantidade de itens faltantes ou de difícil acesso. 3: Sim, porém com uma quantidade intermediária de itens faltantes ou de difícil acesso. 4: Sim, porém com uma pequena quantidade de itens faltantes ou de difícil acesso. 5: Sim, com fácil acesso a todos os itens.</p> <p>Com quem verificar: Organização produtora do indicador</p>	<p>Não existe nenhuma estratégia de promoção com estrutura que facilite a disseminação do indicador e seus conteúdos associados.</p> <p>Avaliação: 1 – Não existe.</p>

Fonte: elaborado pelo autor (2021).

Cabe destacar que, para uma mais satisfatória aplicação do *framework*, o processo deve ser conduzido em conjunto, com consentimento e participação das instituições envolvidas (o portal fonte do Big Data, dos insumos, e o produtor estatístico, do indicador de turismo), diferentemente da simulação experimental aqui realizada – que não contou com a participação da organização fonte dos insumos.

Assim, com base na simulação da aplicação do *framework* proposto, é possível classificar o indicador analisado em termos de sua qualidade estatística, conforme apresenta a Tabela 1, com níveis calculados a partir da média dos escores registrados nos requisitos – pressupondo nível de mensuração intervalar para a escala de avaliação dos requisitos (equidistância numérica, com diferenças constantes, entre as possíveis respostas, arbitrariamente pontuadas com valores de 1 a 5).

É reconhecido aqui que alternativas de aprofundamento do instrumento de mensuração da qualidade podem se dar, por exemplo, através da adaptação de ferramentas como a *Quality Function Deployment* (QFD) ajustada para o contexto do registro de variáveis qualitativas e propriedades subjetivas com auxílio da lógica da teoria dos conjuntos *fuzzy* (*Fuzzy Set Theory*) – contribuindo também para o tratamento à problemática da escala de avaliação da qualidade em nível originalmente ordinal. Entretanto, em virtude de não fazerem parte dos objetivos centrais desta pesquisa, tais questões foram somente adereçadas no âmbito das recomendações de trabalhos futuros (Seção 5.1).

Tabela 1 – Qualidade estatística do indicador de turismo, por domínios e dimensões.

Escopo	Abertura	Requisitos Avaliados	Nível de Qualidade Estatística
Domínio	Insumos (A)	12	3,8
	Produção (B)	6	4,0
	Produto (C)	9	2,6
Dimensão	Relevância	9	3,1
	Acurácia e Confiabilidade	9	4,0
	Oportunismo temporal e Pontualidade	8	3,8
	Coerência e Comparabilidade	9	3,3
	Acessibilidade e Clareza	9	2,2
Geral		27	3,4

Fonte: elaborado pelo autor (2021).

Na presente simulação, a importância – peso – de todos os requisitos para a composição da qualidade estatística geral e de cada dimensão e domínio foi considerada igual. No caso da construção do parecer geral, destaca-se que o nível de qualidade não é obtido a partir da média aritmética simples do nível médio observado em cada domínio ou dimensão, mas sim pela média das mensurações realizadas nos 27 requisitos – contornando eventuais problemas de duplicidade de requisitos que poderiam ser gerados pelo fato de se considerar um mesmo requisito em mais de uma dimensão de qualidade. Da mesma forma, os pareceres parciais por domínio ou dimensão consideram a média dos níveis de atendimento dos requisitos que àqueles componentes pertencem ou impactam – observando-se que, como alguns requisitos estão associados a mais de uma dimensão, a soma de requisitos considerados nas avaliações segmentadas por dimensões supera o total de 27 requisitos. Cabe reforçar aqui que esse formato acaba por atribuir, naturalmente, no âmbito da avaliação geral, pesos maiores para os domínios e dimensões que possuem uma quantidade maior de requisitos associados.

Ainda, reconhece-se que alguns requisitos podem ser mais ou menos determinantes da qualidade de um produto estatístico, geral e por dimensões, necessitando uma especificação diferenciada de pesos para os mesmos, dependendo inclusive do produto e do contexto em questão, cabendo essa diferenciação ser explorada em estudos futuros delineados para tal finalidade.

No caso, a simulação estaria apontando para um indicador com nível de qualidade estatística geral de 3,4, em uma escala intervalar de 1 (pior) a 5 (melhor). Em termos de domínios de qualidade, o que apresentou pior desempenho foi o associado ao produto final, com nível de 2,6. Já para análise por dimensões de qualidade, a acessibilidade e clareza e a relevância apresentaram pior desempenho, com níveis de 2,2 e 3,1, respectivamente. Dessa forma, o *framework* permite indicar que, para melhorar a qualidade do indicador referente ao nível de satisfação do turista com as atrações locais, uma estratégia possível estaria baseada em focar na melhoria dos requisitos associados ao domínio de produto e às dimensões acessibilidade e clareza e relevância.

Nesses termos, a análise sugere uma aplicabilidade factível do ponto de vista prático para o *framework*, desde que as propostas de diagnóstico e mensuração estejam bem formuladas e alinhadas com as definições dos requisitos.

Cabendo destacar que, para o modelo de diagnóstico aqui proposto, permanece a dificuldade associada ao componente subjetivo da interpretação quanto ao atendimento de cada requisito por parte do indicador, quando da atribuição de um nível na escala de mensuração. Como alternativa de contorno, sugere-se que a aplicação do *framework* para análise da qualidade estatística de um indicador de turismo seja feita por um profissional experiente e acostumado a conduzir tal processo – ou até mesmo por um grupo ou comitê de especialistas.

De qualquer forma, não houve, durante o processo de consulta aos especialistas, manifestações que sugerissem inviabilidade completa quanto a aplicabilidade ou operacionalização prática do *framework* – conforme demonstra a Subseção 4.8.2. Reconhecendo-se, apesar disso, que algumas manifestações providencialmente indicaram possíveis melhorias e a necessidade de uma série de avanços em termos de validações e aprofundamentos de pesquisa, apresentados na subseção a seguir.

4.8.2 Verificação da estrutura *framework*

A verificação da estrutura do *framework*, principal finalidade da etapa de avaliação do artefato desenvolvido, teve como intuito confirmar a estrutura de dimensões, domínios e requisitos proposta – incluindo a matriz de relacionamentos. Sua operacionalização consistiu na apresentação do *framework* proposto para os especialistas, seguida de coleta de opiniões e implementação de melhorias, de forma sequencial e em rodadas até que o consenso fosse alcançado.

Mais especificamente, duas rodadas de aplicação de questionários – o primeiro menos e o segundo mais estruturado, apresentados respectivamente nos Apêndices C e E – foram realizadas com o grupo de 9 especialistas participantes em momentos diferentes, separadas pelos processos de consolidação e compartilhamento das manifestações e de reformulação do *framework*, seguindo os passos da técnica Delphi, conforme descreve a Subseção 3.1.2. O Quadro 25 apresenta as solicitações de modificações ou reflexões trazidas pelos especialistas consultados.

Quadro 25 – Solicitações de modificações e reflexões para o framework.

Referência	Solicitação
Estrutura Geral	<p><i>“Fico na dúvida nesse caso se as dimensões de fato se aplicam a todos os requisitos. Exemplo: tenho dificuldade de entender Oportunidade temporal e pontual com padronização de conceitos.”</i></p> <p><i>“Avalio que as dimensões de qualidade escolhidas não fazem sentido para os 3 domínios, por estarem mais voltados para avaliação da qualidade do produto (output) e não tanto dos inputs e dos processos. Por exemplo, não vejo relevância como uma dimensão importante no nível de processos, mas sim no nível de produto. “Relevância” tem a ver com atendimento das necessidades dos usuários, diretamente associada ao produto e, neste aspecto, não vejo muito relação com o domínio Processo. Outras dimensões no domínio Processo poderiam ser consideradas, como robustez/solidez metodológica ou a existência de processos de verificação de qualidade, por exemplo. Outro exemplo, Credibilidade do produtor (assim como Base legal e Percepção ética) está mais relacionado com Ambiente institucional do que com relevância.”</i></p>
Dimensões	<p><i>“Gostaria de no caso de indicadores Big Data ter uma dimensão que apresentasse um, ou mais, indicadores de pesquisas tradicionais que abordem o mesmo problema (necessidade de informação).</i></p> <p><i>Caso esse seja um indicador inovador que trabalha temas ainda não apresentados de forma tradicional, essa informação deve constar como uma forma de aumentar a relevância desse dado.”</i></p> <p><i>“Talvez uma avaliação ‘nacional’ de algum quesito possa ser sugerido, ou seja, o hardcore está definido e nuances de uma avaliação de qualidade dentro da região ou país pudesse ser avaliado.”</i></p>
Domínios	<p><i>“Poderia ser adicionado um quarto domínio de disseminação (p.ex. comunidade de usuários de estatísticas, outros stakeholders, etc.) e até mesmo um outro domínio de capacitação para o uso (capacity building).”</i></p> <p><i>“O que surge de novo que talvez possa ser explorado é com relação a insumos: explorar bases de dados federados que possam ser integradas (interconectadas, interoperáveis) e geograficamente descentralizadas para melhoria da qualidade de análise futura e melhor resultado para apoio à decisão.”</i></p>
Requisitos	<p><i>“Minha única questão é quanto a ‘representatividade’ dos dados de Big Data. Tais fontes de dados nem sempre permitem identificar com clareza o quanto representam a população alvo associada a uma determinada demanda. Gostaria que esse requisito pudesse incluir especificamente a população que é efetivamente alcançada pela coleta Big Data do referido insumo, da forma mais detalhada possível (processos, coleta, etc.). Dessa forma fica clara para todos os usuários e possíveis estudiosos que tiverem acesso a base a que população se refere a base.”</i></p> <p><i>“Minha dúvida fica com relação a questão de heterogeneidade semântica na origem dos dados para garantir o significado, a interpretação e a intenção de uso dos dados na geração de produtos.”</i></p> <p><i>“Não percebo com clareza itens como custos e prazos. Imaginar sempre que não há custos envolvidos na busca de grandes repositórios de dados não me parece aplicável além da Academia. A monetização seria interessante.”</i></p> <p><i>“Depois de dois ou três rounds a gente poderia utilizar métodos estatísticos para verificar redundâncias.”</i></p> <p><i>“Avalio que são muitos requisitos, com algumas redundâncias (principalmente na dimensão Acessibilidade) e com oportunidades de fusão. Por outro lado, sinto falta de alguns requisitos relacionados mais diretamente, como por exemplo, a existência de procedimentos de qualidade/validação de dados realizados ao longo do domínio Processo (produção). O B4 chamaria de Adequação de processos.</i></p> <p><i>No caso do domínio Produto, vejo redundância entre (C5 e C6) e (C7 e C8). As duplas poderiam ser um indicador só. Senti falta também de um requisito de Atendimento às necessidades dos usuários, que é o mais alinhado com Relevância. Talvez Adequação Conceitual tenha esta intenção, mas não está claro o atingimento do objetivo principal, ou seja, que produto satisfaz à necessidade dos usuários.</i></p> <p><i>No caso de Insumos, A9 e A10 poderiam ser fundidos em um só, Validade/Confiabilidade, na medida em que ambos pretendem avaliar o quão bem os dados representam a realidade, seja por erros amostrais ou não amostrais. Além disso, senti falta de algo ligado à facilidade de acesso. Está contemplado no baixo custo operacional de A6?</i></p> <p><i>Vejo redundância entre Credibilidade dos insumos e Credibilidade do produtor. Ambos se referem à credibilidade da instituição.</i></p> <p><i>Senti falta também de um requisito diretamente relacionado com relevância do Insumo, indicando o quanto os dados contribuem/são aderentes ao conceito/indicador de interesse, que se deseja construir para atender às necessidades dos usuários.”</i></p>

Relacionamentos	<p><i>“A base legal é muito relevante e precisa ser garantida não somente com relação a coerência e comparabilidade.”</i></p> <p><i>“Me parece adequada a estrutura de relacionamento, com a observância da questão de custos e prazos para a execução do projeto.”</i></p> <p><i>“Avalio que em cada domínio deveria ter pelo menos um requisito para cada uma das dimensões de qualidade. Isso não ocorre por exemplo no domínio “Produto” para a dimensão “Acurácia e confiabilidade”. No caso do requisito C2 poderia estar associada a dimensão “Acurácia e confiabilidade”. Outra questão está associada com a atribuição de um requisito a mais de uma dimensão de qualidade, uma vez que o diagnóstico pode estar mais voltado para uma das dimensões. Sugiro uma revisão desses casos minimizando ou evitando essas ocorrências. Por exemplo, como sugestão, retirar “Relevância” de C7 e C8. A atribuição de um mesmo requisito para mais de uma dimensão de qualidade pode interferir nos resultados do parecer, uma vez que o requisito pode ter o diagnóstico mais voltado para uma das dimensões e a nota é considerada a mesma para todas as dimensões que o requisito foi atribuído. Para o requisito associado a mais de uma dimensão, seria o caso de definir perguntas distintas para cada dimensão de qualidade? Recomenda-se uma revisão na estrutura de relacionamento entre dimensões, domínios e requisitos do framework.”</i></p> <p><i>“Não vejo relação entre:</i></p> <ul style="list-style-type: none"> <i>- A1 Base legal e Coerência e Comparabilidade. Tem mais a ver com uma dimensão de Ambiente Institucional.</i> <i>- A4 Acordo formal de cooperação e dimensões oportunidade e coerência. Tem mais a ver com uma dimensão de Ambiente Institucional.</i> <i>- A8 Metadados não tem a ver com acurácia, só com acessibilidade e clareza e coerência/comp.</i> <i>- A11 Complexidade Estrutural e Oportunidade temporal. Tem a ver com acessibilidade.</i> <i>- B1 Credibilidade e Relevância. Tem mais a ver com uma dimensão ambiente institucional.</i> <i>- Em produto, apenas Adequação Conceitual está relacionado à Relevância. Os demais requisitos, C3 e C8 não tem a ver com relevância.</i> <i>- C3 Percepção ética tem mais a ver com uma dimensão de Ambiente Institucional.”</i>
Operacionalização	<p><i>“Gostaria de propor um “peso” maior para o domínio INSUMOS. Esse domínio constitui um conjunto de características que são muito importantes quando da utilização de fontes de dado do tipo Big Data, fontes que nem sempre estão sob administração de organizações produtoras de estatísticas públicas/oficiais.”</i></p> <p><i>“Necessidade de avanço com garantia da qualidade deste insumo. Estudos robustos podem auxiliar a depurar a operacionalização dos insumos propostos.”</i></p> <p><i>“Acho que cada dimensão deveria ter um peso (ponderação) somando ao final 100%.”</i></p> <p><i>“Acredito que apesar de válida, se possível seria mais interessante uma comparação com algum benchmark ou algum indicador “padrão ouro” de institutos oficiais. Tenho alguma dificuldade em me conectar com o framework sem alguns benchmarks de indicadores e como eles se enquadrariam no framework proposto. Alguns institutos produzem estatísticas experimentais (exemplo: ONS-UK e FGV-RIO) e talvez uma abordagem interessante poderia ser enquadrar tais estatísticas dentro do framework sugerido. Acho que é um bom começo, mas acho difícil conseguir avaliar algumas dimensões na prática. Acredito que de maneira prática as dimensões “Oportunismo temporal e pontualidade” e “Acurácia e Confiabilidade” parecem bastante complicadas de medir, principalmente em casos de webscraping. Muitos desses requisitos não são trivialmente obtidos e muito provavelmente apenas conhecidos pelo produtor.”</i></p> <p><i>“Outro aspecto está na arbitrariedade da associação da definição do requisito (quadro 23) com a proposta de diagnóstico (quadro 24) sob a forma de uma pergunta, além da dificuldade identificada associada ao componente subjetivo de interpretação quanto ao atendimento de cada requisito por parte do indicador.”</i></p> <p><i>“O único aspecto negativo que eu visualizo seria a utilização dessa ferramenta sem conhecimento adequado do fenômeno ou dos requisitos sob avaliação. Fiquei em dúvidas quanto ao cálculo do nível de qualidade estatística, especialmente com o valor global (que me parece ser a média das três aberturas do domínio, e se for o caso, desconsidera o escopo dimensão). Essa construção deve estar mais bem explicada no texto.”</i></p> <p><i>“Carece de diversas validações, a começar pela validade de face, de construto e a validade das escalas de modo geral. Há todo um processo de validação das escalas propostas, que, por si só, daria uma outra tese, com diversas etapas de validação e confirmação. Há muitas perguntas com interpretação dúbia, com 2 perguntas em uma, que ensejariam leituras diferentes.”</i></p>

Fonte: elaborado pelo autor (2021) com base nas manifestações dos especialistas.

Tais apontamentos contribuíram para uma série de reflexões e alterações na estrutura do *framework*, cabendo destacar que, no âmbito da estrutura geral do artefato proposto:

- i. Ratificou-se a ideia de que esta proposta de *framework* parte do princípio de ser voltado à percepção do usuário, compreendendo a qualidade estatística como a capacidade do produto em atender as necessidades dos usuários, sendo, portanto, composta e conseqüentemente avaliada no âmbito das dimensões de qualidade associadas ao produto – Relevância, Acurácia e Confiabilidade, Oportunismo temporal e Pontualidade, Coerência e Comparabilidade, Acessibilidade e Clareza. Na impossibilidade prática de se medir os conceitos abstratos definidores das dimensões de qualidade, se propõe observar a presença ou o atendimento de uma série de atributos ou características do indicador de turismo, aqui ditos requisitos, que podem estar associados à diferentes etapas ou ambientes do processo de produção estatística, os ditos domínios – Insumos, Produção e Produto. Dessa forma, não é de interesse da proposta forçar a relação entre todos os requisitos, domínios e dimensões. Acredita-se aqui que um requisito é um atributo definidor da qualidade de um produto estatístico, caracterizando-a no âmbito de uma ou mais dimensões – mas não necessariamente todas – e que se faz inserido em algum domínio do processo de produção estatística. Ou seja, entende-se que um determinado requisito não pode estar presente em mais de um domínio, mas que pode sim impactar em mais de uma dimensão de qualidade do indicador de turismo. Naturalmente, os requisitos de um mesmo domínio tendem a se associar predominantemente, mas não necessariamente, com uma mesma dimensão de qualidade do produto. Por exemplo, é natural que os requisitos dos domínios de insumos e de processo estejam mais associados às dimensões acurácia e confiabilidade, oportunismo temporal e pontualidade, coerência e comparabilidade, enquanto os requisitos do domínio de produto impactam mais nas dimensões relevância e acessibilidade e clareza.

Quanto às dimensões da qualidade estatística consideradas, ressalta-se que:

- i. Optou-se por manter a estrutura de cinco dimensões de qualidade de produto para fins de compatibilização com padrões tradicionalmente utilizados por organizações produtoras estatísticas, internacionais e nacionais;
- ii. Em contrapartida, as solicitações de inclusão de uma dimensão que se referisse à inovatividade do indicador, no âmbito da localidade para a qual está sendo proposto, foram incorporadas sob a forma de um requisito, Grau de inovação (C3), inserido no domínio de produto e associado à qualidade estatística no âmbito da dimensão relevância.

No que se refere aos domínios do fluxo da produção de estatísticas:

- i. Assim como nas dimensões de qualidade, optou-se por seguir a padronização predominantemente utilizada pelas organizações produtoras de estatísticas e manter a estrutura de três domínios (Insumos, Produção e Produto), por questões de redução de complexidade do *framework*, uma vez que esses domínios apresentam delimitação clara e disjunta, podendo serem vistos como etapas consecutivas no fluxo da produção estatística. Cabendo destacar que o “ambiente institucional”, por vezes encontrado como domínio nos *frameworks* tradicionais de qualidade estatística e hiperdomínio nos protocolos desenhados para o contexto de Big Data, está contemplado sob a forma de requisitos, perpassando de forma transversal e fragmentada todos os domínios propostos – de insumos (quando se referir à organização geradora do Big Data e seu ambiente externo), de produção (quando se referir ao produtor estatístico e seu ambiente) e de produto (quando se referir ao espectro do usuário final);
- ii. As solicitações por novos domínios – disseminação/divulgação, capacitação para uso e uso/integração de dados federados – foram atendidas por meio da inclusão ou modificação dos requisitos Disseminação e promoção (C9), Capacitação para uso (C7) e Credibilidade dos insumos (A2), inseridos nos domínios de produto (2x) e

de insumos, respectivamente, entendidas as duas primeiras como componentes do nível de acessibilidade e clareza de um produto estatístico, enquanto para a terceira a associação se dá com a relevância e a acurácia e confiabilidade.

Quanto aos requisitos considerados, por sua vez, cabe destacar que, a partir das manifestações dos especialistas:

- i. Foram expandidas ou reformuladas as definições dos requisitos Representatividade da amostra (A9), Padronização dos conceitos (A7), Complexidade estrutural (A11), Adequação conceitual (C2), Facilidade de aquisição (A6) e Disponibilidade do indicador (C8), com os dois últimos passando a considerar explicitamente o fator dos custos, com a obtenção dos insumos e do indicador final, respectivamente, partindo do entendimento de que dados de entrada (Big Data) ou indicadores finais de alto custo – para a organização produtora estatística e para o usuário final, respectivamente – culminam em uma menor acessibilidade ao produto. Da mesma forma, os requisitos Rigor dos processos (B4) e Aderência dos registros (A10) foram modificados. O primeiro, para incorporar a importante necessidade de se considerar a existência ou não de procedimentos de controle, checagem e verificação de consistência do ajuste e funcionamento das etapas do processo de produção do indicador estatístico. Já o segundo, para explicitar a necessidade de que os insumos sejam válidos em termos de refletirem o contexto de forma aderente às expectativas de utilização do usuário final;
- ii. Foram realizadas reflexões sobre possíveis fusões entre diversos pares de requisitos por razões de redundância ou mesma finalidade, tais quais Metadados do indicador (C6) e Capacitação para uso (C7), Disponibilidade do indicador (C8) e Disseminação e promoção (C9), Representatividade da amostra (A9) e Aderência dos registros (A10), Credibilidade dos insumos (A2) e Credibilidade do produtor (B1). Ao final, optou-se por não concretizar essas quatro fusões – tendo presente, além dos raciocínios expostos a seguir, que o objetivo maior desta tese não está na conclusão da proposta de operacionalização da mensuração da qualidade, mas sim na estruturação de um *framework* conceitual que

perceba, elenque e relacione todos os elementos componentes de um sistema. No caso de C6 e C7, se partiu do princípio de que um indicador que estiver acompanhado de, além dos metadados, um material de capacitação, terá maior qualidade pela maior acessibilidade e clareza. Entende-se aqui que os metadados estão mais associados ao registro, documentação e transparência do indicador, enquanto a capacitação para uso está mais associada à preparação, à orientação, à instrução e ao suporte do usuário. Para C8 e C9, o entendimento foi de que o primeiro se propõe a representar o quão disponível é o indicador ao usuário, em termos de possibilidade de aquisição e utilização, pela gratuidade ou baixo custo e pelo formato aberto, enquanto o segundo se refere aos esforços praticados pela instituição produtora para promulgar, divulgar, promover e disseminar o produto estatístico e sua utilização. Em relação à A9 e A10, a opção por mantê-los separados foi pela grande importância de cada um desses aspectos na qualidade dos insumos. Por mais que os dois requisitos se refiram a possíveis erros presentes no Big Data, os primeiros essencialmente amostrais e os seguintes predominantemente não amostrais, ambos congregam uma gama densa de aspectos a serem observados, de grande importância e difícil avaliação. Juntá-los em um único requisito poderia aumentar a complexidade da mensuração e reduzir a importância do conteúdo na aferição da qualidade estatística do indicador. Sobre A2 e B1, a razão principal está associada ao fato de que não necessariamente a organização geradora do Big Data é a instituição produtora do indicador estatístico. Além disso, a credibilidade dos insumos possui também relação com a veracidade e a verificabilidade dos dados de entrada, enquanto a credibilidade do produtor se volta para a reputação e a autoridade da instituição que está conduzindo os processos de produção estatística para geração do indicador, fazendo-se importante considerar tais aspectos separadamente.

