

FUNDAÇÃO GETULIO VARGAS
ESCOLA DE ADMINISTRAÇÃO DE EMPRESAS DE SÃO PAULO

SIMONE DA SILVA LUVIZAN

REÚSO DE DADOS NA ERA DO BIG DATA
Uma Jornada Rumo a Novos Paradigmas no Setor Financeiro

SÃO PAULO
2019

SIMONE DA SILVA LUVIZAN

REÚSO DE DADOS NA ERA DO BIG DATA
Uma Jornada Rumo a Novos Paradigmas no Setor Financeiro

Tese apresentada à Escola de Administração de Empresas de São Paulo da Fundação Getúlio Vargas, como requisito para obtenção do título de Doutor em Administração de Empresas.

Campo do Conhecimento: Tecnologia e Data Science

Orientador: Prof. Dr. Eduardo Henrique Diniz

SÃO PAULO
2019

Luvizan, Simone da Silva.

Reúso de dados na era do Big Data : uma jornada rumo a novos paradigmas no setor financeiro / Simone da Silva Luvizan. - 2019.
178 f.

Orientador: Eduardo Henrique Diniz.

Tese (doutorado CDAE) – Fundação Getulio Vargas, Escola de Administração de Empresas de São Paulo.

1. Big Data. 2. Mineração de dados (Computação). 3. Mercado financeiro - Inovações tecnológicas. I. Diniz, Eduardo Henrique. II. Tese (doutorado CDAE) – Escola de Administração de Empresas de São Paulo. III. Fundação Getulio Vargas. IV. Título.

CDU 007.522.2

SIMONE DA SILVA LUVIZAN

REÚSO DE DADOS NA ERA DO BIG DATA

Uma Jornada Rumo a Novos Paradigmas no Setor Financeiro

Tese apresentada à Escola de Administração de Empresas de São Paulo da Fundação Getúlio Vargas, como requisito para obtenção do título de Doutor em Administração de Empresas.

Campo do Conhecimento: Tecnologia e Data Science

Data da Aprovação:

____/____/____

Banca Examinadora:

Prof. Dr. Eduardo Henrique Diniz (orientador)
FGV-EAESP

Prof. Dr. Eduardo de Rezende Francisco
FGV-EAESP

Profa. Dra. Elaine Tavares
Coppead-UFRJ

Prof. Dr. João Luiz Becker
FGV-EAESP

Prof. Dr. Norberto Hoppen
UNISINOS

Ao meu amado Marco e à minha querida
família, que trazem sentido, cor e sabor a
todos os dias.

Agradecimentos

Estes agradecimentos estão recheados de clichês. Não consegui fugir deles, pois, em meu coração, todos são sinceros e únicos. Aos que o recebem, peço que acreditem na força e na verdade de cada um.

A jornada do doutorado não se percorre só e, ao final dela, há muitos a quem agradecer. A Deus, por iluminar o caminho. Aos amigos e familiares, pelo apoio, incentivo e compreensão constantes. Aos colegas de turma, pelo companheirismo, colaboração e amizade. À FGV e todos os seus funcionários, pela acolhida e suporte. À HEC, seus funcionários, professores, alunos e pesquisadores visitantes com os quais tive a oportunidade de interagir e que tornaram minha experiência em Montreal enriquecedora, para a tese e para a vida.

Aos queridos professores da FGV, minha profunda gratidão e respeito. De maneira especial, ao prof. Dr. Eduardo Henrique Diniz, orientador dessa tese e meu grande companheiro nessa caminhada. Sagaz em suas orientações, generoso ao compartilhar seus conhecimentos, inspirador ao falar de suas ideias. À prof. Dra. Marlei Pozzebon que iluminou essa tese com seus saberes e tornou possível a estada na HEC. Aos professores das bancas de qualificação e defesa por sua disponibilidade e suas contribuições para o avanço desse trabalho.

Um grande agradecimento às empresas que abriram suas portas para a presente pesquisa. Também a todos que colaboraram para o desenvolvimento dessa tese com seus questionamentos e sugestões. Um reconhecimento especial aos autores cujos estudos sustentaram os passos desse caminho.

RESUMO

A crescente oferta de dados e de tecnologias para processá-los alavancada pelo Big Data (BD) fomenta as expectativas de geração de inteligência e valor a partir da informação. Nesse cenário, é cada vez mais comum que os dados sejam utilizados para novos fins, distintos daqueles para os quais foram gerados. Essa prática, denominada reuso de dados, oferece grandes oportunidades de desenvolvimento, mas também pode trazer implicações pouco discutidas na academia e na sociedade como um todo. Para contribuir com esse debate, esta tese visa caracterizar o reuso de dados, identificar seus desafios e propor um modelo para explicar como os atores estão lidando com tais desafios na prática. O trabalho concentrou-se no setor financeiro e foi realizado em 3 etapas, dando origem a 3 artigos. O primeiro realiza uma revisão bibliográfica para refletir sobre a literatura existente e uma meta análise de casos reportando reuso de dados em artigos empíricos. A partir dessa meta-análise, é proposto um esquema classificatório para o reuso de dados baseado em duas dimensões: fonte dos dados (pública, privada interna ou externa à organização) e finalidade do reuso (repropósito ou recontextualização). A reflexão sobre os potenciais desafios dos diferentes tipos de reuso sugere que eles podem variar em função da distância entre a geração e o uso dos dados, dando origem a 3 níveis de desafios: direto, intermedial e extremo. O segundo artigo propõe um modelo conceitual para explorar casos de reuso de dados na prática. Inspirado no Multilevel Framework (Pozzebon, Diniz & Jayo, 2009), o modelo utiliza os elementos do contextualismo (Pettigrew, 1985), da SST (Social Shaping of Technology), da abordagem processual (Langley, 1999; Pettigrew, 1997) e do reuso de dados. Nesta etapa, o modelo foi exercitado em um estudo de caso exploratório na empresa Alpha, que oferece uma ferramenta para análise de risco de crédito baseada em perfis comportamentais gerados a partir de dados da rede de celulares. A terceira etapa da pesquisa realizou um estudo de 3 casos, representando diferentes tipos de reuso de dados: um banco (repropósito e recontextualização de dados internos - nível de desafio direto), um bureau de crédito (repropósito de dados externos públicos e privados - nível de desafio intermedial) e fintechs (recontextualização de dados externos públicos e privados - nível de desafio extremo). A análise dos casos identificou as temáticas principais envolvidas em seus desafios e permitiu refletir sobre suas dimensões legais, tecnológicas, intrínsecas, societárias e econômicas. Nos casos, o processo pelo qual os atores lidam com os desafios do reuso de dados na prática revelou uma dinâmica cíclica de ajuste/aprendizado e de mobilização. Tal processo é animado por grupos que mobilizam recursos por meio de uma abordagem de posse ou de prática de poder, como também observado em outro estudo de casos envolvendo colaboração complexa. Esse trabalho contribui para o avanço das pesquisas sobre reuso de dados discutindo conceitos e questões relevantes, especialmente para o campo de SI. Suas escolhas metodológicas e teóricas podem inspirar outras pesquisas que compartilhem o desafio de explorar fenômenos complexos.

Palavras-chave: Big Data; Analytics; Reuso de Dados; Multilevel Framework; Ecossistemas de Inovação.

ABSTRACT

Data and technology increasing availability, intensified by Big Data (BD), stimulate expectations for more intelligence and value creation using existing information. In this scenario, it's becoming more frequent the use of data for new purposes, different than the ones they were generated for. This growing practice, called reuse of data, brings great development opportunities, but also implications that were poorly discussed by academy and society as a whole. To collaborate with this debate, this thesis aims to characterize the reuse of data, identify its challenges and propose a model to explain how the actors are dealing with those challenges in practice. The study was concentrated in finance sector and was done in 3 steps, generating 3 papers. The first one performed a bibliographic review to ponder about the existing literatures and also a meta-analysis of data reuse empirical cases. Considering this meta-analysis, it's proposed a classification schema based on two dimensions: data source (internal, external public or private) and reuse purpose (repurpose or recontextualization). The discussion about potential challenges of different reuse of data categories indicates they are influenced by the distance between data generation and use, suggesting 3 levels of challenges: direct, intermediate and extreme. The second paper proposes a theoretical model to explore reuse of data cases in practice. Inspired on Multilevel Framework (Pozzebon, Diniz & Jayo, 2009), the model catches elements of contextualism (Pettigrew, 1985), SST (Social Shaping of Technology), process approach (Langley, 1999; Pettigrew, 1997) and reuse of data. At this stage, the model was exercised in an exploratory case at Alpha, a company that offer a tool for credit risk analysis based on behavioral profiles generated over mobile network data. The third research phase investigated 3 cases, representing different reuse of data types: a bank (repurposing and recontextualization of internal data – level of challenge direct), a credit bureau (repurposing of public or private external data – level of challenge intermediate) and fintechs (recontextualization of public or private external data – level of challenge extreme). The cases analysis identified the main themes involved on challenges and allow the discussion of their legal, technological, intrinsic, societal and economical dimensions. The cases disclosed that the process by which the actors are dealing with the reuse of data challenges in practice is based on a cyclic dynamic of adjustment/learning and mobilization. Such process is animated by groups that mobilize resources through a possession or practice approach, like also showed in other case study involving complex collaboration. This work contributes to advance the research about reuse of data bringing concepts and relevant questions, especially in IS field. Its methodological and theoretical choices can inspire other researches that share the challenge of exploring complex phenomena.

Key-words: Big Data; Analytics; Reuse of data; Multilevel Framework; Innovation Ecosystem.

LISTAS DE FIGURAS

Figura 1.1: Abordagem da pesquisa ao seu Quintain.....	22
Figura 2.1: Questões ilustrativas e aspectos críticos das quatro dimensões do reúso de dados...	37
Figura 2.2: Sequência de buscas e seleções.....	38
Figura 2.3: Diagrama do processo e meta-análise da pesquisa.....	48
Figura 3.1: Reúso de dados em um ecossistema de negócios	78
Figura 3.2: Modelo multilevel framework original.....	82
Figura 3.3: Ecossistema do reúso de dados pelas lentes da prática.....	83
Figura 3.4: Processo do caso Alpha.....	93
Figura 4.1: Matriz do esquema classificatório do reúso de dados por fonte x finalidade.....	111
Figura 4.2: Dimensões do reúso de dados que implicam em seus desafios.....	112
Figura 4.3: Ecossistema do reúso de dados pelas lentes da prática.....	113
Figura 4.4: Dinâmica do reúso de dados para lidar com os desafios de seu ecossistema.....	141
Figura 5.1: Ecossistema do reúso de dados pelas lentes da prática.....	172
Figura 5.2: Dinâmica do reúso de dados para lidar com os desafios de seu ecossistema.....	173

LISTA DE GRÁFICOS

Gráfico 2.1: Estudos sobre reuso de dados em periódicos acadêmicos por ano.....	39
Gráfico 2.2: Distribuição dos casos de reuso de dados por finalidade x fonte.....	54
Gráfico 4.1: Desafios do reuso de dados percebidos por caso.....	131
Gráfico 4.2: Desafios ligados à mudança por dimensão.....	133

LISTA DE QUADROS

Quadro 2.1: Potenciais desafios do reuso de dados em diferentes situações.....	55
Quadro 3.1: Matriz finalidade x fonte do reuso de dados.....	75
Quadro 3.2: Contexto do caso Alpha.....	91
Quadro 3.3: Conteúdo – o ecossistema do caso da Alpha.....	95
Quadro 4.1: Casos selecionados por tipo de desafio do reuso de dados.....	117
Quadro 4.2: Quadro de entrevistas dos casos.....	119
Quadro 4.3: Produtos de análise dos dados por etapa.....	121
Quadro 5.1: Nível de desafios potenciais do reuso de dados por categoria.....	171

LISTAS DE TABELAS

Tabela 2.1: Distribuição das publicações por área de periódicos.....	40
Tabela 2.2: Temas abordados nos artigos sobre reúso de dados por área de periódicos.....	40
Tabela 2.3: Resultado das buscas.....	50

SUMÁRIO

1	INTRODUÇÃO	14
1.1	Contexto e motivação da pesquisa	16
1.2	Posicionamento, metodologia e objetivos da pesquisa	19
1.3	Declaração das colaborações	23
	Referências bibliográficas	25
2	ARTIGO I – BIG DATA E REÚSO DE DADOS: DESAFIOS E OPORTUNIDADES	29
2.1	Resumo	29
2.2	Introdução	30
2.3	Big data e seus tantos “v’s”	31
2.4	Reúso de dados	34
2.4.1	O que é reuso de dados e porque isso importa	34
2.4.2	Pesquisa acadêmica sobre reuso de dados	38
2.4.3	Algumas implicações do reúso de dados relevantes para SI	42
2.5	Classificação do reúso de dados	47
2.5.1	Metodologia da meta-análise	47
2.5.2	Resultados	52
2.5.3	Desafios emergentes do reúso de dados em suas distintas categorias	54
2.6	Conclusão	57
2.7	Limites e sugestões para estudos futuros	58
	Referências bibliográficas	58
	Apêndices	66
3	ARTIGO II – REÚSO-DE-DADOS-NA-PRÁTICA: RECONTEXTUALIZAÇÃO DE DADOS DA TELEFONIA CELULAR PARA AVALIAÇÃO DE CRÉDITO	70
3.1	Resumo	70
3.2	Introdução	71
3.3	Reúso de dados na era do <i>big data</i>	73
3.4	Ecossistema de reúso de dados pelas lentes da prática	75
3.4.1	Conceitos do ecossistema e sua aplicação ao reúso de dados	76
3.4.2	Teorizando a Prática	78
3.4.3	Multilevel Framework aplicado ao ecossistema de reúso de dados	80

3.5	Metodologia	84
3.6	O Caso da análise de crédito usando dados da rede de celulares	87
3.6.1	O nascimento da Alpha	87
3.6.2	Breve histórico da avaliação de risco de crédito	88
3.6.3	A Solução da Alpha	89
3.6.4	O caso da Alpha no Brasil	90
3.6.4.1	Contexto	90
3.6.4.2	Processo	92
3.6.4.3	Conteúdo	94
3.7	Discussão	95
3.8	Conclusões, limites e oportunidades	99
	Referências Bibliográficas	100
4	ARTIGO III – REÚSO DE DADOS: UMA JORNADA DE MUDANÇAS RUMO A NOVOS PARADIGMAS NO SETOR FINANCEIRO	106
4.1	Resumo	106
4.2	Introdução	107
4.3	Revisão teórica	110
4.3.1	Reúso de dados	110
4.3.2	Multilevel Framework adaptado	113
4.4	Metodologia	114
4.4.1	Design da pesquisa	115
4.4.2	Seleção dos casos	116
4.4.3	Coleta dados	118
4.4.4	Análise dos dados	120
4.5	Resultado dos casos	121
4.5.1	O caso do Banco Ômega	121
4.5.2	O caso do Bureau Sigma	122
4.5.3	O caso das fintechs	124
4.5.3.1	Minicaso da fintech Zeta	124
4.5.3.2	Minicaso da fintech Delta	126
4.5.3.3	Minicaso da fintech Gama	127
4.6	Análise cross-case	128
4.6.1	O contexto	129

4.6.2	O processo e o conteúdo que dele emerge	130
4.6.3	Os desafios do reuso de dados	133
4.7	Discussão dos resultados	139
4.7.1	Potenciais desafios do reuso de dados	139
4.7.2	Como os atores estão lidando com os desafios que emergem do reuso de dados	141
4.8	Conclusão	144
4.9	Limites e sugestões para futuros estudos	145
	Referências bibliográficas	146
	Apêndices	151
	5 CONCLUSÃO DA TESE	170
5.1	Conteúdo e resultados da tese	170
5.2	Contribuições da Tese	173
5.2.1	Para o reuso de dados	173
5.2.2	Para a área financeira	175
5.2.3	Para a pesquisa	175
5.3	Limites e sugestões para estudos futuros	176
	Referências bibliográficas	177

1 INTRODUÇÃO

Não se usam mais os dados como antigamente. As aplicações não ficam mais restritas ao uso dos dados que elas próprias geram. As ferramentas do *big data* (BD) permitem que elas cruzem suas fronteiras e capturem dados de outras fontes, em formatos diversos e com grande agilidade (Baesens et al., 2016; Martin, 2015; Muller et al., 2016). Diante dessas possibilidades, as organizações não se concentram mais nas informações que produzem em seus processos e sistemas internos (Martin, 2015; Muller et al., 2016; Newell & Marabelli, 2015). Os proprietários não guardam mais seus próprios dados. Agora, eles ficam nas nuvens (Abbasi, Sarker & Chiang, 2016). E como esses dados transitam! Até as coisas trocam informações de forma autônoma, sem a intervenção dos humanos (Baesens et al., 2016).

O contexto de mais dados e mais tecnologia para manejá-los tornou-se solo fértil, no qual florescem soluções inovadoras, gerando crescentes expectativas de criação de valor a partir dos dados (Baesens et al., 2016; Goes, 2015; Ketter et al., 2016; Martin, 2015; Muller et al., 2016; Newell & Marabelli, 2015). Mais conhecimento, novos modelos de negócios, maior eficiência em negócios tradicionais, novas soluções para problemas do cotidiano de indivíduos e organizações públicas ou privadas (Baesens et al., 2016; Letouzé, 2012; Ketter et al., 2016; Martin, 2015; Muller et al., 2016). Em busca desses benefícios, é cada vez mais comum que a informação gerada seja utilizada para novas finalidades. Essa prática, denominada “reúso de dados” (Curty, 2015; Zimmerman, 2003, 2008), abre um leque de oportunidades de conhecimento e desenvolvimento de soluções, mas também pode gerar distorções e efeitos indesejados e nocivos (Borgman, 2015; Newell & Marabelli, 2015; Martin, 2015; Samarajiva & Lokanathan, 2016; Zuboff, 2015).

Esse cenário repleto de desafios e oportunidades evidencia a necessidade de estudar o reúso de dados. Assim como a criação de novas soluções baseadas nesse fenômeno requer a quebra de paradigmas, a discussão sobre ele e suas implicações também urge por um novo olhar (Custers & Ursic, 2016) capaz de capturar a urgência e a relevância do tema para as organizações e a sociedade como um todo (Martin, 2015; Muller et al., 2016; Samarajiva & Lokanathan, 2016; Zuboff, 2015). Um olhar que compreenda sua multidisciplinaridade e dinâmica emergente e veloz (Baesens et al., 2016; Custers & Bachlechner, 2017), disposto a abrir mão de padrões tradicionais sem deixar de, humilde e inteligentemente, aprender com eles (Baesens et al., 2016; Borgman, 2015; Boyd & Crawford, 2012; Goes, 2015; Ketter et al., 2016).

Embora o reúso de dados e a consciência sobre suas implicações estejam crescendo na prática, ele tem sido pouco explorado na academia (Custers & Bachlechner, 2017; Newell & Marabelli, 2015; Samarajiva & Lokanathan, 2016; Yoon, 2015). No campo de Sistemas de Informação (SI), há um significativo aumento de estudos abordando aplicações do BD nos últimos anos (Abbasi, Sarker & Chiang, 2016; Luvizan, Meirelles & Diniz, 2015), sendo que diversos deles reportam situações em que há reúso de dados. No entanto, a maioria não discute essa questão em si e não nomina o fenômeno nesses termos, dificultando o diálogo com a literatura sobre o tema. Muitos artigos sequer refletem sobre o fato de que os dados estão sendo usados para novos fins ou sobre as implicações desse repropósito. Entre os trabalhos que o fazem, não há unanimidade em sua denominação. Alguns se referem ao fenômeno como “uso secundário de dados”, enquanto outros falam em “reúso de dados” (Curty, 2015). Nessa tese, foi adotado o termo “reúso de dados” porque ele aparece com mais frequência na revisão bibliográfica. Diversos autores alertam para a necessidade do debate sobre o tema, refletindo sobre seus desafios e implicações de forma holística (Custers & Bachlechner, 2017; Newell, & Marabelli, 2015; Martin, 2015; Samarajiva & Lokanathan, 2016).

Para colaborar modestamente com a grandiosa empreitada de preencher essa lacuna de conhecimento e instrumentar o debate sobre o reúso de dados, o objetivo da presente pesquisa é **caracterizar o reúso de dados, identificar os desafios que dele emergem e propor um modelo para explicar como os atores lidam com tais desafios na prática**. O trabalho empírico concentrou-se no setor financeiro e foi realizado por meio de estudos de casos múltiplos. A área financeira foi escolhida por seu histórico de intensivo uso de tecnologia e dados, o que a levou a visualizar o potencial do BD e adotá-lo rapidamente (Manyika et al., 2011), aumentando a probabilidade de se encontrar diversos casos concretos de reúso de dados em suas diferentes categorias.

A pesquisa foi dividida em três etapas, as quais deram origem aos três artigos apresentados nesta tese, que está organizada da seguinte forma:

Capítulo de introdução – refere-se ao presente capítulo, que visa delinear a estrutura da pesquisa, seus componentes e escolhas metodológicas e teóricas, assim como apresentar a organização desse documento, posicionando os artigos e seus objetivos específicos. Oferecer essa visão geral do estudo contribui para a compreensão de suas partes e da articulação entre elas.

Artigo I - *Big data* e reúso de dados: desafios e oportunidades – caracteriza o reúso de dados e propõe um esquema classificatório, refletindo sobre os potenciais níveis de desafio de suas

categorias propostas. Esse estudo foi baseado na revisão bibliográfica sobre o reúso de dados e na meta-análise de casos empíricos em que se identificou ocorrências dessa prática.

Artigo II - Reúso de dados na prática: Recontextualização de dados da telefonia celular para avaliação de crédito – propõe um modelo teórico para investigar casos empíricos de reúso de dados e o exercita em um estudo exploratório, analisando uma situação de reúso de dados de telefonia celular para avaliação de crédito. O caso investigado enquadra-se na categoria que apresenta maior potencial de desafios, de acordo com os critérios do artigo I.

Artigo III - Reúso de dados: uma jornada de mudanças rumo a novos paradigmas no setor financeiro – realiza um estudo de múltiplos casos, a partir do qual identifica-se os desafios do reúso de dados, reflete sobre suas dimensões legais, tecnológicas, intrínsecas, societárias e econômicas, e propõe um modelo para explicar como os atores envolvidos estão lidando com tais desafios na prática.

Capítulo de conclusão da tese – sumariza os resultados e contribuições da tese, trazendo não apenas um resumo das contribuições apresentadas em cada artigo, mas também aquelas que resultam do conjunto da pesquisa. Este sumário final fornece uma visão integrada dos resultados e reflexões gerados nas diferentes etapas do estudo, facilitando a compreensão da tese como um todo.

A seguir será apresentado o contexto e a motivação da pesquisa, seu posicionamento, metodologia e objetivos, assim como sua distribuição nas diferentes etapas do estudo e seus respectivos artigos. Logo após, a declaração das colaborações recebidas neste trabalho encerra o capítulo introdutório.

1.1 Contexto e motivação da pesquisa

Embora o reúso de dados não seja novidade, o contexto do BD o tornou mais frequente e relevante. Primeiramente, pelo aumento da oferta de dados – capturados de forma intensiva ao longo dos processos e rotinas de indivíduos e organizações, e gerados por pessoas ou objetos conectados (Abbasi, Sarker & Chiang, 2016; Baesens et al., 2016). Em segundo lugar, pela disponibilidade tecnológica de atrair e processar dados de diferentes fontes e formatos, que marca o fenômeno do BD, viabilizando e/ou facilitando o acesso e o cruzamento de informações de forma mais ampla (Abbasi, Sarker & Chiang, 2016; Baesens et al., 2016; Martin, 2015). Nesse cenário, o paradigma de análise de dados tem se tornado cada vez mais

marcado pela expectativa da atração de dados de diferentes origens e formas, com agilidade e produzindo análises profundas, usando técnicas e métodos nem sempre triviais (Abbasi, Sarker & Chiang, 2016; Baesens et al., 2016; Ketter et al., 2016). Tais possibilidades de cruzamento e conexão de informações também abriram caminhos para automação de processos e novas aplicações que seriam inviáveis se os dados tivessem de ser coletados diretamente (Custers & Bachlechner, 2017; Wasesa, Stam & Van Heck, 2017). Todas essas possibilidades e expectativas de benefícios explicam a disseminação acelerada do reúso de dados e a relevância de estudá-lo e discutí-lo nesse novo contexto em que ele envolve uma gama de dados e de operações sem precedentes.

Ainda que esteja cada vez mais presente na vida de organizações e indivíduos, o reúso de dados tem sido pouco debatido pela sociedade e pela academia (Custers & Bachlechner, 2017; Custers & Ursic, 2016; Grande et al., 2014; Newell, & Marabelli, 2015; Samarajiva & Lokanathan, 2016; Zuboff, 2015). Pesquisadores alertam para a importância da participação acadêmica nesse debate, contribuindo para a abordagem mais ampla do tema, abarcando dimensões tecnológicas, éticas, sociais, políticas, econômicas e legais (Custers & Bachlechner, 2017, Custers & Ursic, 2016; Grande, et al., 2014; Newell, & Marabelli, 2015; Samarajiva & Lokanathan, 2016, Yoon, 2015; Zuboff, 2015). Esse nível de reflexão pode contribuir para a identificação de implicações e desafios, assim como para a construção de caminhos para superá-los, alavancando os benefícios e mitigando os efeitos indesejados do reúso de dados. Por isso, torna-se cada vez mais relevante e urgente a realização de estudos que se engajem em tal missão, especialmente no campo de SI.

A abordagem sóciotécnica e sua capacidade de se conectar a elementos de outras áreas têm marcado presença nos estudos de SI há algumas décadas (Mumford, 2006). Esse perfil de pesquisa e o corpo de conhecimento que ela produz pode contribuir para a temática complexa e multifacetada do reúso de dados. Além disso, a escassa articulação da área nesse debate não representa apenas uma lacuna em sua literatura, mas também um problema para sua agenda de pesquisa. É cada vez mais comum, por exemplo, que organizações privadas e públicas armazenem dados sem um objetivo de uso pré-definido (Muller et al., 2016). Tal prática estimula o crescimento do reúso de dados para objetivos não previstos em sua geração, colocando em xeque modelos clássicos orientados pelos propósitos do uso para o planejamento, a qualidade e a gestão de informações e seus impactos organizacionais, por exemplo. Na esfera pública, essa prática levanta questões sobre ética, privacidade, direitos individuais e coletivos, como demonstrou o caso que expôs a abrangência da coleta de dados da Agência Nacional de

Segurança (NSA) dos Estados Unidos, cuja forma e justificativa ainda requerem um amplo debate na sociedade (Ball, 2013). Isso ilustra como o reuso de dados tem levantado novas questões e transformado a lógica de alguns temas que fizeram e/ou fazem parte da agenda de pesquisa de SI (Hirschheim & Klein, 2012; Hoppen & Meireles, 2005). Revisitá-los na perspectiva do cenário atual é fundamental para manter seu conhecimento aderente às práticas e questões contemporâneas.

Esta pode ser também uma grande oportunidade para os estudos de SI. Além de atualizar o corpo de conhecimento da área, explorar o reuso de dados pode produzir pesquisas empíricas, modelos teóricos e metodológicos que avancem a discussão sobre o tema em SI e iluminem os caminhos para sua exploração em outras áreas. A vocação do campo para integrar múltiplas dimensões sociais aos fenômenos tecnológicos é uma capacidade habilitadora de tal potência, colocando a área em uma posição privilegiada nessa arena.