No âmbito da estrutura de relacionamentos entre requisitos, domínios e dimensões, e com base nas considerações da pesquisa:

- i. Incluiu-se uma relação entre o requisito Base legal (A1) e a dimensão acessibilidade e clareza, pois, concorda-se com a ideia de que a inexistência desse requisito para um determinado indicador pode interferir diretamente na sua disponibilidade (Acessibilidade e Clareza), assim como interfere na sua consistência temporal (Coerência e Comparabilidade), mas por si só, por exemplo, não modifica a capacidade do indicador em representar a realidade corretamente (Acurácia e Confiabilidade), ou o quanto seu conteúdo satisfaz a necessidade do usuário (Relevância), ou a rapidez com que pode ser processado e divulgado (Oportunismo Temporal e Pontualidade). Cabe reforçar que o critério utilizado para definir se há relação entre um requisito e uma dimensão é o que coloca como requisitos de uma dimensão os atributos primordiais a serem contemplados por um produto estatístico no âmbito daquele princípio desejado – para que então se possa dizer qual o nível de qualidade estatística na dimensão em questão. Em outras palavras, para que um produto estatístico apresente boa qualidade em uma determinada dimensão, quais seriam as características que deve obrigatoriamente possuir? Ou seja, caso um requisito específico não seja satisfeito, causará impacto direto nas dimensões as quais se relaciona, prejudicando a qualidade estatística do indicador nesses aspectos;
- ii. Partindo da mesma linha de raciocínio do item anterior, norteadora das decisões sobre a existência ou não de relação entre requisitos e dimensões, optou-se por não desfazer as relações entre o requisito Metadados dos insumos (A8) e a dimensão acurácia e confiabilidade, da mesma forma que entre Complexidade estrutural (A11) e oportunidade temporal e pontualidade. Sobre a primeira das decisões, o entendimento é o de que dados do tipo Big Data não acompanhados de metadados facilitarão o cometimento de erros de interpretação, culminando em erros não amostrais por vezes desconhecidos, prejudicando a acurácia e a confiabilidade do produto estatístico. Para a decisão seguinte, acredita-se que insumos com maior complexidade estrutural deverão demandar uma quantidade maior de recursos, inclusive tempo de processamento,

impactando na qualidade do produto em relação ao aspecto do oportunismo temporal;

- iii. Ainda na linha dos itens imediatamente anteriores, decidiu-se, após análise e reflexão, por estabelecer uma relação entre a Agilidade da produção (B5) e a dimensão relevância, e desfazer a relação entre os requisitos Capacitação para uso (C7) e Disseminação e promoção (C9) com a mesma dimensão. Além disso, não foram modificados os vínculos dos requisitos Percepção ética (C4), Periodicidade praticada (C5), Metadados do indicador (C6) e Disponibilidade do Indicador (C8), com a dimensão Relevância. O entendimento aqui é o de que a relevância de um produto estatístico, associada ao quão bem o indicador satisfaz a necessidade do usuário, não diz respeito somente ao conteúdo refletido pelo produto. A expectativa do usuário se refere também, por exemplo, aos aspectos de custo, temporalidade, facilidade de acesso, credibilidade e compreensibilidade do produto estatístico, entre outros. De nada serve, para o usuário, um indicador que carregue a informação necessária, mas que seja inviável por questões de alto custo ou inútil do ponto de vista prático por defasagem temporal, por exemplo. Da mesma forma, pior será um indicador que transmita pouca credibilidade por seus insumos ou reputação do produtor, que não seja divulgado regularmente conforme uma periodicidade estabelecida, que seja malvisto do ponto de vista ético ou que não esteja transparecido através dos seus metadados;
- iv. Também se optou por não relacionar o requisito Adequação conceitual (C2) com a dimensão acurácia e confiabilidade pelo entendimento de que esse requisito faz referência única e exclusivamente à expectativa ou necessidade do usuário, impactando na qualidade estatística somente por meio da relevância do indicador. Ou seja, um indicador pode ter boa acurácia e confiabilidade mesmo que não atenda ao requisito C2 – acontecendo quando o indicador conseguir representar o contexto de interesse pensado para o produto estatístico, mas não estiver de acordo com o que espera um determinado usuário. Essa percepção faz retomar o princípio de que a qualidade estatística é relativa, pois um mesmo

indicador pode ser relevante do ponto de vista de adequação conceitual para um tipo de usuário, mas irrelevante para outro. Em ambas as situações, a acurácia e confiabilidade do indicador permanece a mesma;

- v. Refletiu-se sobre a vinculação dos requisitos Base legal (A1), Acordo formal de cooperação (A4), Credibilidade do produtor (B1) e Percepção ética (A4) com uma possível nova dimensão “Ambiente Institucional”, modificando a estrutura inicial pensada para a composição da qualidade estatística de um produto. Por questões de se partir aqui do princípio de que o conceito da qualidade estatística está associado à expectativa do usuário, optou-se por manter a proposição inicial com as cinco dimensões de qualidade de produtos estatísticos padronizadas nacional e internacionalmente – sem a inclusão do ambiente institucional como uma nova dimensão. Entretanto, é reconhecido aqui que o ambiente institucional poderia ser considerado um domínio, um ambiente de pertencimento dos requisitos a serem observados, como acontece na maioria dos *frameworks* de qualidade para o contexto de estatísticas públicas tradicionalmente produzidas. Para o contexto de Big Data, porém, o domínio ambiente institucional aumenta em complexidade pelo fato de que duas podem ser as instituições envolvidas na produção de um indicador – a fornecedora do Big Data e a produtora estatística. Além disso, durante o fluxo de produção estatística, alguns aspectos se relacionam com o ambiente institucional no âmbito dos insumos, outros durante a produção e outros no contexto do produto finalizado. Então, ao se considerar a segmentação do fluxo do processo em insumos, produção e produto, o ambiente institucional não encaixaria como uma etapa nessa sequência linear, pois possuiria aspectos a serem observados durante todas as etapas. Ou seja, é como se o ambiente institucional fosse um domínio transversal no fluxo de construção de produtos estatísticos a partir de Big Data – por vezes chamado de hiperdomínio. Aqui, para fins de redução de complexidade, não foram considerados hiperdomínios no artefato proposto. O entendimento é o de que este encaminhamento não acarretaria perda de qualidade para o *framework*, uma vez que todos os requisitos associados ao ambiente institucional continuam sendo

considerados e observados, porém de forma espalhada pelos domínios de insumos (quando se referir à organização geradora do Big Data e seu ambiente externo), de produção (quando se referir ao produtor estatístico e seu ambiente) e de produto (quando se referir à percepção do usuário final). Reforçando que, apesar de estarem inseridos no hiperdomínio ambiente institucional – oculto, nesta proposta –, tais requisitos, caso não atendidos por um indicador, impactam em alguma – podendo ser em mais de uma – dimensão de qualidade deste produto estatístico – Relevância, Acurácia e Confiabilidade, Oportunismo temporal e Pontualidade, Coerência e Comparabilidade e/ou Acessibilidade e Clareza;

vi. Por fim, mas não menos importante, foram feitas reflexões sobre o aspecto estrutural que diz respeito à necessidade de todas as dimensões possuírem pelo menos um requisito associado no âmbito de cada um dos três domínios e, também, sobre a necessidade de minimizar a ocorrência de relacionamento entre um requisito e mais de uma dimensão de qualidade ao mesmo tempo. Sobre a primeira questão, a proposta atual de fato compreende que nenhum dos requisitos elencados como pertencentes ao domínio de produto impactam ou interferem na acurácia e confiabilidade daquele indicador. Isso ocorre pelo entendimento de que o produto corresponde ao indicador já finalizado, que já não pode ser modificado em termos de conteúdo ou estrutura. Então, nada mais poderia alterar sua qualidade no que diz respeito à acurácia e à confiabilidade que possui. Os requisitos do domínio de produto terão relação, portanto, com o quanto o produto finalizado consegue atender às necessidades do usuário (Relevância) e é efetivamente disponibilizado nos tempos pactuados (Oportunismo temporal e Pontualidade), de forma padronizada (Coerência e Comparabilidade) e acessível (Acessibilidade e Clareza). Sobre a segunda questão, o estabelecimento de relacionamento entre um requisito e mais de uma dimensão foi proposta por motivos de se permitir que o parecer de qualidade gerado pelo *framework* seja emitido de forma fragmentada, por dimensões em separado. Dessa forma, partiu-se de tentar perceber, no âmbito de cada uma das dimensões,

quais seriam os requisitos determinantes da sua qualidade. Foi compreendido, nesses termos, que alguns requisitos estariam impactando em mais de uma dimensão. Ou seja, quando da construção dos pareceres de qualidade segmentados por dimensão, tais requisitos seriam considerados mais de uma vez. Entretanto, quando for de interesse a construção do parecer geral de qualidade do produto estatístico, o nível de qualidade do indicador será obtido pela média dos escores de todos os requisitos – e não pela média dos escores médios de cada dimensão. Seguindo este procedimento, os requisitos são contabilizados uma única vez quando da emissão do nível de qualidade estatística de um produto, contornando o problema da duplicidade – mesmo que sejam considerados em mais de duas dimensões, quando da avaliação fragmentada. Cabe destacar que, dessa forma, a operacionalização proposta para a obtenção da média geral fornece, mesmo que implicitamente, maior peso para os domínios e dimensões que possuem uma quantidade maior de requisitos. Para ambas as situações expostas, é reconhecido que aprofundamentos de experimentação de uso do *framework* deverão ser conduzidas – com objetivo de se analisar o impacto que causam na avaliação final da qualidade dos produtos.

Por fim, em referência ao último aspecto analisado, a proposta de operacionalização prática do *framework*, a consulta aos especialistas apontou para a necessidade de reflexão e possível alteração em uma série de itens, tais quais considerar:

- i. Utilizar pesos diferentes por dimensões e domínios, com uma eventual maior importância atribuída aos requisitos do domínio de insumos;
- ii. Incorporar algum procedimento de análise da qualidade estatística que compare os resultados do indicador inovador, construído a partir de Big Data, com os resultados de algum indicador tradicionalmente produzido – indicador “padrão ouro”, um *benchmark* – para que se perceba o quanto se distanciam – como forma de estimação do viés;
- iii. Evoluir em testes de simulação e na análise da viabilidade de mensuração das dimensões "Oportunismo temporal e Pontualidade" e "Acurácia e Confiabilidade", principalmente em dados de *webscraping*,

pela dificuldade de confirmação quanto ao atendimento dos requisitos, por vezes desconhecidos pelo produtor;

- iv. A limitação trazida pela arbitrariedade da proposta de diagnóstico do requisito, que considera uma única pergunta, que por vezes apresentam interpretação dúbia, com duas perguntas em uma, acrescido do componente subjetivo de interpretação quanto ao atendimento de cada requisito por parte do indicador, exigindo alta expertise prática e teórica do indivíduo ou grupo avaliador;
- v. Apresentar maior clareza na forma de cálculo do nível de qualidade estatística do parecer final do *framework*;
- vi. Realizar estudos robustos para depurar e atestar sobre a qualidade do artefato proposto, com validações e confirmações de face, de construto e das escalas de modo geral.

Sobre o aspecto da operacionalização prática, cabe demarcar que a proposta aqui apresentada tem por utilidade maior a apresentação de uma possibilidade de caminho a ser seguido por pesquisas futuras. Não está no escopo do objetivo geral desta tese a construção de um protocolo definitivo para operacionalização prática do artefato, mas sim a proposição de um *framework* de cunho conceitual, conforme definição trazida no Capítulo 3.

Assim, reforça-se aqui que o protocolo de operacionalização prática apresentado teve por finalidade, essencialmente: fomentar a pesquisa científica sobre o assunto, servindo de ponto de partida para o desenvolvimento de uma mensuração mais factível da qualidade estatística; além de viabilizar, através de um exercício de simulação, uma primeira análise sobre a aplicabilidade do *framework* conceitual proposto – para fins de atendimento do último objetivo desta tese.

Ou seja, se reconhece aqui a necessidade de muitos avanços no protocolo de operacionalização prática utilizado, assim como de testes e melhorias na proposta do *framework* conceitual em si, sendo todos os pontos até então percebidos adereçados no Capítulo 5, de considerações finais.

Realizadas as modificações no *framework*, motivadas pelas manifestações dos especialistas, a aplicação da técnica Delphi foi encerrada após a segunda rodada de aplicação dos questionários em virtude de alto nível de concordância do

grupo consultado com a proposta aqui trazida, indicativo de convergência de opiniões, conforme escores resumidos pela Tabela 2 e comentários finais apresentados no Quadro 26.

Tabela 2 – Escore médio e distribuição dos especialistas por nível de concordância com os pontos avaliados do *framework* nas diferentes rodadas do Delphi.

Rodada	Referência	Nível de concordância					Escore médio
		Discordo totalmente	Discordo parcialm.	Indiferente (neutro)	Concordo parcialm.	Concordo totalmente	
1	1. Estruturação geral	0	0	0	4	5	4,6
	2. Dimensões	0	0	0	2	7	4,8
	3. Domínios	0	0	0	0	9	5,0
	4. Requisitos	0	1	0	3	5	4,3
	5. Relacionamentos	0	2	0	2	5	4,1
	6. Operacionalização	0	0	1	5	3	4,2
	Total	0	3	1	16	34	4,5
2	1. Estruturação geral	0	0	0	1	8	4,9
	2. Dimensões	0	0	0	1	8	4,9
	3. Domínios	0	0	0	1	8	4,9
	4. Requisitos	0	0	0	2	7	4,8
	5. Relacionamentos	0	0	0	2	7	4,8
	6. Operacionalização	0	0	0	2	7	4,8
	Total	0	0	0	9	45	4,8

Fonte: elaborado pelo autor (2021) com base nas manifestações dos especialistas.

Quadro 26 – Manifestações dos especialistas após a segunda rodada de questionários.

Especialista	Manifestação
1	<i>“Parabéns pela consolidação da primeira rodada do Delphi e pela forma como você considerou, analisou e incorporou as contribuições dos especialistas. O framework foi refinado e, na minha avaliação os requisitos de domínios e os relacionamentos estão bem fundamentados. O tema de pesquisa é de fato complexo e o framework tenta cercar todos as dimensões de qualidade para a produção de estatísticas de qualidade. Acho que ainda temos uma quantidade de requisitos grandes, sobretudo no domínio de ‘insumos’, mas também não sei como reduzir. Portanto, melhor seguir com todos os requisitos.”</i>
2	<i>“O framework proposto é um excelente ponto de partida para a construção de uma ferramenta que avalie a qualidade de indicadores com base em fontes de Big Data. Como apresentado, não é uma proposta final, mas traz em si um conjunto bem amplo de dimensões, requisitos e definições que já permitem antever a qualidade de uma estatística calculada com fontes alternativas. O framework proposto, inclusive, é bastante geral, podendo ser aplicado como fonte de qualificação de registros e dados que não são coletados com a intenção de produção de indicadores (mesmo os que não sejam imediatamente classificados como Big Data).”</i>
3	<i>“As observações fornecidas na primeira fase foram trabalhadas com detalhes. O conteúdo proposto está claro.”</i>
4	<i>“O trabalho evoluiu e isso foi muito bom.”</i>
5	<i>“Excelente trabalho. Gostaria de ler a tese final.”</i>
6	<i>“Sem comentários adicionais.”</i>
7	<i>“É nítida a melhoria com os ajustes efetuados no framework reformulado, em especial no que diz respeito a expansão ou reformulação da definição de requisitos e as alterações efetuadas na estrutura de relacionamento entre requisitos, domínios e dimensões. Cabe reforçar o reconhecimento de que há campo para possíveis melhorias e avanços em termos de validações e operacionalização prática do framework.”</i>
8	<i>“Sem comentários adicionais.”</i>
9	<i>“Parabéns pelo trabalho!”</i>

Fonte: elaborado pelo autor (2021) com base nas manifestações dos especialistas.

Destaca-se que a construção dos escores médios de concordância dos especialistas parte do pressuposto de que o nível de mensuração da variável correspondente ao nível de concordância com os aspectos avaliados, originalmente ordinal, pode ser considerado intervalar – assumindo equidistância numérica (diferenças constantes) entre os cinco níveis, arbitrariamente pontuados com valores de 1 a 5.

4.9 ESPECIFICAÇÃO DAS APRENDIZAGENS

O alcance da versão final do *framework* aqui apresentado é fruto de um processo longo de pesquisa que envolveu pelo menos três grandes momentos: a revisão bibliográfica, para compreensão do que já estava sendo pesquisado e feito; a criação do artefato em si, utilizando-se de uma abordagem predominantemente abductiva; e a avaliação da proposta, com verificação de estrutura junto a especialistas e simulação de aplicabilidade em um contexto real.

Destaca-se, entretanto, que, além do trabalho operacional de pesquisa, que obviamente necessita ser rigorosamente bem-feito, a experiência prática prévia deste autor, de seus orientadores e do grupo de especialistas, foi elemento imprescindível para que o caminho metodológico trilhado conseguisse ser percorrido por completo.

Isso porque a Design Science (DS) possui, em sua essência, por conta da sua base filosófica associada a um contexto cujo objetivo de pesquisa está na proposição de um artefato, a necessidade do emprego da criatividade durante a aplicação do método para construção de conhecimento. Ou seja, percebidos os problemas com os artefatos até então existentes, ou as necessidades de preenchimento de lacunas, o pesquisador deverá refletir e propor possíveis soluções para alterar e melhorar a situação presente. Nesse contexto, a criatividade se faz crucial e mais disponível estará quanto maior for a experiência – vivência prática – dos proponentes com o tema.

Outra questão, em linha semelhante, diz respeito à dificuldade de aplicação da técnica Delphi em contextos de construção de um conhecimento muito específico

e/ou complexo. Primeiramente, por uma maior dificuldade em encontrar um grupo de especialistas, em quantidade aceitável, que pudesse contribuir satisfatoriamente – com, inclusive, maior chance de virem a desistir durante o processo. Em segundo lugar, pelo maior tempo necessário para exposição e compreensão da solução, reflexão e coleta de opiniões, e modificações e alcance de um consenso.

A saber, treze (13) especialistas foram convidados a participar desta pesquisa, observando-se duas recusas, duas desistências no decorrer do processo e nove conclusões. Além disso, pensou-se inicialmente em um tempo de quinze (15) dias para aplicação da primeira rodada dos questionários e de dez (10) dias para a segunda. Na prática, os tempos foram de vinte e sete (27) e de quinze (15) dias, respectivamente.

Em adição, ainda sobre a técnica Delphi, destaca-se que, por conta da realização à distância, via internet, é imprescindível que a abordagem e a comunicação com os especialistas sejam planejadas e realizadas de modo a alcançar um satisfatório convencimento e a manter o engajamento do grupo para uma completa jornada de consulta.

Saindo do campo prático e adentrando no campo teórico, é importante também demarcar o fato de que os principais construtos aqui trabalhados são conceitos abstratos, teóricos, mentalmente construídos e intangíveis, sendo esse talvez o ponto central a carecer de atenção para trabalhos futuros semelhantes, quando conduzidos, uma vez que há forte presença de componente subjetivo no tema explorado e, se definições não estiverem bem estabelecidas, diferentes resultados e conclusões de pesquisa poderão ser alcançados.

O conceito de qualidade de um produto estatístico, ou simplesmente da qualidade estatística, por exemplo, foi aqui definido partindo-se da premissa de que se trata de uma construção multidimensional composta por cinco (5) princípios desejados pelos usuários para um determinado produto estatístico. Tais princípios, aqui ditos dimensões de qualidade, também são conceitos abstratos que carregam definição teórica e caráter intangível, estando também à mercê da qualidade e da concordância das suas definições.

Da mesma forma, apresentam também forte componente subjetivo envolvido: as tentativas de operacionalização prática de cada dimensão de qualidade, através da observação ou mensuração dos requisitos; a subdivisão do

fluxo da produção estatística em domínios, ou ambientes de pertencimento dos requisitos; e a própria estrutura de relacionamento – pertencimento e/ou influência – entre dimensões, domínios e requisitos.

Nesses termos, *frameworks* para avaliação e pareceres sobre a qualidade de um produto estatístico se mostram relativos não somente por conta dos aspectos que caracterizam o contexto prático da sua utilização, percebidos ao longo da Subseção 4.2.3, mas também pelas premissas e definições que carregam. Dessa forma, um mesmo grupo de renomados especialistas no assunto poderia concordar com diferentes protocolos de qualidade, dependendo do contexto prático de aplicação e do paradigma teórico de base.

5 CONSIDERAÇÕES FINAIS

As considerações finais destacam os resultados e avanços alcançados com a presente tese, em resposta às questões e lacunas de pesquisa, com relação aos objetivos estabelecidos, seguidos das conclusões, do apontamento das limitações às quais estão submetidos e da apresentação de recomendações e sugestões para trabalhos futuros, organizadas de modo a preencher as etapas de x – conclusões – até xii – comunicação dos resultados – propostas pelo método DSR (DRESCH; LACERDA; ANTUNES JR., 2015).

5.1 CONCLUSÕES

A presente tese, em atendimento ao seu objetivo geral de propor um *framework* conceitual para a avaliação da qualidade estatística de indicadores de turismo produzidos a partir de Big Data, alcançou como principal resultado, ao final da trajetória da pesquisa, um artefato baseado na qualidade estatística enquanto conceito multidimensional que reflete capacidade de um produto estatístico – indicador de turismo – em atender as necessidades dos usuários no âmbito da combinação de cinco (5) princípios desejados, as dimensões, conceitos abstratos, operacionalizados de forma prática por um conjunto de vinte e sete (27) requisitos, observáveis ou mensuráveis, presentes em alguma das três (3) etapas, ou ambientes, os domínios, do fluxo de produção estatística – conforme apresentaram as Seções 4.5, 4.6 e 4.7. Em adição, o atendimento ao objetivo geral foi concluído com a proposição das definições conceituais e da estrutura de relacionamento entre os elementos constituintes do *framework* (dimensões, domínios e requisitos).

A partir da estrutura proposta, o artefato orienta a construção de pareceres sobre o nível de qualidade estatística de um indicador de turismo construído a partir de Big Data, gerais ou parciais (por domínios ou dimensões), baseado na observação ou mensuração quanto ao atendimento dos requisitos de qualidade por parte do produto analisado. Em posse dos pareceres, produtores estatísticos poderão direcionar o emprego de esforços para melhoramento e correção de falhas nos domínios e dimensões mais críticos, de menor qualidade, e usuários do produto

estatístico conseguirão vislumbrar cuidados e potenciais riscos que permeiam o indicador – devendo serem considerados durante sua utilização e interpretação.

Destaca-se que o artefato aqui desenvolvido é o mais abrangente (*comprehensive*) em termos de aspectos de qualidade que considera quando da avaliação de estatísticas públicas e indicadores socioeconômicos construídos a partir de Big Data, na comparação com os outros *frameworks* analisados (Seção 4.4), tendo definido como requisitos todos os fatores determinantes da qualidade estatística mencionados por pesquisadores percebidos ao longo da construção desta tese. Além disso, é o único até então percebido em nível nacional, assim como o único construído especificamente para o contexto dos indicadores de turismo, entregando também, mesmo que não tenha sido um objetivo da pesquisa, um protótipo de protocolo a servir de ponto de partida para uma operacionalização prática que forneça a quantificação do nível de qualidade estatística de forma mais viável (Subseção 4.8.2), até então também não considerado por outros *frameworks*.

Dessa forma, acredita-se que a entrega aqui realizada esteja fornecendo sua parcela de contribuição no avanço do preenchimento das lacunas de pesquisa identificadas (Seção 4.2). Além disso, espera-se que este trabalho tenha conseguido atender as expectativas de promover a temática e a sua relevância, apresentando oportunidades de geração de conhecimento no âmbito dos indicadores socioeconômicos a partir de novas fontes de dados e contribuindo com soluções que podem impulsionar o contorno às dificuldades enfrentadas pelos processos tradicionais de produção de estatísticas públicas, cooperando inclusive com o reposicionamento estratégico dos atuais produtores (Seção 4.1).