Por fim, vale lembrar a relevância deste debate para a sociedade como um todo. Escândalos recentes de vazamento de dados e de mau uso de informações públicas e privadas trouxeram à tona o esgotamento dos modelos atuais de autorização, gestão e controle de uso dos dados. Casos como o da Cambridge Analytica, que expôs práticas comerciais obscuras e falhas na segurança envolvendo dados dos usuários do Facebook (Sherr, 2018), têm demonstrado os efeitos nocivos que a má utilização da informação pode trazer, não apenas para as empresas e os cidadãos envolvidos diretamente no evento, mas para toda a sociedade. Manipular o resultado eleitoral utilizando dados de alguns implica uma consequência que afeta todos. Nesse contexto, a pesquisa acadêmica sobre o reuso de dados pode colaborar para a construção de instrumentos que ajudem na compreensão desse fenômeno em sua amplitude, corroborando o debate informado sobre ele. Assim, o presente trabalho é movido pelo desejo de contribuir, ainda que modestamente, para essa importante missão. Ele busca dialogar com a lacuna de estudos sobre reuso de dados, realizando uma pesquisa empírica que utiliza lentes teóricas de SI para explorar o fenômeno em suas múltiplas dimensões, propor conceitos e revelar questões relevantes no cenário atual, que poderão ser exploradas em futuros estudos.

1.2 Posicionamento, metodologia e objetivos da pesquisa

Alinhada à ideia de que a realidade é socialmente construída a partir da interação viva e recursiva entre as pessoas e o mundo (Giddens, 1984; Sandberg, 2005) e à abordagem construtivista para geração de conhecimento, esta pesquisa posiciona-se no paradigma interpretativo (Cunliffe, 2011; Guba & Lincoln, 1994; Sandberg, 2005; Stake, 2008, Walsham, 1995). Tal posicionamento é fundamental para estabelecer a coerência metodológica da pesquisa (Pozzebon, Rodriguez & Petrini, 2014). Um trabalho interpretativo deve mergulhar no universo de seu objeto de estudo, buscando capturar as visões de seus atores, evitando impor a eles significados prévios e priorizando abordagens que permitam ao pesquisador estudar o fenômeno em seu ambiente natural (Burrell & Morgan, 1979; Orlikowski & Baroud, 1991; Walsham, 1995). Tal proposta, portanto, está coerente com a estratégia qualitativa adotada no trabalho, a qual foi operacionalizada por meio de estudos de casos, revisões bibliográficas e meta-análise da literatura. Para apresentar esses resultados, em alguns momentos foram utilizados gráficos e representações numéricas. Porém, isso não teve a intenção de propor análises quantitativas, foi apenas uma forma de organizar os conteúdos para compartilhá-los com o leitor.

A organização da pesquisa foi inspirada em Stake (2006), que propõe a análise de casos múltiplos para a investigação do que ele denomina *quintain*. “Um *quintain* é um objeto ou um fenômeno ou uma condição a ser estudada” (Stake, 2006, p. 6). O autor adota um termo incomum para destacar a complexidade da questão, maior do que as palavras como “fenômeno” ou “programa” poderiam expressar (Stake, 2006). Essa complexidade está ligada às diversas faces do objeto de estudo e ao entendimento profundo que sua pesquisa almeja. A abordagem de múltiplos casos pode ser uma boa estratégia para explorar as diferentes questões desse objeto multifacetado e sobre o qual busca-se uma compreensão ampla e profunda. Por isso, o autor sugere que o estudo comece pela delimitação do seu *quintain*. Pode-se dizer que o reúso de dados no setor financeiro é o *quintain* dessa tese.

Na proposta de Stake (2006), os casos e todos os demais procedimentos realizados no estudo devem colaborar para a compreensão do *quintain* (Stake, 2006). Essa tese foi produzida em três etapas que visam oferecer contribuições específicas para responder o objetivo central de seu *quintain*: caracterizar o reúso de dados no setor financeiro, identificar seus

desafios e propor um modelo para explicar como os atores estão lidando com tais dificuldades na prática. Em resumo, essas etapas e seus objetivos para elucidar a questão do *quintain* são:

– **Etapa 1 (artigo I):** seu objetivo é **caracterizar o reúso de dados**. Foi realizada uma revisão bibliográfica para identificar os conceitos de reúso de dados e uma meta-análise de casos empíricos, a partir da qual se propôs um esquema classificatório para o reúso. Com base nos conceitos e debates identificados na literatura, o artigo refletiu sobre os potenciais desafios das categorias de reúso propostas: repropósito ou recontextualização¹ de dados internos ou externos, sejam eles públicos ou privados. Corroborando a literatura sobre o tema, a pesquisa disserta sobre o impacto da distância entre a geração do dado e seu reúso. Considerando que essa distância – que pode ser espacial, temporal, cognitiva, etc. – pode dificultar o reúso dos dados, o artigo propõe distintos níveis de desafios potenciais para as categorias de reúso, alocando-as em grupos denominados direto, intermedial e extremo para distinguir esses patamares.

– **Etapa 2 (artigo II):** com a missão de **definir um modelo conceitual para abordar casos empíricos de reúso de dados**, o estudo propôs uma adaptação do modelo Multilevel Framework, incorporando um mapa visual (Langley, 1999) na representação do processo e os elementos do reúso de dados discutidos no artigo I para demonstrar o conteúdo que emerge do reúso de dados na prática. O caso retrata o processo da empresa Alpha para implantar uma solução de avaliação de crédito baseada em perfis comportamentais gerados a partir de dados da rede de celulares. Ao recontextualizar dados da telefonia para operações financeiras, o caso reflete o nível extremo de desafios do reúso de dados, de acordo com a proposta do artigo I. O modelo teórico apresentado contribuiu para a compreensão do caso e mostrou-se uma boa alternativa para explorar o reúso na prática.

– **Etapa 3 (artigo III):** esta etapa refere-se ao estudo de múltiplos casos do tema pesquisado. Ela visa **identificar os desafios do reúso de dados e propor um modelo para explicar como os atores estão lidando com tais desafios na prática**. Seu desafio inicial foi selecionar casos que, em conjunto, oferecessem uma visão ampla do *quintain*. A ideia, nesta abordagem, não é definir uma amostra representativa dos casos em termos estatísticos, mas selecionar aqueles que ofereçam diversidade de cenários e oportunidade de aprender sobre a complexidade e os contextos do *quintain* (Stake, 2006). Para Stake (2006), cada caso deve explorar questões específicas, também chamadas de *issues*. Os casos e suas questões não

¹ Conforme a denominação apresentada em detalhe no artigo 1, o repropósito refere-se ao reúso de dados para finalidade relacionada à temática do propósito original do dado. Já a recontextualização diz respeito ao reúso de dados para uma finalidade não relacionada à sua temática original.

precisam ser necessariamente comparados, mas devem revelar algo sobre o *quintain*. A seleção dos casos deve considerar, portanto, sua potencial contribuição para a compreensão do *quintain*, procurando formar um conjunto que ofereça maior cobertura das diferentes questões e contextos relevantes do *quintain* (Stake, 2006). Visando formar um grupo de casos que trouxesse uma visão ampla do reuso de dados no setor financeiro, o estudo utiliza a classificação e os conceitos propostos no artigo I, selecionando casos referentes a distintas categorias e níveis de desafios potenciais:

- a) **caso Banco Ômega**: investiga a situação do reuso de dados no Banco Ômega, especialmente o repropósito e a recontextualização de dados internos, explorando como a prática começou e se desenvolveu ao longo do tempo e identificando os desafios encontrados para o reuso de dados e o processo pelo qual o banco lidou ou está lidando com tais desafios. As categorias de reuso de dados envolvidas no caso pertencem ao grupo de desafios denominado “direto” no artigo I;
- b) **caso Bureau Sigma**: investiga a situação do reuso de dados no Bureau Sigma, especialmente o repropósito de dados externos públicos ou privados, abordando como a prática começou e se desenvolveu ao longo do tempo e identificando os desafios encontrados para o reuso de dados no Bureau e em seus clientes, assim como os processos pelos quais a empresa e seus usuários lidaram ou estão lidando com tais desafios. As categorias de reuso de dados envolvidas no caso pertencem ao grupo de desafios denominado “intermedial” no artigo I;
- c) **caso fintechs**: investiga a situação do reuso de dados em três *fintechs* cujo serviço está relacionado à essa área, especialmente a recontextualização de dados externos públicos ou privados, explorando como essa prática começou e se desenvolveu ao longo do tempo e identificando os desafios encontrados para o reuso de dados em seus mercados, assim como os processos pelos quais cada *fintech* lidou ou está lidando com tais desafios. As categorias de reuso de dados envolvidas no caso pertencem ao grupo de desafios denominado “extremo” no artigo I.

A figura 1.1 sumariza o *quintain* e as três etapas da pesquisa para abordá-lo.

O QUINTAIN: REUSO DE DADOS NO SETOR FINANCEIRO

Objetivo Principal da Pesquisa para o Quintain:

Caracterizar o reuso de dados no setor financeiro, identificar seus potenciais desafios e propor um modelo para explicar o processo pelo qual os atores envolvidos lidam com tais desafios.



Artigo I – Caracterizar o Reuso de Dados

Rev. Bibliográfica: definição conceitos e posicionamento literatura existente

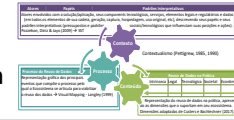


Meta-análise: identificação características, proposta esquema classificatório e níveis de potenciais desafios

		Fonte dos Dados		
		Dados Internos	Dados Públicos	Dados Privados Externos
Finalidade do Reuso dos Dados	Repropósito	●	●	●
	Recontextualização	●	●	●

Artigo II – Modelo p/ caso empírico

Proposta de modelo conceitual: apresentação de um modelo teórico para investigar casos empíricos de reuso de dados, exercitando-o em um caso exploratório.



Artigo III – Identificar desafios e explicar como atores lidam com eles na prática

Estudo de Casos Múltiplos: estudo de diferentes tipos de reuso de dados, conforme categorias e níveis de desafios discutidos no artigo I.

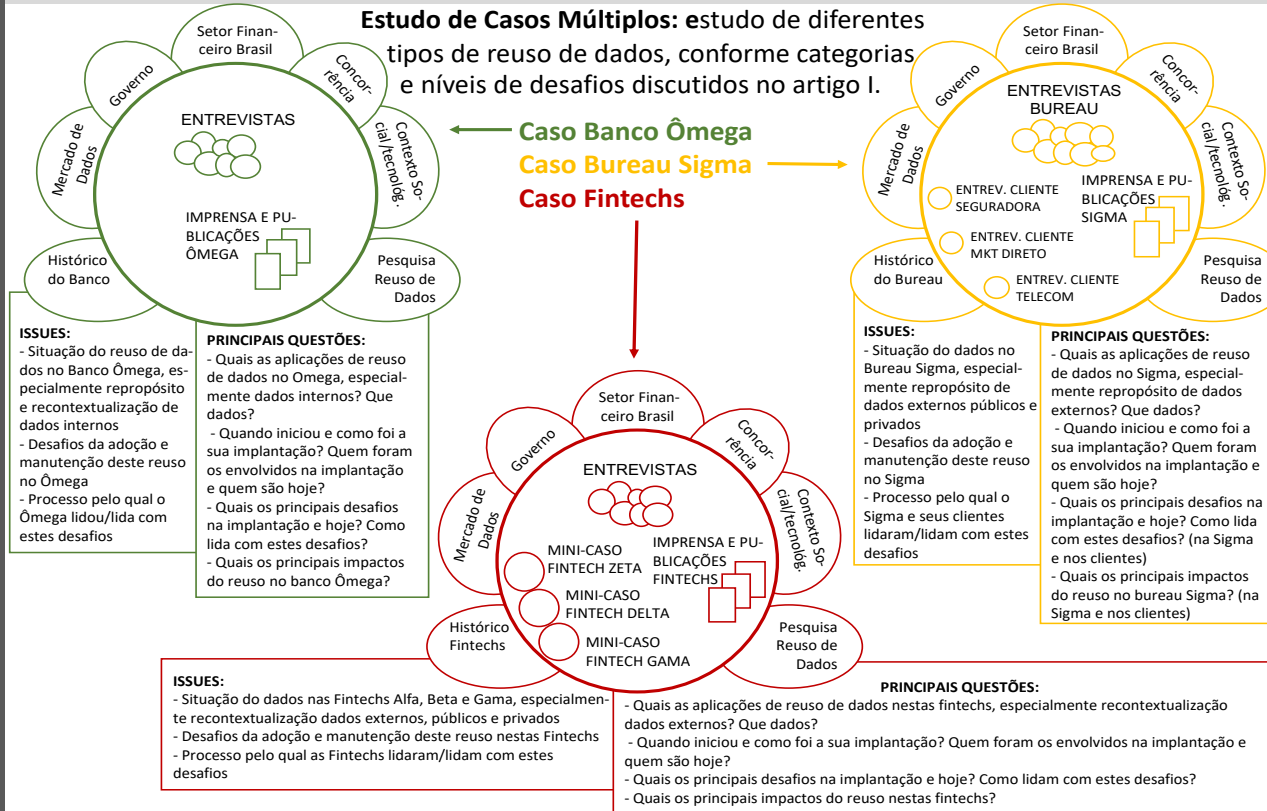


Figura 1.1: Abordagem da Pesquisa ao seu Quintain

Fonte: elaborado pela autora

Por fim, vale ressaltar que, na proposta Stake (2006), um estudo de casos múltiplos requer um amplo esforço de pesquisa, envolvendo uma equipe de pesquisadores dedicada por um longo período, dividindo atividades de planejamento, trabalho de campo, análise de dados, produção de diferentes relatórios, e análise dos casos individuais e *cross-case* (Stake, 2006). No entanto, o autor considera possível a realização desse tipo de trabalho por um pesquisador único e em um período mais curto, como no caso de um projeto de doutorado: “Estou confortável com a ideia que um bom estudo multicase poderia ser completado muito mais rápido (talvez em alguns meses e por um único pesquisador), mas eu esperaria múltiplos revisores do relatório da pesquisa para julgar sua meticulosidade e profundidade” (Stake, 2006, p. 17). Para o autor, um estudo de múltiplos casos pode ser bem realizado em uma pesquisa de doutorado, pois o doutorando contará com seu orientador e outros professores para ajudá-lo na interpretação das observações e no refinamento das questões da pesquisa, que devem ter o “tamanho” adequado para caber nas condições de seu projeto (Stake 2006). Tais cuidados foram adotados nesta pesquisa para mantê-la adequada a essa proposta metodológica. O detalhamento da metodologia, assim como as referências e o suporte teórico utilizado em cada etapa do estudo, consta nos respectivos artigos.

1.3 Declaração das colaborações

Este trabalho recebeu a colaboração de diversos professores e colegas. Algumas vezes, o apoio aconteceu na forma de inspiração e incentivo, que, embora valiosos, não poderiam ser enumerados ou descritos. Houve também contribuições diretas à pesquisa e ao desenvolvimento dos artigos, conforme descrito a seguir.

O primeiro artigo teve a contribuição do professor Dr. Eduardo Henrique Diniz, orientador desta tese. A coleta e análise de dados, assim como a classificação proposta e a produção do texto, foram realizadas pela autora do artigo sob a supervisão do professor. Ele contribuiu com a ideia original do artigo e sua organização, a indicação de literatura, validação da metodologia e a discussão dos achados da análise, que foi fundamental na composição do esquema classificatório proposto no trabalho. O professor também revisou o artigo, colaborando no texto desta primeira versão, que foi apresentada no Encontro da Associação Nacional de Pós-Graduação e Pesquisa em Administração (EnANPAD) de 2017 e teve sua coautoria.

A evolução do artigo contou com as contribuições recebidas na apresentação no EnANPAD e na banca de qualificação da tese, composta pelos professores Dr. Eduardo de Rezende Francisco e Dr. José Luiz Kugler. O trabalho também incorporou uma significativa contribuição na revisão bibliográfica sobre reúso de dados, a qual foi realizada no período de estágio doutoral na Hautes Études Commerciales (HEC), em Montreal, utilizando seu acervo. Além disso, as discussões com a professora Dra. Marlei Pozzebon, da HEC, sobre a metodologia e termos adotados também foram de significativa relevância para a produção da versão final do artigo I apresentada nesta tese.

Uma primeira versão do artigo II teve a contribuição dos professores Dr. Paulo Trombini de Souza Nascimento e Dr. Abraham Yu, da Faculdade de Economia e Administração da Universidade de São Paulo (FEA-USP). A coleta e análise dos dados, concepção do modelo e a produção do texto foram realizadas pela autora da tese. A primeira versão do artigo foi escrita como um trabalho da disciplina Gestão da Inovação, ministrada pelos professores Dr. Paulo e Dr. Abraham na FEA. Ela foi apresentada no Congresso Portland International Center for Management of Engineering and Technology (PICMET) em 2016, no Havaí. Os professores Dr. Paulo e Dr. Abraham contribuíram com a validação da ideia do estudo, a indicação da bibliografia relacionada à gestão de inovação e na discussão dos resultados sob a perspectiva das teorias da inovação, que foi mais explorada nesta versão inicial do artigo. Após o Congresso, o trabalho foi revisitado, dando origem à nova versão que foi apresentada na qualificação da tese. Após as contribuições da banca de qualificação e as novas ideias e colaborações que surgiram durante o doutorado-sanduíche no Canadá, nasceu a terceira versão do artigo que é apresentada nessa tese. Nela foi realizada uma mudança na abordagem teórica, incorporando a ideia do mapa visual para representar o processo e os conceitos do reúso de dados, apresentados no artigo I para mostrar o conteúdo no Multilevel Framework. Também foi feita uma nova revisão bibliográfica e análise dos dados, reposicionando as discussões do caso, sob a orientação do professor Dr. Diniz. Quando esta nova versão for publicada não contará mais com as coautorias dos professores Dr. Paulo e Dr. Abraham, uma vez que suas contribuições constavam no debate sobre inovação, as quais estavam na versão inicial do artigo e foram eliminadas nesta tese.

O artigo III é totalmente inédito e começou a ser produzido juntamente com o estudo de casos múltiplos. Suas ideias iniciais já haviam sido esboçadas quando a autora do artigo foi para a HEC, onde ficou junto ao grupo de doutorandos da universidade e pesquisadores visitantes. As atividades desse período trouxeram profundas transformações para a tese. As discussões com a professora Marlei e com outros professores da HEC, assim como a

participação no *atelier* de pesquisa, workshops e apresentações de artigos de professores e pesquisadores convidados pela universidade, trouxeram novas ideias para a estrutura metodológica e teórica da tese. Além das mudanças ocorridas nos artigos I e II e a organização do caso apresentado no artigo III, os procedimentos de análise dos dados e reportes dos resultados do estudo de casos múltiplos também foram concebidos nesse período, beneficiando-se desses novos conhecimentos e influências.

Referências bibliográficas

- Abbasi, A., Sarker, S., & Chiang, R. H. (2016). Big data research in information systems: Toward an inclusive research agenda. *Journal of the Association for Information Systems*, 17(2).
- Baesens, B., Bapna, R., Marsden, J. R., Vanthienen, J., & Zhao, J. L. (2016). *Transformational Issues of Big Data and Analytics in Networked Business*. *MIS quarterly*, 38(2), 629-631.
- Ball, J. (2013). NSA stores metadata of millions of web users for up to a year, secret files show. *The Guardian*. <https://www.theguardian.com/world/2013/sep/30/nsa-americans-metadata-year-documents>
- Boyd, D., & Crawford, K. (2012). *Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon*. *Information, Communication & Society*, 15, 662-679.
- Borgman, C. L. (2015). *Big data, little data, no data: Scholarship in the networked world*. MIT press.
- Burrell, G., & Morgan, G. (1979). *Social paradigms and organizational analysis. Elements of the sociology of corporate life*.
- Cunliffe, A. L. (2011). Crafting qualitative research: Morgan and Smircich 30 years on. *Organizational Research Methods*, 14(4), 647-673.
- Curty, R. G. (2015). *Beyond "Data Thrifting": An Investigation of Factors Influencing Research Data Reuse In The Social Sciences* (3713677). Available from ProQuest Dissertations & Theses Global. (1698463092).
- Custers, B., & Bachlechner, D. (2017). *Advancing the EU Data Economy: Conditions for Realizing the Full Potential of Data Reuse*. *Information Polity*, 22(4), 291.

- Custers, B., & Ursic, H. (2016). *Big Data and Data Reuse: A Taxonomy of Data Reuse for Balancing Big Data Benefits and Personal Data Protection*. *International Data Privacy Law*, 6(1), 4-15.
- Giddens, A. (1984). *The constitution of society*. Cambridge: Polity.
- Goes, P. B. (2015). *Editor's Comments: Inflection Point: Looking Back or Looking Forward?*. *MIS Quarterly*, 39(3), iii-viii.
- Grande, D., Mitra, N., Shah, A., Wan, F., & Asch, D. A. (2014). *The Importance of Purpose: Moving Beyond Consent in the Societal Use of Personal Health Information*. *Annals of internal medicine*, 161(12), 855-862.
- Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. *Handbook of qualitative research*, 2(163-194), 105.
- Hirschheim, R., & Klein, H. K. (2012). A glorious and not-so-short history of the information systems field. *Journal of the Association for Information Systems*, 13(4), 188.
- Hoppen, N., & Meirelles, F. (2005). Sistemas de informação: um panorama da pesquisa científica entre 1990 e 2003. *RAE-revista de administração de empresas*, 45(1), 24-35.
- Ketter, W., Peters, M., Collins, J., & Gupta, A. (2016). Competitive Benchmarking: An IS Research Approach To Address Wicked Problems With Big Data And Analytics. *MIS Quarterly*, 40(4), 1057.
- Langley, A. (1999). Strategies for theorizing from process data. *Academy of Management review*, 24(4), 691-710.
- Letouzé, E. (2012). *Big Data for Development: Challenges & Opportunities*. UN Global Pulse.
- Luvizan, S., Meirelles, F., & Diniz, E. (2015). *Big Data: Evolução Das Publicações E Oportunidades De Pesquisa*. *Revista Eletrônica de Sistemas de Informação*, 14(3).
- Manyika, J et al.,(2011). *Big Data: The Next Frontier For Innovation, Competition, And Productivity*. McKinsey Global Institute.
- Martin, K. (2015). *Ethical Issues In The Big Data Industry*. *MIS Quarterly Executive*, 14(2), 67-85.
- Müller, O., Junglas, I., Brocke, J. V., & Debortoli, S. (2016). Utilizing big data analytics for information systems research: challenges, promises and guidelines. *European Journal of Information Systems*, 25(4), 289-302.

Mumford, E. (2006). The story of socio-technical design: Reflections on its successes, failures and potential. *Information Systems Journal*, 16(4), 317-342.

Newell, S., & Marabelli, M. (2015). *Strategic opportunities (and challenges) of algorithmic decision-making: A call for action on the long-term societal effects of 'datification'*. The Journal of Strategic Information Systems, 24(1), 3-14.

Orlikowski, W. J., & Baroudi, J. J. (1991). Studying information technology in organizations: Research approaches and assumptions. *Information systems research*, 2(1), 1-28.

Pozzebon, M., Rodriguez, C., & Petrini, M. (2014). *Dialogical Principles For Qualitative Inquiry: A Nonfoundational Path*. International Journal of Qualitative Methods, 13, 293-317.

Samarajiva, R., & Lokanathan, S. (2016). *Using Behavioral Big Data For Public Purposes: Exploring Frontier Issues Of An Emerging Policy Arena*.

Sandberg, J. (2005). How do we justify knowledge produced within interpretive approaches? *Organizational research methods*, 8(1), 41-68.

Sherr, I. (2018). Facebook, Cambridge Analytica and data mining: What you need to know. CBNET. In: <https://www.cnet.com/news/facebook-cambridge-analytica-data-mining-and-trump-what-you-need-to-know/>

Stake, R. E. (2006). *Multiple Case Study Analysis*. Guilford Press.

Stake, R. E. (2008). Qualitative case studies. In N. K. Denzin & Y. S. Lincoln (Eds.), *Strategies of qualitative inquiry* (pp. 119-149). Thousand Oaks, CA, US: Sage Publications, Inc.

Walsham, G. (1995). Interpretive case studies in IS research: nature and method. *European Journal of information systems*, 4(2), 74-81.

Wasesa, M., Stam, A., & van Heck, E. (2017). *The seaport service rate prediction system: Using drayage truck trajectory data to predict seaport service rates*. Decision Support Systems, 95, 37-48.

Yoon, A. (2015). *Data reuse and users' trust judgments: Toward trusted data curation*. Available from ProQuest Central; ProQuest Dissertations & Theses Global.

Zimmerman, A. S. (2003). *Data Sharing And Secondary Use Of Scientific Data: Experiences Of Ecologists*. Available from ProQuest Dissertations & Theses Global: Science & Technology; ProQuest Dissertations & Theses Global: Social Sciences.

Zimmerman, Ann. 2008. "New Knowledge from Old Data: The Role of Standards in the Sharing and Reuse of Ecological Data." *Science, Technology & Human Values* 631-652.

Zuboff, S. (2015). Big other: surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology*, 30(1), 75-89.

2 ARTIGO I – *BIG DATA* E REÚSO DE DADOS: DESAFIOS E OPORTUNIDADES²

2.1 Resumo

A digitalização de atividades rotineiras e o intenso uso da tecnologia por empresas e indivíduos, gerando um imenso e progressivo volume de dados, deu origem ao fenômeno denominado *big data* (BD). Diante do avanço das ferramentas para processar essa miríade de dados e da busca pela criação de valor a partir deles, é cada vez mais frequente que os dados gerados para uma finalidade sejam, posteriormente, utilizados para outros fins. Esse reuso dos dados pode trazer grandes benefícios, mas também está cercado de desafios relevantes e pouco explorados pela literatura acadêmica. Visando colaborar para essa discussão, o presente artigo apresenta uma revisão bibliográfica sobre o reuso de dados, iluminando conceitos, posicionando estudos e discutindo desafios e implicações que ilustram a relevância do engajamento da área de Sistemas de Informação (SI) neste debate. Realizou-se também uma meta-análise dos casos empíricos de reuso de dados identificados nas publicações relacionadas a BD em periódicos de SI. A partir de sua análise, propôs-se um esquema classificatório baseado na fonte dos dados (públicas, privadas internas ou externas) e na finalidade do reuso (repropósito – reuso relacionado aos fins originais - ou recontextualização – reuso não relacionado aos fins originais dos dados). A discussão dos potenciais desafios do reuso de dados em cada categoria sugere que, embora todas enfrentem questões de ordem tecnológica, social, legal e econômica, sua intensidade pode aumentar à medida que a distância entre a geração e o reuso dos dados cresce. Tal distância pode implicar maior diversidade do ecossistema no qual se dá o reuso, envolvendo atores com diferentes valores, interesses e objetivos. Por isso, torna-se relevante reconhecer as diversas situações de reuso de dados para focalizar melhor as questões pertinentes a cada uma. Para destacar os vários níveis de potenciais desafios, as categorias da matriz foram agrupadas em: direto, intermedial e extremo. As reflexões deste trabalho pretendem inspirar o debate sobre o reuso de dados em SI, colaborando para potencializar benefícios e mitigar riscos e consequências indesejadas dessa prática promissora.

Palavras-chave: analytics; *big data*; reuso de dados; uso secundário de dados;

² Versão original deste artigo foi apresentada no congresso Enanpad 2017: LUVIZAN, S. & DINIZ, E. (2017). *Big Data e o Uso Secundário de Dados: Desafios para a Qualidade de Dados e a Inovação*. In XLI Encontro da ANPAD – Enanpad. São Paulo, SP, Brasil.

2.2 Introdução

O desenvolvimento das tecnologias, de suas aplicações e da geração de dados deu origem ao fenômeno denominado *big data* (BD). Nuvens, redes sociais, internet das coisas, tecnologias móveis e de vestir, entre outras tantas ferramentas ligadas ao universo do BD já são uma realidade no dia a dia de indivíduos e organizações (Baesens, Bapna, Marsden, Vanthienen & Zhao, 2016; Letouzé, 2012; Martin, 2015; Manyika et al., 2011; Mcaffee & Brynjolfsson, 2012; Newell, & Marabelli, 2015). Portanto, a sociedade já convive, em algum nível, com seus impactos positivos e negativos, tornando urgente discuti-los e encontrar caminhos para alcançar seus ganhos e evitar suas consequências indesejadas (Baesens et al., 2014; Boyd & Crawford, 2012; Newell, & Marabelli, 2015; Park & Leydesdorff, 2013; Zuboff, 2015).