Para que o resultado maior – o *framework* conceitual – fosse alcançado, outros processos intermediários foram percorridos, em linha com os objetivos específicos desta tese, tais quais:

- i. a identificação de potenciais fontes de dados do tipo Big Data e correspondentes indicadores de turismo que podem auxiliar a gerar, com percepção de que as principais fontes são aquelas ligadas à *World Wide Web*, enquanto as aplicações práticas principais se referem à estimação ou à projeção da quantidade ou demanda de turistas de entrada ou saída (Subseção 2.6.2);

- ii. a análise da estrutura de *frameworks* para a qualidade de estatísticas públicas produzidas a partir de fontes tradicionais (Subseção 2.3.2) e de Big Data (Seções 4.2 e 4.4), permitindo a identificação das principais lacunas nos artefatos até então construídos para o contexto de Big Data, como a necessidade de incorporação de novos aspectos de qualidade, de organização de uma mensuração mais factível e de estruturação menos rígida ou direcionada para algum contexto específico;
- iii. a definição de um arcabouço de requisitos, dimensões e domínios de qualidade aplicável aos indicadores de turismo produzidos a partir de Big Data, possibilitando se chegar às quantias de vinte e sete requisitos, três domínios e cinco dimensões (Subseção 4.2.1 e Seções 4.5, 4.6 e 4.7);
- iv. a verificação do *framework* proposto junto a produtores de estatísticas públicas e usuários de indicadores socioeconômicos, conduzida utilizando-se da técnica Delphi com um grupo heterogêneo de especialistas e que permitiu evoluir a proposta para seu formato final (Subseção 4.8.1);
- v. e a análise da aplicabilidade do *framework* em uma situação real, viabilizada por meio de uma proposta básica e preliminar de protocolo de operacionalização prática, utilizado para fins de simulação, aplicado a um indicador de turismo construído a partir de dados coletados do portal TripAdvisor (Subseção 4.8.2).

Em relação a todos os resultados alcançados, entretanto, cabe reforçar algumas possíveis limitações às quais podem estar submetidos, aqui subdivididas em (1) metodológicas ou associadas às características do (2) produto em si. Em relação ao primeiro grupo de eventuais limitações, destaca-se que:

- i. o paradigma (*Design Science*) e o método de pesquisa (*Design Science Research*) conduziram a resultados altamente determinados pela criatividade, pelo conhecimento e pela experiência prática do autor;
- ii. a técnica Delphi foi executada, por questões de incipiência, especificidade e complexidade do tema, com um grupo de, apesar de heterogêneo e altamente qualificado, apenas 9 especialistas selecionados de maneira não probabilística, intencional (Subseção 3.1.2);

- iii. as revisões bibliográficas da pesquisa tiveram delimitação inicial influenciada pelos critérios considerados nos protocolos das buscas sistemáticas e integrativas da literatura (Subseção 3.1.1); e que
- iv. o indicador de turismo construído a partir de Big Data (Subseção 3.1.2.1), de papel secundário nos objetivos da tese, utilizado durante a análise da aplicabilidade do *framework*, pode conter vieses por desalinhamento entre a população alvo e a amostra, por seletividade na amostra, por má especificação das quantidades e dos critérios de seleção de atrações locais, pela utilização de registros de avaliações falsas, pela invalidez da suposição de que as avaliações (insumos) estejam em escala intervalar, pela baixa quantidade de turistas e avaliações realizadas no período da pandemia de Covid 19, entre outros.

Além dessas, quando do projeto e do desenvolvimento do artefato, durante o processo de especificação das suas características e funcionalidades, algumas decisões podem ter também gerado possíveis limitações no (2) produto em si, tais quais:

- i. o cunho apenas conceitual e a delimitação do contexto de aplicação do *framework* (Seção 4.5), voltado primordialmente para a identificação dos elementos componentes do universo da avaliação da qualidade de indicadores de turismo produzidos a partir de Big Data, e suas relações, a fim de nortear análises, compreendendo os processos e prevendo os resultados – restritivo quanto a imediata operacionalização prática do artefato e sua aplicação em diferentes contextos, aspectos a serem abordados por trabalhos futuros;
- ii. uma estrutura final vinculada às premissas teóricas, aos paradigmas de base e, conseqüentemente, às definições conceituais adotadas para os construtos e elementos componentes do sistema (qualidade estatística, dimensão, domínio, requisito) e para o entendimento de relacionamento, pertencimento e influência entre eles (Seções 4.5, 4.6 e 4.7), que, fossem outros, provavelmente levariam a resultados diferentes;
- iii. a proposta de operacionalização prática do *framework* que, em virtude do seu caráter preliminar e de sua contribuição, nesta tese, apenas auxiliar,

secundária, não passou por diversos tipos de validação, como de face, construto e escala, tampouco foi analisada em relação aos impactos, no parecer final de qualidade, do componente subjetivo intrínseco ao formato do protocolo de perguntas e da fórmula de cálculo que considerou, não avaliando uma eventual atribuição de pesos diferentes aos requisitos, domínios e dimensões de qualidade – temas para trabalhos futuros, também; e

- iv. novamente a proposta de operacionalização, que necessita, para fins de pleno funcionamento, da participação da organização fonte dos insumos (Big Data) e da instituição produtora do indicador de turismo durante o processo de análise dos requisitos, além de um grupo de avaliadores qualificado e experiente no assunto – objetivando contornar uma eventual subjetividade presente no protocolo de perguntas.

Dito isso, ao passo que se acredita aqui – dadas as características, funcionalidades e diferenciais que o *framework* carrega – no satisfatório atendimento aos objetivos traçados, na contribuição para o avanço no preenchimento das lacunas de pesquisa sobre o tema e no sucesso associado à promoção da temática e sua relevância, é reconhecido também que a proposta – dada a incipiência e a complexidade do tema e as limitações que carrega – não possui caráter definitivo, não havendo pretensão aqui de se dar o artefato por concluído. Nesse sentido, entende-se que o *framework* construído nesta tese corresponde a uma proposta inicial, um ponto de partida, com valor maior relacionado à demarcação dos princípios fundamentais e da movimentação da investigação científica sobre a temática, consolidando as bases teóricas e sinalizando os possíveis caminhos futuros a serem percorridos, tanto em termos de dar sequência às próximas etapas de investigação quanto da necessidade de se validar e aprimorar o que já está posto, tais quais realizar estudos e experimentos práticos para:

- i. no âmbito dos requisitos de qualidade considerados pelo *framework*, perceber se ainda há algum aspecto de qualidade a ser incluído, se existe alguma redundância que precisa ser corrigida com fusão, se algum está muito abrangente e necessita ser subdividido, se algum apresenta mensuração inviável, e se a quantidade atual de requisitos está adequada;

- ii. ainda sobre os requisitos, construir um modelo de referência para os graus de atendimento para cada um, ou seja, um referencial a servir de parâmetro e que indique como se caracteriza um indicador de turismo que atende satisfatoriamente ou não cada um dos requisitos de qualidade do *framework*;
- iii. em termos da estrutura de relacionamentos dos requisitos com as dimensões e os domínios, confirmar se existe adequação nos vínculos, se a quantidade de requisitos por domínios e dimensões está apropriada, e avaliar o impacto, no parecer final de qualidade, de se possibilitar que um requisito esteja relacionado com mais de uma dimensão e que alguma dimensão esteja sem nenhum requisito associado no âmbito de algum dos domínios;
- iv. sobre o protótipo de operacionalização prática, considerar sua estruturação a partir de uma adaptação da ferramenta *fuzzy Quality Function Deployment (fuzzy-QFD)* ou similares, avaliar o impacto da mensuração de um requisito através de uma única pergunta, confirmar viabilidade da mensuração, considerar incorporar a análise comparativa dos resultados do indicador inovador com os de um indicador tradicionalmente produzido, revisar as perguntas para detecção de dubiedades e subjetividade e confirmação de alinhamento com as definições dos requisitos, realizar validações de face, construto e escala, e analisar o impacto da utilização de pesos diferentes para os requisitos, domínios e dimensões na geração dos pareceres de qualidade; e
- v. por fim, em relação ao indicador de turismo construído a partir de Big Data utilizado durante a análise da aplicabilidade do *framework*, para fins de se dar continuidade ao mesmo, avaliar o tamanho de um eventual desalinhamento entre a população alvo e a amostra, assim como de seletividade na amostra, validar a quantidade e os critérios de seleção das atrações consideradas, e mensurar a quantidade e o impacto da eventual presença de registros falsos.

5.1.1 Reflexão final

Diante do exposto ao longo de toda a tese, cabe fazer uma última reflexão sobre a temática trabalhada em termos de implementações e implicações, no contexto mais amplo da produção de estatísticas públicas.

No âmbito das sociedades democráticas, estatísticas públicas objetivas e independentes são bens públicos essenciais. Além disso, para a gestão pública, em virtude do caráter institucional das estatísticas, há necessidade de que sejam periódicas e abrangentes, de construção planejada a partir das demandas por problemas socioeconômicos da população, para fins de subsídio à formulação e à gestão de políticas, permitindo ao Estado um papel propositivo. Nesses termos, faz-se difícil vislumbrar a possibilidade de que indicadores socioeconômicos produzidos a partir de Big Data por instituições externas ao sistema estatístico nacional tradicional venham a tornar obsoleta e substituir a produção estatística vigente, em um contexto dominado por uma lógica anárquica ou de livre mercado. Para que isso aconteça, seria necessário dispor de uma grande estrutura comprometida com o obrigatório fornecimento regular de uma gama expressiva de estatísticas e indicadores, garantidos em termos de qualidade – o que faz levantar, nesses termos, tamanha a relevância dos produtos estatísticos, a eventual necessidade de constituição de órgãos independentes e políticas de certificação e regulação dos produtos, sobretudo face a subjetividade presente na proposta de *framework* desenvolvida.

De qualquer forma, crê-se que os atuais produtores podem se utilizar da solução e dos conhecimentos aqui trabalhados para avançar na complementação e na expansão da atual oferta de estatísticas, ou mesmo na substituição do método de produção de algumas, tirando proveito das oportunidades trazidas pelo contexto de Revolução dos Dados – e, inclusive, resguardando-se frente uma retórica liberal pós-moderna.

Do ponto de vista do setor privado, por sua vez, posto que as informações estatísticas servirão a interesses particulares, organizações poderão desenvolver suas formas alternativas de construção de conhecimento, sob a forma explícita dos indicadores socioeconômicos produzidos a partir de Big Data, vislumbrando

finalidades próprias ou de atendimento de mercado, baseando-se para tal no conteúdo aqui explorado e construído.

5.2 GENERALIZAÇÃO PARA UMA CLASSE DE PROBLEMAS

Apesar de aqui construído para o contexto da produção de indicadores de turismo, muito em virtude da percepção – pelas revisões de literatura realizadas – de que *frameworks* para a qualidade de produtos estatísticos gerados a partir de Big Data necessitam ser desenhados para contextos específicos ou flexíveis a ponto de se adaptarem, não foram detectados entraves que inviabilizem a utilização do *framework* proposto em outros contextos da produção estatística a partir de Big Data.

Dessa forma, acredita-se que o artefato resultante desta tese possa ser expandido e/ou adaptado, se aplicando ao espectro da análise da qualidade de indicadores socioeconômicos produzidos a partir de Big Data como um todo, ou, da mesma forma, a todo o tipo de estatística pública e/ou oficial construída a partir de fontes alternativas de dados.

Nesse sentido, considera-se que o *framework* desenvolvido fornece uma satisfatória base, com pilares comuns que podem ser compartilhados por protocolos desenvolvidos para outras finalidades específicas.

5.3 COMUNICAÇÃO DOS RESULTADOS

Por fim, destaca-se que muitos dos resultados gerados durante o processo de pesquisa conduzido nesta tese, associados às etapas intermediárias, executadas em atendimento aos objetivos específicos do trabalho, já foram publicados. A saber, no livro “População, Território e Estatísticas Públicas”, comemorativo aos 20 anos da pós-graduação da Escola Nacional de Ciências Estatísticas, do Instituto Brasileiro de Geografia e Estatística (CAUMO, 2020), no periódico Planejamento e Políticas Públicas do Instituto de Pesquisa Econômica Aplicada (CAUMO; SOUZA, 2021), na revista E-Tech, do Serviço Nacional de Aprendizagem Industrial (CAUMO; BIZ; SOUZA, 2020), e nos anais da 15ª Conferência Internacional de Gestão do

Conhecimento e do 2º Seminário Internacional de Universidades Corporativas e Escolas de Governo (CAUMO; BIZ; SOUZA, 2019), conforme descreve a Seção 4.3.

Em adição, está sendo construído mais um artigo, este com a versão final do *framework* aqui proposto, acompanhado do detalhamento do processo de verificação por meio da aplicação da técnica Delphi junto aos especialistas, a ser publicado em revista científica de alcance internacional.

REFERÊNCIAS

ABNT. **NBR ISO 9000:2005**. Sistemas de gestão da qualidade – Fundamentos e vocabulário. Rio de Janeiro: ABNT, 2005.

ABIDIN, S. Z.; KHAIRI, Basyirah M.; LING, C. T. **Harnessing the value of administrative data**: A Central Bank's experience. *In*: ISI WORLD STATISTICS CONGRESS, 62., 2019 aug 18-23; Kuala Lumpur, 2019.

ABS. **The ABS Data Quality Framework**. 2009. Disponível em: <http://www.abs.gov.au/websitedbs/D3310114.nsf/home/Quality:+The+ABS+data+quality+framework>. Acesso em: 28 jan. 2019.

AGAFITEI, M.; GRAS, F.; KLOEK, W.; REIS, F.; VHATAJU, S. Measuring output quality for multisource statistics in official statistics: Some directions, **Statistical Journal of the IAOS**, v.31, n. 2, p. 203-211. 2015.

AHAS, R.; AASA, A.; ROOSE, A.; MARK, Ü.; SILM, S. Evaluating passive mobile positioning data for tourism surveys: An Estonian case study. **Tourism Management**, v.29, n.3, p. 469-486, 2008.

AHAS, R.; SILM, S.; JÄRV, O.; SALUVEER, E.; TIRU, M. Using Mobile Positioning Data to Model Locations Meaningful to Users of Mobile Phones. **Journal of Urban Technology**, 17, p. 3-27. 2010.

AHMAD, M. K. S.; SAKRI, M. S. M.; SULAIMAN, I. M.; SALLEH, S. M.; LUKOSE, D.; ISMAIL, O.; LATIP, A. A.; MAZLAN, M. A. **MyHarmony**: Generating statistics from clinical text for monitoring clinical quality indicators. *In*: ISI WORLD STATISTICS CONGRESS, 62., 2019 aug 18-23; Kuala Lumpur, 2019.

AL-HAJJAR, D.; JAAFAR, N.; AL-JADAAN, M.; ALNUTAIFI, R. **Framework for Social Media Big Data Quality Analysis**, AISC, v. 312, p 301-314, 2015.

ALIS, C.; LETCHFORD, A.; MOAT, H. S.; PREIS, T. **Estimating tourism statistics with Wikipedia page views**. p. 1-2. 2015.

ALQADHI, Z.; ELKABBANY, M. F.; SHEHHI, H. A.; ALI, N. **Smart meters data as a source of household and farms statistics a case study from the United Arab Emirates**. *In*: ISI WORLD STATISTICS CONGRESS, 62., 2019 aug 18-23; Kuala Lumpur, 2019.

ALVAREZ, G. M. **Análise de agrupamentos e mineração de opinião como suporte à gestão de ideias**. 2018. Dissertação (Mestrado em Engenharia e Gestão do Conhecimento) – Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento, Universidade Federal de Santa Catarina, Florianópolis, 2018.

AMAYA, A.; BIEMER, P. P.; KINYON, D. Total Error in a Big Data World: Adapting the TSE Framework to Big Data, **Journal of Survey Statistics and Methodology**, v. 8, Issue 1, p. 89–119, 2020. DOI: <https://doi.org/10.1093/jssam/smz056>.

ANDRÉ, I.; ABREU, A. **Dimensões e espaços da inovação social**. Finisterra, v.41 n.81, p. 121-141, 2006.

ARACRI, R. M.; BIANCO, A. M.; RADINI, R.; SCANNAPIECO, M.; TOSCO, L.; CROCE, F.; SAVO, D. F.; LENZERINI, M. **On the Experimental Usage of Ontology-based Data Management for the Italian Integrated System of Statistical Registers: Quality Issues**. Paper for the European conference on Quality in Official Statistics, Krakow, Poland, 2018.

ASA. **Discovery With Data: Leveraging Statistics with Computer Science to Transform Science and Society**. July 2. 2014.

ASKITAS, N.; ZIMMERMANN, K. F. The Internet as a data source for advancement in social sciences. **International Journal of Manpower**, v.36, n.1, p. 2-12, 2015.

ASQUER, A. **The Governance of Big Data: Perspectives and Issues**. Department of Financial and Management Studies SOAS, University of London, 2013.

BAGGIO, R. Measuring Tourism: Methods, Indicators, and Needs. *In*: FAYOS-SOLÀ, E.; COOPER, C. (eds). **The Future of Tourism**. Springer, Cham, 2019.

BAKER, R. Big Data. A Survey Research Perspective. **Total Survey Error in Practice**, Wiley series in Survey Methodology, Hoboken, New Jersey, 2017.

BARBIERI, J. C. **Produção e transferência de tecnologia**. São Paulo: Ática S.A., 1990.

BARBOSA, D. M.; BAX, M. A. Design Science como metodologia para a criação de um modelo de Gestão da Informação para o contexto da avaliação de cursos de graduação. **Revista Ibero-Americana de Ciência da Informação**, v. 10, n. 1, p. 32-48, 2017. DOI: 10.26512/rici.v10.n1.2017.2471. Disponível em: <https://periodicos.unb.br/index.php/RICI/article/view/2471>. Acesso em: 11 jun. 2021.

BARCAROLI, G.; SCANNAPIECO, M.; SUMMA, D. On the use of internet as a data source for official statistics: a strategy for identifying enterprises on the web. **Rivista italiana di economia, demografia e statistica**, v. 70, p.20-41, 2016.

BARCAROLI, G.; GOLINI, N.; RIGHI, P. **Quality evaluation of experimental statistics produced by making use of Big Data**. *In*: European conference on Quality in Official Statistics, Krakow, Poland. 2018.

BATINI, C.; CAPPIELLO, C.; FANCALANCI, C.; MAURINO, A. Methodologies for data quality assessment and improvement. **ACM Computing Surveys (CSUR)**, v.41, n.3, p.16, 2009.

BATINI, C.; RULA, A.; SCANNAPIECO, M.; VISCUSI, G. From Data Quality to Big Data Quality. **Journal of Database Management**, v.26, p. 60-82, 2015.

BATINI, C.; SCANNAPIECO, M. **Data Quality: Concepts, Methodologies and Techniques**. 2006. DOI: 10.1007/3-540-33173-5.

BEAHAN, M. **A case study of data integration in Australia**. In: ISI WORLD STATISTICS CONGRESS, 62., 2019 aug 18-23; Kuala Lumpur, 2019.

BELL, D. **The Coming of the Post-Industrial Society**. New York: Basic Books (first ed. 1973), 1999.

BERGAMASCHI, S.; CARLINI, E.; CECI, M.; FURLETTI, B.; GIANNOTTI, F.; MALERBA, D.; MEZZANZANICA, M.; MONREALE, A.; PASI, G.; PEDRESCHI, D.; PEREGO, R.; RUGGIERI, S. Big Data Research in Italy: A Perspective. **Engineering**, 2, p.163-170, 2016.

BERZINA, I.; LAUBERTE, I. **The model of automation and extension of tourism economic impact assessment in specific regions**, v.2, p.195-202, 2018.

BIEMER, P. Total Survey Error: Design, Implementation, and Evaluation. **The Public Opinion Quarterly**, v.74, n.5, p. 817-848, 2010.

BIEMER, P.; TREWIN, D.; BERGDAHL, H.; LILLI, J. A system for managing the quality of official statistics, with discussion. **Journal of Official Statistics**, v.30, n. 3, Sept., p. 381-442, 2014.

BIEMER, P.; DE LEEUW, E.; ECKMAN, S.; EDWARDS, B.; KREUTER, F.; LYBERG, L.; WEST, B. (Eds.). **Total Survey Error in Practice**. Wiley Series in Survey Methodology. Wiley, 2017.

BIFFIGNANDI, S.; SIGNORELLI, S. **From Big Data to information: statistical issues through examples**, 2015.

BIN, C.; GU, T.; SUN, Y.; CHANG, L.; SUN, W.; SUN, L. Personalized POIs Travel Route Recommendation System Based on Tourism Big Data: 15th Pacific Rim International Conference on Artificial Intelligence, Nanjing, China, August, p.28-31, 2018, **Proceedings**, Part II.

BIZ, A. A. **Avaliação dos portais turísticos governamentais quanto ao suporte à estão do conhecimento**. 2009. Tese (Doutorado em Engenharia e Gestão do Conhecimento) – Universidade Federal de Santa Catarina, Florianópolis, 2009.

BLUMENSTOCK, J. E. Estimating Economic Characteristics with Phone Data. **AEA Papers and Proceedings**, 108I, p.72-76, 2018.

BOSSSEL, H. **Indicators for Sustainable Development: Theory, Method, Applications: A report to the Balaton Group.** Winnipeg: IISD, 1999.

BOSSOI, R. A. C. **A proteção dos dados pessoais face às novas tecnologias.** Direito e novas tecnologias, Florianópolis: CONPEDI, 2014.

BOULAALAM, O.; AGHOUTANE, B.; EL OUADGHIRI, D.; MOUMEN, A.; MALININE, M. L. C. Proposal of a Big data System Based on the Recommendation and Profiling Techniques for an Intelligent Management of Moroccan Tourism. **Procedia Computer Science**, 134, p. 346-351, 2018.

BRAAKSMA, B.; DAAS, P.; OFFERMANS, M.; PUTS, M.; TENNEKES, M. **Big Data and official statistics: local experiences and international initiatives.** Paper for the 47th Scientific Meeting of the Italian Statistical Society, p.11-13 june, Cagliari, Italy. 2014.

BRAAKSMA, B.; ZEELENBERG, K. 'Re-make/Re-Model': Should Big Data Change the Modelling Paradigm in Official Statistics? **Statistical Journal of the IAOS**, v.31, n.2, p.193-202, 2015.

BRACKSTONE, G. Managing data quality in a statistical agency. **Survey Methodology**, v.25, n.2, p.1-23,1999.

BRANCATO, G.; ASCARI, G. **Guidelines on the quality of multisource statistics.** Paper for the European conference on Quality in Official Statistics, Krakow, Poland. 2018.

BRANCATO, G.; BOGGIA, A; ASCARI, G.; ZHANG, L-C.; ALMÁDI, A.; DE WALL, T.; VAN DELDEN, A.; SCHOULTUS, S.; BURG, T.; KRPAVICKAITÉ, D.; O' CONNOR, E.; STOLTZE, P. **Guidelines on the quality of multisource statistics - Outline of the quality guidelines document.** Technical report WP1. Essnet on Quality of Multisource, 2017. Disponível em: https://ec.europa.eu/eurostat/cros/content/work-package-1-guidelines-quality -multisource-statistics_en. Acesso em: 28 jan. 2019.

BRANCATO, G.; DI CONSIGLIO, L. **Conceptualising quality for big data.** Paper for the European conference on Quality in Official Statistics, Krakow, Poland. 2018.

BRASIL. Ministério do Turismo. Fundação Getúlio Vargas – FGV. **Plano Nacional Estratégico de Estatísticas Turísticas.** Brasília: Ministério do Turismo, 2016.

BRASIL. Secretaria de Turismo do Estado do Espírito Santo. **Pesquisa de Fluxo Turístico por Big Data - Descritivo Metodológico.** 2017. Disponível em: <https://observatoriodoturismo.es.gov.br/descritivo-metodologico>. Acesso em: 11 jun. 2021.

BRASIL. **Plano Nacional de Turismo 2018-2022: mais emprego e renda para o Brasil.** Brasília: Ministério do turismo, 2018a.

BRASIL. **Anuário Estatístico de Turismo – 2018**. Ano base 2017. Vol. 45 - 1ª Edição. Brasília. Distrito Federal. 2018b.

BREZZI, M.; SANCHEZ-SERRA; D. **Breathing the Same Air?** Measuring Air Pollution in Cities and Regions. OECD Regional Development Working Papers No. 2014/11, 2014. DOI: 10.1787/5jxrb7rkxf21-en.

BROWN, J. S.; DUGUID, P. Organizing Knowledge, **California Management Review**, v.40, n.3, p. 90-111, 1998. DOI: 10.2307/41165945.

BROWN, S. L.; EISENHARDT, K. M. Product development - past research, present findings, and future-directions. **Academy of Management Review**, v.20, n.2, p. 343-378, 1995.

BUCKMANN, D.; CAGALA, T.; WABITSCH, A. **Completing the securities picture: Integrating official securities Statistics with regulatory trading data**. In: ISI WORLD STATISTICS CONGRESS, 62., 2019 aug 18-23; Kuala Lumpur, 2019.

BUELENS, B.; BOONSTRA, H. J.; VAN DEN BRAKEL, J.; DAAS, P. **Shifting paradigms in official statistics**: from design-based to model-based to algorithmic inference. Discussion paper 201218, Statistics Netherlands, The Hague/Heerlen. 2012.