O uso das tecnologias atuais na rotina diária de organizações e indivíduos está transformando os atores sociais em “geradores de dados ambulantes” (Mcaffee & Brynjolfsson, 2012, P. 5). A grande questão agora é como essa imensa e crescente quantidade de dados será utilizada para gerar conhecimento e benefícios concretos, e quem colherá esses frutos (Newell & Marabelli 2015; Samarajiva & Lokanathan, 2016; Zuboff, 2015). Cada vez mais os dados existentes são usados para fins diferentes daqueles para os quais foram gerados (Martin, 2015; Newell & Marabelli, 2015), o que chamamos de reúso de dados (Faniel & Jacobensen, 2010; Yoon, 2015; Zimmerman, 2003). Esses novos usos nem sempre são conhecidos pelos geradores dos dados e seus algoritmos e métodos de processamento nem sempre são conhecidos pelos usuários que realizam o novo uso (que podem ser ou não os criadores dos dados) (Martin, 2015; Newell & Marabelli, 2015). Isso oculta e agrava as tensões entre os interesses dos grupos envolvidos nesse processo (Martin, 2015; Newell & Marabelli, 2015; Samarajiva & Lokanathan, 2016; Zuboff, 2015). Ao mesmo tempo em que apontam para as oportunidades de desenvolvimento que o reúso de dados pode proporcionar, alguns estudiosos alertam para suas implicações e riscos, chamando a atenção para a importância da participação acadêmica nesse debate ainda pouco explorado (Boyd & Crawford, 2012; Newell & Marabelli 2015; Samarajiva & Lokanathan, 2016; Zuboff, 2015).

O objetivo deste trabalho é caracterizar o reúso de dados e os potenciais desafios da prática nas diferentes situações em que ela ocorre. Diversos estudos vêm chamando a atenção para a necessidade de pesquisas que abordem esse tema com perspectivas de análise mais amplas, que incluam suas dimensões tecnológicas, sociais, legais e econômicas (Boyd & Crawford, 2012; Custers & Bachlechner, 2017, Newell & Marabelli 2015). É importante

reconhecer, porém, a dificuldade da empreitada diante de um fenômeno que pode ocorrer de maneira tão diversa (Clarke, 2016). Encontrar uma forma de segmentar tal fenômeno pode ser uma alternativa para viabilizar estudos que focalizem tipos específicos de ocorrências e consigam explorar todas as suas dimensões, permitindo o aprofundamento das discussões. Ao reconhecer o posicionamento dessas partes, é possível conectar suas questões ao todo, ampliando a visão e o conhecimento geral sobre o reúso de dados.

Ao apresentar a produção acadêmica sobre o reúso de dados, o artigo também traz informações que podem estimular o desenvolvimento dessa temática em diferentes áreas. Especialmente no campo de SI, o trabalho demonstra a necessidade de mais engajamento, problematizando a baixa participação da área na pesquisa desse assunto que impacta tantos outros que fazem parte de sua agenda de pesquisa tradicional. Alguns desses temas e seus desafios na perspectiva do reúso de dados são abordados para inspirar o desenvolvimento de novos trabalhos na área. Além disso, o atual artigo também ressalta a importância da participação ativa de SI no debate devido à capacidade da área e de suas teorias explorarem temas multifacetados envolvendo tecnologia, evidenciando seu potencial de contribuir para a evolução de conhecimentos relevantes para a teoria e a prática.

Primeiramente, este trabalho apresenta o cenário informacional do BD e a revisão bibliográfica sobre o reúso de dados, expondo seus conceitos e os temas abordados nos estudos das diversas áreas que se engajaram no debate. Também discute as implicações desse processo que poderiam ser exploradas pelos estudos acadêmicos, especialmente da área de SI. Em seguida, apresenta a metodologia e os resultados de uma meta-análise de casos empíricos de reúso de dados encontrados nos artigos acadêmicos de BD em periódicos de SI, analisando suas características e propondo um esquema classificatório para o reúso de dados. Faz, então, uma reflexão sobre os potenciais desafios de cada categoria descrita no esquema, de acordo com as dimensões envolvidas nesse fenômeno, segundo a literatura pesquisada. O artigo é encerrado com a apresentação de conclusões e sugestões de novos estudos que possam contribuir para o amadurecimento do debate do tema.

2.3 *Big data* e seus tantos “v’s”

A atenção ao BD vem crescendo significativamente na academia e na prática nos últimos anos (Luvizan, Meirelles & Diniz, 2015; Park & Leydesdorff, 2013). Embora a

academia tenha se engajado no tema, com um significativo aumento no volume de publicações, diversos pesquisadores ainda destacam a necessidade de estudos sobre várias questões e aspectos relevantes do BD (Abbasi, Sarker & Chiang, 2016; Boyd & Crawford, 2012; Martin, 2015; Muller, Jungas, Brocke & Debortoli, 2016; Newell & Marabelli, 2015; Park & Leydesdorff, 2013). Entre os trabalhos existentes, o próprio conceito de BD não é unânime, tornando seu perímetro confuso e sua abordagem acadêmica ainda mais desafiadora (Luvizan, Meirelles & Diniz, 2015; Martin, 2015). Há diversas definições de seu conceito: são ferramentas capazes de manejar volumes de dados superiores aos limites das ferramentas tradicionais; é a capacidade analítica de construir inteligência a partir de grandes volumes de dados; é um novo cenário em que há disponibilidade de dados e de ferramentas para manejá-los sem precedentes; entre outras (Demchenko et al., 2013; Letouzé, 2012; Manyika et al., 2011; Martin, 2015; Park & Leydesdorff, 2013). Martin (2015) propõem abordar o BD como uma indústria para destacar a complexidade do ecossistema no qual ele está inserido. Tal ecossistema envolve não apenas as ferramentas em si, mas também aspectos relacionados à sua infraestrutura, aos dados e à obtenção destes, bem como aos impactos para aqueles que as implementam e as usam — seus parceiros e modelos de negócios, entre outros atores e elementos do cenário geral (Martin, 2015). Sem entrar no mérito da terminologia ou na discussão conceitual, que não é o foco deste artigo, será adotada esta última abordagem do BD, cujo enfoque amplo atende melhor aos objetivos do estudo e aos interesses de pesquisa da autora.

O BD é marcado pelos desafios que, inicialmente, ficaram conhecidos como “3 v’s” - volume, velocidade e variedade - enfatizando as diversas questões para viabilizar o manejo dos imensos volumes de dados, origens e formatos variados, com grande agilidade (Martin, 2015; McAfee & Brynjolfsson, 2012). Posteriormente, o aumento da compreensão sobre os desafios do BD na prática agregou duas novas questões: veracidade e valor, alertando para a criticidade da qualidade e da confiabilidade da informação nesse novo cenário e para o desafio da geração de valor por meio do uso dos dados (Baesens Et Al., 2016; Demchenko Et Al., 2013). Assim, os desafios do BD passaram então a ser conhecidos como “5 v’s”. Atualmente, alguns autores falam em “7 v’s”, adicionando variabilidade e visualização para destacar o desafio da dinâmica dos dados e da representação dos grandes volumes destes (Seddon & Currie, 2017; Uddin & Gupta, 2014). Outros autores falam em “10 v’s”, agregando a volatilidade, vulnerabilidade e validação dos dados (Mohd & Biswas, 2018). Enfim, no decorrer de suas jornadas pelo universo do BD, pesquisadores e praticantes têm gerado uma coleção de “v’s”. Em buscas na internet é possível encontrar mais deles e em diferentes arranjos. Este trabalho não pretende se aprofundar na discussão de tais rótulos, mas reconhece que sua

proliferação demonstra o avanço do reconhecimento dos desafios do BD, sua pluralidade e a ânsia de pesquisadores e praticantes de encontrar uma forma de estruturar o discurso sobre tais desafios. Com tantos “v`s”, a missão parece não ser fácil!

As expectativas sobre investimentos e benefícios que o BD pode trazer também têm gerado grande discussão e polêmica. Entusiastas tentam mostrar seu potencial de trazer novas soluções para os problemas da humanidade, alavancar o desenvolvimento e promover negócios em diversas áreas (Ketter et al., 2016; Letouzé, 2012; Martin, 2015; Smolan, 2013). No entanto, alguns estudos alertam que ele também pode trazer uma série de riscos e problemas sociais e éticos (Boyd & Crawford, 2012; Clarke, 2016; Martin, 2015; Newell & Marabelli 2015; Samarajiva & Lokanathan, 2016; Zuboff, 2015). A possibilidade de gerar, armazenar e processar dados muito diversos e volumosos, de forma cada vez mais ágil e sofisticada, permite criar mais conhecimento, aumentar a capacidade preditiva, modelar as ofertas às necessidades de grupos/indivíduos e criar produtos, serviços e negócios inovadores (Ketter et al, 2016; Letouzé, 2012; Martin, 2015; Mcaffee & Brynjolfsson, 2012; Newell & Marabelli 2015; Wasesa, Stam, & Van Heck, 2017). Por outro lado, isso é uma constante ameaça a direitos individuais e coletivos, podendo ser fonte de violação de privacidade e segurança da informação, discriminação, manipulação da informação, desequilíbrio de oportunidades e distorção nas relações de poder (Boyd & Crawford, 2012; Clarke, 2016; Custers & Bachlechner, 2017; Newell & Marabelli 2015; Samarajiva & Lokanathan, 2016; Zuboff, 2015).

No atual cenário, em que já produzimos constantemente uma enorme quantidade de dados, as atenções têm se voltado para as oportunidades de avanço e inovação a partir do uso das informações existentes, encontrando novas aplicações para elas (Ketter et al., 2016; Martin, 2015; Mcaffee & Brynjolfsson, 2012; Newell & Marabelli 2015). Ainda que o reuso de dados seja cada vez mais praticado e que alguns autores estejam alertando sobre suas implicações há algum tempo (Boyd & Crawford, 2012), esse tema ainda é pouco explorado na literatura acadêmica (Clarke, 2016; Custers & Bachlechner, 2017; Custers & Ursic, 2016; Newell & Marabelli 2015; Samarajiva & Lokanathan, 2016; Yoon, 2015). Visando contribuir para o debate, a próxima seção apresenta conceitos e estudos sobre o reuso de dados, trazendo um panorama sobre a produção acadêmica e as implicações relevantes que poderiam inspirar novos estudos, especialmente em SI.

2.4 Reúso de dados

2.4.1 O que é reuso de dados e porque isso importa

Poucos estudos definem formalmente o conceito de reúso de dados (Faniel & Jacobsen, 2010; Yoon, 2015). O próprio termo não é unânime e alguns estudos utilizam a expressão “uso secundário de dados” para se referir ao fenômeno (Curty, 2015; Zimmerman, 2003). Os dois termos são normalmente utilizados de forma intercambiável (Zimmerman, 2003), não havendo na literatura uma definição que diferencie seus racionais ou que estabeleça uma preferência entre eles (Curty, 2015). O presente trabalho adota o termo “reúso de dados”, pois este apareceu mais vezes na revisão bibliográfica, sugerindo que a expressão tem se tornado mais popular entre os pesquisadores.

Geralmente, as pesquisas que mencionam o reúso de dados referem-se a ele como a utilização de dados para fins distintos daqueles para os quais foram gerados (Faniel & Jacobsen, 2010; Solove, 2008, Yoon, 2015) e esse foi o conceito adotado neste estudo. Embora não seja um tema novo e nem esteja presente exclusivamente no contexto do BD, é no cenário informacional que sua discussão tem se tornado mais relevante (Borgman, 2012; Curty, 2015; Custers & Bachlechner, 2017; Custers & Ursic, 2016; Safran et al., 2007). O volume e a diversidade de dados existentes, a abundante oferta de tecnologias para acesso e processamento das informações, e as expectativas crescentes de geração de conhecimento e valor a partir destas estimulam a busca por novos usos dos dados (Ketter et al., 2016; Safran et al., 2007). Para Custers e Ursic (2016), “entre os muitos desafios que o BD traz, o reúso de dados é um dos mais urgentes” (p. 5).

De um lado, essa urgência é fomentada pela expectativa de benefícios que o reúso de dados pode trazer ao cenário do BD. Considerando a imensa quantidade de informações existentes, hoje seu reúso tem um potencial sem precedentes de viabilizar soluções e negócios inovadores e trazer significativa economia de tempo, dinheiro e esforços em processos tradicionais (Curty, 2015; Custers & Bachlechner, 2017, Ketter et al., 2016; Martin, 2015). A atual disponibilidade de dados pode permitir soluções que seriam inviáveis se os dados tivessem de ser coletados diretamente (Custers & Bachlechner, 2017; Wasesa, Stam & Van Heck, 2017). Nesse contexto, a tendência é que cada vez mais as organizações façam reflexões do tipo *make or buy* ao criar novas aplicações e negócios, ponderando quando fará sentido gerar novos dados

ou capturar dados existentes dentro ou fora de seus domínios (Custers & Bachlechner, 2017). Além disso, em algumas situações, coletar os dados novamente ao invés de reutilizá-los de bases existentes pode incomodar o consumidor final de um serviço ou produto, o que poderia ser evitado (Custers & Ursic, 2016). O reúso de dados em tais situações é visto como uma alternativa para melhorar a experiência do cliente (Custers & Ursic, 2016). Finalmente, é uma potente ferramenta na esfera pública, viabilizando mecanismos alternativos de participação social, eficiência de serviços, transparência e gestão pública (Martin, 2015).

Outra razão para a urgência de abordar o desafio do reúso de dados é a sua complexidade. À medida que os casos práticos e a investigação teórica avançam, mais se compreende que, ao permitir aplicações mais sofisticadas e inovadoras por meio do reúso de dados, o cenário tecnológico e de dados da era do BD eleva seus desafios e implicações a patamares igualmente grandiosos. Potenciais impactos e efeitos colaterais indesejáveis, conflitos éticos, mudanças na lógica dos processos, demandas legais e financeiras são alguns exemplos dos muitos impactos que emergem do reúso de dados e do leque de possibilidades que se abre a partir dele (Clarke, 2016; Custers & Ursic, 2016; Martin, 2015; Newell & Marabelli, 2015; Samarajiva & Lokanathan, 2016). Cada vez mais os estudos relatam que não basta ter dados e tecnologia disponíveis para o reúso (Barbera, 2013; Borgman et al., 2015; Curty, 2015; Poole, 2015). Embora as ferramentas de dados tenham avançado muito e sua oferta também, reutilizá-los ainda não é trivial, de modo que pode demandar recursos metodológicos e revisão de práticas operacionais em diversas áreas (Barbera, 2013; Borgman et al., 2015; Curty, 2015; Ketter et al., 2016; Poole, 2015; Muller et al., 2016). Assim, o reúso que poderia colaborar com a solução de problemas importantes pode não acontecer diante dessas barreiras, que precisam ser discutidas e transpostas (Barbera, 2013; Borgman et al., 2015; Curty, 2015; Ketter et al., 2016; Poole, 2015; Muller et al., 2016). Abre-se ainda grandes discussões sobre privacidade, novos modelos de negócios ou revisão de modelos existentes, direitos, deveres e papéis dos atores envolvidos, lembrando que a maior oferta de dados e de reúso significa mais diversidade de tais atores, o que agrega uma nova camada de complexidade à discussão (Custers & Ursic, 2016). Pesquisadores e praticantes concordam que as questões, além de numerosas, são profundas e multidisciplinares (Barbera, 2013; Borgman et al., 2015; Curty, 2015; Ketter et al., 2016; Poole, 2015; Muller et al., 2016).

Convencidos dessa pluralidade, Custers e Bachlechner (2017) propõem que as condições para o reúso de dados sejam analisadas em uma perspectiva ampla, que considere quatro dimensões:

Legal – refere-se às barreiras e aos dispositivos legais habilitadores do reúso de dados. Atualmente, os principais desafios estão relacionados à privacidade, proteção de dados e propriedade intelectual (Custers & Bachlechner, 2017). Diversas questões desdobram-se a partir desses pontos e podem, ao longo de sua jornada evolutiva, somar-se a essas temáticas principais (Custers & Bachlechner, 2017; Newell & Marabelli 2015; Samarajiva & Lokanathan, 2016).

Tecnológica – relacionada à viabilidade técnica para operacionalizar o reúso, essa dimensão preocupa-se fortemente com a interoperabilidade (capacidade de interação entre os componentes técnicos envolvidos) e portabilidade (capacidade de compreender os dados em diferentes ambientes) (Custers & Bachlechner, 2017). Tais questões estão frequentemente ligadas a padrões e formatos de dados, sendo temas muito presentes na discussão dessa perspectiva (Borgman *et al.*, 2015; Custers & Bachlechner, 2017; Curty, 2015; Demchenko *et al.*, 2013, Poole, 2015; Zimmerman, 2008).

Societal – refere-se à aceitação do reúso de dados pela sociedade (Custers & Bachlechner, 2017). A informação ou conscientização dos envolvidos sobre o novo uso é considerada um dos pilares essenciais desta dimensão. Ainda que não seja suficiente para validá-lo, a transparência sobre o reúso é o ponto de partida para a formação da opinião dos atores sociais (Custers & Bachlechner, 2017). Tal opinião também será influenciada pelas expectativas que eles nutrem sobre o reúso e por padrões culturais, éticos e de práticas socialmente aceitas (Custers & Bachlechner, 2017; Newell & Marabelli 2015; Samarajiva & Lokanathan, 2016; Zuboff, 2015).

Econômico – reflete sobre a capacidade do reúso de dados gerar um valor que efetivamente compense o esforço e os recursos que ele requer (Custers & Bachlechner, 2017). Nesta dimensão residem algumas armadilhas. Primeiramente, o valor do reúso nem sempre é monetário, podendo incluir benefícios de desenvolvimento humano, bem-estar, saúde, conhecimento, entre outros (Custers & Bachlechner, 2017; Malin, Karp & Scheuermann, 2010, Safra *et al.*, 2007, Zimmerman, 2008). Além disso, os benefícios do reúso e os recursos para implementá-lo podem recair sobre os diferentes atores envolvidos, ou seja, aquele que arca com o ônus nem sempre é o que colhe os frutos do reúso (Poole, 2015). Por isso, essa perspectiva também deve abordar a construção do ecossistema favorável para sustentar o reúso e do modelo de negócios capaz de entregar valor, monetário ou não, de forma sustentável (Custers & Bachlechner, 2017; Malin *et al.*, 2010, Poole, 2015).

Para Custers e Bachlechner (2017), o reúso de dados somente é viável se for compatível com as questões que emergem dele nas quatro perspectivas citadas. A figura 2.1 apresenta alguns aspectos críticos de tais dimensões.

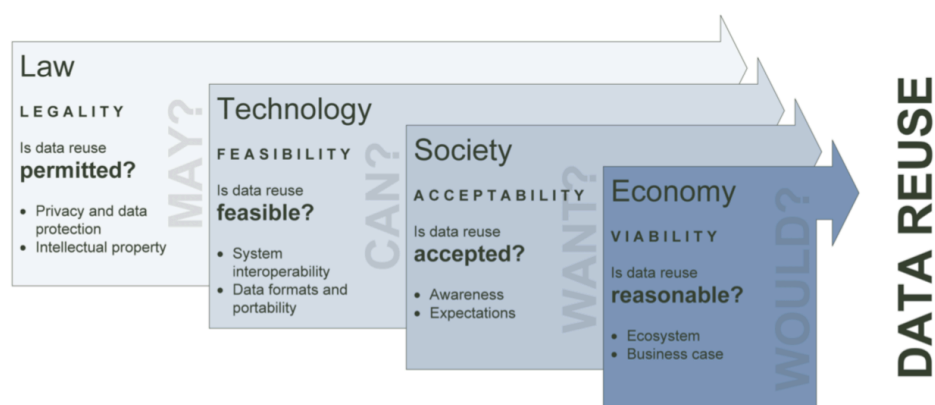


Figura 2.1 - Questões ilustrativas e aspectos críticos das quatro dimensões do reúso de dados
Fonte: Custers & Bachlechner, 2017

As diversas questões mencionadas nas pesquisas que exploram o reúso de dados corroboram a visão de Custers e Bachlechner (2017) sobre a diversidade de dimensões envolvidas nesse fenômeno (ex. Curty, 2015; Zimmerman, 2003). Também é importante destacar a interação entre tais perspectivas: “por exemplo, a lei pode regular tecnologias, mas tecnologias podem também ser desenhadas para reforçar ou impulsionar certos comportamentos” (Custers & Bachlechner, 2017, p. 296). Além disso, mesmo que o reúso não seja inviabilizado, seguir adiante ignorando pontos mal resolvidos em uma das dimensões pode gerar efeitos colaterais desfavoráveis para algum participante ou elemento do ecossistema no qual se dá o reúso, o que pode trazer desequilíbrios a esse ecossistema e ameaçar sua sustentabilidade no longo prazo (Custers & Bachlechner, 2017).

Os possíveis conflitos de interesses entre os atores envolvidos podem inviabilizar o reúso de dados se não houver alternativas para conciliá-los. Por isso, sugiro incluir uma dimensão “intrínseca”, dedicada a refletir sobre a dinâmica entre tais atores. Essa proposta é uma sugestão a ser explorada em futuros estudos.

Diante da diversidade de questões apresentadas, torna-se relevante explorar esse fenômeno em sua amplitude e os variados pontos que dele emergem. Compreender os conceitos, as implicações e os desafios envolvidos no reúso de dados é um passo importante para avançar na jornada rumo a seus benefícios (Faniel & Jacobsen, 2010; Yoon, 2015). Visando identificar tal abordagem na pesquisa acadêmica, o presente trabalho apresenta uma análise da produção sobre esse tema.

2.4.2 Pesquisa acadêmica sobre reuso de dados

Apesar de reconhecer o grande potencial do BD e a importância do reuso de dados para alcançá-lo, as pesquisas acadêmicas sobre esse fenômeno ainda são escassas — embora estejam aumentando nos últimos anos (Yoon, 2015). Para posicionar a literatura acadêmica em relação ao tema, apresenta-se abaixo sua análise bibliométrica, discutindo o conteúdo da produção.

Metodologia do estudo bibliométrico: usando a plataforma ProQuest, pesquisou-se o banco de dados Proquest Central entre os dias 15 e 16/10/2018, buscando em periódicos científicos os termos “secondary use of data” ou “reuse of data”, os quais foram encontrados em 463 artigos. A primeira verificação revelou que muitos desses artigos mencionavam os termos pontualmente ao longo do texto, não sendo trabalhos que exploram a questão do reuso de dados diretamente. Para tentar maximizar as ocorrências de trabalhos que abordam o tema de forma mais central, a busca foi restrita aos artigos que traziam os termos no título, nas palavras-chave ou no resumo do artigo. Essa busca identificou 50 artigos, 46 com “reuse of data” e 4 com “secondary use of data”. Depois da leitura de seus resumos, foram eliminados 20 artigos que se referiam a outro tipo de reuso de dados (por exemplo, trabalhos técnicos que discutem o tema em memória para melhorar a performance das aplicações) e que não eram artigos completos (como um texto editorial curto). Finalmente, 26 artigos foram selecionados para análise, que foi baseada na leitura dos trabalhos completos. A sequência de buscas e seleções está resumida na figura 2.2

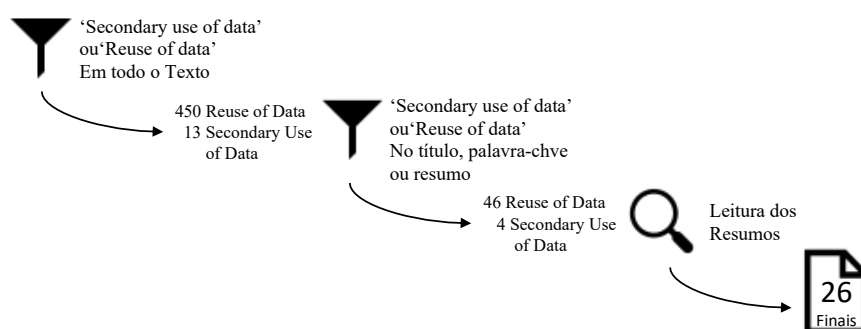


Figura 2.2 - Sequência de buscas e seleções
Fonte: Elaborada pela autora

Análise dos artigos selecionados: A distribuição anual dos 463 artigos que mencionavam os termos relacionados ao reuso de dados indica que houve aumento consistente e acentuado das referências a eles nos trabalhos acadêmicos a partir de 2013 (apêndice 1).

Embora não seja possível afirmar que esses trabalhos estejam se referindo ao reúso de dados nos termos deste artigo ou que façam parte de suas questões centrais, tal evolução sugere o aumento do interesse sobre o tema, corroborando as afirmações de alguns pesquisadores do assunto (Yoon, 2015). Também verificou-se a maior incidência do termo “reúso de dados”, 450 trabalhos com o termo “reuse of data” e 13 com “secondary use of data”.

Ao analisar a evolução anual dos 26 trabalhos publicados em periódicos acadêmicos que abordam o reúso de dados de forma mais central (gráfico 2.1), confirmamos o início de uma curva de interesse no tema nos últimos anos. Estudos demonstram que a pesquisa acadêmica sobre o BD acentuou-se a partir de 2011 (Luvizan, Diniz & Meirelles, 2015). Considerando que as ferramentas do BD ampliaram as possibilidades do reúso de dados, é plausível imaginar que, após alguns anos de estudo, a evolução das discussões e questões mais específicas do BD, como o reúso de dados, tenha emergido. Alguns estudiosos alertam que, apesar da alta expectativa sobre o potencial do BD, muitos de seus benefícios ainda não foram alcançados, e o reúso de dados poderia ser uma via para acessá-los, sugerindo a necessidade de estudos que colaborem com o seu avanço na teoria e na prática (Martin, 2015). O aumento do interesse acadêmico pode estar atendendo a esse chamado.

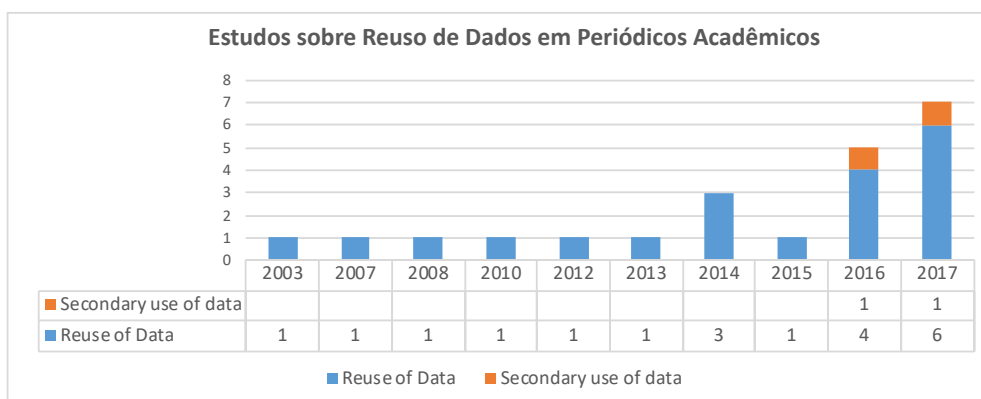


Gráfico 2.1 - Estudos sobre reúso de dados em periódicos acadêmicos por ano

Fonte: elaborado pela autora

Nota: Gráfico não inclui 2018 para não distorcer a análise, já que o ano não estava completo.

Agrupando os periódicos nas áreas as quais declararam-se pertencer (detalhes no apêndice B), observa-se a diversidade de segmentos interessados em reúso de dados, conforme apresentado tabela 1. Os 26 artigos publicados estão distribuídos em periódicos de 10 áreas diferentes. Nota-se que o aumento das publicações nos últimos anos se dá não apenas pela intensificação dos trabalhos nos campos precursores dessa temática (Saúde e Pesquisa/Ciência/Biblioteca), mas também pela publicação pontual de trabalhos em outras áreas. Essa diversidade é coerente com os estudos que indicam o potencial do BD em vários

segmentos de aplicação, destacando que alguns já estão mais sensíveis a tais expectativas e, portanto, têm buscado formas de explorá-las (Manyika et al., 2011).