BUENOS AIRES. Observatorio Turístico. **Sistema de Inteligencia Turística**. 2020. Página inicial. Disponível em: <https://turismo.buenosaires.gov.ar/es/observatorio>. Acesso em: 26 mar. 2020.

CAI, L.; ZHU, Y. 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.

CAJAIBA-SANTANA, G. Social innovation: Moving the field forward. A conceptual framework, **Technol. Forecast. Soc. Change**, 2014.

CAMACHOA, M.; DOMÉNECHB, R. **Nowcasting and forecasting with Dynamic Factors Models**: Some experiences and lessons. In: ISI WORLD STATISTICS CONGRESS, 62., 2019 aug 18-23; Kuala Lumpur, 2019.

CARVALHO, M. A. **Framework conceitual para ambiente virtual colaborativo das comunidades virtuais de prática nas universidades no contexto de e-gov**. 432 f. Tese (Doutorado em Engenharia e Gestão do Conhecimento) – Universidade Federal de Santa Catarina, Florianópolis, 2013.

CASTELFRANCHI, C. Six critical remarks on science and the construction of the knowledge society. **Journal of Science Communication**, v.6, n.4, p.1-3, 2007.

CAUMO, R. B. Big Data na produção de estatísticas públicas: conceitos, oportunidades e desafios. In: MARQUES, César. (org.). **População, território e**

estatísticas públicas: 20 anos do Programa de Pós-Graduação da ENCE. 1ed.Rio de Janeiro: IBGE, v. 26, p. 63-84, 2020.

CAUMO, R. B.; BIZ, A. A.; SOUZA, J. A. de. **Big Data para a geração de indicadores de turismo:** fontes de dados e aplicações, paper, December 4, 2019. Disponível em: <https://digital.library.unt.edu/ark:/67531/metadc1609205>. Acesso em: 13 jun. 2021.

CAUMO, R. B.; BIZ, A. A. ; SOUZA, J. A. de. Big Data para a geração de indicadores de turismo: Fontes de dados e aplicações. **E-tech: tecnologias para competitividade industrial**, v. 13, p. 53-70, 2020.

CAUMO, R. B.; SOUZA, J. A. de. A qualidade de indicadores socioeconômicos produzidos a partir de Big Data. **Planejamento e Políticas Públicas**, 57, 2021. DOI: <http://dx.doi.org/10.38116/ppp57art9>.

CAVALLO, A. Online and Official Price Indexes: Measuring Argentina's Inflation. **Journal of Monetary Economics**, v. 60, n.2, p.152-65, 2013.

CAVALLO, A.; RIGOBON, R. The Billion Prices Project: Using Online Data for Measurement and Research. **Journal of Economic Perspectives**, v.31, n.1, 2016.

CEBR. **Data equity, unlocking the value of big data.** *In:* SAS Reports, p. 1-44, 2012.

CENTOBELLI, P.; NDOU, V. **Managing customer knowledge through the use of big data analytics in tourism research.** *Current Issues in Tourism.* p.1-22. 2019

CERVERA, J. L.; VOTTA, P.; FAZIO, D.; SCANNAPIECO, M.; BRENNENRAEDTS, R.; VAN DER VORST, T. **Big data in official statistics.** ESS Big Data Event in Rome 2014, Technical Workshop Report. 2014.

CHEN, C.; CHEN, H. **Construction of Local Tourist Information Platform Based on Cloud Computing**, 2017.

CHEONG, F. C. **Internet Agents.** Spiders, Wanderers, Brokers and Bots (em inglês). Indianapolis: New Riders, 1996.

CHEW, C.; EYSENBACH, G. Pandemics in the age of Twitter: Content analysis of tweets during the 2009 H1N1 outbreak. **PLoS One**, v.5, n.11, e14118, 2010.

CHRISTENSEN, C. M. **The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail.** Boston, MA: Harvard Business School Press, 1997.

CISCO. **Visual Networking Index:** Global Mobile Data Traffic Forecast Update, 2017–2022 White Paper. Disponível em: <https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-738429.html>. Acesso em: 28 jan. 2019.

CITRO, C. F. From multiple modes for surveys to multiple data sources for estimates. **Survey Methodology**, 40, p.137-161, 2014.

CLASTER, W.; GHOTBI, N.; SHANMUGANATHAN, S. *In: Current Issues and Emerging Trends in Medical Tourism*. Hershey, PA: IGI Global. p.173-188. 2015

COCHRAN, W. G. **Sampling Techniques**. 3rd Edition, John Wiley and Sons, New York, 1977.

COOLEY, R., MOBASHER, B.; SRIVASTAVA, J. Web Mining: Information and pattern discovery on the world wide web. **IEEE International Conference on Tools with AI**, 1997.

COSTA, J.; FERREIRA, J. C.; DOMINGUES, L.; TAVARES, T.; DIEGUES, V.; COUTINHO, C. **Conhecer e utilizar a Web 2.0**: Um estudo com professores do 2.º, 3.º Ciclos e Secundário. X CONGRESSO INTERNACIONAL GALEGO-PORTUGUÊS DE PSICOPEDAGOGIA (p. 5614-5630). Braga: Universidade do Minho, 2009.

COUDIN, E.; POULHES, M. **Experimenting data science with the French SSP Lab**. *In: ISI WORLD STATISTICS CONGRESS*, 62., 2019 aug 18-23; Kuala Lumpur, 2019.

COUPER, M. Is the sky falling? New technology, changing media, and the future of surveys. **Surv. Res. Meth.**, 7, p.145-156, 2013.

CRESPI, G.; ARIAS-ORTIZ, E.; TACSIR, E.; VARGAS, F.; ZU-IGA, P. Innovation for economic performance: the case of Latin American firms. **Eurasian Business Review**, v. 4, n. 1, p.31-50, 2014.

CRESWELL, J. W. **Projeto de pesquisa**: métodos qualitativo, quantitativo e misto. 3. ed. Porto Alegre: Artmed, 2010.

CRISES. **An Introduction to Crises**. Montréal: Université du Québec au Montréal. 2004.

CROSBY, P. B. **Quality is Free**: The Art of Making Quality Certain. McGraw-Hill, New York. 1979.

DAAS, P. Big Data and official statistics. **Sharing Advisory Board, Software Sharing Newsletter**, 7, p.2-3. 2012.

DAAS, P.; OSSEN, S.; VIS-VISSCHERS, R.; ARENDS-TOTH, J. **Checklist for the Quality Evaluation of Administrative Data Sources**. Statistics Netherlands, 2009.

DAAS, P. J. H.; PUTS, M. J. H. Big data as a Source of Statistical Information. **The Survey Statistician**, 69, p. 22-31, 2014.

DAAS, P. J. H.; PUTS, M. J.; BUELENS, B.; VAN DEN HURK, P. A. M. **Big Data and Official Statistics**. NTTS conference, Brussels, Belgium. 2013.

DAAS, P. J. H.; PUTS, M. J.; BUELENS, B.; VAN DEN HURK, P. A. M. Big Data as a Source for Official Statistics. **Journal of Official Statistics**, v.31, n.2, p. 249-262, 2015.

DAAS, P.; TENNEKES, M.; DE JONGE, E.; PRIEM, A.; BUELENS, B.; VAN PELT, M.; VAN DEN HURK, P. **Data Science and the future of statistics Presentation for the first Data Science NL meetup**, Utrecht, Nederland. 2012.

DALKEY, N. C; HELMER, O. An experimental application of the Delphi method to the use of experts. **Management Science**, v.9, n.3, p.458-467, 1963.

DATTILO, B.; RADINI, R; SABATO, M. How many SIM cards in your luggage? A strategy to make mobile phone data usable. **Tourism statistics**, paper for the 14th Global Forum on Tourism Statistics, 2016.

DAVENPORT, T.; PATIL, D. Data scientist: the sexiest job of the 21st century. **Harvard business review**, v. 90, n.10, p. 70-77, 2012.

DAVENPORT, T. H.; PRUSAK, L. **Conhecimento empresarial: como as empresas gerenciam seu capital intelectual**. Rio de Janeiro: Campus, 1998.

DE BROE, S.; SNIJKERS, G.; SCHOUTEN, B. **Sensor data at the heart of innovation in official statistics**. *In*: ISI WORLD STATISTICS CONGRESS, 62., 2019 aug 18-23; Kuala Lumpur, 2019.

DE HAAN, J; HENDRIKS, R. **Online data, fixed effects and the construction of high-frequency price indexes**, paper presented at the 13th Economic Measurement Group, Sydney, 28 to 29 November 2013.

DE JONGE, E.; VAN PELT, M.; ROOS, M. **Time patterns, geospatial clustering and mobility statistics based on mobile phone network data**. Discussion paper 201214, Statistics Netherlands. 2012.

DE LEEUW, E. D.; DE HEER, W. Trends in Household Survey Nonresponse: A Longitudinal and International Comparison. *In*: GROVES, R. M.; DILLMAN, D. A.; ELTINGE, J. L.; LITTLE, R. J. A., (eds). **Survey Nonresponse**. New York: Wiley. p.41-45. 2002.

DE VAUS, D. **Surveys in Social Research**. 5th Edition, Routledge, London, 2002.

DE WAAL, T.; DELDEN, A. Van; SCHOLTUS, S. **Work Package 3: Framework for the quality evaluation of statistical output based on multiple sources**. 2017.

DE WAAL, T., DELDEN, A. Van; SCHOLTUS, S. Quality measures for multisource statistics. **Statistical Journal of the IAOS**, v.35, n.2, p.179-192, 2019. DOI: <https://doi.org/10.3233/SJI-180468>.

DE WAAL, T., DELDEN, A. Van; SCHOLTUS, S. Multi-source Statistics: Basic Situations and Methods. **International Statistical Review**, 88, p. 203–228, 2020. DOI: 10.1111/insr.12352.

DEMUNTER, C. **Tourism Statistics: Early Adopters Of Big Data?**, s.l.: UNWTO. Sixth UNWTO International Conference on Tourism Statistics. 2017.

DEZFOULI, M. B.; SHAHRAKI, M. H.; ZAMANI, H. **A Novel Tour Planning Model using Big Data**. *In*: INTERNATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE AND DATA PROCESSING (IDAP). p.1-6. 2018.

DONOVAN, C.; FLAHERTY, E. T.; HEALY, E. Q. Using big data from Wikipedia page views for official tourism statistics. **Statistical Journal of the IAOS**, 33, p.1-7, 2017.

DRESCH, A.; LACERDA, D. P.; ANTUNES JÚNIOR, J. A. V. **Design science research: método de pesquisa para avanço da ciência e tecnologia**. Porto Alegre: Bookman, 2015.

DRESCH, A. **Desenvolvimento científico em design science para a engenharia de produção: formulações conceituais e análise empírica**. 2018. Tese (Doutorado em Engenharia de Produção) - Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal de Santa Catarina, Florianópolis, 2018.

DRUCKER, P. **Post-Capitalist Society**. Oxford: Butterworth-Heinemann, 1993.

DRUCKER, P.F. **Inovação e espírito empreendedor (entrepreneurship): práticas e princípios**. São Paulo: Pioneira Thomson, 2002.

DUFOUR, S.; BOWLBY, G.; LAFLAMME, F.; BONHOMME, S.; MULLIN, H.; SAINT-PIERRE, E.; BARZYK, F.; ERMAN, S. **Modernizing data collection in Canada**. *In*: ISI WORLD STATISTICS CONGRESS, 62., 2019 aug 18-23; Kuala Lumpur, 2019.

DUMMEE, J.; TRONGTHAMAKIT, K. **Assessment of condominium occupancy rate in Bangkok and its vicinity from electricity meter data analytics**. *In*: ISI WORLD STATISTICS CONGRESS, 62., 2019 aug 18-23; Kuala Lumpur, 2019.

DUNNE, J. **Big Data coming soon..... to an NSI near you**. *In*: ISI WORLD STATISTICS CONGRESS, 59., 2013 aug 25-30, Hong Kong. 2013.

DZIEKANIAK, G. V. **Método para Inclusão de Conhecimento Presente em Mídias Sociais no Aprimoramento de Plataformas de Governo Eletrônico**. 412f. Tese (Doutorado em Engenharia do Conhecimento) - Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento, Universidade Federal de Santa Catarina, Florianópolis, 2012

ECB. **ECB Statistics Quality Framework (SQF)**, April, 2008. Disponível em: <http://www.ecb.int/stats/html/sqf.en.html>. Acesso em: 28 jan. 2019.

EINAV, L.; LEVIN, J. **The data revolution and economic analysis**. NBER Conference, 2013.

ESSC. **ESS Big Data Action Plan and Roadmap 1.0**. European Statistical System Committee, 26th September, 2014.

ESSNET. **Use of Administrative and Accounts Data in Business Statistics. Deliverable 6.3**: "Guidance on the accuracy of mixed-source statistics". 2013.

ESSNET. **Some Quality Aspects and Future Prospects for the Production of Official Statistics with Mobile Phone Data**. 2018. Disponível em: https://webgate.ec.europa.eu/fpfis/mwikis/essnetbigdata/index.php/WP5_Reports,_milestones_and_deliverables. Acesso em: 28 jan. 2019.

EUROSTAT. **Scheveningen Memorandum on Big Data and Official Statistics**. 2013. Disponível em: <http://ec.europa.eu/eurostat/documents/42577/43315/Scheveningen-memorandum27-09-13>. Acesso em: 28 jan. 2019.

EUROSTAT. **Accreditation procedure for statistical data from non-official sources**. 2014a. Disponível em: https://ec.europa.eu/eurostat/cros/content/accreditation-big-data-sources-input-data-official-statistics_en. Acesso em: 28 jan. 2019.

EUROSTAT. **Big data – an opportunity or a threat to official statistics?** Paper presented at the Conference of European Statisticians, 62nd plenary session, Paris. 2014b.

EUROSTAT. **Methodological manual for tourism statistics**. Luxembourg: Office for the Official Publications of the European Communities. 2014c.

EUROSTAT. **Report to the ESSC on cooperation and integration within the European Statistical System**. ESSC document 2016/28/7/EN. 2016.

EUROSTAT. **European Statistics Code of Practices**. 2017. Disponível em: <http://ec.europa.eu/eurostat/web/quality/overview>. Acesso em: 28 jan. 2019.

EUROSTAT. **The Quality Assurance Framework of the European Statistical System (ESS QAF)**. v. 2.0. 2019. Disponível em: <https://ec.europa.eu/eurostat/documents/64157/4392716/ESS-QAF-V1-2final.pdf/bbf5970c-1adf-46c8-afc3-58ce177a0646>. Acesso em: 15 jun. 2021.

FAGUNDES, P.; MACEDO, D.; FREUND, G. 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**, Campinas, SP, v. 16, n. 1, p. 194-210, 2017. DOI: 10.20396/rdbci.v16i1.8650412.

FAN, R.; DU, J.; ZHOU, Y. **Study and Implementation of Accurate Retrieval System Based on Attractions Interest Model**. 336, p.477-485, 2015.

FAO. **The FAO Statistics Quality Assurance Framework (SQAF) - Implementation Strategy and Plan (Draft)**. Rome. 2014.

FAYYAD, U. M. Data mining and knowledge discovery: Making sense out of data. **IEEE Expert: Intelligent Systems and Their Applications**, v.11, n.5, p.20-25, 1996.

FEIJÓ, C.; VALENTE, E. As estatísticas oficiais e o interesse público. **Bahia Análise & Dados**, Salvador, v. 15, n. 1, p. 43-54, 2005.

FENG, N.; LI, J.; ZHANG, G. A Study of the Structure of China's Mainstream Online Tourism Information Network Based on SNA. *In: Communications in Computer and Information Science*, 2013.

FERNANDES, R. F. **Uma proposta de modelo de aquisição do conhecimento para identificação de oportunidades de negócio nas redes sociais**, 2012. Mestrado (Engenharia e Gestão do Conhecimento) – Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento, UFSC, Florianópolis, 2012.

FIGUEROA, S.Y. R.; MAYORGA, S. L. M. **Big data for the National Agricultural Census**, Colombia 2014. Unpublished data, 2015.

FLEKOVA, L.; GUREVYCH, I. **Can We Hide in the Web? Large Scale Simultaneous Age and Gender Author Profiling in Social Media**. Paper for the evaluation lab on uncovering plagiarism, authorship, and social software misuse at Conference and Labs Evaluation Forum 2013, September 23-26, Valencia, Spain. 2013.

FLORESCU, D.; KARLBERG, M.; REIS, F.; DEL CASTILLO, P. R.; SKALIOTIS, M.; WIRTHMANN, A. **Will 'big data' transform official statistics?** 2014.

FMI. **The Data Quality Assessment Framework (DQAF)**. 2003. Disponível em: <https://www.imf.org/external/np/sta/dsbb/2003/eng/dqaf.htm#II>. Acesso em: 28 jan. 2019.

FORNARO, P.; LUOMARANTA, H. **Nowcasting finnish real economic activity: A machine learning approach**. *In: ISI WORLD STATISTICS CONGRESS*, 62., 2019 aug 18-23; Kuala Lumpur, 2019.

FACIONE, P. A. **Critical thinking**: a statement of expert consensus for purposes of educational assessment and instruction. Research findings and recommendations (Report). Newark: American Philosophical Association, 1990.

FRICKER, S.; TOURANGEAU, R. Examining the relationship between nonresponse propensity and data quality in two national household surveys. **Public Opinion Quarterly**, v.74, n.5, p. 935-955, 2010.

FROLOVA, M. **Comparative analysis of methodologies used for collecting data on formal and informal employment and tertiary education efficiency**. *In: ISI WORLD STATISTICS CONGRESS*, 62., 2019 aug 18-23; Kuala Lumpur, 2019.

FRY, B. **Visualizing Data: Exploring and Explaining Data with the Processing Environment**. Sebastopol, CA: O'Reilly Media Inc. 2008.

GARCÍA, F. C.; VALVERDE, M. I.; MASCUÑANO, J. P.; GIMENO, M. V. Quality implications of the use of big data in tourism statistics: Three exploratory examples. **Boletín de Estadística e Investigación Operativa**, 34. p.138-148, 2018.

GIBBONS, M.; TROW, M.; SCOTT, P.; SCHWARTZMAN, S.; NOWOTNY, H.; LIMOGES, C. **The new production of knowledge: The dynamics of science and research in contemporary societies**. Los Angeles, CA: Sage Publications, 1994.

GINSBERG, J.; MOHEBBI, M.; PATEL, R. et al. Detecting influenza epidemics using search engine query data. **Nature**, 457, p. 1012-1014, 2009.
<https://doi.org/10.1038/nature07634>

GRAVESTIJN, R.; JAIN, M. K.; LEE, J. **Examining customer journeys at financial institutions in Cambodia**. *In: ISI WORLD STATISTICS CONGRESS*, 62., 2019 aug 18-23; Kuala Lumpur, 2019.

GROVES, R. M. **Survey Errors and Survey Costs**, New York: Wiley-Interscience, 1989.

GROVES, R. M. Three Eras of Survey Research. **Public Opinion Quarterly**, 75, p.861-71, 2011.

GROVES, R.; LYBERG, L. Total Survey Error: Past, present and future. **Public Opinion Quarterly**, v.74, n.5, p. 849-879, 2010.

GROVES, R. M.; PEYTCHEVA, E. The impact of nonresponse rates on nonresponse bias: A metaanalysis. **Public Opinion Quarterly**, v. 72, n.2, p.167-189. 2008.

GUARDA, T.; HAZ, L.; AUGUSTO, M.; VICTOR, J. **Pervasive Smart Destinations**. 2018.

GUDIVADA, V.; APON, A.; DING, J. Data Quality Considerations for Big Data and Machine Learning: Going Beyond Data Cleaning and Transformations. **International Journal on Advances in Software**, 10, p. 1-20, 2017.

GUO, T.; GUO, B.; ZHANG, J.; YU, Z.; ZHOU, X. **CrowdTravel: Leveraging Heterogeneous Crowdsourced Data for Scenic Spot Profiling and Recommendation**. 9917, p.617-628, 2016a.

GUO, Y., BARNES, S., JIA, Q. Mining Meaning from Online Ratings and Reviews: Tourist Satisfaction Analysis Using Latent Dirichlet Allocation. **Tourism Management**, 59, p. 467-483, 2016b.

GUO, T.; GUO, B.; OUYANG, Y.; YU, Z.; LAM, J.; LI, V. CrowdTravel: scenic spot profiling by using heterogeneous crowdsourced data. **Journal of Ambient Intelligence and Humanized Computing**, 9, 2017a.

GUO, T.; GUO, B.; ZHANG, J.; YU, Z.; ZHOU, X. CrowdTravel: leveraging heterogeneous crowdsourced data for scenic spot profiling. **Journal of Zhejiang University (Engineering Science)**, v.51, n. 4, p. 663-668, 2017b

GURWITZ, P. **Combining data mines and attitude research. In Leading edge marketing research: 21st century tools and practices.** Washington, DC: Sage, 2012.

HACKL, P. Big Data: What Can Official Statistics expect? **Statistical Journal of the IAOS**, 32, p. 43-52, 2016.

HAJNOVIC, F. **Measuring the quality of commercial and big data sources for official statistics.** Paper for the European conference on Quality in Official Statistics, Krakow, Poland. 2018.

HAN, X.; LI, J.; HU, Y.; YUAN, J.; YE, S. **Tourism Information Push System Based on Convolutional Neural Network.** E3S Web of Conferences. 53, 2018.

HAND, D.J. **Official Statistics in the New Data Ecosystem.** Paper presented at the New Techniques and Technologies in Statistics conference, Brussels, March 10-12. 2015.

HAND, D. J. Statistical challenges of administrative and transaction data (with discussion). **J R Stat Soc Series A**, 181, p. 555-605. 2018.

HARWOOD, A.; MAYER, A. Big data and semantic technology: A future for data integration, exploration and visualization. **Statistical Journal of the IAOS**, 32, p. 613-626, 2016.

HASSANI, H.; SAPORTA, G.; SIRIMAL, S. E. Data Mining and Official Statistics: The Past, the Present and the Future. **Big Data**, 2, p.1-10, 2014.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The Elements of Statistical Learning: Data Mining, Inference, and Prediction.** 2nd ed. New York: Springer Science Business Media, LLC. 2009.

HAWELKA, B.; SITKO, I.; BEINAT, E.; SOBOLEVSKY, S.; KAZAKOPOULOS, P.; RATTI, C. Geo-Located Twitter as Proxy for Global Mobility Patterns. **Cartography and Geographic Information Science**, 41, 2013.

HELBING, D.; FREY, B. S.; GIGERENZER, G.; HAFEN, E.; HAGNER, M.; HOFSTETTER, Y.; VAN DEN HOVEN, J.; ZICARI, R. V.; ZWITTER, A. Behavioural Control or Digital Democracy? - **A Digital Manifesto**, 2016.

HEERSCHAP, N.; ORTEGA, S.; PRIEM, A.; OFFERMANS, M. **Innovation of tourism statistics through the use of new big data sources**. In: 12th GLOBAL FORUM ON TOURISM STATISTICS, Prague, CZ, 2014.

HEVNER, A. R. *et al.* Design Science in information systems research. **MIS Quarterly**, v. 28, n. 1, p. 75-105, 2004.

HEVNER, A. R. A Three Cycle View of Design Science Research. **Scandinavian Journal of Information Systems**, v. 19, n. 2, p. 87-92, 2007.

HILBERT, M. Big data for development: a review of promises and challenges. **Dev Policy Rev**, v.34, n.1, p.135-174, 2016.

HIQA - Health Information and Quality Authority. **Background paper to support guidance for a data quality framework for health and social care**. 2018. Disponível em: <https://www.hiqa.ie/sites/default/files/2018-10/Background-to-support-guidance-on-data-quality-framework.pdf>. Acesso em: 28 jan. 2019.

HOEKSTRA, R.; TEN BOSCH, O.; HARTEVELD, F. Automated data collection from web sources for official statistics: First experiences. **Statistical Journal of the IAOS**, 28, p. 99-111, 2012. 10.3233/SJI-2012-0750.

HOLT, D. T. The official statistics Olympics challenge: Wider, deeper, quicker, better, cheaper. **The American Statistician**, 61 (1, February), p.1-8. With commentary by G. Brackstone and J.L. Norwood. 2007.

HOWE, J. **The Rise of Crowdsourcing**, Wired, 2006.

HSIEH Y. P.; MURPHY, J. Total Twitter Error. **Decomposing Public Opinion Measurement on Twitter from a Total Survey Error Perspective**. In Total Survey Error in Practice, Wiley series in Survey Methodology, Hoboken, New Jersey. 2017.