Tabela 1. Distribuição das publicações por área de periódicos

Ano Publicação	Área do Periódico										Total
	Saúde	Pesq/Ciência/Biblioteca	Política/Legal	Arqueologia	Multidisciplinar	Astronomia	Biologia	Comunicação	Economia Trabalho	Química	
2003	1										1
2007		1									1
2008		1									1
2010	1										1
2012				1							1
2013		1									1
2014			2							1	3
2015		1									1
2016	1		1			1	1		1		5
2017	2	1	1		2			1			7
2018	2	1		1							4
Total	7	6	4	2	2	1	1	1	1	1	26

Fonte: elaborada pela autora

A tabela 2 apresenta os temas abordados pelas áreas dos periódicos, demonstrando que não há grande diversidade nesses trabalhos, os quais orbitam por poucos assuntos, algumas vezes relacionados entre si. Adicionalmente, a maioria dos temas é abordada pontualmente e apenas três aparecem em vários trabalhos, sugerindo haver, de fato, um esforço de pesquisa mais articulado sobre eles. Dos 26 estudos analisados, 14 referem-se a reuso de dados em Pesquisa/Ciência e Curadoria de Dados. Eles estão distribuídos em oito áreas, o que indica uma preocupação em relação a essa temática em diversos campos de estudo. O reuso de dados médicos também foi explorado em vários trabalhos da área de Saúde, corroborando a afirmação de pesquisadores sobre a importância do assunto nessa área (Grande et al., 2014; Krumm et al., 2014; Safran et al., 2007).

Tabela 2.2 - Temas abordados nos artigos sobre reuso de dados por área de periódicos

Tema	Saúde	Pesq/Ciência/Biblioteca	Política/Legal	Multi-disciplinar	Arqueologia	Biologia	Astronomia	Química	Comunicação	Economia do Trabalho	Total
Reuso de dados em pesquisa/ciência	1	2	1	1	2	1			1		9
Curadoria de dados	1	2		1				1			5
Reuso de dados médicos	4										4
Digital library - eScience		2									2
Reuso de dados p/ indicadores trabalho										1	1
Condições p/ expansão reuso			1								1
Dimensão legal reuso de dados			1								1
Prog. gestão conhecim. pesq. espacial							1				1
Serviços usando open data			1								1
Regulamentos Genoma e Biobanking	1										1
Total	7	6	4	2	2	1	1	1	1	1	26

Fonte: elaborada pela autora

A ausência de SI também chama a atenção nesses quadros. Entretanto, isso não significa que a área não tenha produzido nenhum artigo sobre o tema. Inclusive, algumas citações relevantes deste artigo são de trabalhos publicados em periódicos de SI. Tais trabalhos, no entanto, não fazem referência ao termo “reuso de dados” ou “uso secundário de dados” e,

por isso, não foram capturados nas buscas desta revisão. Diante desse cenário, vale lembrar que, ao abordar questões específicas sem usar o termo que denomina o reúso de dados, fica mais difícil capturá-lo quando buscamos pesquisas sobre esse fenômeno, dificultando a conexão de suas questões com o conjunto de discussões que integra a literatura sobre o tema. Por outro lado, ainda que haja trabalhos pontuais, a produção da área sobre o assunto não é compreensível e existem diversas citações indicando a necessidade de mais estudos acadêmicos sobre os pontos envolvidos no reúso de dados também referem-se a trabalhos publicados em SI - por exemplo, Clarke (2016) e Newell & Marabelli (2015) - confirmando a necessidade de mais engajamento da área na discussão.

Acredito que uma atuação mais significativa da área de SI nessa discussão poderia contribuir para ampliar sua cobertura temática. Primeiramente, algumas questões pertinentes ao reúso de dados estão ligadas a temas clássicos de SI, como qualidade de dados, segurança da informação, gestão de recursos de tecnologia, integração, adoção de tecnologias e diversos assuntos referentes à dimensão social do fenômeno tecnológico (Ketter et al., 2016). O corpo de conhecimento já desenvolvido em SI sobre tais questões poderia colaborar para a investigação desse novo cenário do reúso de dados na era do BD. Além disso, se o atual padrão de tecnologia e oferta de dados está delineando novos contornos para essas questões, é papel da área de SI revisar os impactos dessas transformações sobre elas, atualizando a literatura da área. Vale lembrar também que, assim como aconteceu em fenômenos tecnológicos estudados no passado, os desafios do reúso na era do BD demandam a investigação de múltiplos desdobramentos, como destacam Custers e Ursic (2016):

“Os desenvolvimentos na era do *big data* demandam novos modelos tecnológicos (considerando padronização e infraestrutura de TI adequada), novos modelos econômicos (considerando sigilo corporativo e direitos de privacidade individual), novos modelos sociais (considerando suporte público) e novos modelos legais (considerando a proteção de dados pessoais) em que o reúso de dados seja encorajado ao invés de atravancado” (Custers & Ursic, 2016, p. 14). Além da capacidade de navegar pelos temas sugeridos por Custers e Ursic, 2016, a área de SI também poderia ser uma alternativa para a produção de perspectivas de análise mais amplas e pluralistas, já que possui vertentes que buscam abordagens mais holísticas nos estudos dos fenômenos tecnológicos (Baesens et al., 2014; Ketter et al., 2016; Feldman & Orlikowski, 2011). Para colaborar com o engajamento de SI nesse debate, a seguir destaca-se os principais desafios do reúso de dados discutidos nos artigos da revisão de literatura e que são relacionados a temas relevantes para a agenda de SI.

2.4.3 Algumas implicações do reúso de dados relevantes para SI

São muitas as implicações relevantes do reúso de dados para SI. Não existe a pretensão de apresentar uma lista exaustiva, mas sim destacar alguns pontos que apareceram de forma mais consistente nos artigos sobre o tema na revisão de literatura. Além de reforçar a importância da área, acredito que este debate também possa inspirar novos estudos. As discussões que apareceram de forma bastante consistente nos artigos estavam relacionadas aos seguintes temas: capacidade de compartilhamento e reúso de dados; qualidade de dados; acesso e autorização de uso de dados; controle de dados, privacidade e direitos individuais; engajamento individual na segurança de dados e educação digital; e riscos de desigualdades digitais. Apresenta-se abaixo uma breve reflexão de cada um.

Capacidade de compartilhamento e reúso de dados: a existência de mais dados não é, por si só, garantia de melhores informações ou de compartilhamento e reúso dos dados (Barbera, 2013; Borgman et al., 2015; Boyd & Crawford, 2012; Curty, 2015; Custers & Bachlechner, 2017; Custers & Ursic, 2016; Kansa & Kansa, 2018; Poole, 2015). É necessário haver o desenvolvimento da capacidade de explorar e analisar dados e da infraestrutura tecnológica para viabilizar o acesso e o processamento de grandes volumes de informações (Barbera, 2013; Borgman et al., 2015; Curty, 2015; Custers & Bachlechner, 2017). Ainda que a tecnologia tenha avançado bastante nas últimas décadas, a localização de dados relevantes para fins específicos ainda é um desafio (Borgman et al., 2015; Curty, 2015). Em diversas situações, é preciso acessar informações adicionais sobre os dados para garantir que sejam encontrados e interpretados corretamente (Aleksynska & Cazes, 2016; Borgman et al., 2015; Curty, 2015; Naudet et al., 2018; Poole, 2015). Também é necessário que as condições de coleta e armazenamento possam ser verificadas para garantir que o reúso seja adequado do ponto de vista legal e ético (Curty, 2015). Tais demandas ficam mais complexas à medida que a geração do dado fica mais distante de seu reúso (Aleksynska & Cazes, 2016; Yoon, 2015; Zimmerman, 2008;). Elas têm sido endereçadas por meio de documentação ou pela participação de intermediários humanos ou virtuais nesse processo de curadoria dos dados (Borgman et al., 2015, Curty, 2015; Newell & Marabelli 2015; Palmer et al., 2017, Poole, 2015; Poole, 2017). Tal intermediação, no entanto, abre um outro leque de discussões e preocupações (Newell & Marabelli 2015, Samarajiva & Lokanathan, 2016; Poole, 2015). Em diversas áreas, torna-se cada vez mais necessário rever papéis, políticas e procedimentos para desenvolver práticas de compartilhamento e reúso de dados que sejam social e economicamente sustentáveis (Custers

& Ursic, 2016; Jones, 2012; Kansa & Kansa, 2018; Newell & Marabelli 2015; Poole, 2015; Wallis et al., 2013).

Qualidade dos dados e do que se produz a partir deles: sempre foi um desafio para o uso de dados, porém torna-se ainda mais complexa no contexto em que as informações devem atender a critérios de qualidade que viabilizem usos que não eram previstos quando foram geradas (Curty, 2015). A qualidade dos dados não é um critério único e estático, mas sim algo dinâmico e contextual, que pode variar de acordo com a demanda de uso (Baesens et al., 2016). Por isso, garantir a qualidade dos dados para demandas desconhecidas representa uma mudança na lógica de planejamento e gestão das informações, e um grande desafio para todos os envolvidos nesse processo (Muller et al., 2016; Poole, 2015). Além disso, para quem for reutilizar os dados poder confiar neles, é preciso que ele tenha condições de verificar sua qualidade, o que nem sempre é simples ou viável (Curty, 2015). Aqui, além da questão da confiança, novamente aparecem a documentação e os intermediários como alternativas a serem debatidas (Baesens et al., 2016; Borgman et al., 2015, Curty, 2015; Newell & Marabelli 2015; Palmer et al., 2017, Poole, 2017).

Além de pensar na qualidade dos dados, é preciso atentar-se para a qualidade do que está sendo produzido a partir do reuso desses dados (Aleksynska & Cazes, 2016). Acesso a grandes volumes de informações e ferramentas de análise relativamente amigáveis podem esconder os riscos da má classificação e interpretação dos dados, assim como as falhas conceituais e metodológicas na composição de indicadores e outros produtos analíticos (Aleksynska & Cazes, 2016; Muller et al., 2016). Pesquisadores alertam para as consequências nocivas de tais distorções para os negócios e para a sociedade como um todo (Boyd & Crawford, 2012; Samarajiva & Lokanathan, 2016)

Acesso e autorização do uso de dados: o mercado de dados ainda está em desenvolvimento e diversos aspectos legais e éticos relacionados ao seu reuso não foram amplamente discutidos pela sociedade (Custers & Bachlechner, 2017; Custers & Ursic, 2016; Grande et al., 2014; Hsueh et al., 2017; Safran et al., 2007; Samarajiva & Lokanathan, 2016). Governos e organizações de diversas naturezas têm se engajado nesse debate. A General Data Protection Regulation (GDPR) – Regulamento Geral sobre a Proteção de Dados, criada pela União Europeia e o Marco Civil da Internet no Brasil são exemplos desses esforços. Tal empreitada, no entanto, não pertence apenas ao poder público, mas requer a participação da sociedade como um todo. Nos últimos anos, escândalos de vazamento de dados de redes sociais, manipulações eleitorais e fraudes colaboraram para que algumas discussões emergissem, fazendo avançar o nível de informação e engajamento da sociedade quanto ao tema. Apesar

disso, ainda estamos longe de atingir um alto nível de transparência e maturidade nesse cenário permeado por inúmeras tensões. A discussão sobre a coleta de metadados pela Agência Nacional de Segurança Americana (NSA) ilustra a polêmica (Ball, 2013). A divulgação de documentos internos da agência por Edward Snowden intensificou o debate sobre quais dados estão sendo coletados, de quem, quando e para o quê. A agência possui uma ferramenta para coleta e consulta de metadados que captura informações de boa parte do tráfego da internet nos EUA e fora dele. Enquanto muitos acreditam que esse é um processo invasivo e inapropriado, a entidade argumenta que a captura é necessária para a composição de perfis, além de mitigar o risco de não ter dados de alguém que entre na lista de perfis a serem monitorados. Também afirma que a simples captura dos dados não significa vigilância do cidadão comum, já que ele só será de fato analisado individualmente se a pessoa entrar em um perfil de risco. Mais clareza e transparência sobre esses procedimentos ajudaria na discussão, mas contraria as regras de segurança nacional do órgão.

Controle de dados: as discussões, políticas e práticas nesse campo direcionaram grande atenção aos controles na coleta de dados. Porém, esse modelo mostra-se cada vez mais inviável e insuficiente. “Hoje, há tantos dados sendo coletados, de tantas formas, que é praticamente impossível dar às pessoas uma forma significativa de rastrear toda a informação existente sobre elas por aí e, menos ainda, de consentir sua coleta em primeiro lugar” (Mundie, 2014). Além disso, o fato de consentir a coleta de um dado para um determinado fim não significa estar de acordo com seu uso para outras finalidades. Especialistas sugerem mover o foco do controle da coleta e propriedade para o acesso e uso dos dados, como uma alternativa mais efetiva para lidar com as questões éticas, de segurança e de direitos no cenário atual (Grande et al., 2014; Mundie, 2014; Safran et al., 2007). Entretanto, esse novo foco está recheado de desafios de ordem prática. Nem todo reuso de dados é conhecido no momento em que estes são capturados, de forma que não poderia constar claramente nos termos de autorização assinados pelos usuários na coleta de dados. Pedir novas autorizações ou atualizar os termos originais a cada novo reuso é complicado e muitas vezes inviável (Custers & Ursic, 2016). Além disso, esse uso pode envolver modelos sofisticados e difíceis de explicar aos leigos (Baesens et al., 2016; Custers & Ursic, 2016). Como alternativa, os termos de autorização têm expressado os possíveis usos de maneira bem ampla, tentando resguardar-se para o futuro desconhecido. Em consequência, o usuário está cada vez menos informado sobre o que, de fato, está sendo feito com seus dados, o que tem gerado certa polêmica sobre a ética e a legalidade dessa prática (Custers & Ursic, 2016).

Privacidade e direitos individuais: se por um lado é necessário criar mecanismos para proteger a privacidade e os direitos individuais, requerimentos de minimização e declaração de propósito de uso para consentimento prévio do usuário podem limitar o reuso de dados, reduzindo possibilidades de ganhos financeiros e sociais que eles poderiam gerar (Baesens et al., 2016; Custers & Ursic, 2016; Mundie, 2014). Algumas regras estabelecidas na GDPR, por exemplo, buscam regular essa questão, porém criam situações que podem restringir reusos de dados que interessariam à sociedade, evidenciando mais uma vez a necessidade de um debate social mais amplo (Custers & Ursic, 2016).

Educação digital e engajamento individual na segurança de dados: vale lembrar também que a conscientização e o engajamento dos indivíduos nesta questão devem ir além do debate sobre autorização de uso de dados. O usuário tem um papel fundamental em toda a cadeia de segurança de seus dados. Um estudo demonstrou que pessoas que se diziam preocupadas com sua privacidade se comportavam de forma insegura ao compartilhar informações pessoais quando estimuladas por descontos, brindes ou outras vantagens, corroborando demais ações que também apontam desconexões entre o discurso e a prática dos usuários (Acquisti & Gross, 2006; Custers & Ursic, 2016). Isso demonstra que tal questão não poderá ser equacionada apenas com instrumentos legais que não venham acompanhados de transformações culturais e de educação digital. Além disso, pesquisas alertam que, embora muitas empresas que operam no cenário do BD tenham políticas de privacidade bem estruturadas, seus usuários continuam preocupados com questões de privacidade (Custers & Ursic, 2016). Recentes polêmicas em torno do Facebook e as práticas comerciais da empresa e de seus parceiros baseadas no reuso de dados, além de sua abordagem “take it or leave it” – que não oferece alternativas de *opt-out* aos usuários (Agrela, 2018; Martin, 2015) – ilustram alguns fundamentos desse sentimento.

Venda de dados dos usuários, definição de perfis de comportamento, algoritmos que organizam e filtram as informações que lemos e por aí vai. Tudo isso faz parte do debate. Logo, para que todos os atores interessados possam participar dele de forma ativa e embasada é necessário que as empresas atuem com mais transparência e simplifiquem seus discursos, assumindo a responsabilidade de se fazerem entender pelo cidadão comum. Entendo, porém, que também é necessário expandir o horizonte e a ambição em relação à inclusão digital para que a discussão possa incluir, de fato, toda a sociedade e que todos tenham condições de compreender e exercer os direitos que lhes vêm sendo assegurados pelos avanços regulatórios. Custers e Ursic (2016) destacam a complexidade dos novos instrumentos legais, alertando para as controvérsias conceituais e os desafios de operacionalizar esses direitos na prática. Se

quisermos uma sociedade engajada no debate e na vigilância de seus direitos, é fundamental que ela seja capaz de interagir com suas diversas questões que devem, portanto, integrar a agenda da inclusão/educação digital.

Desigualdades digitais: enquanto os entusiastas do BD destacam sua contribuição para democratizar o uso de dados entre as organizações de diversos tamanhos, alguns estudos chamam a atenção para o risco de exclusão de pequenas e médias empresas desse cenário (Barbera, 2013; Custers & Bachlechner, 2017). Na economia de dados tradicional, somente as grandes corporações tinham capacidade de ter acesso às maiores bases de dados. Com a chegada do BD, o contexto não mudou muito, pois ainda são as grandes empresas que têm acesso, majoritariamente, às tecnologias e competências necessárias para o reúso de dados em aplicações mais sofisticadas, como o chamado Data Analytics (Barbera, 2013; Custers & Bachlechner, 2017). O próprio custo dos dados pode ser proibitivo para organizações de menor porte (Custers & Bachlechner, 2017). Ainda que se tenha diversos casos de *startups* e pequenas empresas que prosperam nesse ambiente, a oferta de dados públicos venha crescendo e o mercado de dados se desenvolvendo, a participação de pequenas e médias empresas nessa grande festa ainda é limitada (Barbera, 2013; Custers & Bachlechner, 2017).

Embora essa lista de questões pareça uma visão pessimista e desencorajadora das práticas de reúso de dados, a autora do presente artigo acredita no grande potencial de benefícios do reúso de dados que o BD viabiliza. Por isso, empenha-se em trazer à luz tal problemática e evidenciar a importância da participação de SI nesse debate. Todos os desafios mencionados fazem parte, em alguma medida, da agenda de pesquisa e do corpo de conhecimento de SI. Não participar ativamente de sua discussão significa não colocar o conhecimento a serviço da exploração desse fenômeno. Significa também permitir que tal entendimento fique desatualizado em relação às práticas contemporâneas, já que o reúso de dados só tende a aumentar no cenário atual do BD.

Além disso, vale lembrar que o reúso de dados é um fenômeno de grandes dimensões e impactos em diversas áreas, de forma que sua investigação não cabe somente a SI. O desafio é abordar as múltiplas dimensões desse tema. A identificação das diferentes situações apresentadas a seguir pode colaborar para o desenvolvimento sustentável de pesquisas que abracem tais perspectivas.

2.5 Classificação do reúso de dados

2.5.1 Metodologia da meta-análise

Este trabalho adotou uma estratégia qualitativa construtivista, visando colaborar com a definição do conceito de reúso de dados e a caracterização das situações nas quais ele ocorre. Foi realizada uma revisão bibliográfica, apresentada na seção anterior, que aborda os conceitos, posiciona a literatura existente e expõe diversas implicações do reúso de dados no contexto do BD. Para identificar elementos que possam caracterizar esse fenômeno, foi realizada uma busca nos artigos publicados nos últimos cinco anos em periódicos de SI sobre BD, selecionando aqueles que reportavam casos em que se pode identificar a ocorrência do reúso. A meta-análise desses casos deu origem a um esquema classificatório do reúso de dados. Os potenciais desafios das categorias do esquema proposto foram discutidos à luz das implicações reportadas na literatura. Tal estratégia mostra-se adequada para pesquisas que buscam aprofundar o entendimento sobre um fenômeno a partir de sua própria realidade (Flick, 2004), o que corresponde aos interesses do presente trabalho.

Para Whitemore & Knafl (2005), revisões sistemáticas que combinem as evidências de múltiplos estudos sobre um tema específico podem ser boas alternativas para mapear suas práticas. Tais revisões muitas vezes incluem meta-análises baseadas em estudos quantitativos que combinam dados das pesquisas originais usando técnicas estatísticas (Trikalinos et al., 2008; Whitemore & Knafl, 2005). Pesquisadores destacam, porém, que também é possível utilizar a meta-análise em revisões nas quais não se aplica técnicas quantitativas, de modo que as análises são apresentadas em forma de narrativas, que podem ser associadas a contagens ou outras representações quase-estatísticas (Coper, 1998; Whitemore & Knafl, 2005). Diversos métodos vêm sendo desenvolvidos para a meta-análise de estudos qualitativos, propondo abordagens e níveis de interpretação diferentes (Whitemore & Knafl, 2005). Uma dessas possibilidades é uma abordagem mais interpretativa do que agregativa (Ke, 2009) e que possa ser associada a estudos posicionados no paradigma construtivista (Da Costa, Hall & Spear, 2016). Tal alternativa é adequada ao posicionamento da presente pesquisa e é a orientação que ela persegue.

Visando trazer a transparência necessária para demonstrar o rigor metodológico da pesquisa, as revisões devem partir de uma questão bem definida, explicitar seus métodos e

descrever claramente todas as suas etapas, especialmente no que se refere à seleção e análise de dados (Cooper, 1998; Whittemore & Knafl, 2005). Para conferir rigor e organização para a meta-análise deste estudo, o trabalho segue as fases propostas por Cooper (1998), sumarizadas na figura 2.3 e descritas na sequência.

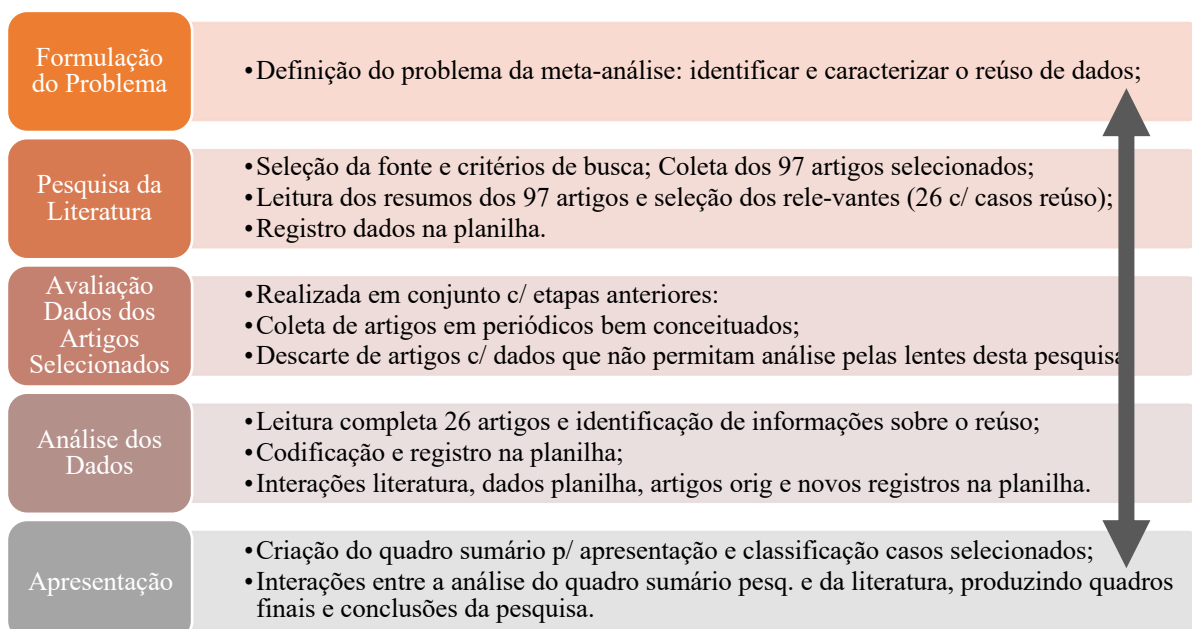


Figura 2.3 - Diagrama do processo de meta-análise da pesquisa

Fonte: elaborado pela autora

- a) **formulação do problema:** refere-se à definição objetiva do problema, variáveis de interesse (conceitos, problemas e população) e quadros amostrais apropriados (tipos de estudos empíricos e literatura teórica) (Whittemore & Knafl, 2005). Essa definição orienta as demais fases da pesquisa e permite que a revisão não produza apenas uma análise descritiva do estudo existente, mas estabeleça um conceito ao redor do qual construirá uma nova análise dos dados (Whittemore & Knafl, 2005). Nessa meta-análise, busca-se identificar e caracterizar o reúso de dados. Portanto, ela está interessada nas informações relacionadas aos dados e ao reúso reportado em cada caso. Não foi definido a priori quais seriam tais informações, o que permitiu capturar o que emergiu na análise dos casos;
- b) **pesquisa da literatura:** a estratégia de pesquisa e todas as decisões de recorte da literatura utilizadas para definir os estudos que serão considerados na meta-análise devem ser bem definidas e justificadas explicitamente para conferir rigor ao trabalho (Cooper, 1998; Whittemore & Knafl, 2005). Atendendo a essa recomendação, os procedimentos de seleção e coleta de dados usados na pesquisa serão descritos a seguir:

● **seleção de fonte e coleta de dados** – Os dados foram coletados entre os dias 08/03/2017 e 12/04/2017 nas bases da EBSCO e Science Direct, selecionando o período de janeiro/2012 a dezembro/2016. Optou-se pelo período de cinco anos, considerando que este propiciaria uma visão adequada, dada a atualidade do fenômeno em questão e a intensificação do interesse acadêmico no tema a partir de 2011 (Luvizan, Meirelles & Diniz, 2015). Os termos de busca utilizados foram *big data* ou *analytics* ou *intelligence*, já que o objetivo era localizar casos empíricos de BD e há uma convergência entre esses termos na literatura da área (Baesens et al., 2016; Chen, Chiang, & Storey, 2012, Martin, 2015). Na seleção, não foram utilizados termos como *reuse of data*, pois já havia sido verificado que os artigos nem sempre identificam que os casos se referem ao reúso de dados. Assim, escolheu-se analisar casos de BD para identificar situações de reúso de dados e selecionar os artigos para a meta-análise. As buscas foram limitadas aos periódicos do grupo J8, que compreende os oito principais periódicos da área de SI, segundo o Colegiado Acadêmico Senior da Association for Information Systems (AIS). Ao todo, foram encontrados 97 artigos, cujos dados foram registrados em uma planilha eletrônica (MS-Excel). Além do registro dos critérios de seleção e quantidades retornadas em cada busca, registrou-se os detalhes dos artigos selecionados, como nome do periódico e do artigo, citação no padrão APA e resumo detalhado;

● **seleção dos artigos relevantes para pesquisa** – realizou-se a leitura dos resumos dos 97 artigos, buscando identificar se traziam casos empíricos e se estes retratavam situações em que houve reúso de dados. Em algumas situações, foi necessário expandir a leitura do artigo para confirmar o entendimento. Nesta etapa, foram desconsiderados os trabalhos que não traziam dados empíricos ou não retratavam situações de reúso de dados, como ensaios teóricos e estudos baseados em respostas de questionários. Dentre os artigos iniciais, 26 foram selecionados de acordo com esse novo critério, conforme mostra a tabela 3 (a lista dos artigos escolhidos está no apêndice C). As informações dessa nova análise foram agregadas à planilha da pesquisa, registrando dados como descrição geral do artigo, indicador de dados empíricos no artigo (sim ou não) e do reúso de dados (sim ou não);

Tabela 2.3 - Resultado das buscas

Journal	Fator de impacto	Resultados da busca	Relevante para a pesquisa
European Journal of Information Systems	2,892	4	0
Information Systems Journal	2,522	8	1
Information Systems Research	3,047	16	7
Journal of AIS	1,774	3	0
Journal of Information Technology	4,775	11	0
Journal of MIS	3,025	20	5
Journal of Strategic Information Systems	2,595	6	1
MIS Quarterly	5,384	29	12
Total		97	26

Fonte: elaborado pela autora

- c) avaliação dos dados:** refere-se aos critérios e procedimentos de verificação da qualidade dos dados utilizados na meta-análise (Whittemore & Knafl, 2005). Diversos pesquisadores reconhecem a complexidade de avaliar a qualidade das informações utilizadas em uma revisão integrativa, que reúna trabalhos de diferentes fontes e métodos (Whittemore & Knafl, 2005). Esta etapa foi realizada em conjunto com as anteriores. A primeira estratégia para garantir a qualidade dos estudos originais foi adotada na fase de definição dos critérios de seleção de dados. Só foram utilizados casos publicados em periódicos do J8, reconhecidos pelos elevados critérios de avaliação dos artigos que publicam, envolvendo avaliação cega por múltiplos membros experientes do campo de SI. A leitura dos artigos buscou também verificar a autenticidade e o rigor metodológico dos estudos, assim como seu valor informacional para a pesquisa. Trabalhos que apresentaram informações insuficientes para a análise pelas lentes deste estudo foram desconsiderados na fase de seleção dos artigos;
- d) Análise dos dados:** esta é a fase em que os dados coletados dos estudos originais são ordenados, codificados, categorizados e sumarizados de acordo com o conceito agregador que rege o problema da meta-análise (Cooper, 1998; Whittemore & Knafl, 2005). Seu objetivo é gerar uma forma inovadora de sintetizar os dados, agregando conhecimento adicional ao que os estudos originais já apresentaram (Whittemore & Knafl, 2005). Esse processo de análise de dados é realizado em etapas e pode ser inspirado em diversos métodos de pesquisas qualitativas ou mistas, especialmente aqueles que permitem a interação entre as fases e a comparação constante entre os dados da meta-análise e os dados originais (Miles & Huberman, 1994; Whittemore & Knafl, 2005). Assim, as etapas referentes à análise da pesquisa podem ser descritas como:
- **redução dos dados dos artigos selecionados** – Após uma nova leitura completa dos artigos, as informações sobre os dados envolvidos e o contexto do reuso retratado em cada

caso foram registradas na planilha, que passou a conter questões como finalidade original do dado e do reuso do dado, entre outras. Algumas dessas informações foram transformadas em códigos. Por exemplo, a fonte dos dados foi sumarizada em três códigos (dados internos, dados públicos e dados privados externos);

● **análise de dados e da literatura** – durante a análise, algumas informações foram ganhando relevância no diálogo dos dados com a literatura, como fonte do dado (pública, privada interna ou externa) e finalidade (relacionada ou não ao uso original), dando origem às categorias centrais, que serão discutidas abaixo junto aos resultados da pesquisa. A execução das etapas não se deu de forma totalmente linear. Já na fase de análise, algumas vezes foi necessário fazer novas consultas aos artigos para buscar mais dados, confirmando a dinâmica interativa entre as etapas de coleta e análise de dados descritas nas propostas metodológicas que inspiram a pesquisa (Flick, 2004; Miles & Huberman, 1994; Strauss & Corbin, 1998; Whitemore & Knafl, 2005);

- e) **apresentação** – consiste em integrar os dados individuais dos casos em um repositório único em que os dados dos diferentes casos possam ser apresentados conjuntamente. Essa base unificada pode conter representações diversas dos dados (matrizes, redes, diagramas, ou outras formas descritivas de apresentação) e é extremamente relevante para a meta-análise de múltiplos casos, sendo o ponto de partida para uma interpretação integrativa (Knafl & Webster, 1988; Whitemore & Knafl, 2005). Esse repositório é a planilha que foi construída ao longo de todas as etapas da pesquisa, conforme descrito anteriormente. Essa planilha deu origem às representações usadas no processo interativo de análise, aos quadros apresentados e ao sumário principal, que resultou no esquema classificatório do reuso de dados proposto neste estudo. Criatividade e análise crítica são fundamentais nesta etapa, pois além de trazer clareza aos resultados da pesquisa, tal representação pode estimular reflexões que ajudem a identificar novos temas relevantes (Whitemore & Knafl, 2005). Para criar um quadro sumário da pesquisa, não basta resumir os dados; é preciso escolher dimensões mais relevantes e um fio condutor para organizar sua apresentação. Os quadros e figuras produzidas, assim como suas descrições e análises, serão apresentados a seguir nos resultados do estudo.

2.5.2 Resultados

A análise dos artigos, individualmente e em conjunto, permitiu refletir sobre o que caracterizou o reúso de dados nos casos, indicando duas dimensões relevantes: sua finalidade e a fonte dos dados. Ambas, de alguma forma, eram significativas em seus casos individuais e, portanto, configuraram uma boa alternativa para caracterizá-los.

A finalidade refere-se ao propósito do reúso do dado. Essa dimensão emergiu ao perceber uma variação da distinção entre o propósito original e o do reúso, sugerindo que essa pode ser uma questão relevante para a análise desse fenômeno. Observando os casos de reúso, nota-se que, em algumas situações, embora os dados estejam sendo utilizados para uma finalidade diferente daquela para a qual foram criados, o novo uso está ligado à temática original. É o caso, por exemplo, dos logs de acesso e navegação dos usuários de uma plataforma de gestão de marketing on-line de um varejista para a análise da efetividade dos anúncios realizados (Ghose & Todri-Adamopoulo, 2016). Embora a análise dos dados históricos da plataforma seja um novo uso para esses logs, está relacionado à mesma temática da geração de dados. Já em outros casos, o novo uso se dá em uma área de aplicação totalmente diferente daquela em que os dados foram criados, podendo envolver novos atores, conflitos de interesses e de valores. O uso de informações postadas na internet por empresas e indivíduos para a determinação de perfis e prevenção de fraudes, por exemplo, está usando dados gerados em contextos diversos para uma finalidade de segurança específica que não está diretamente ligada às temáticas envolvidas na criação desses dados.

Os desafios e implicações do reúso de dados podem variar de acordo com sua relação com seu uso original, resultando em uma dimensão relevante para o modelo. O reúso de dados para fins relacionados aos originais vem sendo referenciado na literatura como “repropósito” (Custers & Ursic, 2016). Já para fins não relacionados aos originais tem sido chamado de “recontextualização” (Custers & Ursic, 2016). Nesta meta-análise, embora os 26 artigos reportem o reúso de dados, a maioria (21) refere-se ao repropósito dos dados e apenas cinco apresentam casos de recontextualização. O universo da pesquisa não permite definir a causa dessa predominância, mas sugere que a exploração dos dados em novos usos ainda está focada nas próprias áreas em que os novos usos se darão.

A dimensão “fonte dos dados” está relacionada à origem da informação, que também apresentou um padrão relevante entre os casos analisados, dando origem a três categorias. A primeira diz respeito às fontes de domínio público, ou seja, dados disponíveis

para livre acesso, como dados abertos em blogs, redes sociais ou informações diversas disponíveis em sites de acesso público, por exemplo. Nesse caso, ainda que os dados estejam disponíveis para todos, é preciso analisar a conformidade ética, metodológica e social, entre outras questões a serem verificadas para validar seu reúso. A segunda refere-se a fontes privadas internas à organização/entidade em que se dá o reúso, ou seja, dados gerados e mantidos pela mesma organização que fará um novo uso deles. Além dos desafios éticos e metodológicos para garantir a adequação desse processo, também há que se lidar com as implicações técnicas, políticas e práticas de disponibilizar os dados para novos grupos e em novas áreas e aplicações da organização. Já a terceira categoria está relacionada às fontes privadas externas à organização/entidade em que se dá o reúso, quer dizer, os dados utilizados são de domínio privado de uma organização diferente daquela que está fazendo o novo uso. Nessa situação, todas as implicações das categorias anteriores estão presentes e são potencializadas pela diversidade de atores envolvidos, vindos de diferentes organizações, que podem ser movidas por diferentes interesses e pressupostos.

Nesta análise, verificou-se que três artigos faziam referência ao reúso de dados da própria organização, 12 citavam o reúso de dados públicos, seis exploravam o reúso de dados privados de outras organizações e cinco retratavam situações mistas (dados internos/públicos, internos/privados externos ou públicos/privados externos). Ainda que não possamos tirar conclusões generalizadas, no universo de dados deste estudo observa-se a maior incidência de reúso de dados públicos, o que pode indicar uma tendência para tal prática. Embora a pesquisa não tenha informações suficientes para definir as causas dessa distribuição, a predominância desse tipo de caso pode indicar uma linha de exploração de fontes acessíveis. A pequena participação de artigos referindo-se ao reúso de dados internos, sem ocorrências de recontextualização, pode ser coerente com o alerta da literatura sobre barreiras internas na exploração de dados das próprias organizações (Meirelles, 2013).

Ao classificar os casos de acordo com as categorias dessas duas dimensões, observou-se o trânsito dos dados entre diversas áreas temáticas e fontes, o que nos ajuda a compreender as diferentes situações em que se dá o reúso e a refletir sobre os desafios específicos de cada uma. A tabela 4 resume a distribuição dos artigos na matriz fonte x finalidade. O apêndice D apresenta o sumário dos casos selecionados mapeados nessa matriz.

de um dado requer que seu dono e portador (que podem ou não ser o mesmo indivíduo/entidade) consigam “chegar” ao reuso, ou seja, compreender o que ele significa e os impactos que pode causar, para avaliar se podem e/ou devem ceder a informação (Custers & Ursic, 2016; Malin et al., 2010; Martin, 2015).

À medida que a distância entre a geração e o reuso dos dados aumenta, cresce também o potencial de desafios (Custers & Ursic, 2016; Zimmerman, 2008). Maior distância indica que o reuso poderá envolver maior quantidade e/ou diversidade de atores e/ou processos e relações mais complexos. Esse ecossistema pode ser marcado por sinergias ou conflitos de interesses e de paradigmas entre geradores, consumidores e intermediários dos dados e demais atores que interagem no ambiente, como legisladores e agentes sociais. Contextos organizacionais marcados por múltiplos objetivos, poder difuso e processos baseados em conhecimento são denominados pluralistas (Denis, Langley & Rouleau, 2007). Podemos dizer, então, que a distância do reuso pode tornar seu ecossistema mais pluralista, o que aumenta o potencial de desafios a serem superados para viabilizá-lo. Tais desafios podem impactar as dimensões tecnológicas, sociais, legais e econômicas do reuso de dados (Custers & Bachlechner, 2017).

Considerando que a finalidade do reuso e a fonte dos dados podem influenciar a distância entre a geração e o reuso dos dados, é plausível que o potencial de desafios varie entre as categorias do esquema classificatório finalidade x fonte proposto. Esses diferentes níveis de desafios foram sumarizados na tabela 5 e discutidos na sequência. Vale lembrar que os desafios do reuso de dados só podem ser discutidos com precisão na perspectiva de cada caso, pois cada um terá sua particularidade. Essa não é a pretensão do artigo. Aqui, aspira-se apenas ilustrar as potenciais questões das categorias, a partir dos desafios que a literatura propõe para o reuso de dados em geral.

		Fonte dos Dados		
		Dados Internos	Dados Públicos	Dados Privados Externos
Finalidade do Reuso dos Dados	Repropósito	Direto Pode haver questões tecnológicas, sociais, legais e econômicas, porém a proximidade entre a geração e o reuso dos dados (ambos na mesma organização) pode facilitar seu encaminhamento.	Intermedial Pode haver questões tecnológicas, sociais, legais e econômicas, que tendem ser mais complexas e seu encaminhamento pode depender de atores externos à organização.	
	Recontextualização		Extremo Pode haver questões tecnológicas, sociais, legais e econômicas, que tendem ser mais complexas e seu encaminhamento pode depender de atores externos à organização e ao contexto do reuso, aumentando a distância destes atores, o que pode dificultar o processo de reuso.	

Quadro 2.1 - Potenciais desafios do reuso de dados em diferentes situações

Fonte: Elaborado pelos autores.

As categorias de repropósito e recontextualização referentes ao reúso de dados internos foram agrupadas sob a denominação “direto” (destacadas em verde na tabela 5). Nessa situação, as questões relacionadas às dimensões técnicas, sociais, legais e econômicas merecem atenção e podem ganhar novos contornos nesse momento em que os dados são utilizados para novos fins. No entanto, nessa conjuntura tende a haver maior proximidade entre os atores e as situações envolvidas na geração e no reúso dos dados. Ambos pertencem à mesma organização, de forma que ela tem acesso direto a eles, aumentando suas possibilidades de manejar as demandas emergentes nesse cenário. Portanto, embora haja desafios a serem enfrentados, a organização pode ampliar sua capacidade de ação, mobilizando recursos humanos e materiais para superar as dificuldades, se assim desejar.

Já as categorias ligadas ao repropósito de dados externos, públicos ou privados, foi reunida no grupo denominado “intermedial” (destacadas em amarelo na tabela 5). Nesse cenário, pode haver questões tecnológicas, sociais, legais e econômicas mais complexas do que as que aparecem no grupo anterior. A distância entre a geração e o reúso dos dados é maior, já que a criação destes envolve, necessariamente, atores externos à organização. O fato de o reúso ser ligado à temática original dos dados (repropósito), no entanto, sugere que esses atores externos já tenham alguma relação com a área. Por isso, a distância entre geração e reúso pode ser maior do que no caso de reúso direto e menor do que nos casos mais extremos, gerando um potencial de desafios intermedial³.

Finalmente, as categorias relacionadas à recontextualização de dados externos, públicos ou privados, foram agrupadas em “extremo” (destacadas em vermelho na tabela 5). Essa é a situação em que há maior distância entre a geração e o reúso de dados. Além de envolver atores externos à organização, a criação de dados visa algo totalmente não relacionado ao seu reúso, o que sugere que os atores envolvidos na geração não pertencem à área em que a organização está promovendo o reúso. Isso significa que o reúso em questão implica um ecossistema com mais diversidade, que pode ser marcado por diferenças de paradigmas, interesses e valores entre os envolvidos. Assim como nos grupos anteriores, pode haver desafios tecnológicos, sociais, legais e econômicos. No entanto, o envolvimento de atores de áreas diferentes, discutindo questões com as quais poderão ter diferentes níveis de interação e interesse, pode tornar as relações e as negociações mais complexas, configurando um potencial extremo de desafios.

³ A palavra intermedial foi escolhida para evitar o uso do termo intermediário que, embora seja mais comum, poderia se confundir com a presença de intermediários no processo, o que pode não existir neste e em outros grupos de reúso de dados.

2.6 Conclusão

Atendendo ao chamado para ampliar o debate sobre o reúso de dados no cenário do BD, o presente artigo explorou a pesquisa acadêmica sobre o tema, posicionando sua produção e capturando conceitos e algumas de suas discussões relevantes. A partir de casos empíricos identificados na literatura de BD, o estudo também propôs um esquema classificatório para o reúso de dados. A análise das situações retratadas pelo esquema permitiu refletir sobre os potenciais desafios dessa prática.

Ao apresentar conceitos e caracterizar o reúso de dados, esta pesquisa contribui para a expansão de sua exploração no campo, pois chama a atenção para questões inerentes a ele, estimulando que sejam abordadas em trabalhos que envolvam casos do tema. Como a própria revisão bibliográfica revelou, embora haja casos empíricos de reúso de dados estudados na literatura de BD, poucos refletem sobre a questão em si. Por isso, disseminar um vocabulário para esse fenômeno e discutir implicações que decorrem de suas características pode estimular a incorporação da temática nas pesquisas, fazendo avançar seu debate no campo.

Pesquisadores também alertam para a necessidade de estudos que abordem o reúso de dados em perspectivas mais holísticas, incorporando suas dimensões tecnológicas, sociais, legais e econômicas. Nesse sentido, o atual artigo contribui de duas maneiras. Primeiramente, oferecendo um critério para o recorte do fenômeno. Como propõe Stake (2006), o entendimento de um fenômeno complexo pode ser viabilizado pela abordagem de suas múltiplas faces, colaborando para a compreensão do todo. O esquema classificatório proposto contribui para a segmentação dessa prática, colaborando para a operacionalização de pesquisas que explorem todas as dimensões, mas sejam focadas em categorias específicas, sem perder a referência de seu posicionamento no todo com o qual dialogam. Em segundo lugar, o artigo contribui estimulando o engajamento de SI nesse esforço. Acredito que o perfil de temas e teorias, assim como o dos pesquisadores desse campo, habilita a área a criar abordagens que integrem tais distintas dimensões do reúso de dados, como já tem feito na exploração de outros fenômenos tecnológicos. Para abrir esse diálogo com o campo de SI, o artigo apresenta questões relevantes do reúso de dados que impactam assuntos inerentes à agenda da área, problematizando sua pequena participação nesse debate e sugerindo caminhos para alavancar seu engajamento.

Finalmente, a reflexão sobre as categorias do esquema classificatório sugeriu a influência da pluralidade do ecossistema de reúso de dados e seus potenciais desafios. Tal visão indica que o estudo desse fenômeno pode se beneficiar de lentes teóricas que abordem contextos

pluralistas, indicando uma trilha de pesquisa que pode colaborar para o avanço do tema na teoria e na prática.

2.7 Limites e sugestões para estudos futuros

O universo de dados deste estudo mostrou-se adequado para seus objetivos e para a proposta de um esquema classificatório do reúso de dados para iluminar uma discussão teórica sobre seus potenciais desafios. Porém, não permite que ele seja considerado uma taxonomia de fato e que seja adequado para outras demandas de segmentação desse fenômeno. Novos trabalhos analisando um espectro maior de dados poderiam confirmar o modelo e verificar a distribuição dos casos entre suas categorias, dando origem a novas reflexões. Indica-se também a realização de estudos empíricos para analisar em profundidade suas reais implicações e desafios, corroborando ou não as potenciais questões sugeridas neste estudo. A utilização de lentes teóricas que integrem a perspectiva sóciotécnica e abordem contextos pluralistas, como as teorias de SI inspiradas na estruturação e na teoria ator-rede, parecem promissoras para a missão de produzir estudos que abracem o reuso de dados em suas múltiplas dimensões.

Referências bibliográficas

Abbasi, A., Albrecht, C., Vance, A., & Hansen, J. (2012). Metafraud: A Meta-Learning Framework For Detecting Financial Fraud. *MIS Quarterly*, 36(4), 1293-A12.

Acquisti, A., & Gross, R. (2006, June). Imagined communities: Awareness, information sharing, and privacy on the Facebook. In *International workshop on privacy enhancing technologies* (pp. 36-58). Springer, Berlin, Heidelberg.

Agrela, L. (2018). O Escândalo de Vazamento de Dados do Facebook é Muito Pior do que Parecia. *Revista Exame*. Ed. Abril. Publicado em 06 de abril de 2018. Disponível em <https://exame.abril.com.br/tecnologia/o-escandalo-de-vazamento-de-dados-do-facebook-e-muito-pior-do-que-parecia/>

AIS, Senior Scholar Committee. *Journal Senior Scholar Basket*. Disponível em <https://aisnet.org/?SeniorScholarBasket> consultado em 08/03/2017 às 15:02h.

Aleksynska, M., & Cazes, S. (2016). Composite indicators of labour market regulations in a comparative perspective. *IZA Journal of Labor Economics*, 5(1), 3.

Baesens, B., Bapna, R., Marsden, J. R., Vanthienen, J., & Zhao, J. L. (2016). *Transformational Issues of Big Data and Analytics in Networked Business*. *MIS quarterly*, 38(2), 629-631.

Ball, J. (2013). NSA stores metadata of millions of web users for up to a year, secret files show. *The Guardian*. <https://www.theguardian.com/world/2013/sep/30/nsa-americans-metadata-year-documents>

Barbera, M. (2013). Linked (open) data at web scale: research, social and engineering challenges in the digital humanities. *JLIS. it*, 4(1), 91.

Bardhan, I., Jeong-ha (Cath), O., Zhiqiang (Eric), Z., & Kirksey, K. (2015). *Predictive Analytics for Readmission of Patients with Congestive Heart Failure*. *Information Systems Research*, 26(1), 19-39.

Benjamin, V., Zhang, B., Nunamaker, J. F., & Chen, H. (2016). *Examining Hacker Participation Length in Cybercriminal Internet-Relay-Chat Communities*. *Journal Of Management Information Systems*, 33(2), 482-510.

Borgman, C. (2012). The Conundrum of Sharing Research Data. *Journal of the American Society for Information Science and Technology*, 63(6), 1059-1078.

Borgman, C. L., Darch, P. T., Sands, A. E., Pasquetto, I. V., Golshan, M. S., Wallis, J. C., & Traweek, S. (2015). Knowledge infrastructures in science: Data, diversity, and digital libraries. *International Journal on Digital*.

Boyd, D., & Crawford, K. (2012). *Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon*. *Information, Communication & Society*, 15, 662-679.

Chau, M., & Xu, J. (2012). *Business Intelligence In Blogs: Understanding Consumer Interactions And Communities*. *MIS Quarterly*, 36(4), 1189-1216.

Chen, H., Chiang, R. H., & Storey, V. C. (2012). *Business Intelligence and Analytics: From Big Data to Big Impact*. *MIS Quarterly*, 36, 1165-1188.

Clarke, R. (2016). Big data, big risks. *Information Systems Journal*, 26(1), 77-90.

Cooper, H. M. (1998). *Synthesizing research: A guide for literature reviews* (Vol. 2). Sage.

Curry, R. G. (2015). *Beyond "Data Thrifting": An Investigation of Factors Influencing Research Data Reuse In The Social Sciences* (3713677). Available from ProQuest Dissertations & Theses Global. (1698463092).

Custers, B., & Bachlechner, D. (2017). *Advancing the EU Data Economy: Conditions for Realizing the Full Potential of Data Reuse*. Information Polity, 22(4), 291.

Custers, B., & Ursic, H. (2016). *Big Data and Data Reuse: A Taxonomy of Data Reuse for Balancing Big Data Benefits and Personal Data Protection*. International Data Privacy Law, 6(1), 4-15.

Costa, R. B. D., Hall, S. M., & Spear, A. (2016). Whose reality? A meta-analysis of qualitative research in international and comparative education.

Danings, H., Zhao, J. L., Zhimin, H., & Wong, M. S. (2012). Network-Based Modeling And Analysis Of Systemic Risk In Banking Systems. MIS Quarterly, 36(4), 1269-1291.

Demchenko, Y., Grosso, P., De Laat, C., & Membrey, P. (2013). *Addressing Big Data Issues in Scientific Data Infrastructure, In: First International Symposium on Big Data and Data Analytics in Collaboration (BDDAC 2013)*. Part of The 2013 International Conference on Collaboration Technologies and Systems (CTS 2013), San Diego, USA.

Deng, X., & Chi, L. (2012). Understanding Postadoptive Behaviors in Information Systems Use: A Longitudinal Analysis of System Use Problems in the Business Intelligence Context. Journal Of Management Information Systems, 29(3), 291-326.

Denis, J. L., Langley, A., & Rouleau, L. (2007). Strategizing in pluralistic contexts: Rethinking theoretical frames. Human relations, 60(1), 179-215.

Faniel, I. M. & Jacobsen, T. E. (2010). Reusing scientific data: How earthquake engineering researchers assess the reusability of colleagues' data. *Computer Supported Cooperative Work (CSCW)*, 19(3), 355-375.

Feldman, M. S., & Orlikowski, W. J. (2011). Theorizing practice and practicing theory. *Organization science*, 22(5), 1240-1253.

Flick, U. (2004). *Uma Introdução à Pesquisa Qualitativa*. Porto Alegre: Bookman.

Ghose, A., & Todri-Adamopoulos, V. (2016). *Toward A Digital Attribution Model: Measuring The Impact Of Display Advertising On Online Consumer Behavior*. MIS Quarterly, 40(4), 889-A30.

Ghoshal, A., Menon, S., & Sarkar, S. (2015). *Recommendations Using Information from Multiple Association Rules: A Probabilistic Approach*. Information Systems Research, 26(3), 532-551.

Grande, D., Mitra, N., Shah, A., Wan, F., & Asch, D. A. (2014). *The Importance of Purpose: Moving Beyond Consent in the Societal Use of Personal Health Information*. *Consent for Secondary Use of Health Information*. Annals of internal medicine, 161(12), 855-862.

Hsueh, P. Y., Cheung, Y. K., Dey, S., Kim, K. K., Martin-Sanchez, F. J., Petersen, S. K., & Wetter, T. (2017). *Added Value from Secondary Use of Person Generated Health Data in Consumer Health Informatics*. IMIA Yearbook.

Jabr, W., Mookerjee, R., Yong, T., & Mookerjee, V. S. (2014). *Leveraging Philanthropic Behavior For Customer Support: The Case Of User Support Forums*. MIS Quarterly, 38(1), 187-A8.

Kansa, S. W., & Kansa, E. C. (2018). Data beyond the archive in digital archaeology: an introduction to the special section. *Advances in Archaeological Practice*, 6(2), 89-92.

Ketter, W., Peters, M., Collins, J., & Gupta, A. (2016). Competitive Benchmarking: An IS Research Approach To Address Wicked Problems With Big Data And Analytics. MIS Quarterly, 40(4), 1057.

Krumm, R., Semjonow, A., Tio, J., Duhme, H., Bürkle, T., Haier, J., ... & Breil, B. (2014). *The need for harmonized structured documentation and chances of secondary use—Results of a systematic analysis with automated form comparison for prostate and breast cancer*. Journal of biomedical informatics, 51, 86-99.

Lash, M. T., & Zhao, K. (2016). *Early Predictions of Movie Success: The Who, What, and When of Profitability*. Journal of Management Information Systems, 33(3), 874-903.

Lau, R. K., Liao, S. Y., Wong, K. F., & Chiu, D. W. (2012). *Web 2.0 Environmental Scanning And Adaptive Decision Support For Business Mergers And Acquisitions*. MIS Quarterly, 36(4), 1239-A6.

Letouzé, E. (2012). *Big Data for Development: Challenges & Opportunities*. UN Global Pulse.

Li, W., Chen, H., & Nunamaker, J. F. (2016). *Identifying and Profiling Key Sellers in Cyber Carding Community: AZSecure Text Mining System*. Journal Of Management Information Systems, 33(4), 1059-1086.

Liang, N., Biros, D. P., & Luse, A. (2016). *An Empirical Validation of Malicious Insider Characteristics*. Journal Of Management Information Systems, 33(2), 361-392.

Luvizan, S., Meirelles, F., & Diniz, E. (2015). *Big Data: Evolução Das Publicações E Oportunidades De Pesquisa*. Revista Eletrônica de Sistemas de Informação, 14(3).

Malin, B., Karp, D., & Scheuermann, R. H. (2010). Technical and policy approaches to balancing patient privacy and data sharing in clinical and translational research. *Journal of Investigative Medicine*, 58(1), 11-18.

Manyika, J et al. (2011). *Big Data: The Next Frontier For Innovation, Competition, And Productivity*. McKinsey Global Institute.