HU, F.; LI, Z.; YANG, C.; JIANG, Y. A graph-based approach to detecting the tourist movement patterns using social media data. **Cartography and Geographic Information Science**, v.46, n.4, p. 368-382, 2018.

HUANG, J. **Age of personal credit score**. In: ISI WORLD STATISTICS CONGRESS, 62., 2019 aug 18-23; Kuala Lumpur, 2019.

HUMBELIN, O.; FARYS, R. **The impact of redistribution through taxes and deductions on the income distribution – A decomposition analysis with administrative tax data from Switzerland**. In: ISI WORLD STATISTICS CONGRESS, 62., 2019 aug 18-23; Kuala Lumpur, 2019.

HUSAIN, W.; DIH, L. Y. A Framework of a Personalized Location-based Traveler Recommendation System in Mobile Application. **International Journal of Multimedia and Ubiquitous Engineering**. 7, 2012.

IBGE. **Metodologia do Censo Demográfico 2000**. Série Relatórios Metodológicos, 25. Rio de Janeiro: IBGE, 2003.

IBGE. **Código de Boas Práticas das Estatísticas do IBGE**. Segunda edição. Rio de Janeiro: IBGE, 2021. Disponível em: <https://biblioteca.ibge.gov.br/visualizacao/livros/liv101744.pdf>. Acesso em 15 jun. 2021.

ICHIFUJI, Y.; MATSUO, Y.; KOIDE, N.; AKASHI, N.; TERAJ, Y.; KOBAYASHI, T. **A study for understanding of tourist person trip pattern based on log data of Wi-Fi access points**. *In: IEEE International Conference on Big Data (Big Data)*. p.2167-2174, 2016.

INDICADOR. **Dicionário online do Dício**, 09 jan. 2020. Disponível em <https://www.dicio.com.br/indicador/>. Acesso em: 09 jan. 2020.

ISMAIL, F.; AMIN, F. MD.; NASIR, W. M. H. M. **Use of web-scraping for the compilation of Consumer Price Index: Malaysia's experience**. *In: ISI WORLD STATISTICS CONGRESS*, 62., 2019 aug 18-23; Kuala Lumpur, 2019.

IWIG, W.; BERNING, M.; MARCK, P.; PRELL, M. **Data Quality Assessment Took for Administrative Data. 2013**. Disponível em: <https://www.bls.gov/osmr/datatool.pdf>. Acesso em: 28 jan. 2019.

JANNUZZI, P. de M. **Indicadores Sociais no Brasil: conceitos, fontes de dados e aplicações**. Campinas: Editora Alínea. 2001.

JANNUZZI, P. de M. Considerações sobre o uso, mau uso e abuso dos indicadores sociais na formulação e avaliação de políticas públicas municipais. **Revista de Administração Pública**, v. 36, n. 1, p. 51-72, 2002.

JANNUZZI, P. de M. **Indicadores sociais no Brasil: conceitos, medidas e aplicações**. 3. ed. Campinas: Alínea; Campinas: PUC, 2004.

JANNUZZI, P. de M. **Informação estatística e políticas públicas no Brasil: contribuições de pesquisas do IBGE para as políticas de desenvolvimento social e combate à fome (2004-2014)**. 2020. 107f. Monografia (Especialização em Política e Sociedade) – Instituto de Estudos Sociais e Políticos, Universidade do Estado do Rio de Janeiro, 2020.

JANNUZZI, P. de M; CARLO, S. de. **Da agenda de desenvolvimento do milênio ao desenvolvimento sustentável: oportunidades e desafios para planejamento e políticas públicas no século XXI**. Bahia Análise e Dados, vol. 28, n. 02, 2018.

JANNUZI, P. de M.; GRACIOSO, L. S. Produção e disseminação da informação estatística: agências estaduais no Brasil. **São Paulo em perspectiva**, São Paulo, v.16, n. 3, p. 92- 103, 2002.

JAVA et al. **Why we twitter**: understanding microblogging usage and communities in Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis. ACM, 2007

JAYA, I. M.; SIDI, F.; ISHAK, I.; AFFENDEY, L. S.; JABAR, M. A.: A review of data quality research in achieving high data quality within organization. **J. Theor. Appl. Inform. Technol**, v.95, n.12, p. 2647-2657. 2017.

JUG, M. **Information revolution from data to policy action in low-income countries: how can innovation help?** PARIS21, Partnership in Statistics for Development in the 21st century. Discussion Paper No. 3. 2014.

JURAN, J.; GRZYNA, F. **Quality Planning and Analysis**. 2nd ed. New York: McGrawHill, 1980.

KARLBERG, M.; SKALIOTIS, M. **Big Data for Official Statistics**: Strategies and Some Initial European Applications, WP30 presented at the Seminar on Statistical Data Collection, 2013.

KASPERIUNIENE, V.; LIMPACH, S. **Development in merchandise trade indicators**. In: ISI WORLD STATISTICS CONGRESS, 62., 2019 aug 18-23; Kuala Lumpur, 2019.

KHEE, N. S.; TAN, J.; HOONG, C. K. **Profiling the internet economy in Singapore**. In: ISI WORLD STATISTICS CONGRESS, 62., 2019 aug 18-23; Kuala Lumpur, 2019.

KIM, H. Y.; CHO, J. **Data Governance Framework for Big Data Implementation with a Case of Korea**. 2017 IEEE International Congress on Big Data (BigData Congress), p. 384-391, 2017.

KITAYAMA, D. **Extraction method for Anaba spots based on name recognition and user's evaluation**. iiWAS, 2016.

KITCHIN, R. Big data and human geography: opportunities, challenges and risks. **Dialogues in Human Geography**, v.3, n.3, p. 262-267, 2013.

KITCHIN, R. **The Data Revolution**: Big Data, Open Data, Data Infrastructures and Their Consequences. Sage, London, 2014.

KITCHIN, R. The opportunities, challenges and risks of big data for official statistics. **Statistical Journal of the IAOS**, v.31, n. 3. p. 471-481, 2015.

KLEIN, V. B. **Uma Proposta de Modelo Conceitual para Uso de Big Data E Open Data para Smart Cities**. 2015. Dissertação (Mestrado em Engenharia do

Conhecimento) - Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento, Universidade Federal de Santa Catarina, Florianópolis, 2015.

KOGUT, B.; ZANDER, U. Knowledge of the firm, combinative capabilities, and the replication of technology. **Organization Science**, 3, p. 383-97, 1992.

KOIDE, N.; ICHIFUJI, Y.; YOSHII, H.; SONEHARA, N. **Estimation of national tourism statistics based on Wi-Fi association log data**. p.2175-2179, 2016.

KRATKE, F.; BYIERS, B. **The political economy of official statistics**: Implications for the data revolution in sub-Saharan Africa. PARIS21 Discussion Papers, No. 5, Partnership in Statistics for Development in the 21st Century, Paris. 2014.

KREUTER, F.; BERG, M.; BIEMER, P.; DECKER, P.; LAMPE, C.; LANE, J.; USHER, A. AAPOR Report on Big Data. **Mathematica Policy Research**. 2015.

KROON, J. **Mobile phone and credit card data: Experience from 10 years of public private partnership**. *In*: ISI WORLD STATISTICS CONGRESS, 62., 2019 aug 18-23; Kuala Lumpur, 2019.

LACERDA, D. P.; DRESCH, A.; PROENÇA, A.; JÚNIOR, J. A. V. A. Design Science Research: método de pesquisa para a engenharia de produção. **Gest. Prod.**, São Carlos, v. 20, n. 4, p. 741-761, 2013.

LAURO, B.; TRAVERSO, R. **Data Fitness for Integration**. Paper for the European conference on Quality in Official Statistics, Krakow, Poland. 2018.

LAVALLE, S. Big data, analytics and the path from insights to value. **MIT Sloan Manage Rev**. 52, 2011.

LEAL, F.; DIAS, J. M.; MALHEIRO, B.; BURGUILLO, J. **Analysis and Visualisation of Crowd-sourced Tourism Data**. p.98-101, 2016.

LEAL, F.; GONZALEZ-VELEZ, H.; MALHEIRO, B.; BURGUILLO, J. **Semantic Profiling and Destination Recommendation based on Crowd-sourced Tourist Reviews**. p.140-147, 2018.

LESTARI, S. T. K. **Combining mobile phone data and survey data for the best result: Experience from Indonesia**. *In*: ISI WORLD STATISTICS CONGRESS, 62., 2019 aug 18-23; Kuala Lumpur, 2019.

LEYVA, G.; CORONADO, A. **The mood of the twitterers in Mexico**. *In*: ISI WORLD STATISTICS CONGRESS, 62., 2019 aug 18-23; Kuala Lumpur, 2019.

LI, J.; LIZHI, X.; TANG, L.; WANG, S.; LI, L.. Big data in tourism research: A literature review. **Tourism Management**, 68, p.301-323, 2018.

LIANG, S; SCHUCKERT, M; LAW, R.; GUO, X. A Retrospective Analysis and Future Trends of E-Tourism Research in China. **Journal of China Tourism Research**, 13, 2017.

LINDBLOM, C. E. **El proceso de elaboracion de politicas públicas**. MAP: Madrid, 1991.

LINSTONE, H. A.; TUROFF, M. **The Delphi Method: techniques and applications**. New Jersey: Listone e Turof, 2002.

LIU, P; ZHANG, H; ZHANG, J; SUN, Y.; QIU, M. Spatial-temporal response patterns of tourist flow under impulse pre-trip information search: From online to arrival, **Tourism Management**, v.73, p.105-114, 2019.

LIU Z.; JIANG B.; HEER J. imMens: real-time visual querying of big data. **Computer Graphics Forum**, v. 32, n.3, p.421-430, 2013.

LIU, Y.; SHI, J. How inter-city high-speed rail influences tourism arrivals: evidence from social media check-in data, **Current Issues in Tourism**, v.22, n.9, p.1025-1042, 2019.

LÓPEZ-GONZALES, J. L.; FUENTES, R. S.; UBAL, C.; NICOLIS, O.; TORRES, R. **Self-organizing ensemble of LSTM to enhance the air pollution estimation in Santiago of Chile**. *In*: ISI WORLD STATISTICS CONGRESS, 62., 2019 aug 18-23; Kuala Lumpur, 2019.

LOSHIN, D. **Understanding big data quality for maximum information usability**. SAS, USA. 2014. Disponível em: <http://www.dataqualitybook.com>. Acesso em: 28 jan. 2019.

LOURENÇO, G. M.; ROMERO, M. Indicadores econômicos. Economia empresarial. Curitiba: Associação Franciscana de Ensino Senhor Bom Jesus. **Fae Business School**, p. 27-41, 2002.

LUGOMER, K.; SOUNDARARAI, B.; MURCIO, R.; CHESHIRE, J.; LINGLEY, P. **Understanding sources of measurement error in Wi-Fi sensor data in the Smart City**. 2017. Disponível em: http://huckg.is/gisruk2017/GISRUK_2017_paper_95.pdf. Acesso em: 28 jan. 2019.

LUZ, N.; ALMEIDA, A.; ANACLETO, R.; SILVA, N. Collective intelligence in toursplan: An online tourism social network with planning and recommendation services. **ACM International Conference Proceeding Series**, 42-48, 2013.

MA, Y.; XIANG, Z.; DU, Q.; FAN, W. Effects of user-provided photos on hotel review helpfulness: An analytical approach with deep learning. **International Journal of Hospitality Management**, 71. 120-131, 2018.

MACFEELY, S. The continuing evolution of official statistics: some challenges and opportunities, **Journal of Official Statistics**, v.32, n.4, p.789-810, 2016.

MANSKI, C. F. **Communicating uncertainty in official economic statistics**. NBER Working Paper No. 20098. Cambridge, MA: National Bureau of Economic Research, 2014.

MANYIKA, J.; CHUI, M.; BROWN, B.; BUGHIN, J.; DOBBS, R.; ROXBURGH, C.; BYERS, A. H. **Big data**: The next frontier for innovation, competition, and productivity. McKinsey Global Institute, 2011.

MARCH, S. T.; SMITH, G. F. Design and natural science research on information technology. **Decision Support Systems**, v. 15, p. 251-266, 1995.

MARIANI, M; BAGGIO, R; FUCHS, M; HÖEPKEN, W. Business intelligence and big data in hospitality and tourism: a systematic literature review. **International Journal of Contemporary Hospitality Management**, 2018.

MARQUES, J. B. V.; FREITAS, D. Método DELPHI: caracterização e potencialidades na pesquisa em Educação. **Pro-Posições [online]**, v.29, n.2, p.389-415, 2018. ISSN 1980-6248. <http://dx.doi.org/10.1590/1980-6248-2015-0140>.

MASLANKOWSKI, J.; NOWICKA, A. **Big Data quality issues regarding multidomain statistical data combining – a survey and case studies**. Paper for the European conference on Quality in Official Statistics, Krakow, Poland, 2018.

MAYER-SCHONBERGER, V.; CUKIER, K. **Big data**: A Revolution that will transform how we live, work, and think. New York. Houghton Mifflin Harcourt, 2013.

MENCHAVEZ, J. C.; ESPINOSA, K. J. P. **Fun in the Philippines**: Automatic Identification and Sentiment Analysis of Tourism-Related Tweets. p.660-667, 2015.

MILLER, H. J. The data avalanche is here. Shouldn't we be digging? **Journal of Regional Science**, 50, p.181-201, 2010.

MILLS, S.; LUCAS, S.; IRAKLIOTIS, L.; RAPPA, M.; CARLSON, T.; PERLOWITZ, B. **Demystifying Big Data**: A Practical Guide to Transforming the Business of Government. Technical report. 2012.

MOKHTAR, N. A.; ZUBAIRI, Y. Z.; HUSSIN, A. G.; YUNUS, R. M. **Modelling Wind Direction Data in Kota Kinabalu Coastal Station Using Simultaneous Linear Functional Relationship Model**. In: ISI WORLD STATISTICS CONGRESS, 62., 2019 aug 18-23; Kuala Lumpur, 2019.

MUNIZ, E. C. L. **Gestão do Conhecimento do Cliente e Destinos Turísticos Inteligentes**: um Framework para a Gestão Inteligente da Experiência Turística – SMARTUR.2020. Tese (Doutorado em Engenharia e Gestão do Conhecimento) - Programa de Pós-Graduação em Engenharia e Gestão do Universidade Federal de Santa Catarina, Florianópolis, 2020.

MYDIN, L.; JAMALUDDIN, J. A.; ZAINI, Z. **Use of big data analytics in targeting right customers for sustainable and social finance.** *In: ISI WORLD STATISTICS CONGRESS, 62., 2019 aug 18-23; Kuala Lumpur, 2019.*

NARDO, M.; SAISANA, M.; SALTELLI, A.; TARANTOLA, S.; HOFFMAN, A., GIOVANNINI, E. **Handbook on constructing composite indicators: methodology and user guide.** OECD Statistic Working Papers, OECD, Paris, 2005.

NAS. **Frontiers in Massive Data Analysis.** Washington, DC: The National Academies Press. 2013.

NASEM. **Innovations in Federal Statistics: Combining Data Sources While Protecting Privacy.** Washington, DC: The National Academies Press. 2017a.

NASEM. **Federal Statistics, Multiple Data Sources, and Privacy Protection: Next Steps.** Washington, DC: The National Academies Press. 2017b.

NEDERPELT, P. **A New Model for Quality Managemen.** Statistics Netherlands. 2010.

NEDERPELT, P.; DAAS, P. **49 Factors that Influence the Quality of Secondary Data Sources,** Statistics Netherlands. 2012.

NELSON, R. R.; WINTER, S. G. **An evolutionary theory of economic change.** Cambridge, MA: Harvard University Press, 1982.

NEWMAN, I.; BENZ, C. R. **Qualitative-quantitative research methodology: exploring the interactive continuum.** Carbondale: Southern Illinois University Press, 1998.

NOLAN, L.; JEREMY ROWE, J.; HOPKINS, S.; WILLIAMS, S. **What can data science do for economic statistics?** *In: ISI WORLD STATISTICS CONGRESS, 62., 2019 aug 18-23; Kuala Lumpur, 2019.*

NOORAENI, R.; PURBA, N. S.; YUDHO, N. P. **Using Google trend data as an initial signal Indonesia unemployment rate.** *In: ISI WORLD STATISTICS CONGRESS, 62., 2019 aug 18-23; Kuala Lumpur, 2019.*

NOZAWA, Y.; ENDO, M.; EHARA, Y.; HIROTA, M.; YOKOYAMA, S.; ISHIKAWA, H. **Inferring Tourist Behavior and Purposes of a Twitter User.** p.101-112, 2017.

NURMI, O.; PIELA, P. **The use of mobile phone data in Tourism Statistics.** *In: ISI WORLD STATISTICS CONGRESS, 62., 2019 aug 18-23; Kuala Lumpur, 2019.*

OCDE. **Rumo a um Ambiente Sustentável: Indicadores Ambientais.** Série Cadernos de Referência Ambiental volume 9. Salvador: Organização para a Cooperação e Desenvolvimento Económico, 2002.

OCDE. Organization for Economic Co-operation and Development. **Oslo Manual: Guidelines for Collecting and Interpreting Innovation Data**. 3rd edition, The Measurement of Scientific and Technological Activities, OECD Publishing, 2005.

OCDE. **Recommendation of the OECD Council on good statistical practice**. 2015. Disponível em: <http://www.oecd.org/statistics/good-practice-toolkit/Brochure-Good-Stat-Practices.pdf>. Acesso em 28 jan. 2019.

O'CONNOR, B.; BALASUBRAMANYAN, R.; ROUTLEDGE, B. R.; SMITH, N. A. **From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series**. Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media, May 23-26, Washington DC, USA. 2010.

OLIVEIRA, L. G. **Sistema de recomendação de meios de hospedagem baseado em filtragem colaborativa e informações contextuais**. 2007. Dissertação (Mestrado em Engenharia e Gestão do Conhecimento) - Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento, Universidade Federal de Santa Catarina, Florianópolis, 2007.

OLIVEIRA, R. A. de. **Extração de dados web como suporte na elaboração de indicadores do turismo de Minas Gerais: uma iniciativa em big data**. 2017. Dissertação (Mestrado em Gestão e Organização do Conhecimento) - Programa de Pós-Graduação em Gestão e Organização do Conhecimento da Universidade Federal de Minas Gerais, Belo Horizonte, 2017.

OLIVEIRA, R. A. de; BARACHO, R. M. A. The development of tourism indicators through the use of social media data: the case of Minas Gerais, **Brazil Information Research**, v. 23, n. 4, 2018.

OLIVEIRA, T. P. S. de. **Sistemas Baseados em Conhecimento e Ferramentas Colaborativas para a Gestão Pública: Uma Proposta ao Planejamento Público Local**. 2009. Dissertação (Mestrado em Engenharia e Gestão do Conhecimento) - Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento, Universidade Federal de Santa Catarina, Florianópolis, 2009.

ONDER, I.; KOERBITZ, W.; HUBMANN-HAIDVOGEL, A. Tracing Tourists by Their Digital Footprints: The Case of Austria. **Journal of Travel Research**, 55, 2014.

ONS. **Guidelines for Measuring Statistical Output Quality**, v 4.1. 2013. Disponível em: <http://www.ons.gov.uk/ons/guide-method/method-quality/quality/guidelines-formeasuring-statistical-quality/index.html>. Acesso em: 28 jan. 2019.

OPHER, A.; CHOU, A.; ONDA, A.; SOUNDERRAJAN, K. **The Rise of the Data Economy: Driving Value through Internet of Things Data Monetization**. USA: IBM Corporation. 2016.

O'REILLY, T. **What Is Web 2.0 - Design Patterns and Business Models for the Next Generation of Software**. O'Reilly Publishing, 2005.

OSBORNE, J.; COLLINS, S.; RATCLIFFE, M.; MILLAR, R.; DUSCHL, R. What "Ideas-about-Science" should be taught in school science? A Delphi study of the expert community. **Journal of Research in science teaching**, v.40, n .7, p. 692-720, 2003.

PACHECO, R. C. S. **Instituto InCommons**: Rede Internacional de P&D em Commons Digitais. Projeto submetido ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), para participação na Chamada Pública INCT-MCTI/CNPq/CAPES/FAPs. 2014.

PACHECO, R. C. S. Coprodução em Ciência, Tecnologia e Inovação: fundamentos e visões. In: PEDRO, Joana Maria; FREIRE Patrícia de Sá. (org.). **Interdisciplinaridade** - Universidade e Inovação Social e Tecnológica. Curitiba: CRV, v. 1, p. 21-62, 2016

PARISE, S.; IYER, B.; VESSET, D. Four strategies to capture and create value from big data. **Ivey Business Journal**, v.76, n.4, n.1-5, 2012.

PARK, E.; KANG, J.; CHOI, D.; HAN, J. **Understanding customers hotel revisiting behaviour**: a sentiment analysis of online feedback reviews. *Current Issues in Tourism*. p.1-7, 2018.

PINCHOT, G.; PELLMAN R. **Intraempreendedorismo na Prática**: um guia de inovação nos negócios. Tradução Márcia Nascentes. Rio de Janeiro: Elsevier, 2004.

PINK, B; BOROWIK, J.; LEE, G. The case for an international statistical innovation program –Transforming national and international statistics systems, **Statistical Journal of the IAOS**, v.26, n.3, p.125-133, 2009.

PIZZOL, L. D. **Uso da Web de Dados como Fonte de Informação no Processo de Inteligência Competitiva Setorial**.2014. Dissertação (Mestrado em Engenharia e Gestão do Conhecimento) - Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento, Universidade Federal de Santa Catarina, Florianópolis, 2014.

PLATEK, R.; SARNDAL, C-E. Can a Statistician Deliver? **Journal of Official Statistics**, v.17, n.1, p.1-20, 2001.

PLENTZ, N. D.; BERNARDES, M. M. e S.; FRAGA, P. G. R. **Sistema de indicadores de inovação, competitividade e Design para empresas desenvolvedoras de produtos**: versão 01. Porto Alegre: Marca visual, 2015.

POL, P.; VILLE, S. Social innovation: Buzz word or enduring term. **The Journal of SocioEconomics**, v. 38, p.878-885, 2009.

POLIDORO, F.; GIANNINI, R.; CONTE, R.; MOSCA, S.; ROSSETTI, F. Web scraping techniques to collect data on consumer electronics and airfares for Italian HICP compilation. **Statistical Journal of the IAOS**, 31, p. 165-176, 2015. 10.3233/sji-150901.

PORTER, M. E. **A Vantagem Competitiva das nações**. Rio de Janeiro, Campus, 1989.

PORTER, M. E. **Estratégia Competitiva**. Rio de Janeiro: Campus, 1991.

PRAMANA, S.; YUNIARTO, B.; KURNIAWAN, R.; YORDANI, R.; LEE, J.; AMIN, I.; SATYANING, Ni Luh Putu P.; RIYADI, Y.; HASYYATI, Nashirah A.; INDRIANI, R. **Big data for government policy**: Potential implementations of bigdata for official statistics in Indonesia. p.17-21, 2017.

PROVENZANO, D.; HAWELKA, B.; BAGGIO, R. The mobility network of European tourists: a longitudinal study and a comparison with geo-located Twitter data, **Tourism Review**, v. 73, n. 1, p. 28-43, 2018.

PURAO, S. *et al.* The Sciences of Design: Observations on an Emerging Field. **Journal of the Association for Information Systems**, v. 23, November 2008, p. 523-546, 2008.

QUEVEDO, M. **Gestão do Conhecimento em Portais Virtuais de Turismo**: Uma Abordagem Empreendedora. 2009. Dissertação (Mestrado em Engenharia e Gestão do Conhecimento) - Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento, Universidade Federal de Santa Catarina, Florianópolis, 2009.

RAI, A. **Evaluating the image of healthcare hubs online** – A cross sectional analysis of google and face book reviews of private hospitals of Kolkata, eastern India. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. XLII-5. p. 893-899, 2018.

REIMSBACH-KOUNATZE, C. **The Proliferation of “Big Data” and Implications for Official Statistics and Statistical Agencies**: A Preliminary Analysis. OECD Digital Economy Papers, No. 245, OECD Publishing. 2015.

REINERT, R.; STOLTZE, P. T. **Checklist for Evaluating the Quality of Input Data**. 2016. Disponível em: https://ec.europa.eu/eurostat/cros/content/work-package-1-checklists-evaluating-quality-input-data_en. Acesso em: 28 jan. 2019.