- Martens, D., Provost, F., Clark, J., & de Fortuny, E. J. (2016). *Mining Massive Fine-Grained Behavior Data To Improve Predictive Analytics*. MIS Quarterly, 40(4), 869-888.
- Martin, K. (2015). *Ethical Issues In The Big Data Industry*. MIS Quarterly Executive, 14(2), 67-85.
- McAfee, A., & Brynjolfsson, E. (2012). *Big Data: The Management Revolution*. Harvard Business Review, 90(10), 60-68.
- Meirelles, F. (2013). *Pesquisa Anual – Administração de Recursos de Informática*. GVcia – Centro de Tecnologia de Informação Aplicada da FGV-EAESP.
- Miles, M., & Huberman, A. (1994). *Qualitative Data Analysis: An Expanded Sourcebook*. Sage.
- Mohd, A. A., & Biswas, R. (2018). Sorted r-train: An improved dynamic data structure for handling *big data*. International Journal of Intelligent Systems and Applications, 10(11), 27.
- Mundie, C. (2014). Privacy Pragmatism: Focus on Data Use, Not Data Collection. Foreign Affairs. 93(2). Disponível em: <http://rational.com/privacy-pragmatism-focus-on-data-use-not-data-collection/>
- Müller, O., Junglas, I., Brocke, J. V., & Debortoli, S. (2016). Utilizing *big data* analytics for information systems research: challenges, promises and guidelines. European Journal of Information Systems, 25(4), 289-302.
- Naudet, F., Sakarovitch, C., Janiaud, P., Cristea, I., Fanelli, D., Moher, D., & Ioannidis, J. P. (2018). Data sharing and reanalysis of randomized controlled trials in leading biomedical journals with a full data sharing policy: survey of studies published in The BMJ and PLOS Medicine. *bmj*, 360, k400.
- Newell, S., & Marabelli, M. (2015). *Strategic opportunities (and challenges) of algorithmic decision-making: A call for action on the long-term societal effects of 'datification'*. The Journal of Strategic Information Systems, 24(1), 3-14.
- Niederman, F., & Lukyanenko, R. (2018, June). Categorization: A Source of Theory and Output of Research. In Proceedings of the 2018 ACM SIGMIS Conference on Computers and People Research (pp. 143-150). ACM.
- Oh, O., Agrawal, M., & Rao, H. R. (2013). *Community Intelligence And Social Media Services: A Rumor Theoretic Analysis Of Tweets During Social Crises*. MIS Quarterly, 37(2), 407-A7.
- Palmer, C. L., Thomer, A. K., Baker, K. S., Wickett, K. M., Hendrix, C. L., Rodman, A., ... & Fouke, B. W. (2017). Site-based data curation based on hot spring geobiology. *PloS one*, 12(3).

Pant, G., & Srinivasan, P. (2013). *Status Locality on the Web: Implications for Building Focused Collections*. Information Systems Research, 24(3), 802-821.

Park, H., & Leydesdorff, L. (2013). *Decomposing Social And Semantic Networks In Emerging "Big Data" Research*. Journal of Informetrics, 7, 756-765.

Pil Han, S., Park, S., & Oh, W. (2016). *Mobile App Analytics: A Multiple Discrete-Continuous Choice Framework*. MIS Quarterly, 40(4), 983-A42.

Poole, A. H. (2015). Forging our cultural commonwealth: The importance of digital curation in the digital humanities (Doctoral dissertation, The University of North Carolina at Chapel Hill).

Poole, A. H. (2017). "A greatly unexplored area": Digital curation and innovation in digital humanities. Journal of the Association for Information Science and Technology, 68(7), 1772-1781.

Provost, F., Martens, D., & Murray, A. (2015). *Finding Similar Mobile Consumers with a Privacy-Friendly Geosocial Design*. Information Systems Research, 26(2), 243-265.

Saboo, A. R., Kumar, V., & Park, I. (2016). *Using Big Data To Model Time-Varying Effects For Marketing Resource (Re)Allocation*. MIS Quarterly, 40(4), 911-940.

Safran, C., Bloomrosen, M., Hammond, W. E., Labkoff, S., Markel-Fox, S., Tang, P. C., & Detmer, D. E. (2007). Toward a national framework for the secondary use of health data: an American Medical Informatics Association White Paper. Journal of the American Medical Informatics Association, 14(1), 1-9.

Samarajiva, R., & Lokanathan, S. (2016). *Using Behavioral Big Data For Public Purposes: Exploring Frontier Issues Of An Emerging Policy Arena*.

Seddon, J. J., & Currie, W. L. (2017). A model for unpacking *big data* analytics in high-frequency trading. Journal of Business Research, 70, 300-307.

Shi, Z., Lee, G. M., & Whinston, A. B. (2016). *Toward A Better Measure Of Business Proximity: Topic Modeling For Industry Intelligence*. MIS Quarterly, 40(4), 1035-A53.

Shollo, A., & Galliers, R. D. (2016). *Towards An Understanding Of The Role Of Business Intelligence Systems In Organisational Knowing*. Information Systems Journal, 26(4), 339-367.

Smolan, R. (2013). The human face of *big data*.

Solove, D. J. (2008). Understanding privacy.

Stake, R. E. (2006). *Multiple Case Study Analysis*. Guilford Press.

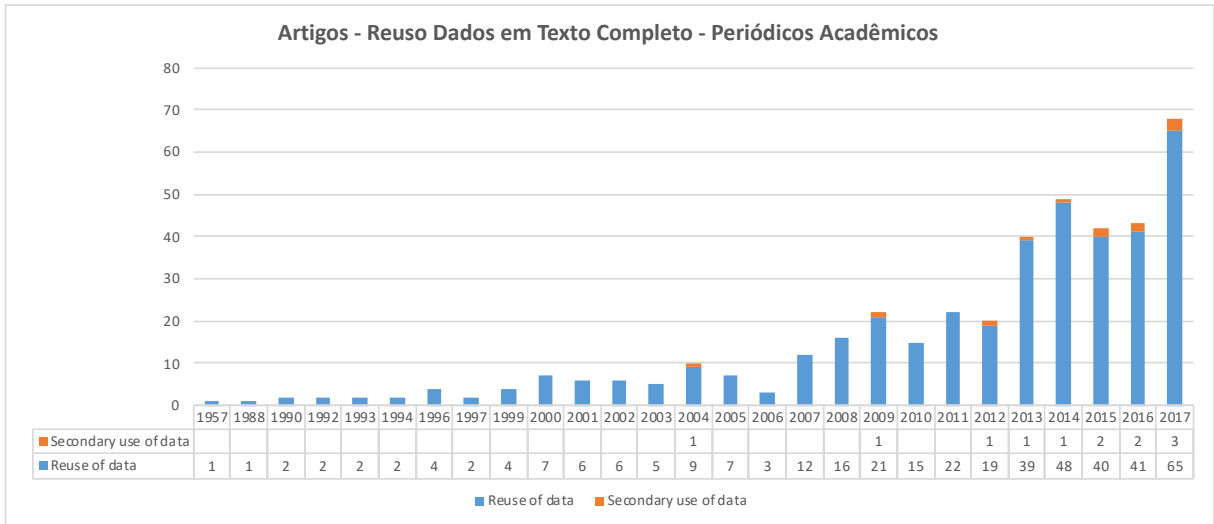
- Strauss, A., & Corbin, J. (1998). *Basics of Qualitative Research: Procedures And Techniques For Developing Grounded Theory*. Sage Publications.
- Sung-Hyuk, P., Soon-Young, H., Wonseok, O., & Sang Pil, H. (2012). *A Social Network-Based Inference Model For Validating Customer Profile Data*. MIS Quarterly, 36(4), 1217-1237.
- Teo, T. S., Nishant, R., & Koh, P. B. (2016). *Do Shareholders Favor Business Analytics Announcements?*. The Journal of Strategic Information Systems, 25(4), 259-276.
- Trikalinos, T. A., Salanti, G., Zintzaras, E., & Ioannidis, J. P. (2008). Meta-analysis methods. *Advances in genetics*, 60, 311-334.
- Uddin, M. F., & Gupta, N. (2014). Seven V's of *Big Data* understanding *Big Data* to extract value. In American Society for Engineering Education (ASEE Zone 1), 2014 Zone 1 Conference of the (pp. 1-5). IEEE.
- Wallis, Jillian C., Elizabeth Rolando, and Christine L. Borgman. 2013. "If We Share Data, Will Anyone Use Them? Data Sharing and Reuse in the Long Tail of Science and Technology." *PLoS One* 1-17.
- Wayne Gregory, R., & Muntermann, J. (2014). *Heuristic Theorizing: Proactively Generating Design Theories*. Information Systems Research, 25(3), 639-653.
- Wasesa, M., Stam, A., & van Heck, E. (2017). *The seaport service rate prediction system: Using drayage truck trajectory data to predict seaport service rates*. Decision Support Systems, 95, 37-48.
- Whittemore, R., & Knafl, K. (2005). The integrative review: updated methodology. *Journal of advanced nursing*, 52(5), 546-553.
- Xixi, L., J. J. Po-An, H., & Rai, A. (2013). *Motivational Differences Across Post-Acceptance Information System Usage Behaviors: An Investigation in the Business Intelligence Systems Context*. Information Systems Research, 24(3), 659-682.
- Yoon, A. (2015). *Data reuse and users' trust judgments: Toward trusted data curation* (3719920). Available from ProQuest Central; ProQuest Dissertations & Theses Global. (1718199510).
- Zhiqiang (Eric), Z., Fader, P., & Padmanabhan, B. (2012). *From Business Intelligence to Competitive Intelligence: Inferring Competitive Measures Using Augmented Site-Centric Data*. Information Systems Research, 23(3), 698-720.
- Zimmerman, A. S. (2003). *Data Sharing And Secondary Use Of Scientific Data: Experiences Of Ecologists* (3079559). Available from ProQuest Dissertations & Theses Global: Science & Technology; ProQuest Dissertations & Theses Global: Social Sciences. (287907131).

Zimmerman, Ann. 2008. "New Knowledge from Old Data: The Role of Standards in the Sharing and Reuse of Ecological Data." *Science, Technology & Human Values* 631-652.

Zuboff, S. (2015). Big other: surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology*, 30(1), 75-89.

Apêndices

APÊNDICE A - Distribuição anual dos artigos acadêmicos mencionando os termos “secondary use of data” ou “reuse of data”



Fonte: elaborado pela autora

Nota: O ano de 2018 não foi incluído no gráfico para não distorcer a análise, já que o ano não estava completo no momento desta pesquisa.

APÊNDICE B - Periódicos agrupados por área

Área	Journal
Arqueologia	Advances in Archaeological Practice
Arqueologia	World Archaeology
Astronomia	Proceedings of the International Astronomical Union
Biologia	PLoS Biology
Comunicação	Canadian Journal of Communication
Economia do Trabalho	IZA Journal of Labor Economics
Multidisciplinar	PLoS One
Pesquisa/Ciência/Biblioteca	Journal of the Association for Information Science and Technology
Pesquisa/Ciência/Biblioteca	International Journal on Digital Libraries
Pesquisa/Ciência/Biblioteca	JLIS.it - Italian Journal of Library, Archives, and Information Science
Pesquisa/Ciência/Biblioteca	Science, Technology, and Human Values
Pesquisa/Ciência/Biblioteca	International Journal on Digital Libraries
Política/Legal	Information Polity
Política/Legal	International Data Privacy Law
Política/Legal	Electronic Journal of E-Government
Política/Legal	The International Journal on Drug Policy
Química	Journal of Chemical Information and Modeling
Saúde	BMJ : British Medical Journal (Online)
Saúde	Health Care Financing Review
Saúde	BMC Medical Ethics
Saúde	Journal of Investigative Medicine
Saúde	BMC Medical Informatics and Decision Making
Saúde	Genetics in Medicine
Saúde	Orphanet Journal of Rare Diseases

Fonte: elaborado pela autora

Nota: A definição da área dos periódicos foi feita com base nas informações dos próprios periódicos em suas páginas oficiais.

APÊNDICE C - Lista dos Artigos Seleccionados para Meta-análise

Artigo	Autor e Ano Publicação
<i>A Social Network-Based Inference Model For Validating Customer Profile Data</i>	Sung-Hyuk, Soon-Young, Wonseok, & Sang Pil (2012).
<i>An Empirical Validation Of Malicious Insider Characteristics</i>	Liang, Biros, & Luse, (2016).
<i>Business Intelligence In Blogs: Understanding Consumer Interactions And Communities</i>	Chau, & Xu, (2012).
<i>Community Intelligence And Social Media Services: A Rumor Theoretic Analysis Of Tweets During Social Crises</i>	Oh, Agrawal, & Rao (2013).
<i>Do Shareholders Favor Business Analytics Announcements?</i>	Teo, Nishant, & Koh (2016).
<i>Early Predictions Of Movie Success: The Who, What, And When Of Profitability</i>	Lash, & Zhao (2016).
<i>Examining Hacker Participation Length In Cybercriminal Internet-Relay-Chat Communities</i>	Benjamin, Zhang, Nunamaker, & Chen (2016).
<i>Finding Similar Mobile Consumers With A Privacy-Friendly Geosocial Design</i>	Provost, Martens, & Murray (2015).
<i>From Business Intelligence To Competitive Intelligence: Inferring Competitive Measures Using Augmented Site-Centric Data.</i>	Zhiqiang, Fader, & Padmanabhan, (2012).
<i>Heuristic Theorizing: Proactively Generating Design Theories.</i>	Wayne Gregory, & Muntermann (2014).
<i>Identifying And Profiling Key Sellers In Cyber Carding Community: Azsecure Text Mining System</i>	Li, Chen & Nunamaker (2016).
<i>Leveraging Philanthropic Behavior For Customer Support: The Case Of User Support Forums.</i>	Jabr, Mookerjee, Yong, & Mookerjee, (2014).
<i>Metafraud: A Meta-Learning Framework For Detecting Financial Fraud</i>	Abbasi, Albrecht, Vance, & Hansen (2012).
<i>Mining Massive Fine-Grained Behavior Data To Improve Predictive Analytics</i>	Martens, Provost, Clark, & de Fortuny (2016).
<i>Mobile App Analytics: A Multiple Discrete-Continuous Choice Framework.</i>	Pil Han, Park, & Oh (2016).
<i>Motivational Differences Across Post-Acceptance Information System Usage Behaviors: An Investigation In The Business Intelligence Systems Context</i>	Xixi, Po-An, & Rai (2013).
<i>Network-Based Modeling And Analysis Of Systemic Risk In Banking Systems</i>	Daning, Zhao, Zhimin, & Wong (2012).
<i>Predictive Analytics For Readmission Of Patients With Congestive Heart Failure</i>	Bardhan, Jeong-ha, Zhiqiang, & Kirksey (2015).
<i>Recommendations Using Information From Multiple Association Rules: A Probabilistic Approach</i>	Ghoshal, Menon, & Sarkar (2015).
<i>Status Locality On The Web: Implications For Building Focused Collections</i>	Pant, & Srinivasan (2013).
<i>Toward A Better Measure Of Business Proximity: Topic Modeling For Industry Intelligence</i>	Shi, Lee, & Whinston (2016).
<i>Toward A Digital Attribution Model: Measuring The Impact Of Display Advertising On Online Consumer Behavior</i>	Ghose, & Todri-Adamopoulos (2016).
<i>Towards An Understanding Of The Role Of Business Intelligence Systems In Organisational Knowing</i>	Shollo, & Galliers (2016).
<i>Understanding Postadoptive Behaviors In Information Systems Use: A Longitudinal Analysis Of System Use Problems In The Business Intelligence Context.</i>	Deng, & Chi (2012).
<i>Using Big Data To Model Time-Varying Effects For Marketing Resource (Re)Allocation</i>	Saboo, Kumar, & Park (2016).
<i>Web 2.0 Environmental Scanning And Adaptive Decision Support For Business Mergers And Acquisitions</i>	Lau, Liao, Wong, & Chiu (2012).

Fonte: elaborado pela autora

APÊNDICE D - Sumário do mapeamento do uso secundário

		Fonte dos Dados		
		Dados Internos	Dados Públicos	Dados Privados Externos
Finalidade do Reuso dos Dados	Repropósito	<p>Dados de sistemas transacionais internos para análise gerencial.</p> <p>Problemas no uso do sistema de BI reportados pelos usuários em sistema de chamados para análise do uso do sistema (pesquisa).</p> <p>Log de acesso e navegação dos usuários da plataforma que gerencia o marketing online do varejista para análise da efetividade dos anúncios.</p> <p>Dados transacionais internos (lojas/online) e dados demográficos de base privada externa p/ análise de variação de comportamento do cliente no tempo p/ direcionar ações de marketing. <small>Mix Priv Ext</small></p> <p>Dados transacionais internos e dados externos p/ análise da performance da empresa, condições de mercado e tendências da indústria p/ definição de estratégia e ações de marketing.</p> <p>Dados transacionais internos e dados externos de mercado para definição de medidas de competitividade das empresas do setor.</p> <p>Dados de páginas WEB, seus links e históricos de acessos para análise de status e posicionamento das páginas em pesquisas.</p>	<p>Dados públicos (websites, portais de finanças, blogs, foruns de discussão, etc) para solução de problemas na área financeira.</p> <p>Redes sociais e sites (cinema) p/ traçar perfil de "sucesso" dos filmes baseado em suas características (elenco, bilheteria, etc).</p> <p>Publicações sobre casos e as pessoas envolvidas (postadas por elas ou por outros) para traçar o perfil de ofensores.</p> <p>Publicações em comunidades virtuais criminosas para analisar o comportamento e perfil dos participantes.</p> <p>Base de dados pública da indústria high-tech (informações de artigos e fontes abertas da web ou por contribuições de usuários validadas por moderadores) para analisar rede de negócios.</p> <p>Publicação em blogs para definição de perfil do cliente e de seu relacionamento na rede de clientes.</p> <p>Mensagens no Twitter em momentos de crise para analisar a situação social e o perfil do reporte social nestes contextos.</p> <p>Relatórios públicos de órgãos regulatórios para criação de cenários para avaliação de riscos do sistema bancário.</p> <p>Relatórios financeiros públicos oficiais p/ detecção de fraudes.</p> <p>Anúncios implem. Analytics e desempenho ações no período.</p> <p>Perguntas e respostas em fóruns de TI p/ análise da relação qualidade/efetividade dos conteúdos e o modelo de reconhecimento do fórum.</p>	<p>Fichas médicas dos pacientes de diferentes hospitais de uma região para análise de reincidência de doenças.</p> <p>Base de dados de entidade privada de monitoramento de audiência para análise do perfil de uso de aplicativos.</p> <p>Bases de dados privadas do ramo varejista para melhorar a qualidade das recomendações de produtos aos clientes p/ aumentar a probabilidade de compra.</p>
	Recontextualização		<p>Dados postados por empresas e indivíduos para identificação de perfil para prevenção de fraudes.</p> <p>Relatórios, notícias e dados do mercado financeiro de fontes públicas e privadas para geração de scorecard para suporte a decisões sobre M&A (Merge&Acquisitions)</p>	<p>Dados dos dispositivos móveis (IP, localidade, etc) para seleção de anúncios na venda de espaço publicitário na WEB (RTB - real-time bidding).</p> <p>Transações de pagamentos ao comércio para definição de perfil dos clientes para direcionar a oferta de produtos bancários</p> <p>Dados de ligações telefônicas (quantidade, duração, frequência, etc) para definição de perfil de relacionamentos para confirmação de dados demográficos</p>

Fonte: elaborado pela autora

3 ARTIGO II – REÚSO-DE-DADOS-NA-PRÁTICA: RECONTEXTUALIZAÇÃO DE DADOS DA TELEFONIA CELULAR PARA AVALIAÇÃO DE CRÉDITO ⁴

3.1 Resumo

Apesar das expectativas sobre o potencial inovador do reúso de dados na era do *big data* (BD), a academia ainda carece de estudos que explorem seus desafios. Para colaborar com esse debate, o presente estudo posiciona o reúso de dados como um fenômeno que ocorre em um contexto pluralista e marcado pela colaboração complexa, sugerindo a abordagem teórica da prática como um caminho promissor para investigá-lo. Esse artigo propõe um modelo teórico baseado no Multilevel Framework para operacionalizar tal perspectiva prática em pesquisas empíricas. O modelo utiliza os elementos do contextualismo para articular conceitos da Social Shaping of Technology (SST) e da tecnologia-na-prática. Ele foi adaptado para demonstrar o processo por meio de um mapa visual e apresentar o reúso-de-dados-na-prática no conteúdo, evidenciando as dimensões intrínseca, legal, tecnológica, societal e econômica do reúso de dados. O modelo ajustado foi exercitado no caso exploratório da empresa Alpha, que oferece uma ferramenta para a avaliação de crédito usando dados da rede de telefonia celular para inclusão financeira. O estudo demonstrou os desafios encontrados pela Alpha para implementar sua solução, decorrentes de assimetrias de acesso a recursos em seu ecossistema. Foram identificadas também as manobras da Alpha para viabilizar a implementação de uma solução alternativa, que não realiza seu objetivo original de inclusão financeira, mas traz sustentabilidade para sua operação no Brasil e abre caminhos para que ela avance na direção desse propósito. A dinâmica desse processo combinou a epistemologia da posse e da prática do poder, corroborando outro estudo envolvendo colaboração complexa.

Palavras-chave: *big data*; reúso de dados; ecossistemas de inovação; inclusão financeira; Multilevel Framework.

⁴ Versão original deste artigo: Luvizan, S. S., Nascimento, P. T., & Yu, A. (2016). *Big Data For Innovation: The Case Of Credit Evaluation Using Mobile Data Analyzed By Innovation Ecosystem Lens*. In Portland International Conference on Management of Engineering and Technology (PICMET), 2016 (pp. 925-936). IEEE.

3.2 Introdução

Com sua crescente oferta de tecnologia e informação, o cenário do *big data* (BD) estimula o reúso dos dados, ou seja, a utilização dos dados para fins diferentes daqueles para os quais foram gerados (Abbasi, Sarker & Chiang, 2016; Baesens, Bapna, Marsden, Vanthienen & Zhao, 2016; Martin, 2015; Yoon, 2015). Embora já existisse anteriormente, essa prática ganha nova dimensão com as possibilidades do BD (Ketter, Peters, Collins & Gupta, 2016; Muller, Junglas, Brocke & Debortoli, 2016). Ela já é utilizada em diversas aplicações, mas ainda não atingiu todo o seu potencial e foi pouco explorada na academia (Clarke, 2016; Custers & Bachlechner, 2017; Custers & Ursic, 2016; Newell & Marabelli, 2015; Samarajiva & Lokanathan, 2016; Yoon, 2015).

Atualmente, há grandes expectativas sobre o reúso de dados. Alguns acreditam que ele trará os benefícios transformacionais esperados do BD que ainda não foram realizados na prática (Ketter *et al.*, 2016; Muller *et al.*, 2016). Pesquisas alertam para a necessidade de discutir as diversas questões envolvidas no reúso de dados para estimular seu desenvolvimento de forma que se possa colher seus benefícios e mitigar seus efeitos indesejados (Clarke, 2016; Curty, 2015, Custers & Bachlechner, 2017; Custers & Ursic, 2016; Newell & Marabelli, 2015, Samarajiva & Lokanathan, 2016; Zuboff, 2015). Visando contribuir nessa jornada, este estudo reflete sobre os conceitos do reúso de dados, propõe um modelo conceitual para investigá-lo em casos empíricos e exercita essa proposta em um estudo de caso exploratório.

O reúso de dados pode se referir à utilização dos dados para fins relacionados aos originais (repropósito) ou para fins totalmente distintos da temática de sua geração (recontextualização) (Custers & Ursic, 2016). Os dados reutilizados podem vir de fontes públicas ou privadas, internas ou externas à organização que os reutilizará (Baesens *et al.*, 2016). A complexidade do ecossistema do qual emerge o reúso de dados pode ser diferente em cada situação, pois implica questões legais, tecnológicas, sociais e econômicas que assumem diferentes contornos à medida que a distância entre a geração e o reúso dos dados aumenta (Aleksynska & Cazes, 2016, Custer & Ursic, 2016; Yoon, 2015, Zimmerman, 2008).

O estudo desse fenômeno requer, portanto, uma lente de análise capaz de explorar o ecossistema de forma ampla e dinâmica. Este artigo sugere que a perspectiva teórica da prática é uma via promissora para investigar fenômenos em contextos pluralistas (Denis, Langley & Rouleau, 2007) e marcados pela colaboração complexa (Faraj, Kolker, Bevilacqua & Özdemir,

2012), como é o caso do reúso de dados. Pesquisadores alertam para a dificuldade de implementar essa abordagem em pesquisas empíricas, ressaltando a necessidade de modelos conceituais que as suportem (Vangen, 2016).

O atual trabalho propõe uma alternativa para a operacionalização de pesquisas que adotem essa perspectiva. O modelo proposto é inspirado no Multilevel Framework (Pozzebon, Diniz & Jayo, 2009), que utiliza os conceitos do contextualismo (Pettigrew, 1985, 1990) para articular elementos do Social Shaping of Technology (SST) (ex. Bijker & Law, 1992; Pinch & Bijker, 1984; Walsham, 2002) e da tecnologia-na-prática (Orlikowski, 2000). Para potencializar a contribuição do modelo na exploração do reúso de dados, alguns elementos de sua configuração original foram substituídos.

O processo passou a ser representado de forma visual para intensificar sua capacidade de apresentar os eventos significativos e a dinâmica entre eles, buscando promover a reflexão sobre esse processo. Para Langley (1999), as representações gráficas, com sua grande capacidade de síntese e engajamento, são uma forma estimulante de apresentar processos em pesquisas acadêmicas, especialmente quando combinadas com outros métodos que permitam aprofundar a análise das questões cuja superfície foi exposta pelo “desenho”.

A tecnologia-na-prática foi substituída pelo reúso-de-dados-na-prática, que preserva a noção original de emergência da estrutura a partir da prática social, mas evidencia as múltiplas dimensões desse fenômeno para reforçar suas diversas facetas e estimular a discussão sobre estas. Tal debate atende ao chamado dos pesquisadores para a ampliação das perspectivas dos estudos sobre o reúso de dados (Custers & Ursic, 2016; Newell & Marabelli, 2015). Foram consideradas as dimensões sugeridas por Custers e Bachlechner (2017) – legal, tecnológica, societal e econômica – e acrescentada a dimensão intrínseca, que representa as questões relacionadas à dinâmica do grupo diretamente envolvido no reúso de dados.

Para exercitar o modelo proposto, esse trabalho explorou o caso da Alpha no Brasil. A empresa oferece uma solução que utiliza dados da rede de telefonia celular para compor perfis comportamentais que são usados para análise de crédito. Tal aplicação visa colaborar para a expansão do acesso ao crédito (especialmente para os excluídos do sistema financeiro tradicional) e viabilizar serviços financeiros inovadores. No entanto, ao utilizar dados de telefonia para um serviço financeiro, seu ecossistema reúne atores que possuem objetivos, interesses e racionalidades diferentes. As assimetrias de poder entre eles geram desequilíbrios que podem inviabilizar todo o ecossistema. Assim, o estudo revelou como os atores lidaram com esses desafios, criando alternativas para seguir em frente.

O estudo oferece três contribuições. Primeiro, oferece *insights* interessantes sobre a dinâmica do ecossistema de reúso de dados para lidar com os desequilíbrios que emergem das assimetrias de poder. No caso da Alpha, a manobra realizada viabilizou o reúso dos dados, ainda que não fosse para seu objetivo inicial. O caso exploratório não pretende fazer generalizações, mas sugere reflexões sobre as duas epistemologias de poder presentes nele – poder como posse ou como prática (Tello-Rozas, Pozzebon & Mailhot, 2015). Tais visões permitem discutir as dificuldades ou oportunidades que os contextos pluralistas oferecem para a agência lidar com os desequilíbrios do ecossistema na prática. Segundo, ao posicionar o reúso de dados como um fenômeno que emerge em um ecossistema pluralista e marcado pela colaboração complexa, destacando suas dimensões no reúso-de-dados-na-prática, este trabalho discute conceitos e elementos que colaboram para a exploração teórica e prática do reúso de dados. Terceiro, ao utilizar o Multilevel Framework para implementar a perspectiva prática, o trabalho apresenta uma alternativa para a realização de estudos empíricos usando a abordagem prática para investigar contextos complexos, cada vez mais frequentes nas organizações contemporâneas.

Após essa introdução, o artigo expõe suas bases teóricas, apresentando os conceitos do reúso de dados e dos ecossistemas e sugerindo as lentes conceituais da prática para estudá-los. Em seguida, o artigo apresenta o Multilevel Framework e propõe ajustes ao modelo para adaptá-lo à pesquisa do reúso de dados. Para exercitar o modelo proposto, realiza um estudo de caso exploratório, em que sua metodologia e seus resultados são demonstrados na sequência. A discussão dos resultados, conclusões finais do estudo, seus limites e sugestões para futuros trabalhos encerram o artigo.

3.3 Reúso de dados na era do *big data*

Poucos estudos definem o conceito de reúso de dados, mas há certa conformidade entre os pesquisadores sobre o fato de que ele se refere ao uso dos dados para fins diferentes daqueles para os quais foram gerados (Curty 2015; Faniel e Jacobsen, 2010; Yoon, 2015; Zimmerman, 2003, 2008). Essa não é uma prática nova, mas o avanço tecnológico e a crescente oferta de dados trazidos pelo BD elevaram as possibilidades de seu reúso a patamares sem precedentes (Baesens et al., 2016; Curty, 2015; Martin, 2015; Muller et al., 2016). Ele está cada

vez mais presente na vida de organizações e cidadãos, porém poucos estudos acadêmicos têm debatido sobre isso, especialmente nas perspectivas amplas que um fenômeno complexo deste requer (Custers & Ursic, 2016; Newell & Marabelli, 2015).

Fontes mais diversificadas de informação implicam maior pluralidade de grupos sociais envolvidos, com diferentes interesses e pressupostos. Por outro lado, potentes radares por todos os lados em busca de informações interessantes para diversos fins pode nos levar a questões éticas, de propriedade, confiabilidade e segurança da informação, privacidade, e direito individual e coletivo que ainda precisam ser discutidas (Borgman, 2015; Boyd & Crawford, 2012; Custers & Ursic, 2016; Newell & Marabelli, 2015; Samarajiva & Lokanathan, 2016; Zuboff, 2015). Além disso, análises sumarizando dados ou produzindo conclusões a partir deles demandam não apenas soluções tecnológicas que suportem um processamento mais sofisticado, mas também bases metodológicas firmes, que confirmem rigor a todo o processo, da captura dos dados iniciais à produção dos resultados finais (Muller et al., 2016; Newell & Marabelli, 2015). Diversos pesquisadores alertam para a importância da participação acadêmica nesse debate, construindo bases teóricas e metodológicas que poderão apoiar seu desenvolvimento no campo da ciência e da prática (Baesens et al., 2016; Borgman, 2015; Newell & Marabelli, 2015).

O reuso de dados pode acontecer em diferentes situações, que foram mapeadas a partir de duas dimensões: fonte e finalidade (Luvizan & Diniz, 2017). A fonte refere-se à origem dos dados, podendo ser pública, privada interna à organização que está os utilizando ou privada de uma organização externa (Baesens et al., 2016). Já a finalidade concerne à relação entre os fins para os quais os dados foram gerados e os fins de seu novo uso, podendo pertencer a temáticas relacionadas (repropósito) ou não (recontextualização) (Custer & Ursic, 2016). As diferentes situações de reuso que resultam do cruzamento dessas duas dimensões está representada na tabela 3 O potencial de desafios cresce à medida que aumenta a distância entre a geração e o reuso de dados (Aleksynska & Cazes, 2016, Custer & Ursic, 2016; Yoon, 2015, Zimmerman, 2008). Isso pode estar ligado à dimensão legal, tecnológica, societal ou econômica do reuso (Custers & Bachnechler, 2017). Considerando o perfil desses desafios, as categorias geradas pela matriz fonte-finalidade foram agrupadas em ‘Direto’, ‘Intermedial’ ou ‘Extremo’, conforme destacado na tabela 1.

		Fonte dos Dados			
		Dados Privados Internos	Dados Públicos	Dados Privados Externos	
Finalidade Reuso dos Dados	Repropósito	Repropósito de Dados Internos	Repropósito de Dados Públicos	Repropósito de Dados Privados Externos	Intermedial
	Recontextualização	Recontextualização de Dados Internos	Recontextualização de Dados Públicos	Recontextualização de Dados Privados Externos	Extremo
		Direto			

Quadro 3.1 - Matriz finalidade x fonte do reúso de dados

Fonte: Elaborada pela autora.

O grupo de reúso ‘Direto’ refere-se ao repropósito ou à recontextualização de dados privados internos, ou seja, da própria organização que faz o reúso. A organização tem acesso direto à geração dos dados (suas ferramentas, seu contexto, seus atores). Ainda que possa enfrentar desafios ao propor o novo uso dos dados, seu potencial tende a ser menor do que o observado nas demais situações. Já o grupo ‘Intermedial’ diz respeito ao repropósito de dados externos públicos ou privados. Nesse grupo, cresce o potencial de desafios, já que a organização que promove o reúso deverá articular-se com atores externos à organização, o que pode trazer diversas questões de ordem legal, tecnológica, societal e econômica. Por fim, o grupo ‘Extremo’ representa a recontextualização de dados externos privados ou públicos. Ele apresenta o maior potencial de desafios. Ao reusar dados para finalidades fora de sua temática original, esse grupo pode demandar interação entre atores que pertencem a áreas distintas, regidas por diferentes racionalidades e objetivos.

A reflexão sobre as diferentes situações dessa matriz sugere que o trânsito entre fontes e áreas temáticas pode oferecer oportunidades de novas soluções e negócios, mas também aumenta a diversidade de atores envolvidos e a complexidade de suas relações. Compreender essa dinâmica, portanto, parece relevante para avançar na teoria e na prática do fenômeno. A analogia com a teoria dos ecossistemas tem sido utilizada para discutir ambientes dinâmicos em diversas áreas, podendo oferecer elementos interessantes para estudar o reúso de dados.

3.4 Ecossistema de reúso de dados pelas lentes da prática

Esta seção explora a ideia dos ecossistemas na pesquisa organizacional, posiciona o reúso de dados como um fenômeno que emerge em um ecossistema pluralista e apresenta as

lentes teóricas da prática como uma alternativa para abordá-lo. Para operacionalizar tal perspectiva, sugere a utilização do modelo Multilevel Framework com algumas adaptações para investigar um caso de reúso de dados, destacando as múltiplas dimensões nele implicadas.

3.4.1 Conceitos do ecossistema e sua aplicação ao reúso de dados

O termo ‘ecossistema’ surgiu na biologia e se aplica a um conjunto de elementos, sejam organismos vivos, materiais não-vivos ou fluxos de energia, reunidos em um ambiente no qual interagem, formando um sistema (Beyers, 1964). A partir da década de 1980, a analogia desse conceito passou a ser utilizada em diversas áreas de estudo (Pilinkiene & Maciulis, 2014), dando origem a teorias mais específicas, como a do ecossistema industrial (Frousch & Gallopoulos, 1989), ecossistema de negócios (Moore, 1993), ecossistema de negócios digitais (Nachira, 2002), ecossistema da inovação (Adner, 2006), ecossistema do empreendedorismo (Isenberg, 2010), ecossistema de inovação digital (Rao & Jimenez, 2011) e ecossistema de inovação aberta (Chesbrough, Vanhaverbeke & West, 2014).

Cada abordagem se apropriou da teoria dos ecossistemas de forma distinta para atender às suas demandas de pesquisa, mas em todas há uma definição dos atores e do ambiente em que interagem e uma clara identificação dos impactos no nível micro e macro do sistema (Pilinkiene & Maciulis, 2014). Tal semelhança sugere que essas analogias nas diferentes áreas compartilham a ambição de produzir análises mais dinâmicas e holísticas, que integram a visão do todo e da parte dos fenômenos que observam.

Moore (1993) propõe o termo ecossistema de negócios para designar o complexo ambiente de negócios em que as organizações e os indivíduos interagem, incluindo clientes, fornecedores, concorrentes, intermediários, fornecedores de produtos complementares, agências reguladoras e distribuidores (Moore, 1996; Iansiti & Levien, 2004). Todo ecossistema precisa de um tempo para sua formação e possui um ciclo de vida dividido em 4 fases: nascimento, expansão, estabilidade e autoridade (Moore, 1996). A interação entre os atores, cooperativa ou competitiva, determina o desenvolvimento dos ecossistemas, e a combinação do conhecimento e das capacidades de seus participantes pode gerar novos negócios (More, 1996). O ecossistema que não evolui e se renova está fadado à morte (Annanpera, Liukkunen & Markkula, 2015). Dialogando com essas ideias, surge o conceito do ecossistema da inovação.

O ecossistema da inovação é um ecossistema de negócios orientado à geração de inovações (Adner & Kapoor, 2010). Ele é altamente dinâmico para acompanhar a evolução das demandas e das condições disponíveis para atendê-las com agilidade (Adner, 2006). Os componentes desse ecossistema também podem variar ao longo do tempo, uma vez que seus participantes não são definidos por uma estrutura fixa, como a de uma indústria, mas pelas condições mutantes do ambiente e de seus participantes (Adner, 2006). O alinhamento dos objetivos, a complementaridade do conhecimento, a confiança e a sintonia entre os atores são alguns dos fatores fundamentais para o desenvolvimento e a sobrevivência do ecossistema da inovação (Annanpera *et al.*, 2015). Também é importante algum tipo de liderança, que normalmente será exercida pelos que buscam os benefícios financeiros que a inovação pode trazer ou pelos que possuem maior conhecimento ou recurso (Annanpera *et al.*, 2015).

A noção de ecossistema é propícia para investigar o reúso de dados na perspectiva ampla que o fenômeno requer. Os novos usos dos dados e as soluções que eles viabilizam emergem em um contexto marcado pela pluralidade. Os geradores de dados podem pertencer a grupos diferentes de seus novos consumidores. O processo também pode envolver outros participantes, como intermediários, técnicos e legisladores. Estes atores, oriundos de diferentes grupos sociais, devem colaborar para que o reúso aconteça. Ao cruzar fronteiras entre áreas, funções e culturas, a colaboração torna-se complexa (Faraj *et al.*, 2012), potencializando ambiguidades, desentendimentos e conflitos de interesses e objetivos, o que dificulta a colaboração (Beyerlein, Johnson & Beyerlein, 2004), que pode acontecer em um contexto pluralista, no qual atividades baseadas em conhecimento são realizadas em meio a múltiplos objetivos e poderes difusos entre os atores (Denis, Langley & Rouleau, 2007). A figura 3.1 abaixo ilustra a articulação desses conceitos no ecossistema de reúso de dados.

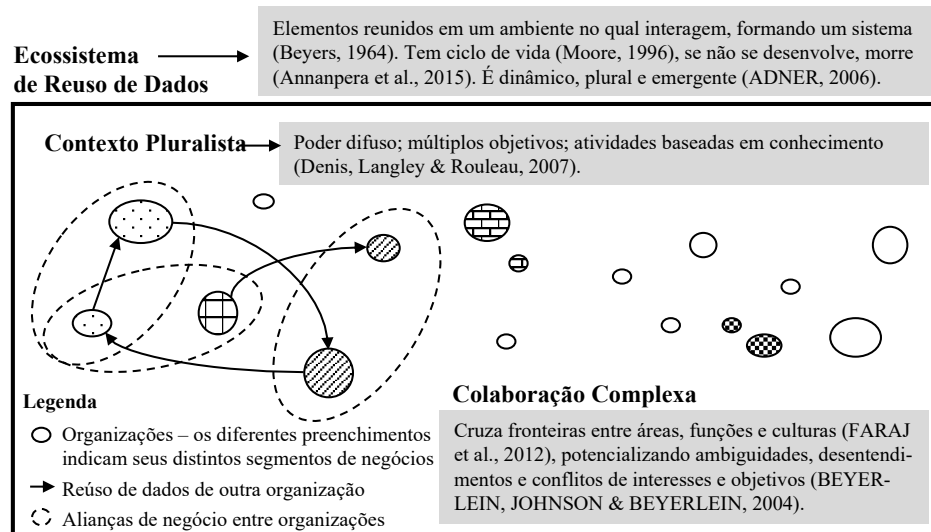


Figura 3.1 - Reuso de dados em um ecossistema de negócios

Fonte: elaborada pela autora

O potencial de desafios do reuso de dados cresce à medida que sua distância da geração de dados aumenta, pois pode implicar na diversidade de atores e pluralidade do contexto em que se dá o reuso (Custers & Ursic, 2016; Zimmerman, 2008). É relevante, portanto, compreender o ecossistema que reúne tais agentes, o contexto e a dinâmica de suas relações e ações. Pesquisadores reconhecem a importância de abordagens amplas, que compreendam todo o ecossistema dos fenômenos, especialmente os mais diversos e complexos (Vangen, 2016). O desafio reside na necessidade de modelos teóricos que permitam operacionalizar essa perspectiva em pesquisas empíricas (Vangen, 2016). As lentes teóricas da prática podem oferecer uma boa alternativa (Tello-Rozas, Pozzebon & Mailhot, 2015; Vangen, 2016), a qual este trabalho irá explorar a seguir.

3.4.2 Teorizando a Prática

As lentes teóricas da prática partem da ideia de que a vida social é construída pela ação recorrente dos atores que dela participam (Denis, Langley & Rouleau, 2007; Feldman & Orlikowski, 2011). Elas vêm ganhando espaço no cenário acadêmico nos últimos anos, de modo que as organizações são cada vez mais vistas como um ambiente “complexo, dinâmico, distribuído, móvel, transitório e sem precedentes” (Feldman & Orlikowski, 2011, p. 1240). Novas composições empresariais, arranjos colaborativos e atividades baseadas no

conhecimento demandam estruturas e processos organizacionais alternativos (Denis, Langley & Rouleau, 2007; Vangen, 2016; Zammuto, Griffith, Majchrzak, Dougherty & Faraj, 2007). Na esfera pública, os problemas e desafios também têm se tornado mais complexos e multifacetados, requerendo novas formas de organização social e envolvendo agentes públicos, privados ou outros tipos de entidades (Vangen, 2016).

Pesquisar tais cenários requer novas abordagens teóricas que sejam capazes de abarcar sua dinâmica e emergência (Denis, Langley & Rouleau, 2007; Feldman & Orlikowski, 2011; Vangen, 2016, Zammuto et al., 2007). As lentes teóricas da prática podem ser uma alternativa interessante para atender a essa demanda (Denis, Langley & Rouleau, 2007; Feldman & Orlikowski, 2011; Vangen, 2016).

Feldman e Orlikowski (2011) destacam que a abordagem da prática nas pesquisas pode adotar um foco empírico, teórico ou filosófico, que podem ser combinados nos estudos, ainda que algum deles seja prevalente. O foco empírico direciona grande atenção à ação das pessoas nos contextos organizacionais, pois reconhece a importância das práticas para a operação e os resultados da organização (Feldman & Orlikowski, 2011). Ele é adequado para abordar questões do tipo “o que”, pois propõe um olhar sobre rotinas, improvisos e emergências que compõem o dia a dia organizacional (Feldman & Orlikowski, 2011). O foco teórico utiliza os elementos das teorias da prática visando buscar explicações para a atividade organizacional, sendo adequado para questões do tipo “como” (Feldman & Orlikowski, 2011). Ele busca compreender as dinâmicas e relações do dia a dia, assim como sua operação em diferentes contextos e ao longo do tempo, engajando-se na lógica de como as práticas são produzidas, reforçadas e modificadas, e suas consequências pretendidas ou não (Feldman & Orlikowski, 2011). Por fim, o foco filosófico propõe a visão do mundo como algo que emerge das atividades diárias, elementos constitutivos da realidade social (Feldman & Orlikowski, 2011). Esse posicionamento ontológico pode estar mais ou menos explícito nas pesquisas e é adequado para questões do tipo “por que” (Feldman & Orlikowski, 2011). Embora os elementos do foco empírico e filosófico estejam presentes nesta pesquisa, o foco teórico destaca-se na busca de compreender como é o processo do qual emerge o reuso-de-dados-na-prática.

As abordagens teóricas da prática compartilham a ideia de que as ações diárias fazem parte da produção da vida social, de forma que o dualismo das teorias deve ser rejeitado e as relações devem ser consideradas mutuamente constitutivas (Feldman & Orlikowski, 2011). Tais ideias dialogam com demandas no campo de SI, dando origem a vertentes de pesquisa que

adotam as teorias da prática, ainda que inicialmente não se intitulassem dessa forma (Zamuto et al., 2007).

A noção da construção social da realidade proposta por Berger e Luckman (1966) inspira a teoria da construção social da tecnologia (SCOT), adotada por diversos pesquisadores de SI a partir dos anos 1980 como uma alternativa ao determinismo tecnológico das teorias dominantes na época (Feldman & Orlikowski, 2011; Pozzebon & Diniz, 2012). A teoria da estruturação de Giddens (1984), propondo que a agência humana produz as estruturas ao mesmo tempo que é habilitada e restrita por elas num processo recursivo e mutuamente constitutivo, também inspirou modelos teóricos que se apropriaram de seus elementos para refletir sobre a adoção e o uso das tecnologias na prática (Orlikowski, 2000). Esses modelos buscavam romper com as visões dicotômicas dos estudos de tecnologia e refletir a importância da agência humana e de suas práticas diárias na construção e no uso das tecnologias (Feldman & Orlikowski, 2011; Orlikowski, 2000; Pozzebon & Diniz, 2012). Na última década, inspirados por autores como Latour (1992), surgiu a proposta da sociomaterialidade, tentando responder à crítica do determinismo social das abordagens anteriores e incorporar a interação dos atores não-humanos às dinâmicas da prática em fenômenos que envolvam a tecnologia (Orlikowski & Scott, 2008).

Apesar da evolução nas discussões teóricas, a aplicação da perspectiva da prática em pesquisas empíricas ainda é desafiadora. Esse tipo de estudo requer uma estrutura capaz de lidar com a complexidade, ambiguidade, pluralidade e emergência que marcam a vida cotidiana (Feldman & Orlikowski, 2011, Vangen, 2016). Demanda também uma lógica e uma linguagem capazes de refletir a essência recorrente e relacional das práticas diárias que constituem o fenômeno social, o que não é o caso das teorias organizacionais que, em sua maioria, mantêm-se focadas em entidades mais estáticas e descontextualizadas (Feldman & Orlikowski, 2011). O modelo Multilevel Framework pode ser uma alternativa para vencer esses desafios, especialmente no estudo de fenômenos que se passam em ecossistemas pluralistas e marcados pela colaboração complexa, como o reúso de dados.

3.4.3 Multilevel Framework aplicado ao ecossistema de reúso de dados

O modelo Multilevel Framework foi concebido para dar suporte a uma pesquisa empírica de SI sobre correspondentes bancários e inclusão financeira, buscando capturar as

perspectivas dos diferentes grupos sociais envolvidos na prática (Pozzebon, Diniz & Jayo, 2009). Ele foi denominado multilevel para expressar sua capacidade de abordar mais de um nível conceitual e analítico, alinhado à abordagem processual e multinível da pesquisa social na tradição construtivista (Pozzebon & Diniz, 2012). O modelo original, apresentado na figura 3.2, combina elementos de três perspectivas teóricas:

Contextualismo – proposto por Pettigrew (1985, 1990), esta perspectiva foi adotada em SI, que necessitava incorporar os aspectos processuais e contextuais dos fenômenos que investigava. O contextualismo considera as ações e percepções dos atores e seu contexto, considerando seu histórico e evolução ao longo do tempo como parte de seu processo constitutivo (Pettigrew, 1985, 1990). Ele é capaz de conectar diversos níveis de análise (individual ou grupal) e o processo pelo qual interagem (Pettigrew, 1990). Nesta perspectiva, contexto, processo e conteúdo são elementos igualmente importantes, uma vez que se influenciam mutua e continuamente, não podendo ser dissociados (Pettigrew, 1985, 1990). O Multilevel Framework utiliza tais elementos como fio condutor para articular os demais conceitos (Pozzebon, Diniz & Jayo, 2009).

Social Shaping of Technology (SST) – esta corrente teórica, com diversos adeptos e abordagens, traz em seus fundamentos a centralidade dos atores e suas ações (Bijker & Law, 1992; Pinch & Bijker, 1984; Pozzebon & Diniz, 2012). Ela também se atenta às relações e dinâmicas entre esses atores no ambiente em que interagem, gerando impacto sobre a tecnologia (Bijker & Law, 1992; Pinch & Bijker, 1984; Pozzebon & Diniz, 2012). O Multilevel Framework apreende da SST os elementos que compõem o contexto, que são os grupos sociais relevantes para o objeto de estudo, seus papéis e pressupostos tecnológicos que interferem em suas ações (Pozzebon, Diniz & Jayo, 2009). Ele também se inspira na SST para descrever o processo com base nos mecanismos de negociação utilizados pelos atores envolvidos (Pozzebon, Diniz & Jayo, 2009).

Tecnologia-na-prática – proposto por Orlikowski (2000), este modelo inspira-se na construção social e emergente da realidade a partir da prática de seus atores, de modo a considerar que a tecnologia se estabelece no contexto social por meio da interação destes, influenciando a construção e o uso da tecnologia, ao mesmo tempo que também são influenciados por ela em uma dinâmica contínua e recursiva (Orlikowski, 2000). No Multilevel Framework, a tecnologia-na-prática representa o conteúdo, resultado do processo ocorrido no contexto do fenômeno em estudo (Pozzebon, Diniz & Jayo, 2009).

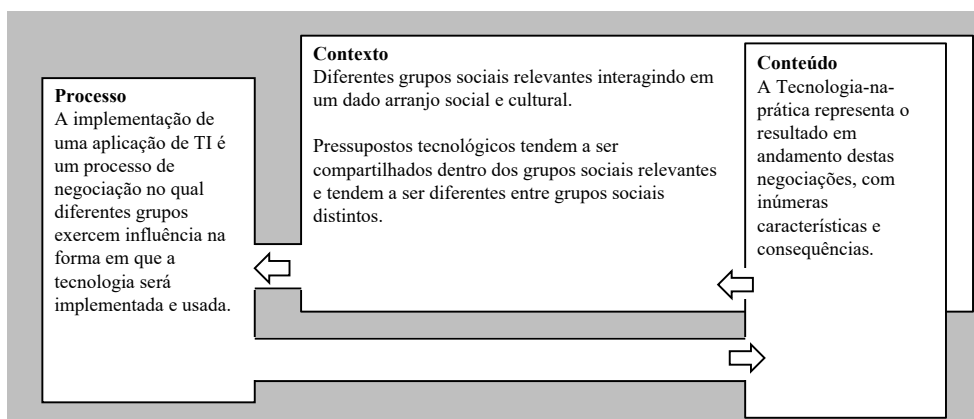


Figura 3.2 - Modelo Multilevel Framework original
 Fonte. Pozzebon, Diniz e Jayo (2009) – traduzido pela autora.

Embora a explicação do modelo possa dar a entender que há uma sequência linear entre seus componentes, todos os elementos interagem simultaneamente em um processo no qual se influenciam mutuamente, como propõe o contextualismo (Pettigrew, 1990). O Multilevel Framework permite que se analise um recorte do caso em um dado momento no tempo, sem esquecer, no entanto, que ele refere-se a um fenômeno dinâmico e recursivo, o que é uma característica inerente à prática (Pozzebon, Diniz & Jayo, 2009). O modelo colabora oferecendo este racional para o recorte do fenômeno, sem perder os elementos necessários para compreensão de suas características contínuas. A forma como os dados são organizados no modelo também contribui para a apresentação do caso, o que é um desafio para estudos com o enfoque da prática devido à diversidade de informações que pode emergir em seus casos. Tais contribuições são relevantes no estudo do reúso de dados, caracterizado pela pluralidade do ambiente e pela complexidade da colaboração envolvida. A capacidade do modelo de focalizar o fenômeno e compilar dados relevantes na apresentação do caso justificam sua adoção nesse estudo.

O Multilevel Framework já foi utilizado em seu formato original e modificado em diversas pesquisas empíricas, demonstrando a utilidade e a flexibilidade do modelo (Pozzebon & Diniz, 2012; Pozzebon, Mackrell & Nielsen, 2014). Sua pluralidade teórica permite que outros elementos sejam utilizados para melhor atender às questões em análise (Pozzebon & Diniz, 2012; Pozzebon, Mackrell & Nielsen, 2014). Tal possibilidade é oportuna para esse estudo, já que permite a incorporação de conceitos sobre o reúso de dados que esse trabalho pretende destacar, sendo esta mais uma justificativa para sua adoção. O modelo proposto é apresentado na figura 3.3 e contém as seguintes alterações em relação ao original:

Processo – será apresentado por meio de uma representação gráfica que destaca a sequência dinâmica entre os principais eventos que compõem o processo no qual os atores lidam

com os desafios do reúso dos dados. A negociação continua presente, mas está diluída no processo que será representado por um mapa visual. A visão do processo nessa proposta é inspirada em Pettigrew (1997) e Langley (1999), buscando compreender a evolução das coisas ao longo do tempo. Visualizar eventos relevantes ordenados no tempo e suas relações é fundamental para compreender a história do processo (Pettigrew, 1997; Langley, 1999). Os mapas visuais podem ajudar nesse entendimento, pois contribuem para a sintetização dos dados, bem como para a demonstração de precedências e outras relações entre os eventos (Langley, 1999). Os mapas visuais podem assumir diversos formatos e, embora tenham a capacidade de expressar grande volume de informações, podem não ser capazes de apresentar todos os detalhes e ambiguidades do processo ilustrado (Langley, 1999). Por essa razão, o processo será apresentado pelo mapa visual e descrito de forma narrativa.

Conteúdo – será representado pela descrição do reúso-de-dados-na-prática, classificado de acordo com as categorias descritas na seção três. Também serão apresentadas as condições das diferentes dimensões que envolvem o fenômeno de reúso de dados propostas por Custers e Bachlechner (2017): legal, tecnológica, societal e econômica. Propõe-se acrescentar a dimensão intrínseca para retratar as questões relativas ao grupo diretamente envolvido no reúso de dados. Nessa representação, continuam presentes as ideias e a lógica do modelo da tecnologia-na-prática, por meio da ação dos atores, da relação recursiva e mutuamente constitutiva entre eles e da estrutura que emerge do uso da tecnologia. Ao apresentar o reúso-de-dados-na-prática ao invés da tecnologia-na-prática, pretende-se destacar o reúso de dados em si e suas múltiplas dimensões, estimulando a discussão sobre seus diversos aspectos.

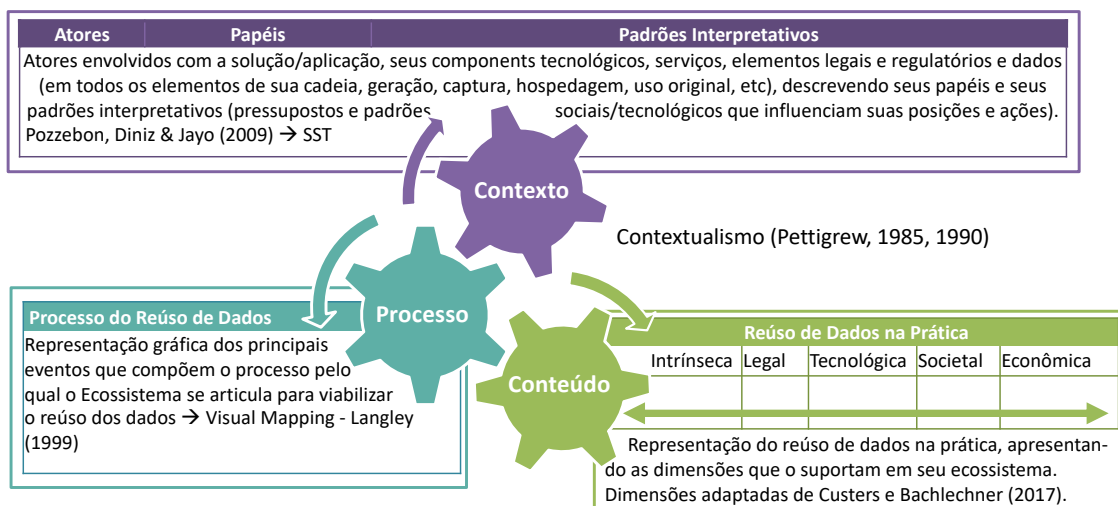


Figura 3.3 - Ecossistema do reúso de dados pelas lentes da prática

Fonte: adaptado do modelo Multilevel Framework de Pozzebon, Diniz e Jayo (2009).

3.5 Metodologia

Para explorar um caso de reúso de dados, este trabalho propõe a abordagem teórica da prática, operacionalizada por um modelo conceitual inspirado no Multilevel Framework (Pozzebon, Diniz & Jayo, 2009). As adaptações do modelo visam destacar o caráter multifacetado do reúso de dados e a necessidade de alinhamento entre suas dimensões. Para exercitar essa proposta e demonstrar a utilização do modelo em um estudo empírico foi realizado um estudo de caso único exploratório.