REIS, F.; DI CONSIGLIO, L.; KOVACHEV, B.; WIRTHMANN, A.; SKALIOTIS, M. **Comparative assessment of three quality frameworks for statistics derived from big data**: the cases of Wikipedia page views and Automatic Identification Systems. Paper for the European Conference on Quality in Official Statistics, Madrid 1-3 June, 2016.

RIBEIRO, A. C. **Modelo de reconhecimento de padrões em ideias usando técnicas de descoberta de conhecimento em textos**. 2018. Dissertação (Mestrado em Engenharia e Gestão do Conhecimento) - Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento, Universidade Federal de Santa Catarina, Florianópolis, 2018.

ROCA SALVATELLA. **Big Data and Tourism: New Indicators for Tourism Management**. Barcelona: Telefónica I+D and RocaSalvatella. 2014.

ROCHA, C. V. **Neoinstitucionalismo como modelo de análise para as políticas públicas**. Civitas, Porto Alegre, 5(1):11-28, 2005.

RONG, F. **Design of Tourism Resources Management Based on Artificial Intelligence**. p.436-439, 2016.

ROSSUM, G. V. **Foreword for "Programming Python" (1st ed.)** » (em inglês). Python Software Foundation, 1996.

ROTELLA, P. **Is data the new oil?** Forbes. 2012.

RUSLAN, M.; TETIANA, M.; KATERYNA, P. **Prospects for the development of higher education based on Big Data technology**. *In: ISI WORLD STATISTICS CONGRESS*, 62., 2019 aug 18-23; Kuala Lumpur, 2019.

SAEBO, H. V. **Quality in statistics – from Q2001 to 2016**. Paper for the European Conference on Quality in Official Statistics, Madrid 1-3 June, 2016.

SÁFADI, C. M. Q. Delphi: um estudo sobre sua aceitação. *In: V SEMEAD*, 2001. São Paulo: **Anais...** São Paulo: FEA-USP, 2001.

SAINT-PIERRE, E. **The integration of the Census of Agriculture with the Business Statistics Program: The keystone for the next generation of Censuses in Canada**. *In: ISI WORLD STATISTICS CONGRESS*, 62., 2019 aug 18-23; Kuala Lumpur, 2019.

SAISANA, M.; TARANTOLA, S. **State-of-the-art report on current methodologies and practices for composite indicator development**. Report EUR 20408 EN. European Commission–Joint Research Centre, Ispra, 2002.

SALEMINK, I.; VAN DER VEN, H.; COENEN, B.; KONEN, R.; LAMMERS, J. **Business registers as the backbone of statistics; Challenges and opportunities regarding business register and outcome quality**. *In: ISI WORLD STATISTICS CONGRESS*, 62., 2019 aug 18-23; Kuala Lumpur, 2019.

SALGADO, D.; ESTEBAN, E.; SALDANA, S.; OANCEA, B.; SAKAROVITCH, B.; DEBUSSCHERE, M.; ALEXANDRU, C.; NURMI, O.; PIELA, P.; RADINI, R.; SANGUIAO-SANDE, L.; TENNEKES, M.; WILLIAMS, S.; ZWICK, M. **Estimation of population counts combining official data and aggregated mobile phone data**. Paper for the European conference on Quality in Official Statistics, Krakow, Poland. 2018.

SALTELLI, A. Composite indicators between analysis and advocacy, **Social Indicators Research**, 81, p.65-77, 2007.

SAMANTA, G. P. **Does Google search index help track and predict inflation rate? An exploratory analysis for India.** *In: ISI WORLD STATISTICS CONGRESS*, 62., 2019 aug 18-23; Kuala Lumpur, 2019.

SAMMAN, E. **10 things to know about development and the data revolution.** ODI, 2015.

SANTOS, A. R. dos. **Gestão do conhecimento: uma experiência para o sucesso empresarial.** Universitária Champagnat, 2001.

SANTOS, G. N. C.; INÁCIO, J.B. Observatório do Turismo e Big Data: a importância da informação e da tecnologia no desenvolvimento de destinos turísticos e sustentáveis. **Caminhos da Geografia (UFU. Online)**, 19, p.286-299, 2018.

SAUNDERS, M.; LEWIS, P.; THORNHILL, A. **Research Methods for Business Students.** 6. ed. London: Pearson Education Limited, 2012.

SCANNAPIECO, M.; VIRGILLITO, A.; ZARDETTO, D. **Placing Big Data in Official Statistics: A Big Challenge?** Paper presented at New Techniques and Technologies in Statistics. 2013.

SHEHABUDEEN, N.; PROBERT, D.; PHAAL, R.; PLATTS, K. **Representing and approaching complex management issues: part 1 - role and definition (Working Paper)**, Institute for Manufacturing, University of Cambridge, UK. CTM2000/03, ISBN: 1-902546-21-0, 2000.

SCHENK, E.; GUITTARD, C.: **Crowdsourcing: What can be Outsourced to the Crowd, and Why?** 2009.

SCHMIDT, E.; COHEN, J. **A nova era digital: como será o futuro das pessoas, das nações e dos negócios.** Rio de Janeiro: Intrínseca, 2013.

SCHMITT, M. T. B. **Inteligência Competitiva na Web: Um Framework Conceitual para Aquisição de Ativos de Conhecimento no Contexto do Front-End da Inovação.** 2013. Dissertação (Mestrado em Engenharia e Gestão do Conhecimento) - Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento, Universidade Federal de Santa Catarina, Florianópolis, 2013.

SCHNORR-BACKER, S. Statistical monitoring systems to inform policy decision-making, and new data sources. **Statistical Journal of the IAOS**, 33, p.1-12, 2016.

SCHUTT, R.; O'NEIL, C. **Doing Data Science: Straight Talk from the Frontline.** Sebastopol, CA: O'Reilly Media, 2013.

SCHREIBER, G.; AKKERMANS, H.; ANJEWIERDEN, A.; DE HOOG, R.; SHADBOLT, N.; VAN DE VELDE, W.; WIELINGA, B. **Knowledge engineering and management: the CommonKADS methodology.** MIT press, 2002.

- SCHUMPETER, J. A. **A teoria do desenvolvimento econômico**. São Paulo: Nova Cultural, 1988.
- SCHWAB, K. **A quarta revolução industrial**. Tradução de Daniel Moreira Miranda. São Paulo: Edipro, 2016.
- SCHWARTZMAN, S. **Legitimidade, controvérsias e traduções em estatísticas públicas**. Teoria e Sociedade, Belo Horizonte, 1997.
- SEID, Y. M. **Cost-effectiveness of remote sensing for Agricultural Statistics in developing and transition countries**. *In*: ISI WORLD STATISTICS CONGRESS, 62., 2019 aug 18-23; Kuala Lumpur, 2019.
- SENRA, N. Informação estatística: política, regulação, coordenação. **IBICT, Ciência da Informação**, v. 28, n. 2, 1999.
- SENRA, N. C. **Estatísticas Formalizadas (C.1972–2002)**. História das estatísticas brasileiras. Rio de Janeiro: IBGE, vol. 4. 2009.
- SETIADI, Y.; ULUWIYAH, A. **Improving data quality through big data**: Case study on big data-mobile positioning data in Indonesia tourism statistics. p.43-48, 2017.
- SHARNOBY, A. E. **Use of GIS and statistics in measuring land consumption rate to population growth**. *In*: ISI WORLD STATISTICS CONGRESS, 62., 2019 aug 18-23; Kuala Lumpur, 2019.
- SHEN, J.; SHEN, J.; MEI, T.; GAO, X. Landmark Reranking for Smart Travel Guide Systems by Combining and Analyzing Diverse Media. **IEEE Transactions on Systems, Man, and Cybernetics: Systems**, 46. p.1-13, 2016.
- SHI, B.; ZHAO, J.; CHEN, PO-JU. Exploring urban tourism crowding in Shanghai via crowdsourcing geospatial data, **Current Issues in Tourism**, v.20, n.11, p.1186-1209, 2017.
- SHIMADA, K.; INOUE, S; MAEDA, H.; ENDO, T. **Analyzing Tourism Information on Twitter for a Local City**, 2011.
- SHOVAL, N.; AHAS, R. The use of tracking technologies in tourism research: the first decade. **Tourism Geographies**, 18, p.1-20, 2016.
- SILVA, E. L. da; MENEZES, E. M. **Metodologia da pesquisa e elaboração de dissertação**. 4. ed. Florianópolis: UFSC, 138 p. 2015.
- SILVA, P. L. N. **Statistical thinking & methodology**: pillars for quality in the big data era. Presentation for the Eighth IFC Conference on “Statistical implications of the new financial landscape”. 2016.

SILVA, P. L. do N.; MIRANDA, V. G. **Challenges and opportunities for cost of living measurement in the Big Data era - Perspectives from Brazil.** *In: ISI WORLD STATISTICS CONGRESS*, 62., 2019 aug 18-23; Kuala Lumpur, 2019.

SIMON, H. A. **The sciences of the artificial.** Cambridge, MA: MIT Press, 1969.

SNZ. **Statistics New Zealand, Using cellphone data to measure population movements experimental analysis** following the 22 february 2011 hristchurchearth quake New Zealand government. ISBN 978-0-478-37759-0, 2012

SPERONI, R. de M. **Modelo de Referência Para Indicadores de Inovação Regional Suportado por Dados Ligados.** 2016. Tese (Doutorado em Engenharia e Gestão do Conhecimento) - Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento, Universidade Federal de Santa Catarina, Florianópolis, 2016.

STATISTICS CANADA. **Statistics Canada Quality Guidelines** (6th ed.). Ottawa: Statistics Canada, 2019. ISSN: 1708-6256.

STRUIJS, P.; BRAAKSMA, B.; DAAS, P. Official statistics and Big Data. **Big Data and Society**, p.1-6, april-june, 2014.

STRUIJS, P.; DAAS, P. J. H. Big Data, **Big Impact?** Paper and presentation for the Seminar on Statistical Data Collection, Geneva, Switzerland, 2013.

STRUIJS, P.; DAAS, P. J. H. **Quality Approaches to Big Data in Official Statistics. Paper for the European conference on Quality in Official Statistics**, Vienna, Austria, 2014.

STRUIJS, P.; DE BROE, S. **Big Data Strategies for Official Statistics.** Paper for the European conference on Quality in Official Statistics, Krakow, Poland, 2018.

STUBBINGS, P.; PESKETT, J. **Mapping the urban forest at street level**, ONS Data Science Campus, Newport, UK, 2018.

SUMMA, D. **Experiment report: Web Scraping.** 2014.

TAM, S-M.; CLARKE, F. **Big Data, Official Statistics and Some Initiatives by the Australian Bureau of Statistics.** Paper presented at International Conference on Big Data for Official Statistics, Beijing, China, 2014.

TAM, S-M.; CLARKE, F. Big data, official statistics and some initiatives by the Australian Bureau of Statistics. **International Statistical Review**, 83, p.436-448, 2015a.

TAM, S-M.; CLARKE, F. **Big Data, Statistical Inference and Official Statistics.** Research Paper No. 1351.0.55.054. Canberra: Australian Bureau of Statistics, 2015b.

TAM, S-M.; KIM, J-K. Big Data ethics and selection-bias: An official statistician's perspective. **Statistical Journal of the IAOS**, 34, p.1-12, 2018.

TAO, C.; GAO, J. **Quality Assurance for Big Data Application – Issues, Challenges, and Needs**. In: Proc of the Twenty-Eighth International Conference on Software Engineering and Knowledge Engineering, Redwood City, San Francisco Bay, California, 2016.

TEIXEIRA, F. P. **Gestão do conhecimento aplicada à gestão sustentável do turismo em parques nacionais**. 2010. Tese (Doutorado em Engenharia e Gestão do Conhecimento) - Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento, Universidade Federal de Santa Catarina, Florianópolis, 2010.

TENNEKES, M.; DE JONGE, E.; DAAS, P. J. H. **Visual Profiling of Large Statistical Datasets**. Statistics Netherlands, 2011.

TENNEKES, M.; OFFERMANS, M. P. W. **Daytime population estimations based on mobile phone metadata**. Paper prepared for the Joint Statistical Meetings, Boston, 2014.

THELWALL, M. **Sentiment Analysis for Tourism: Managerial Approaches, Techniques, and Applications**, 2019.

TIDD, J; BESSANT, J.; PAVITT, K. **Gestão da inovação**. Tradução Elizamari Rodrigues Becker et al. 3. ed. Porto Alegre: Bookman, 2008.

TIRU, M. **Overview of the Sources and Challenges of Mobile Positioning Data for Statistics**, presented in the International Conference on Big Data for Official Statistics. Beijing: UN Trade Statistics, Positium, and University of Tartu, 2014.

THOMPSON, M. Dynamic data science and official statistics. **Canadian Journal of Statistics**, 46, 2017.

TUFECKI, Z. **Big Data: Pitfalls, Methods and Concepts for an Emergent Field**. March 7. 2013.

UKSA. **Code of Practice for Statistics**. Edition 2.0. Ensuring official statistics serve the public. 2018. Disponível em: <https://www.statisticsauthority.gov.uk/wp-content/uploads/2017/07/DRAFT-Code-2.pdf>. Acesso em: 28 jan. 2019.

ULUWIYAH, A. **Trusted big data for official statistics: Study case: Statistics Indonesia (BPS)**. p.1-6. 2016.

UN. United Nations. **International Recommendations for Tourism Statistics 2008**. 2010. ISBN: 978-92-1-161521-0. Disponível em: <http://unstats.un.org/unsd/tradeserv/tourism/manual.html>. Acesso em: 13 jun. 2021.

UN. United Nations. **Big data and modernization of statistical systems**, March 2014. E/CN.3/2014/11, 2014a.

UN. United Nations. **A World That Counts: Mobilising the Data Revolution for Sustainable Development.** A report by the Secretary General's Independent Expert Advisory Group on a Data Revolution. New York: UN. 2014b.

UN. United Nations. **Earth Observations for Official Statistics: Satellite Imagery and Geospatial Data Task.** 2017.

UN. United Nations. UN Statistical Quality Assurance Framework. **Committee of the Chief Statisticians of the United Nations System.** 2018. Disponível em: <https://unstats.un.org/unsd/unsystem/documents/UNSQAF-2018.pdf>. Acesso em 28 jan. 2019.

UN. United Nations. **United Nations National Quality Assurance Frameworks Manual for Official Statistics.** 2019. Disponível em: <https://unstats.un.org/unsd/methodology/dataquality/references/1902216-UNNQAManual-WEB.pdf>. Acesso em: 15 jun. 2021.

UNECE. United Nations Economic Commission for Europe. **Fundamental Principles of Official Statistics in the UNECE region.** Geneva, 1992.

UNECE. United Nations Economic Commission for Europe. **Generic Statistical Business Process Model (GSBPM).** 2013a. Disponível em: <https://statswiki.unece.org/display/GSBPM/GSBPM+v5.0>. Acesso em: 28 jan. 2019.

UNECE. United Nations Economic Commission for Europe. **Classification of Types of Big Data.** 2013b. Disponível em: <https://statswiki.unece.org/display/bigdata/Classification+of+Types+of+Big+Data>. Acesso em: 13 jun. 2021.

UNECE. United Nations Economic Commission for Europe. **A Suggested Framework for National Statistical Offices for assessing the Quality of Big Data.** 2014. Disponível em: <https://statswiki.unece.org/download/attachments/108102944/Big%20Data%20Quality%20Framework%20-%20final-%20Jan08-2015.pdf?version=1&modificationDate=1420725063663&api=v2>. Acesso em: 13 jun. 2021.

UNESCO. United Nations Educational, Scientific and Cultural Organization. **Toward knowledge societies.** UNESCO World Report. Conde-sur-Noireau, France: Imprimerie Corlet, 2005.

UNGP. United Nations Global Pulse. **Big data for development: Challenges and opportunities.** May 2012. UN Global Pulse, New York. 2012.

UNWTO. World Tourism Organisation. **Sustainable tourism indicators and destination management: regional workshop.** Kolasin, Montenegro, 25-27 April 2007. Madrid: World Tourism Organisation, 2017.

UNWTO. World Tourism Organisation. **International Recommendations for Tourism Statistics 2008**. Department of Economic and Social Affairs, v.83, n.1, 2008a.

UNWTO. World Tourism Organisation. **Tourism Satellite Account: Recommended Methodological Framework 2008**. New York: United Nations, Department of Economic and Social Affairs - Statistics Division, 2008b.

UNWTO. World Tourism Organization. **Methodological Notes to the Tourism Statistics Database**, 2020 Edition, UNWTO, Madrid, 2019.

URGAL, B.; QUINTÁS, M. A.; ARÉVALO-TOMÉ, R. Knowledge resources and innovation performance: the mediation of innovation capability moderated by management commitment. **Technology Analysis & Strategic Management**, 25, p.543-65, 2013.

VALDIVIA, A.; LÓPEZ ALCALDE, J.; VICENTE, M.; PICHUUE, M.; RUIZ, M.; ORDOBÁS, María. **Monitoring influenza activity in Europe with Google Flu Trends**: Comparison with the findings of sentinel physician networks - results for 2009-10. Euro surveillance: bulletin européen sur les maladies transmissibles = European communicable disease bulletin. 15, 2010. 10.2807/ese.15.29.19621-en.

VACCARI, C. **Big Data in Official Statistics**. Thesis University of Camerino, Italy. 2014.

VAISHNAVI, V.; KUECHLER, W. **Design Research in Information Systems**. 2009. Disponível em: <http://desrist.org/design-research-in-information-systems>. Acesso em: 16 jun. 2021.

VAJU, S. C.; MESZAROS, M. T. **Administrative data and quality – Guidelines towards better quality of administrative data**. Paper for the European conference on Quality in Official Statistics, Krakow, Poland, 2018.

VALE, S. International collaboration to understand the relevance of Big Data for official statistics. **Statistical Journal of the IAOS**, 31, p.159-163, 2015.

VALENTE, N. T. Z.; FUJINO, A. Atributos e dimensões de qualidade da informação nas Ciências Contábeis e na Ciência da Informação: um estudo comparativo. **Perspectivas em Ciência da Informação**, v. 21, n. 2, p.141-167, 2016.

VAN AKEN, J. E. Management Research Based on the Paradigm of the Design Sciences: The Quest for Field-Tested and Grounded Technological Rules. **Journal of Management Studies**, v. 41, n. 2, p. 219-246, 2004.

VAN DONKELAAR, A.; MARTIN, R. V.; BRAUER, M.; HSU, N. C.; KAHN, R. A.; LEVY, R. C.; LYAPUSTIN, A.; SAYER, A. M.; WINKER, David M. **Environmental Science & Technology**, v.50, n.7, p. 3762-3772, 2016. DOI: 10.1021/acs.est.5b05833

VAN RUTH, F.J., **Traffic intensity as indicator of regional economic activity**, Discussion paper 2014/21, Statistics Netherlands, 2014.

VENZIN, M.; KROGH, G. von; ROOS, J. Future Research into Knowledge Management. KROGH, George von; ROOS, Johan; DIRK, Kleine (Eds.), **Knowing in Firms-Understanding, Managing and Measuring Knowledge**. London: SAGE Publications, 1998.

VIRGILLITO, A., **Experiment report**: Job Vacancies. 2014. Disponível em: <http://www1.unece.org/stat/platform/display/bigdata/Experiment+report%3A++Job+Vacancies>. Acesso em: 11 jun. 2021.

WALLGREN, A.; WALLGREN, B. **Register-based Statistics**: statistical methods for administrative data, 2nd Ed. Chichester: Wiley. 2014.

WANG, D. J.; SHI, X.; MCFARLAND D. A.; LESKOVEC J. **Measurement error in network data**: A re-classification. Social networks. 2012.

WANG, R. Y.; STRONG, D. M. Beyond Accuracy: What Data Quality Means to Data Consumers. **Journal of Management Information System**, v.12, n.4, p.5-34, 1996.

WB. **Open Data for Economic Growth**. Transport & ICT Global Practice. 2014

WEF. **Report about Personal Data**: The Emergence of a New Asset Class, World Economic Forum, 2011.

WELLMAN, J. L. **Organizational learning**: How companies and institutions manage and apply knowledge. 2009. 10.1057/9780230621541.

WIRTHMANN, A.; REIS, F. **Ethical implications of using Big Data for Official Statistics**. Paper for the European conference on Quality in Official Statistics, Krakow, Poland, 2018.

WONG, D. **Data is the next frontier, analytics the new tool**: Five trends in big data and analytics, and their implications for innovation and organizations. Big Innovation Centre, London, 2012.

WOSZEZENKI, C. R. **Modelo para descoberta de conhecimento baseado em associação semântica e temporal entre elementos textuais**. 2016. Tese (Doutorado em Engenharia e Gestão do Conhecimento) - Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento, Universidade Federal de Santa Catarina, Florianópolis, 2016.

WRIGHT, J. T. C.; GIOVINAZZO, R. A. DELPHI - uma ferramenta de apoio ao planejamento prospectivo. **Caderno de Pesquisas em Administração**, São Paulo, v. 1, n. abr./jul 2000, p. 54-65, 2000.

XIONG, H.; ZHOU, Y.; HU, C.; WEI, X.; LI, L. **A novel recommendation algorithm frame for tourist Spots based on multi-clustering bipartite graphs**. p. 276-282, 2017.

YANG, K.; WAN, T.; XIA, T.; HE, X. **Urban tourism research based on the social media check-in data**, 4th International Conference on Smart and Sustainable City (ICSSC 2017), Shanghai, p. 1-3, 2017.

YIFAN, Y.; JUNPING, D.; DAN, F.; LEE, J. **Design and implementation of tourism activity recognition and discovery system**, 12th World Congress on Intelligent Control and Automation (WCICA), Guilin, p. 781-786, 2016.

YOUNG, A.; VERHULST, S. **Jamaica's interactive community mapping: Open data and crowdsourcing for tourism**, 2017.

ZIKOPOULOS *et al.* **Harness the Power of Big Data The IBM Big Data Platform. McGraw Hill Professional**, 2012. ISBN: 9780071808170.

ZHANG, Z.; PAN, H.; XU, G. Research on personalized tourism attractions recommendation model based on user social influence. **Advances in Modelling and Analysis B**. 59, p. 62-75, 2016.

ZHANG, Z.; PAN, H.; XU, G.; WANG, Y.; ZHANG, P. A Context-Awareness Personalized Tourist Attraction Recommendation Algorithm. **Cybernetics and Information Technologies**, 16, 2016.

ZHU, H.; MADNICK, S.; LEE, Y.; WANG, R. **Data and Information Quality Research: Its Evolution and Future**, Working Paper, MIT, USA, 2012.

APÊNDICE A – Carta convite enviada aos especialistas participantes

Prezado Dr. _____,

Faço esse contato em virtude da sua expertise profissional e acadêmica, por indicação da Profa. Dra. Denise Britz do Nascimento Silva (Escola Nacional de Ciências Estatísticas do Instituto Brasileiro de Geografia e Estatística – ENCE/IBGE), com intuito de convidá-lo para ser um dos especialistas a contribuir com a verificação de uma ferramenta que estou propondo em uma tese de doutorado.

Me chamo Rafael Bassegio Caumo (lattes), doutorando do programa de Engenharia e Gestão do Conhecimento (EGC) da Universidade Federal de Santa Catarina – Brasil (UFSC), com uma tese intitulada “Indicadores socioeconômicos produzidos a partir de Big Data: Um framework para avaliação da qualidade estatística aplicado ao turismo”, orientada pelo Prof. Dr. João Artur de Souza (lattes), com coorientação da Profa. Dra. Denise Britz do Nascimento Silva (lattes) e do Prof. Dr. Alexandre Augusto Biz (lattes).

A saber, a proposta já foi aprovada na etapa de qualificação, com banca formada pelos professores Dr. Paulo de Martino Jannuzzi (ENCE/IBGE), Dr. Alexandre Leopoldo Gonçalves (UFSC) e José Leomar Todesco (UFSC).

O tema do trabalho, como o próprio título sugere, é o da produção de estatísticas públicas a partir de fontes alternativas de dados (Big Data), diferentes das tradicionais pesquisas de campo (censos e levantamentos amostrais).

Nesse contexto, percebidas na literatura algumas lacunas referentes a aferição da qualidade dessas estatísticas públicas geradas a partir de Big Data, meu objetivo é o de propor uma ferramenta, aqui chamada de framework, que possa orientar a avaliação da qualidade de tais estatísticas – utilizando a Design Science Research (DSR) como método de pesquisa.

Para tanto, tenho como ponto de partida estruturas já conhecidas para tal finalidade, como o UNECE Big Data Quality Framework (UNECE, 2014), o Eurostat Accreditation Procedure for statistical data from Non-Official Sources (EUROSTAT, 2014a), o AAPOR Big Data Total Error Framework (KREUTER et al., 2015), o Conceptual Framework for Quality in Big Data (BATINI et al., 2015) e o Big Data Quality Assessment Framework (CAI; ZHU, 2015). Em adição, realizei um extenso trabalho de revisão sistemática e integrativa de literatura para que conseguisse então construir a proposta trazida na tese.

Essencialmente, trata-se de um protocolo norteador das questões que devem ser consideradas e observadas quando da avaliação da qualidade de uma estatística pública construída a partir de fontes alternativas de dados (Big Data).

Dito isso, gostaria de convidá-lo a ser um dos especialistas que participarão do processo de verificação do instrumento que estou propondo, contribuindo com a etapa de “Avaliação do Artefato” da pesquisa. Sua contribuição será essencialmente fornecer sua opinião sobre a proposta, direcionando modificações e melhorias. Para auxiliar neste processo, apresentarei, em conjunto com o framework, uma simulação de sua aplicação prática em um contexto real.

Esse processo será conduzido seguindo o protocolo da técnica Delphi. Ou seja, um grupo previamente formado de especialistas será consultado e fornecerá sua opinião de forma anônima através do preenchimento de questionários, à distância (online). Na sequência, de forma interativa e em rodadas, as opiniões serão sintetizadas e compartilhadas novamente (feedbacks) com os especialistas até que seja alcançado um consenso sobre a melhor estrutura para o artefato que está sendo proposto. Espera-se que consigamos iniciar este processo até o final deste mês de janeiro, concluindo-o até o final de março/2021.

Seria um prazer e uma honra poder contar com sua contribuição neste processo. Assim, caso haja interesse e disponibilidade de sua parte, solicito manifestação para poder trazer maiores informações sobre os próximos passos.

Desde já, muitíssimo obrigado pela atenção.

Atenciosamente,

APÊNDICE C – Primeiro questionário aplicado aos especialistas

Rodada Delphi 1
Prezado especialista, esta primeira rodada visa fomentar uma discussão mais ampla e aberta sobre a proposta de framework. As respostas serão compiladas, compartilhadas (anonimamente) com os demais especialistas e então utilizadas para nortear ajustes no framework. Caso necessário, uma nova rodada será conduzida.

Vamos começar. Pra fins de identificação, qual seu e-mail?*

Parte 1: Estrutura Geral

1. Estrutura Geral
O framework proposto entende a qualidade estatística como um conceito multidimensional, sendo aqui considerada como o resultado da composição das duas dimensões de qualidade, operacionalizadas na prática por conjuntos de requisitos, cada um pertencente a um domínio.
Ou seja, para analisar a qualidade estatística geral e por dimensões, o framework propõe a investigação de um produto estatístico em termos do quanto atendem a requisitos de qualidade no âmbito de diferentes domínios.

Indicadores Socioeconômicos produzidos a partir de Big Data
Insuportos Produção Produto

1.a) Qual seu nível de concordância com a proposta de estruturação geral do framework?*

LEGENDA:
Um - discordo totalmente
Dois - discordo parcialmente
Tres - indiferente/neutral
Quatro - concordo parcialmente
Cinco - concordo totalmente

1.b) Qual sua opinião sobre a proposta de estruturação geral do framework?*
Considere-se: (i) Concorda com a utilização de dimensões, domínios e requisitos e suas definições? (ii) Visualiza algum problema, risco, entrave ou aspecto negativo? (iii) Sugere alguma alteração? (iv) Possui algum outro comentário a fazer?

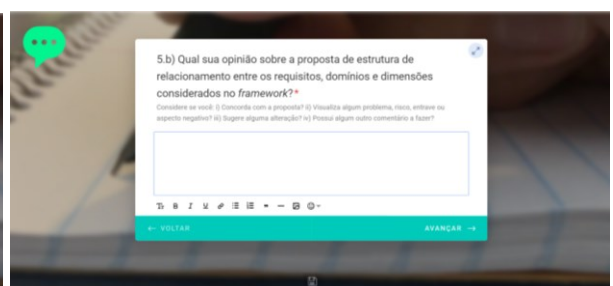
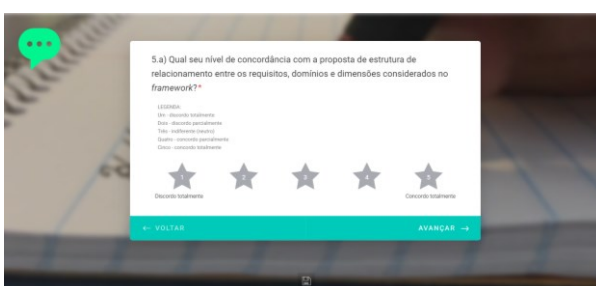
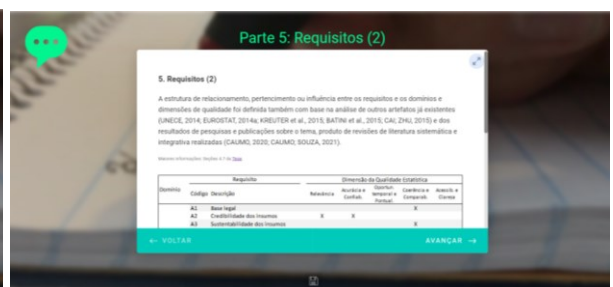
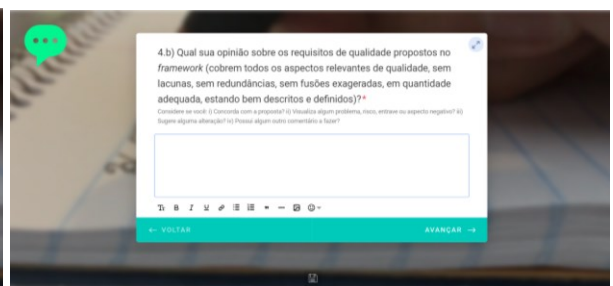
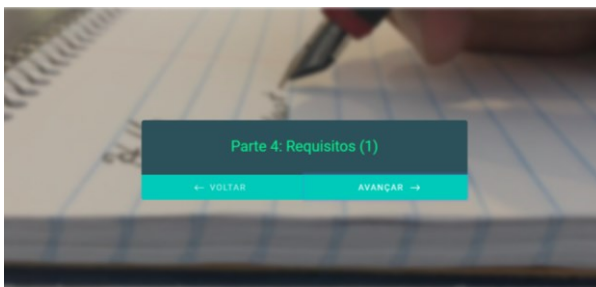
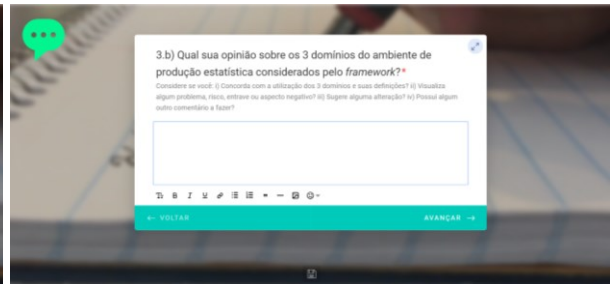
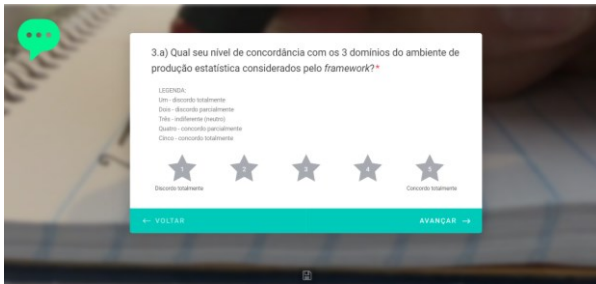
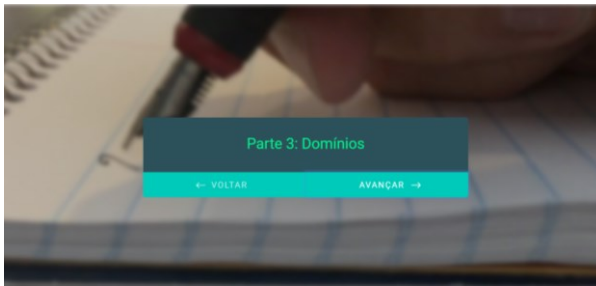
Parte 2: Dimensões

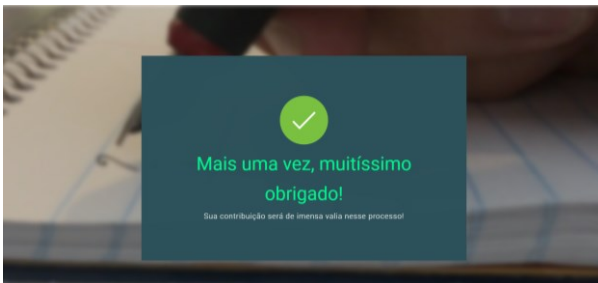
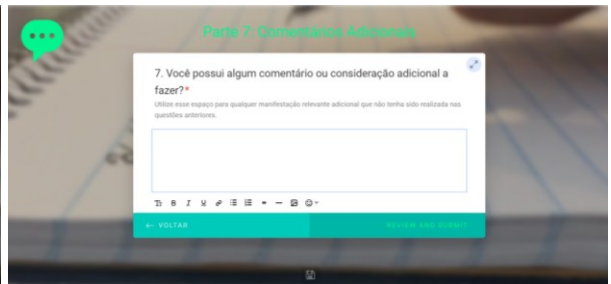
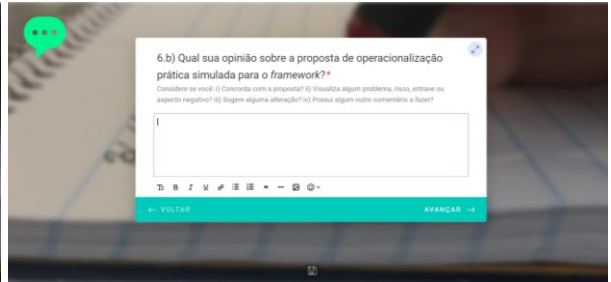
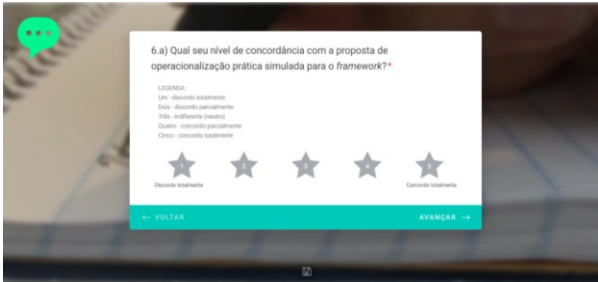
2. Dimensões
O framework proposto considera que a qualidade de um produto estatístico é composta por 5 (cinco) dimensões, as mesmas utilizadas por Eurostat (2017) e IN (2019): Relevância, Acurácia e Confiabilidade, Oportunismo temporal e Personalidade, Coerência e Comparabilidade, Acessibilidade e Clareza.

2.a) Qual seu nível de concordância com as 5 dimensões de qualidade consideradas pelo framework?*

LEGENDA:
Um - discordo totalmente
Dois - discordo parcialmente
Tres - indiferente/neutral
Quatro - concordo parcialmente
Cinco - concordo totalmente

2.b) Qual sua opinião sobre as 5 dimensões de qualidade consideradas pelo framework?*
Considere-se: (i) Concorda com a utilização das 5 dimensões e suas definições? (ii) Visualiza algum problema, risco, entrave ou aspecto negativo? (iii) Sugere alguma alteração? (iv) Possui algum outro comentário a fazer?





QUADRO: DESCRIÇÃO DOS REQUISITOS DE QUALIDADE DO DOMÍNIO "PRODUÇÃO"

Código Referência	Definição (resumida)
B1	Credibilidade do produtor Indicador de turismo seja produzido e disponibilizado por uma organização especializada nacional ou internacionalmente, ou no ramo, com autoridade e boa reputação, transparente, íntegra e independente, com produtos assinados por profissionais especializados reconhecidos.
B2	Infraestrutura e recursos humanos Organização produtora do indicador de turismo disponha de recursos humanos e não humanos para propor, conduzir e manter os processos conforme o esperado. Em relação aos recursos humanos, se faz necessário que haja profissionais com habilidades e recursos para manter os processos sem erros. Na parte de infraestrutura de tecnologia da informação, pode ser necessário que sistemas e componentes apresentem performance suficiente para lidar com dados complexos.
B3	Padronização dos processos Os métodos, técnicas, protocolos e procedimentos utilizados pelo produtor durante todas as etapas da concepção do indicador de turismo sigam padrões reconhecidos e aceitos perante os stakeholders, nacional e internacionalmente, claros e consistentes temporariamente, que não modifiquem os conceitos e domínios inerentes aos dados de entrada.
B4	Rigor dos processos Os métodos, técnicas, protocolos e procedimentos utilizados durante a produção sejam robustos e sólidos, com procedimentos de controle, checagem e verificação para que não causem erros de registro ou mensuração, assim como com amostragem ou redução de dados ou dimensionalidade que não diminuam a representatividade e algoritmos e modelos de inferência, ponderação, imputação e mineração que não causem erros de estimação e de pressupostos.
B5	Agilidade da produção Indicador de turismo seja processado (controlado) e divulgado no menor tempo possível (se possível com todas as etapas automatizadas) imediatamente após o recebimento dos dados de entrada (Big Data/insumos) pela organização produtora.
B6	Documentação dos processos Existência de documentação, seguindo padrões e classificações consistentes e uniformes de sustentabilidade, que descreva por completo e de forma clara, inclusive com apresentação de indicadores de qualidade, o processo de produção do indicador de turismo produzido a partir de Big Data em todas as suas etapas, destacando como a eventual complexidade dos dados de entrada foi tratada.

QUADRO: DESCRIÇÃO DOS REQUISITOS DE QUALIDADE DO DOMÍNIO "PRODUTO"

Código Referência	Definição (resumida)
C1	Padronização do indicador Indicador de turismo calculado reflita um construto e siga padrões de formato reconhecidos e aceitos perante os stakeholders, compatível e integrado com indicadores equivalentes em âmbito nacional e/ou internacional, claros e consistentes temporariamente.
C2	Adequação conceitual Indicador final reflita um conceito (domínios, conceitos, construtos, elementos, classificações, escopo e cobertura, período de referência, detalhamento geográfico, granularidade, tipo de indicador, etc.) alinhado com as expectativas do usuário, carregando em si o conteúdo adequado para subsistir insight para fundamentar mais do interesse, sendo válido do ponto de vista de sua utilidade conceitual.
C3	Grau de inovação Indicador final carregue um diferencial frente ao que já existe e é produzido no ambiente tradicional de produção estatística, seja com novo conteúdo, refletindo domínios e construtos até então não explorados, ou reduzindo a quantidade de recursos necessários na sua produção.
C4	Percepção ética Indicador de turismo seja positivamente percebido pelos stakeholders, com ações conduzidas pela organização produtora do indicador que transponham a não reflexão à questões éticas e morais de privacidade e segurança da informação tanto no que diz respeito à origem dos insumos quanto aos processos (práticas, métodos, técnicas, etc.) utilizados durante a produção do indicador final.
C5	Periodicidade praticada Indicador de turismo seja disponibilizado rigorosamente em uma periodicidade regular preestabelecida, necessitando, para tanto, que os dados de entrada também cheguem ao produtor estatístico seguindo um cronograma rígido.
C6	Metadados do indicador Existência de metadados, segundo padrões e classificações consistentes e uniformes, que descrevam por completo e de forma clara tanto o contexto, definições, hierarquias agregadas, intervalos de valores válidos, escopo e cobertura, período de referência, especificação geográfica, etc., quanto o formato do indicador de turismo produzido a partir de Big Data, permitindo ao usuário facilmente perceber se atende às suas expectativas no uso.
C7	Capacidade para uso Existência de oferta de capacidade para uso do indicador, com instruções detalhadas e claras sobre restrições, incertezas, limitações e cuidados operacionais e legais (localizados por questões de complexidade nos insumos, especificidades do processo de produção ou por aspectos de legislação ou de acordos de cooperação), que o usuário deve ter em atenção e seguir o indicador de turismo produzido a partir de Big Data.
C8	Disponibilidade do indicador Indicador e suas informações de apoio (metadados, documentações, etc.) estejam disponíveis ao grande público de forma gratuita e em formato que atende as especificações dos dados ditos abertos.
C9	Disseminação e promoção Existência de uma política com estratégia de divulgação e disseminação do uso do indicador, com sistema de fácil acesso (plataforma, interface, aplicativo, website ou outra ferramenta) aos dados e informações complementares, para fins de facilitação e promoção.

QUADRO: REQUISITOS DE QUALIDADE DOS INDICADORES

Domínio	Requisito	Dimensão da Qualidade Estatística			
		Relevância	Atualidade e Confiança	Opportunidade e Precisão	Compreensão e Acesso e Clareza
Insumos	A1			X	
	A2	X	X		X
	A3				X
	A4		X	X	
	A5			X	
	A6			X	X
	A7		X	X	X
	A8		X		X
	A9		X		X
	A10		X		X
	A11		X	X	X
A12		X		X	
Produção	B1	X	X		
	B2	X	X	X	X
	B3		X	X	X
	B4	X	X	X	X
	B5	X	X	X	X
Produto	C1	X	X		
	C2	X			
	C3	X			
	C4	X			
	C5	X	X	X	X
	C6	X			X
	C7	X			X
	C8	X			X
	C9	X			X

REQUISITOS DE QUALIDADE

Os requisitos do framework foram definidos inicialmente mediante análise de outros artefatos já existentes, complementados com os resultados de pesquisas e publicações sobre o tema, produto de revisões de literatura sistemática e integrativa realizadas (CAUMO, 2020; CAUMO, SOUZA, 2023), com posterior ampliação e ajustes baseados nas manifestações dos especialistas durante a primeira rodada de consulta via Delphi.

Questionário 2
Link: <https://urtesa.in.gov.br/q2>

Tese (novo texto)
Link: <https://urtesa.in.gov.br/t2>

POSSO OUVIR SUA OPINIÃO NOVAMENTE?

Mais uma vez, muitíssimo obrigado por ter chegado até aqui!

Para concluirmos sua participação, um questionário novo, mais objetivo e sintético precisa ser respondido (primeiro link acima).

Caso necessário, é possível consultar o texto da tese (atualizado) no segundo link.

APÊNDICE E – Segundo questionário aplicado aos especialistas

The screenshots illustrate the user interface of a survey application, organized into three main sections:

- Rodada Delphi 2:** The first screen displays the title and a message: "Prezado, esta segunda rodada tem por objetivo ser mais sintética e quantitativa. Ao final, em caso de avaliação com nível global satisfatório, o processo de consulta aos especialistas será encerrado." A "COMEÇAR" button is at the bottom.
- Parte 1: Estrutura Geral:** The second screen shows a form for email identification: "Vamos começar. Pra fins de identificação, qual seu e-mail?*" with a text input field containing "exemplo@exemplo.com.br" and an "AVANÇAR" button.
- Parte 1: Estrutura Geral (Content):** The third screen displays the title and introductory text: "A lógica de estruturação do framework mantém-se a mesma inicialmente proposta, partindo do entendimento de qualidade estatística enquanto um conceito multidimensional, sendo aqui considerada como o resultado da composição das duas dimensões de qualidade, operacionalizadas na prática por conjuntos de requisitos, cada um pertencente a um domínio." It includes a "VOLTAR" button on the left and an "AVANÇAR" button on the right.
- Parte 1: Estrutura Geral (Question):** The fourth screen asks: "1.a) Qual seu nível de concordância com a proposta de estruturação geral do framework?*" It features a legend with five levels: "Um - discordo totalmente", "Dois - discordo parcialmente", "Três - indiferente (neutro)", "Quatro - concordo parcialmente", and "Cinco - concordo totalmente". Below the legend are five star icons, with the first one highlighted.
- Parte 2: Dimensões:** The fifth screen shows the title and a "VOLTAR" button on the left and an "AVANÇAR" button on the right.
- Parte 2: Dimensões (Content):** The sixth screen displays the title and text: "Foram mantidas as dimensões preconizadas pelo European Statistics Code of Practice (EUROSTAT, 2017) e pelo United Nations National Quality Assurance Frameworks Manual for Official Statistics (UN, 2018), segundo padronização internacional: Relevância, Análise e Confiabilidade; Oportunismo temporal e Pontualidade; Coerência e Comparabilidade; Acessibilidade e Clareza." It includes a legend and five star icons, with the second and third stars highlighted.
- Parte 2: Dimensões (Question):** The seventh screen asks: "2.a) Qual seu nível de concordância com as 5 dimensões de qualidade consideradas pelo framework?*" It features a legend with five levels and five star icons, with the second and third stars highlighted.
- Parte 3: Domínios:** The eighth screen shows the title and a "VOLTAR" button on the left and an "AVANÇAR" button on the right.
- Parte 3: Domínios (Content):** The ninth screen displays the title and text: "Foram mantidos os três domínios inicialmente considerados, Insumos (entrada), Produção (processo) e Produtos, segundo uma estrutura já consolidada e aceita quando de atenção à qualidade estatística." It includes a legend and three flag icons representing the domains: "Insumos", "Produção", and "Produtos".
- Parte 3: Domínios (Question):** The tenth screen asks: "3. Domínios" and features a legend with three levels and three flag icons.

3.a) Qual seu nível de concordância com os 3 domínios do ambiente de produção estatística considerados pelo framework?*

LEGENDA:
Um - discordo totalmente
Dois - discordo parcialmente
Três - suficiente (neutro)
Quatro - concordo parcialmente
Cinco - concordo totalmente

★ ★ ★ ★ ★
Discordo totalmente Concordo totalmente

← VOLTAR AVANÇAR →

Parte 4: Requisitos

← VOLTAR AVANÇAR →

Parte 4: Requisitos

4. Requisitos

Os requisitos do framework foram definidos inicialmente mediante análise de outros artefatos já existentes, complementados com os resultados de pesquisas e publicações sobre o tema, produto de revisões de literatura sistemática e integrativa realizadas (CALAM, 2020; CALAM, SOUZA, 2023), com posterior ampliação e ajustes baseados nas manifestações dos especialistas durante a primeira rodada de consulta via Delphi.

Comentário: Seção 4 do capítulo de primeira rodada Delphi.
Mostrar informações sobre a 1ª Rodada

REQUISITOS DO DOMÍNIO "INSUMOS":

Código	Referência	Definição (resumida)
A1	Barral Aguiar	Condição de trabalho que estabelece como os dados de contabilidade, produção de prestação de serviços de uma empresa são gerados, armazenados, analisados e compartilhados com os demais setores da empresa (Ribeiro, 2019; Calam, 2020).

← VOLTAR AVANÇAR →

4.a) Qual seu nível de concordância com os requisitos de qualidade propostos no framework?*

LEGENDA:
Um - discordo totalmente
Dois - discordo parcialmente
Três - suficiente (neutro)
Quatro - concordo parcialmente
Cinco - concordo totalmente

★ ★ ★ ★ ★
Discordo totalmente Concordo totalmente

← VOLTAR AVANÇAR →

Parte 5: Relacionamentos

← VOLTAR AVANÇAR →

Parte 5: Relacionamentos

5. Relacionamentos

A estrutura de relacionamento, pertencimento ou influência entre os requisitos e os domínios e dimensões de qualidade foi definida também com base na análise de outros artefatos já existentes (LUNCE, 2014; EURÓSTAT, 2014a; KRÜSTER et al., 2015; BATINI et al., 2015; CAI, ZHU, 2015), nos resultados de pesquisas e publicações sobre o tema, produto de revisões de literatura sistemática e integrativa realizadas (CALAM, 2020; CALAM, SOUZA, 2023), com posterior ampliação e ajustes baseados nas manifestações dos especialistas durante a primeira rodada de consulta via Delphi.

Comentário: Seção 5 do capítulo de primeira rodada Delphi.
Mostrar informações sobre a 1ª Rodada

Domínio	Requisito	Dimensão de Qualidade Estática				
		Atividade e Qualidade	Operativ. e Recursos e Processos	Confiabilidade e Acesso e Dados	Atividade e Qualidade	Operativ. e Recursos e Processos
A1	Barral Aguiar					

← VOLTAR AVANÇAR →

5.a) Qual seu nível de concordância com a proposta de estrutura de relacionamento entre os requisitos, domínios e dimensões considerados no framework?*

LEGENDA:
Um - discordo totalmente
Dois - discordo parcialmente
Três - suficiente (neutro)
Quatro - concordo parcialmente
Cinco - concordo totalmente

★ ★ ★ ★ ★
Discordo totalmente Concordo totalmente

← VOLTAR AVANÇAR →

Parte 6: Operacionalização

← VOLTAR AVANÇAR →

Parte 6: Operacionalização

6. Operacionalização

Sobre este aspecto da operacionalização prática, cabe destacar que a proposta aqui apresentada tem por finalidade visar a apresentação de uma possibilidade de controle a ser seguida por pesquisas futuras. Não está no escopo do objetivo geral desta tese a construção de um protocolo definitivo para operacionalização prática do artefato, mas sim a proposição de um framework de cunho conceitual. Assim, reforça-se aqui que o protocolo de operacionalização apresentado teve por finalidade, essencialmente, fomentar a pesquisa científica sobre o assunto, servindo de ponto de partida, além de viabilizar, através de um exercício de simulação, uma primeira análise da aplicabilidade do framework conceitual proposto – para fins de atendimento do último objetivo específico desta tese.

Ou seja, não foram feitas alterações na estrutura de operacionalização, reconhecendo-se aqui a necessidade de muitos avanços no protocolo que hoje consiste na análise do indicador em relação à

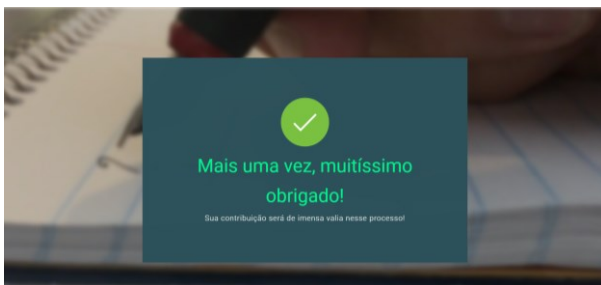
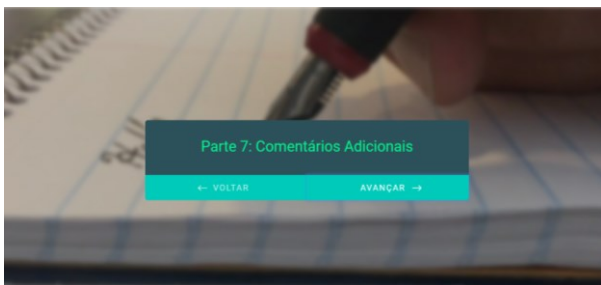
← VOLTAR AVANÇAR →

6.a) Qual seu nível de concordância com a proposta de operacionalização prática simulada para o framework?*

LEGENDA:
Um - discordo totalmente
Dois - discordo parcialmente
Três - suficiente (neutro)
Quatro - concordo parcialmente
Cinco - concordo totalmente

★ ★ ★ ★ ★
Discordo totalmente Concordo totalmente

← VOLTAR AVANÇAR →



APÊNDICE F – Script (Python) do *crawler* utilizado na mineração web

```

import sys
import csv
from selenium import webdriver
import time
import random
from selenium.webdriver.support.ui import WebDriverWait
from selenium.webdriver.common.by import By
from selenium.webdriver.support import expected_conditions as EC
path_to_file = "c:/Users/rbcau/OneDrive/Desktop/barradalagoa.csv"
num_page = 544
url = "https://www.tripadvisor.com.br/Attraction_Review-g303576-d312316-Reviews-
Barra_da_Lagoa_Beach-Florianopolis_State_of_Santa_Catarina.html#REVIEWS"
if (len(sys.argv) == 4):
    path_to_file = sys.argv[1]
    num_page = int(sys.argv[2])
    url = sys.argv[3]
driver = webdriver.Firefox()
driver.get(url)
csvFile = open(path_to_file, 'a')
fieldnames = ["Nota", "Data_Exp", "Data_Ava"]
csvWriter = csv.DictWriter(csvFile, fieldnames=fieldnames,
delimiter=';', lineterminator='\r')
csvWriter.writeheader()
for i in range(0, num_page):
    time.sleep(random.randint(5,10))
    container = driver.find_elements_by_xpath("//div[@data-reviewid]")
    dates = driver.find_elements_by_xpath("//div[@class='_2fxQ4TOx']")
    datesrw = driver.find_elements_by_xpath("//div[@class='_27JpaCjI']")
    for j in range(len(container)):
        rating = container[j].find_element_by_xpath("//span[contains(@class,
'ui_bubble_rating_bubble_')]").get_attribute("class").split("_")[3]
        data = ".join(dates[j].text.split(" ")[-3:])
        data1 = ".join(datesrw[j].text.split(" ")[-3:])
        csvWriter.writerow( { "Nota" : rating , "Data_Exp" : data1, "Data_Ava" : data} )
    button = driver.find_element_by_xpath('//a[@class="ui_button nav next primary
"]')
    driver.execute_script("arguments[0].click();", button)
driver.quit()

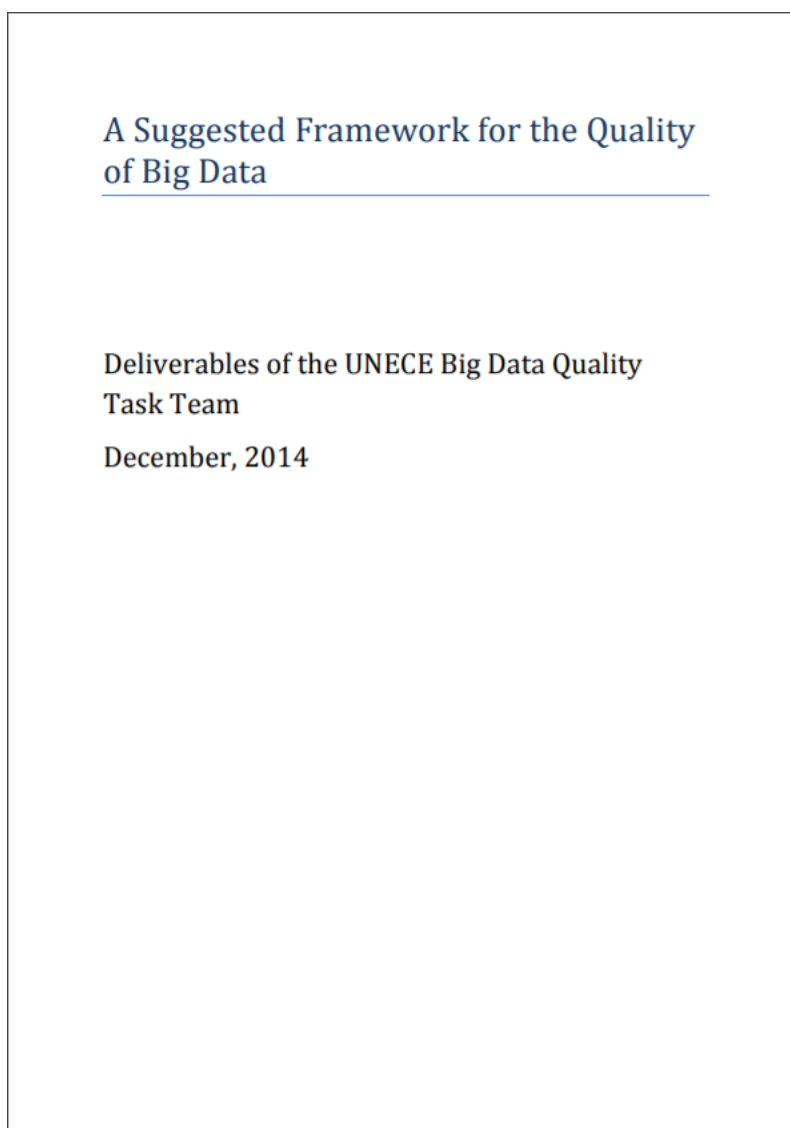
```

ANEXO A – Relação de fontes de dados sobre o turismo no Brasil

Estudos e Pesquisas	Descrição	Periodicidade	Abrangência	Divulgação	Instituição responsável
Estudo da demanda turística internacional	Dados sobre o perfil do turista internacional em visita ao Brasil e do brasileiro em viagem ao exterior.	Anual	Brasil	Internet	Ministério do Turismo Executor - Fundação Instituto de Pesquisa Econômica - FIPE
Chegadas de turistas não residentes ao Brasil	Dados estimados sobre a chegada de turistas não residentes ao Brasil, por portão de entrada no Brasil, com a utilização de dados de migração coletados e cedidos pela Polícia Federal.	Anual	Brasil	Internet	Departamento de Polícia Federal - DPF Ministério do Turismo
Estudo da demanda turística doméstica	Dados sobre o dimensionamento e caracterização do mercado interno de viagens no Brasil.	Irregular	Brasil	Internet	Ministério do Turismo Executor - Fundação Instituto de Pesquisa Econômica - FIPE
Receita cambial turística	Dados sobre o fluxo de divisas (Receita e Despesa Cambial) na Conta Viagens Internacionais do Balanço de Pagamentos, a partir de dados compilados junto ao Banco Central	Mensal/Anual	Brasil	Internet,	Banco Central do Brasil - BACEN
Sondagem de expectativas do consumidor - Análise do setor de viagens	Dados mensais sobre intenção das famílias em realizar viagens nacionais e internacionais nos próximos seis meses.	Mensal	Brasil	Internet	Ministério do Turismo Executor - Fundação Getúlio Vargas - FGV
Boletim de desempenho econômico do turismo	Dados trimestrais sobre o mapeamento da opinião de empresários a respeito da evolução e as perspectivas do setor, nas áreas de agências de viagem, hotelaria, restaurantes, turismo receptivo, eventos, operadoras, parques temáticos, atrações turísticas e transporte aéreo.	Trimestral	Brasil	Internet	Ministério do Turismo Executor - Fundação Getúlio Vargas - FGV
Pesquisa anual de conjuntura econômica do turismo brasileiro	Dados anuais sobre o cenário econômico do setor de turismo no Brasil, isto é, negócios realizados e perspectivas futuras para o ano, traçado a partir da percepção dos 80 principais executivos dos setores de agências de viagem, receptivo e operadores de turismo e promotores de feiras, promotores de eventos, locadoras de automóveis, meios de hospedagem, transporte aéreo e rodoviário.	Anual	Brasil	Internet	Ministério do Turismo Executor - Fundação Getúlio Vargas - FGV
Movimento de passageiros nos aeroportos brasileiros	Dados sobre o desembarque e embarque de passageiros nacionais e internacionais, por UF, em voos regulares e não-regulares, utilizando-se de base de dados da INFRAERO, ANAC e concessionárias.	Mensal/Anual	Brasil	Internet	Empresa Brasileira de Infraestrutura Aeroportuária - INFRAERO Agência Nacional de Aviação Civil - ANAC
Movimento de passageiros em voos regulares	Dados sobre o movimento de passageiros em voos nacionais e internacionais, por companhia aérea, origem e destino, assentos oferecidos e utilizados.	Anual	Brasil	Internet	Agência Nacional de Aviação Civil - ANAC
Movimentação de passageiros em rodoviárias do Brasil	Dados sobre a movimentação de passageiros nas rodoviárias que recebem linhas nacionais e internacionais.	Anual	Brasil	Internet	Agência Nacional de Transportes Terrestres - ANTT
Equipamentos e prestadores de serviços turísticos cadastrados no Ministério do Turismo	Dados sobre o número de equipamentos e prestadores de serviços turísticos cadastrados no Ministério do Turismo: agências de turismo, meios de hospedagem, transportadoras turísticas, instituições de ensino, bacheleiros em turismo e guias de turismo.	Anual	Brasil	Internet	Ministério do Turismo
Desembolso de recursos por instituições financeiras federais para o financiamento do turismo no Brasil.	Dados sobre o desembolso de recursos por instituições financeiras federais (Banco do Brasil, Caixa Econômica Federal, Banco do Nordeste, Banco da Amazônia e BNDES) para o financiamento do turismo no Brasil.	Mensal/Anual	Brasil	Internet	Ministério do Turismo
Mercado de trabalho no setor de turismo	Dados sobre a ocupação formal nas Atividades Características do Turismo - ACTs, a partir da combinação de dados da Relação Anual de Informações Sociais - RAIS e Cadastro Geral de Emprego e Desemprego - CAGED.	Anual	Brasil	Internet	Instituto de Pesquisa Econômica Aplicada - IPEA Ministério do Trabalho e Emprego Ministério do Turismo
Economia do turismo - Uma perspectiva macroeconômica - 2003-2009	Principais agregados macroeconômicos das atividades características do turismo e sua participação na economia do país, a partir de dados compilados no Sistema de Contas Nacionais.	Anual	Brasil	Internet e impresso	Instituto Brasileiro de Geografia e Estatística - IBGE

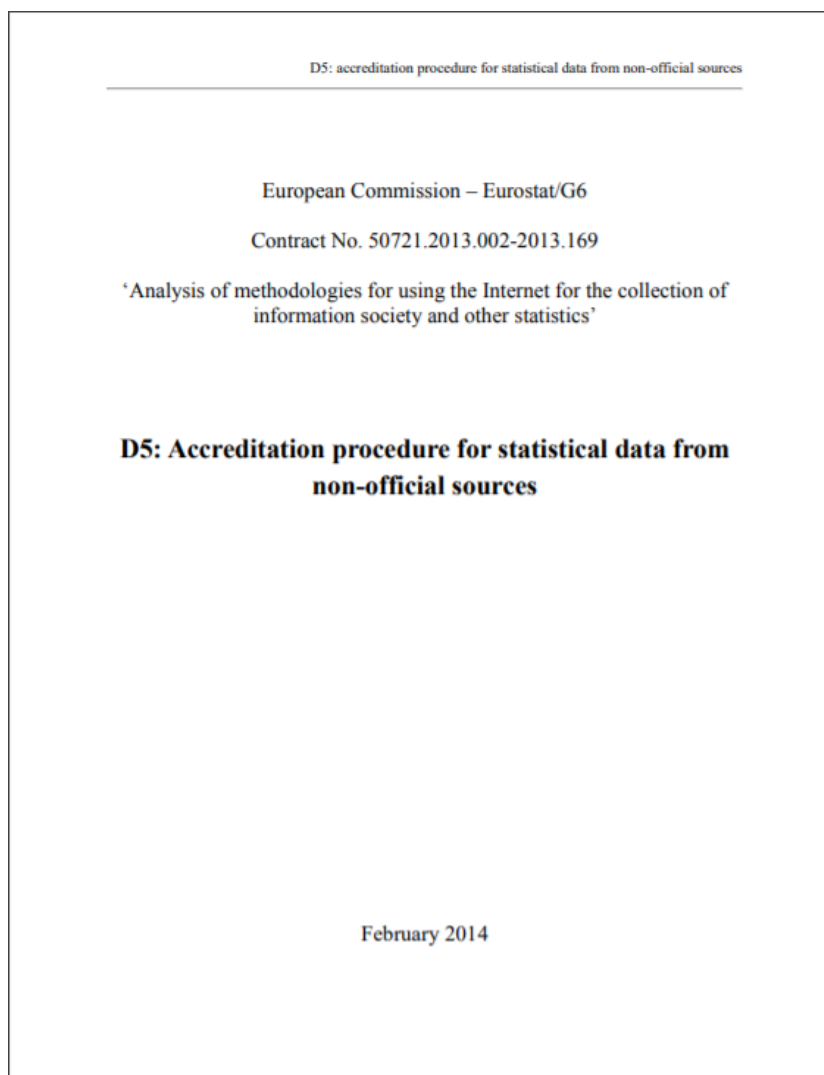
Fonte: Brasil (2018b).

ANEXO B – *UNECE Big Data Quality Framework*



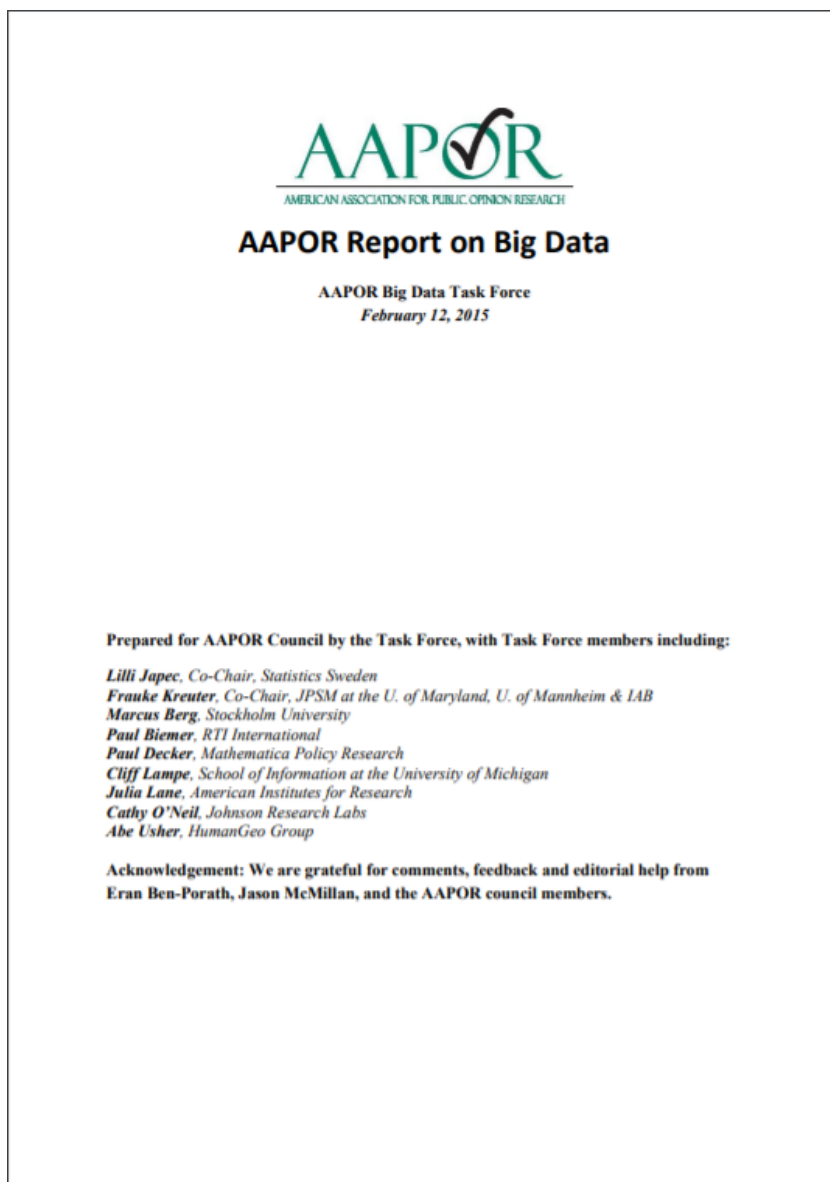
Em virtude de sua extensão – 32 páginas – e da falta de alguma figura ou quadro síntese que pudesse ser aqui apresentado, o *UNECE Big Data Quality Framework* (UNECE, 2014) foi suprimido, podendo ser obtido (acesso verificado em 22 jan. 2021) através do link <https://statswiki.unece.org/download/attachments/108102944/Big%20Data%20Quality%20Framework%20-%20final-%20Jan082015.pdf?version=1&modificationDate=1420725063663&api=v2>.

ANEXO C – Eurostat Accreditation Procedure for statistical data from Non-Official Sources



Em virtude de sua extensão – 34 páginas – e da falta de uma figura ou quadro síntese único que pudesse ser aqui apresentado, o Eurostat *Accreditation Procedure for statistical data from Non-Official Sources* (EUROSTAT, 2014a) foi suprimido, podendo ser obtido (acesso verificado em 22 jan. 2021) através do link https://ec.europa.eu/eurostat/cros/system/files/D5_Accreditation%20procedure%20for%20statistical%20data%20from%20non-official%20sources_20140206_0.pdf.

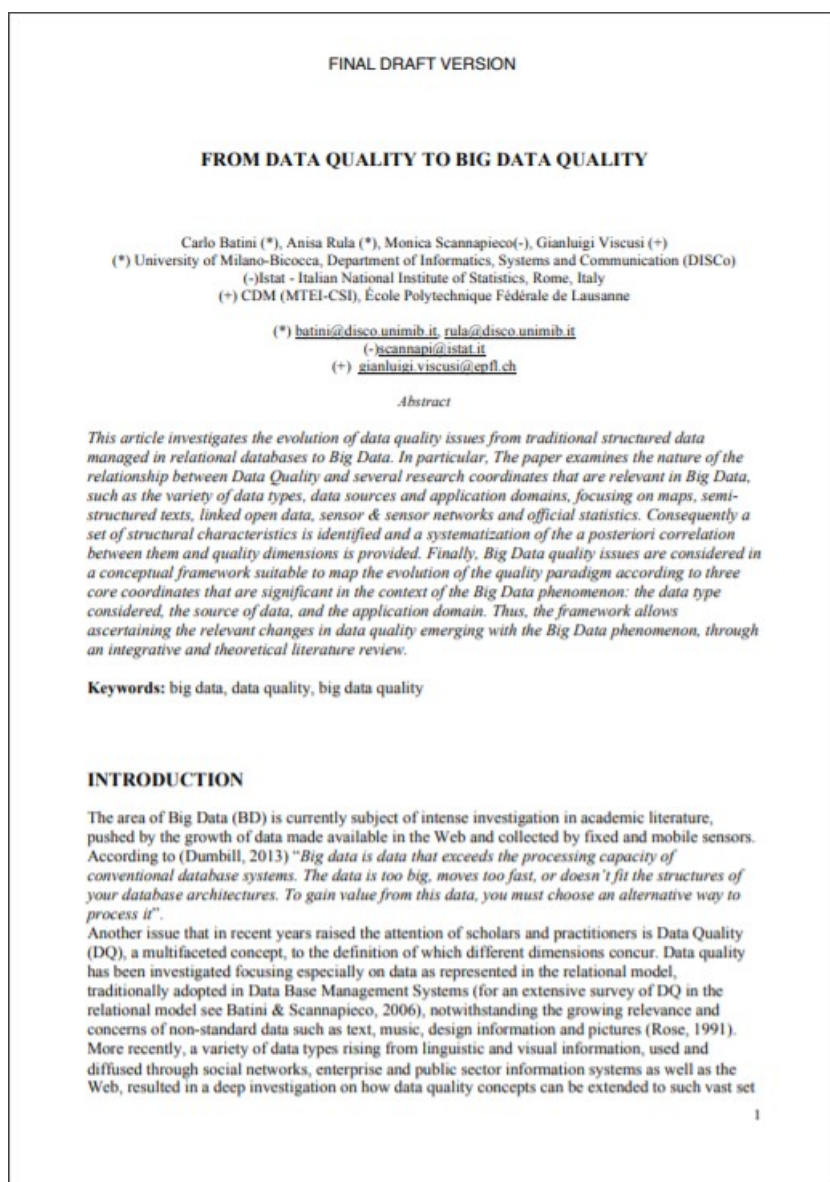
ANEXO D – AAPOR Big Data Total Error



Em virtude de sua extensão – 50 páginas – e da falta de uma figura ou quadro síntese único que pudesse ser aqui apresentado, o AAPOR *Big Data Total Error Framework* (KREUTER et al., 2015)²⁰ foi suprimido, podendo ser obtido (acesso verificado em 22 jan. 2021) através do link https://www.aapor.org/AAPOR_Main/media/Task-Force-Reports/BigDataTaskForceReport_FINAL_2_12_15_b.pdf.

²⁰ Atualizado e operacionalizado em Amaya, Biemer e Kinyon (2020).

ANEXO E – Conceptual Framework for Quality in Big Data



Em virtude de sua extensão – 22 páginas – e da falta de uma figura ou quadro síntese único que pudesse ser aqui apresentado, o *Conceptual Framework for Quality in Big Data* (BATINI et al., 2015) foi suprimido, podendo ser obtido (acesso verificado em 22 jan. 2021) através do link <https://www.researchgate.net/publication/283681085> From Data Quality to Big Data Quality.

ANEXO F – *Big Data Quality Assessment Framework*



Em virtude de sua extensão – 11 páginas – e da falta de uma figura ou quadro síntese único que pudesse ser aqui apresentado, o *Big Data Quality Assessment Framework* (CAI; ZHU, 2015) foi suprimido, podendo ser obtido (acesso verificado em 22 jan. 2021) através do link <https://www.researchgate.net/publication/277943983> The Challenges of Data Quality and Data Quality Assessment in the Big Data Era.