A estratégia qualitativa e abordagem interpretativista, adotadas na pesquisa, são apropriadas para estudos que busquem compreender a construção de uma realidade social, explorando os significados produzidos em seus contextos (Flick, 2004; Pozzebon, Rodriguez & Petrini, 2014), o que está alinhado à abordagem teórica e aos objetivos do trabalho. O estudo de caso também é adequado para tratar temas nos quais a pesquisa e a teoria são iniciais e/ou se baseiam na prática, em que a experiência dos atores é importante e o contexto de ação é crítico (Pozzebon & Freitas, 1998; Stake, 1995).

A investigação de um único caso, bem selecionado, pode ser uma alternativa interessante para refletir sobre teorias emergentes (Stake, 1995). Para Stake (2008), o estudo de caso não é uma escolha metodológica, mas uma escolha sobre o que deve ser estudado, de modo que a seleção do caso precisa estar alinhada com o que se pode aprender a partir dele. Stake (1995) define três tipos de estudos de caso: intrínseco (quando há um interesse particular no caso), instrumental (quando o caso é estudado como um instrumento para a compreensão de algo específico) e coletivo (quando um conjunto de casos é estudado). A escolha do caso, assim como os métodos de coleta e análise dos dados, deve ser coerente com o posicionamento do caso.

Nesta pesquisa, o estudo de caso é instrumental, pois visa: 1) exercitar o modelo Multilevel Framework adaptado, verificando sua capacidade de exploração do reúso-de-dados-na-prática; 2) revelar questões sobre o caso que colaborem para a compreensão dos desafios que emergem do ecossistema de reúso de dados e como seus atores lidam com eles. Não se pretende aqui esgotar a investigação do fenômeno, mas exercitar empiricamente o modelo conceitual, de forma que sua aplicação seja demonstrada e aponte questões que iluminem novos trabalhos. Vale ressaltar que este trabalho faz parte de uma pesquisa mais ampla e o foco do presente estudo é demonstrar a aplicação do modelo teórico proposto, que será utilizado em

etapas empíricas posteriores. É possível dizer, portanto, que o presente caso tem uma função ilustrativa do modelo teórico dentro da pesquisa maior à qual esse estudo pertence.

Mesmo sendo um caso ilustrativo, é fundamental estabelecer critérios de validade do estudo. Esse trabalho procurou observar os critérios de Pozzebon *et al.* (2014): autenticidade (interação com o material empírico, presença no campo); plausibilidade (construção de uma interpretação lógica, de uma história que faça sentido e soe plausível); criticalidade (engajamento com a interpretação crítica); reflexividade (posicionamento claro dos pesquisadores, seus critérios para a seleção de entrevistados e outras decisões ao longo da coleta de dados, além da identificação explícita de visões pessoais e conteúdos confessionais); arte (criatividade e cultura na expressão das ideias, em favor da transparência e da clareza do trabalho).

Iluminados pelos princípios da plausibilidade, criticalidade e arte, buscou-se apresentar o estudo, suas colocações teóricas e o conteúdo empírico de forma clara e coerente, trazendo reflexões críticas e oferecendo elementos para que os leitores possam se engajar nelas. Para atender aos critérios da autenticidade e reflexividade, os procedimentos de coleta e análise de dados, assim como os critérios para a seleção do caso, serão descritos na sequência.

Seleção do caso: o caso da Alpha foi escolhido por representar uma situação prática de recontextualização de dados privados externos, envolvendo atores de grupos totalmente distintos, situação de reuso de dados com potencial extremo de desafios. Selecionar um caso em que o acesso aos dados não fosse público aumentou o potencial de revelar desafios do reuso de dados. A Alpha demonstrou grande experiência em dificuldades de acesso aos dados, relatando passagens ocorridas também em suas operações em outros países. Embora o caso estudado seja limitado à operação no Brasil, a visão do cenário global pareceu oportuna para ajudar a identificar se suas implicações são idiossincrasias do cenário brasileiro ou estão alinhadas ao que a empresa encontra nos demais países. Em suma, o caso apresentou grandes perspectivas de aprendizado e, por isso, foi selecionado.

Coleta de dados: os dados foram coletados por meio de entrevistas, análises de informações públicas e material fornecido pela Alpha. As entrevistas foram realizadas com o diretor geral da Alpha no Brasil, responsável pela implantação da empresa no país. Elas foram realizadas de setembro a novembro de 2015, totalizando 6 horas. A primeira foi conduzida como uma conversa, na qual o entrevistado falou sobre a história da criação da Alpha, bem como sua solução, seu encontro com a empresa, a decisão de instalar uma operação no Brasil, a experiência desta implantação, a estrutura e os serviços locais. A segunda foi semi-estruturada

e buscou compreender melhor o funcionamento da solução e dos negócios no Brasil, explorando de forma particular as dificuldades encontradas em suas primeiras tentativas de implantação. Na terceira, também semi-estruturada, o modelo teórico já estava em fase de elaboração e o objetivo era esclarecer os detalhes necessários para completar a interpretação do caso. Em todas as entrevistas, o personagem foi estimulado a contar as histórias, possibilitando identificar a sequência e relação entre os eventos, fundamental para o modelo conceitual adotado. Elas foram conduzidas pela autora deste artigo e registradas em notas e relatórios produzidos logo após a coleta das informações, capturando algumas falas completas do entrevistado. A Alpha também forneceu um documento relatando um caso brasileiro e indicou locais para consulta de material divulgado pela empresa, como entrevistas do chief executive officer (CEO) global e outros casos. Material do próprio site da empresa e outros conteúdos encontrados na internet sobre a Alpha também ajudaram a compreender sua história, confirmando e enriquecendo as informações obtidas nas entrevistas. Stake (1995) sugere que buscar a acuracidade dos dados e a emergência de explicações alternativas por meio do protocolo de pesquisa é uma forma de triangulação. A utilização de dados de diversas fontes busca acompanhar essa sugestão, lembrando que esta triangulação ocorre na perspectiva interpretativa do estudo e não como uma validação de dados objetiva proposta pelas abordagens positivistas. Vale também destacar que o objetivo principal desse estudo é ilustrar a aplicação empírica do modelo teórico proposto, o que pode ser feito com um conjunto de dados mais limitado como o que temos nesse caso. Etapas posteriores de pesquisa serão dedicadas a estudos de casos com maior profundidade.

Análise de dados: a coleta e análise dos dados foi realizada em fases, como proposto por Miles and Huberman (1994): coleta, redução, apresentação e conclusão (desenho/verificação). Nesse esquema, os autores admitem que possa haver interação entre as etapas e que, chegando à etapa de conclusão, é possível voltar à coleta para obter novos dados e, novamente, iniciar o ciclo de análise de todo o conjunto, o que aconteceu nesta pesquisa. Durante a redução, que é uma fase interpretativa dos dados, as informações coletadas foram analisadas e alocadas no modelo teórico, já procedendo também à etapa da apresentação. Na conclusão, foram realizadas reflexões sobre o resultado produzido, gerando dúvidas sanadas em uma última entrevista para a composição dos resultados finais.

3.6 O Caso da análise de crédito usando dados da rede de celulares

3.6.1 O nascimento da Alpha

A solução da Alpha é baseada em um algoritmo de análise comportamental, que utiliza combinações matemáticas para gerar perfis de comportamento a partir de grandes bases de dados. A ideia de usar esses perfis na análise de risco de crédito surgiu do desejo do criador do algoritmo aplicar seu estudo em alguma solução prática que pudesse contribuir para o desenvolvimento social, bem como do seu encontro com um empreendedor, já envolvido com iniciativas de emancipação financeira. Da troca de experiências entre eles surgiu a Alpha, cujo objetivo é colaborar para a inclusão financeira no mundo, oferecendo ferramentas para avaliação do risco de crédito usando dados presentes no cotidiano das pessoas, independentemente de seu histórico no sistema financeiro tradicional. Considerando a grande penetração da telefonia celular em muitos países, a empresa optou pela utilização dos dados das redes de celulares.

Após uma prova de conceito realizada na Tanzânia, cruzando dados de um banco com os dados da operadora de telefonia móvel local, a ideia confirmou-se viável. Captando o apoio financeiro de investidores e reunindo profissionais altamente qualificados (cientistas de dados, programadores, estatísticos, etc.), a solução foi desenvolvida e apresentada ao mercado como uma inovadora alternativa para a análise de crédito de consumidores emergentes. Como confirma uma declaração do CEO global da Alpha à imprensa: “Nosso intento é monetizar pessoas que vocês tradicionalmente renegaram... Bancos negam grande parte das solicitações recebidas. Nossa pontuação permite atribuir valor para dezenas de milhares de pessoas que estão fora do radar pela falta de informações. Esse é um grande novo mercado a ser atendido”. Fiel ao propósito, a empresa não se propõe a ser um agente financeiro, financiando ou emprestando dinheiro pessoalmente, mas sim a prover uma solução que permita que outros o façam. Desde sua fundação, ela tem apoiado diversas iniciativas envolvendo modelos alternativos para fornecimento de crédito e outros serviços financeiros pelo mundo, que são viabilizados pela sua ferramenta.

3.6.2 Breve histórico da avaliação de risco de crédito

A avaliação do risco de não-pagamento é tema central em operações de crédito, impactando a seleção dos clientes a quem se concederá crédito e taxas de juros envolvidas nessas operações (Stiglitz & Weiss, 1981). A participação acadêmica no debate é antiga, como confirmam pesquisas de Smith e Winakor (1935) e Mervin (1942). O fim da década de 1960 foi marcado por diversos estudos, como o pioneiro trabalho de Altman (1968), propondo um modelo de análise discriminante, que iniciou uma sequência de pesquisas sobre análises preditivas na área (Bruni, Murray & Famá, 1998). Bancos e instituições financeiras têm realizado grandes investimentos na criação de ferramentas para fazer avaliações dos clientes que solicitam crédito, buscando identificar bons e maus pagadores (Stiglitz & Weiss, 1981) e colocando em prática diversos desses modelos acadêmicos.

Normalmente, os sistemas de avaliação de risco de crédito convencionais utilizam, entre outras informações, dados de operações financeiras anteriores dos indivíduos, detalhes sobre seus bens e outras garantias, o que limita o acesso ao crédito para uma parcela significativa da população, que não tem esse histórico. Um estudo realizado nos Estados Unidos com dados de 8 milhões de pessoas propôs a composição do *credit score* (pontuação de crédito) dos indivíduos a partir de dados alternativos, como aluguel, conta de energia elétrica e outras despesas recorrentes (Turner, Lee, Schnare, Varghese & Walker, 2006). Ele argumenta que a ferramenta poderia contribuir para a diminuição da pobreza no país, já que preencheria a lacuna de informações sobre o risco financeiro de milhões de americanos, expandindo o acesso ao crédito da população mais pobre e excluída do sistema financeiro tradicional (Turner *et al.*, 2006). Um estudo brasileiro sobre a relação do consumo de energia elétrica e o nível de renda (De Rezende Francisco, Aranha, Zambaldi & Goldszmidt, 2007), sugere que esse tipo de informação também poderia ser uma alternativa para a construção de modelos de avaliação de crédito mais inclusivos no Brasil.

Muitos programas alternativos de concessão de crédito brasileiros estão ligados a iniciativas de microcrédito voltadas ao empréstimo de pequenos valores para financiar atividades produtivas e pequenos negócios (Diniz, 2007). Apesar da expansão consistente do microcrédito em todo o mundo nas últimas décadas e da entrada dos bancos comerciais nesse mercado a partir dos anos 1990, as operações de microcrédito ainda se concentram em bancos públicos e iniciativas de fomento ao desenvolvimento social públicas ou privadas (Diniz, 2007).

Os bancos veem nessa parcela da população uma alternativa para compensar a saturação de serviços financeiros em seu público histórico e têm desenvolvido programas para expandir sua atuação nesse novo nicho. Porém, a falta de informações reconhecidas pelos sistemas financeiros tradicionais eleva o risco e o custo de transação dos bancos comerciais, o que desestimula sua abordagem mais agressiva.

Enquanto isso, proliferam-se pelo mundo iniciativas de novos arranjos e modelos de negócios para esse mercado. Movidos por objetivos mais amplos e munidos de novas tecnologias, tais iniciativas têm sido lideradas pelas chamadas *fintechs* (Siqueira, Diniz & Albino, 2018). Suas soluções visam democratizar o acesso aos serviços financeiros e permitem a atuação de novos agentes de financiamento e crédito, sejam eles empresas ou indivíduos trabalhando em colaboração (Mackenzie, 2015; Siqueira, Diniz & Albino, 2018). Atenta a esses movimentos, a Alpha acredita que soluções para análise de risco, que não dependam de bases históricas de domínio exclusivo de grandes instituições financeiras e que não sejam limitadas a uma parcela da população que já participa do sistema financeiro tradicional, podem ajudar a viabilizar o crescimento e a consolidação dos modelos emergentes.

3.6.3 A Solução da Alpha

A solução da Alpha representa uma alternativa para preencher a lacuna apresentada acima, uma vez que utiliza dados da rede de telefonia celular para traçar o perfil comportamental do cliente. Como nos sistemas preditivos tradicionais, essa ferramenta aplica um modelo de análise discriminante, no qual o cliente é comparado a uma base de perfis e sua pontuação é definida a partir do histórico de perfis semelhantes ao dele.

Atenta à privacidade dos indivíduos, a ferramenta não acessa informações sobre o conteúdo das ligações ou mensagens de texto enviadas e recebidas pelos usuários. Os dados utilizados são restritos aos registros capturados pela rede de telefonia celular, conhecidos como Call Detail Record (CDR). Tais dados, normalmente utilizados para bilhetagem e faturamento, são guardados por tempo determinado e depois descartados pelas operadoras devido ao seu gigantesco volume. Os dados contidos nos CDRs podem variar entre as operadoras, mas, basicamente, contêm informações como número de origem e destino das chamadas, data e hora de início da ligação, duração e o meio pelo qual a chamada foi iniciada e interrompida. A

ferramenta da Alpha não copia esses dados para sua base, apenas os lê para definição de comportamentos que eles representam. Tais comportamentos ficam armazenados na base da solução da Alpha sem qualquer identificação individual, como o número do telefone ou dados de localização.

3.6.4 O caso da Alpha no Brasil

Acreditando no potencial do mercado brasileiro para esse tipo de solução e vislumbrando os benefícios que ela poderia trazer para o desenvolvimento local, a Alpha abriu um escritório no Brasil. Para analisar tal experiência, usando a lente conceitual proposta neste trabalho, mapeou-se o contexto no qual o processo entre os atores deu origem ao reuso-de-dados-na-prática como conteúdo resultante dessa dinâmica. Tais elementos serão apresentados a seguir.

3.6.4.1 Contexto

O contexto inicial encontrado pela Alpha no Brasil era promissor, mas também desafiador. De acordo com os dados levantados pela empresa, quase metade da população adulta do país não possui conta em bancos. Já a telefonia celular chega a um contingente bem maior, confirmando sua tese de que muitos cidadãos que não participam do sistema financeiro convencional poderiam ser beneficiados por sua solução. O mercado de telefonia celular no Brasil está concentrado, majoritariamente, em quatro grandes operadoras. Tais empresas, detentoras de estruturas robustas estão atentas às possibilidades de novos serviços aos seus clientes, já que esse mercado é extremamente competitivo. No entanto, operam em um ambiente regulado por agências públicas, que determinam e monitoram o cumprimento de normas de serviço e mercado.

A imensa e crescente base de usuários de telefonia móvel, gera um gigantesco volume de dados de operação da rede. Essas informações são pouco utilizadas pelas empresas para ações de inteligência, normalmente limitadas a demandas internas. Os usuários, por sua

vez, embora tenham aderido rápida e fortemente ao uso de celulares, desconhecem detalhes mais técnicos sobre o funcionamento da rede e, em sua maioria, ignoram a existência e o conteúdo dos CDRs. Em contrapartida, estão cada vez mais atentos a questões de privacidade e segurança da informação, adotando, aos poucos, uma postura mais exigente junto a operadoras e órgãos reguladores.

A Alpha encontrou também um sistema financeiro altamente informatizado no Brasil, liderado por grandes bancos que realizaram investimentos significativos no desenvolvimento de seus sistemas. Eles armazenam grande volume de informações sobre seus clientes e transações, e possuem ferramentas para análise de risco baseadas em tecnologia de ponta e modelos estatísticos amplamente reconhecidos no meio. Operando em um ambiente no qual a informação é um bem muito valioso, possuem rígidas normas de controle e segurança de dados. A regulamentação desse mercado é realizada por um conjunto de órgãos públicos que visam organizar a operação e os serviços prestados à sociedade, considerando também questões legais, políticas, macroeconômicas e de segurança.

Para realizar seu objetivo de contribuir para a democratização do crédito no Brasil, a Alpha deveria oferecer sua solução para empresas que possam prover qualquer tipo de crédito, especialmente para a parcela da população que não seria qualificada pelos sistemas de avaliação de risco tradicionais. Nesse cenário, utilizar os dados da rede de telefonia móvel, fortemente difundida nas diversas camadas sociais do país, apresenta-se como uma alternativa adequada aos propósitos da Alpha. Por outro lado, tal opção coloca a solução no quadrante mais desafiador na matriz de reúso de dados: recontextualização de dados privados externos. A tabela 2 sumariza esse contexto.

Contexto		
Atores	Papéis	Padrões interpretativos
Companhias de telefonia	Captura e armazena dados da rede de telefonia — CDR.	Progressistas; têm visão limitada do potencial do CDR; sensíveis às pressões da legislação de telecom e da opinião pública.
Usuários de celulares	Gera dados, sendo seu legítimo proprietário.	Numerosos e distribuídos em todas as classes sociais; desconhecem o conteúdo do CDR; sensíveis à privacidade dos dados; ignoram como perfil comportamental pode ser capturado dos dados.
Agências reguladoras: telecom — governo	Definem regras e monitoram a conformidade das empresas Telecom.	Responsáveis pelas regras e procedimentos a serem seguidos pelos provedores de telecom para regular o mercado e proteger a privacidade dos usuários; sensível à opinião pública; regulam o uso comercial dos dados dos usuários.
Bancos	Capturam e armazenam registros das transações financeiras e de crédito dos clientes.	Conservadores; sensíveis às pressões da legislação financeira e da opinião pública; fortemente atentos à segurança da informação e evasão de dados; ligados a modelos existentes de análise de crédito; detentores de grandes volumes históricos financeiros; usuários de sistemas proprietários.
Agências reguladoras: bancos — governo	Definem regras e monitoram a conformidade de bancos e empresas financeiras.	Responsáveis por regras e procedimentos a serem seguidos pelos provedores de serviços financeiros para regular o mercado e proteger a privacidade das pessoas; sensíveis às pressões políticas, macroeconômicas e à opinião pública; Regulam o uso comercial dos dados dos clientes.
Provedores de crédito	Operam no mercado de crédito (financiam, emprestam, etc.)	Perfil diverso: pode ser uma loja financiando seus clientes, um provedor de serviços financeiros, um banco ou outra entidade provendo crédito.
Alpha	Provê solução para avaliação de risco de crédito baseada em dados de CDR.	Progressista; orientada à tecnologia e ao conhecimento; clara visão sobre o potencial do CDR, comprometida em gerar alternativas para o desenvolvimento baseadas em inteligência de dados.

Quadro 3.2 - Contexto do caso Alpha

Fonte: elaborado pela autora

3.6.4.2 Processo

Embora a Alpha esteja convicta de que sua ferramenta não ofereça riscos à privacidade dos indivíduos ou à segurança da informação das empresas envolvidas, ainda falta informação entre os atores e uma legislação clara que suporte esse tipo de atividade. Tal situação inibe o interesse das empresas de celular em dar acesso às informações para geração de perfis e dos bancos em apoiar a validação do *credit score* da Alpha com base em sua experiência no mercado brasileiro, ainda novo para a Alpha.

Com pouca convicção sobre os reais benefícios que a solução pode trazer aos seus negócios e muitas preocupações e riscos no horizonte, esses atores não liberam o acesso às informações que a Alpha precisa para a operação de sua solução. A restrição de acesso à informação e a diferença da racionalidade e de interesses entre os atores envolvidos geram

assimetrias que ameaçam o ecossistema da Alpha, situação que a empresa já havia enfrentado em sua entrada em outros países. Ainda que sua ferramenta tenha condições técnicas, os desequilíbrios em seu ambiente inviabilizam o serviço proposto pela Alpha.

Confiante na qualidade de sua solução e no potencial de sua atuação no Brasil, a Alpha mudou sua estratégia, ampliando a oferta de serviços para incorporar o atendimento de demandas internas das empresas. Embora esteja aberta a atender empresas de qualquer segmento, seu foco principal são os bancos e as operadoras de celular, visando acelerar o conhecimento desses atores sobre sua solução para viabilizar outros negócios no futuro.

A estratégia também pretende atrair negócios que possam sustentar financeiramente o funcionamento da Alpha enquanto os serviços planejados originalmente não se viabilizam. Além disso, os objetivos dessa mudança estratégica estão relacionados à divulgação da ferramenta da Alpha e à conquista de sua legitimidade, criando casos reais que demonstrem seu potencial de aplicação, funcionamento e forma de acesso, e uso de dados. Tal experiência pretende gerar confiança entre os envolvidos (com especial atenção aos bancos e às operadoras de celular), concentrando-se em negócios claramente permitidos pela legislação e ajudando a fortalecer a credibilidade da Alpha. A empresa acredita que esses elementos irão contribuir para viabilizar sua oferta de soluções voltadas ao reúso de dados de celulares para avaliação de crédito, visando a inclusão financeira. A figura 3.4 sumariza este processo.

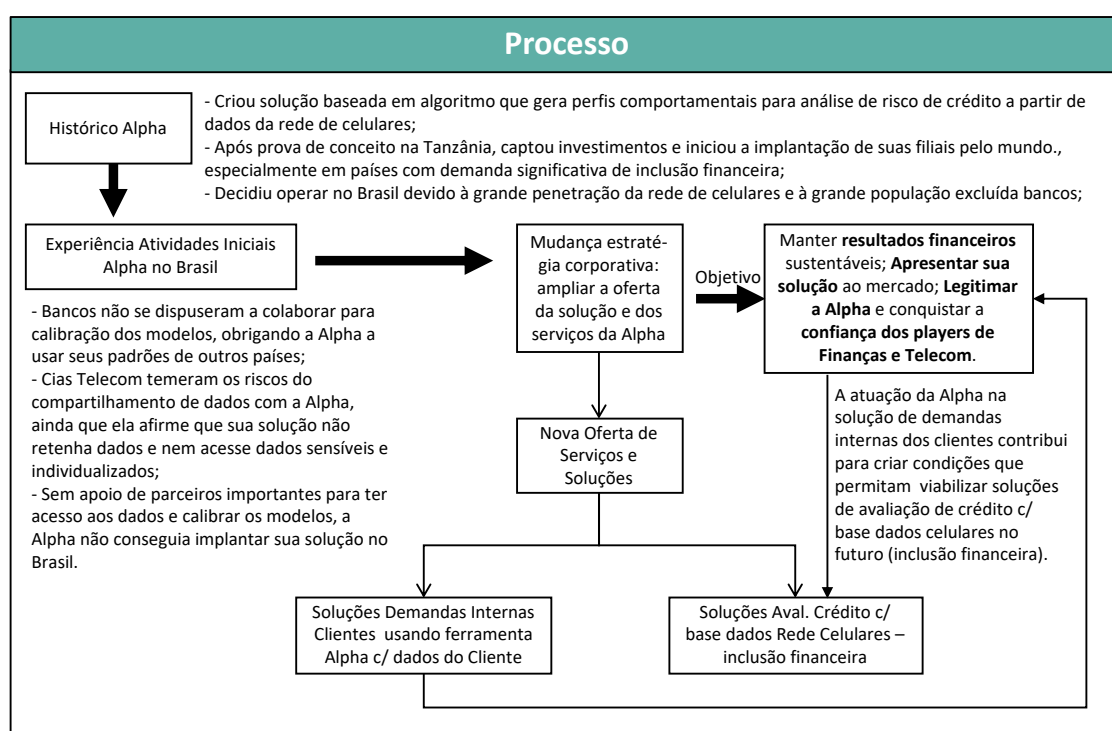


Figura 3.4 - Processo do caso da Alpha

Fonte: elaborada pela autora

3.6.4.3 Conteúdo

A Alpha abraça a oportunidade de oferecer seus serviços para a *joint venture* de um banco com uma operadora de telefonia celular, negócio oportuno para a sua entrada no mercado. O objetivo da parceria é oferecer crédito aos clientes da operadora para financiar a compra de seus produtos e serviços. Por razões legais, o banco não pode compartilhar informações dos clientes com a operadora e vice versa. A solução da Alpha acessa os dados da operadora e gera os modelos comportamentais que permitem indicar o risco financeiro da oferta de crédito ao cliente, assim como sua propensão a adquirir serviços da operadora, direcionando as ações comerciais. Nesse caso, a Alpha colabora para que a operação conjunta atenda aos requisitos legais dos dois segmentos, financeiro e de telecomunicações. Além disso, por meio de sua relação com a *joint venture*, oferece ao banco e à operadora a experiência sobre o funcionamento de sua solução e a robustez de seus mecanismos de segurança e privacidade. Com o mesmo propósito, a Alpha possui outros projetos tanto em empresas de telefonia quanto em segmentos distintos, sempre focados na leitura de informações da empresa para atender suas próprias demandas internas.

Ao adotar a estratégia, a Alpha simplificou seu ecossistema, ao mesmo tempo que mantém atividades que exercitam o uso do conceito de sua solução, limitada aos dados disponibilizados por seus contratantes. Mesmo no caso em que utiliza informação do banco e da operadora de celular, a Alpha o faz sob a demanda de uma *joint venture*, que contrata e monitora a aplicação da ferramenta, sendo também a usuária da informação produzida, operação que obedece ao regime legal estabelecido. Também é essa *joint venture* que orquestra os interlocutores do banco e da operadora para prover os acessos e as interfaces necessárias para viabilizar a ação, o que requer tecnologias já disponíveis na estrutura dos envolvidos e com as quais já estão familiarizados. A operação é economicamente atrativa, pois os custos são pagos com folga pela receita que o negócio gera, oriunda do serviço de crédito e da venda dos produtos e serviços da operadora. A mecânica da operação é restrita aos parceiros, reduzindo a exposição pública do modelo e o risco de questionamento popular. Os envolvidos também não têm dúvidas sobre a adequação ética do negócio, já que não expõem os dados dos clientes a nenhum tipo de risco e utilizam os dados para ampliar a oferta aos próprios consumidores, que podem se beneficiar das linhas de crédito. O reúso de dados nessa operação pode ser classificado como recontextualização de dados internos, que representa o uso de dados

existentes no ambiente da própria empresa, para aplicação não relacionada aos fins originais para os quais foram gerados. O quadro 3.3 sumariza esse conteúdo.

Conteúdo				
Reúso-de-dados-na-prática: recontextualização de dados internos Reúso de dados da empresa de telefonia celular para elaborar perfis de crédito que são usados pela joint venture telecom-banco na concessão de crédito aos clientes de telecom para comprarem seus equipamentos e serviços				
Intrínseca	Legal	Tecnológica	Societal	Econômica
As três partes têm interesse no caso: Alpha – apresentar sua solução ao mercado e aproximar-se de operadoras de telecom e bancos; Operadora – viabilizar o uso dos dados da rede para avaliar o crédito de seus clientes que será concedido por sua <i>joint venture</i> ; Banco – ter parâmetros para conceder crédito aos clientes da operadora por meio de sua <i>joint venture</i> .	Não há restrição legal: dados estão sendo acessados por parceiro da operadora (Alpha) para gerar perfil de crédito. A <i>joint venture</i> tem acesso ao perfil de crédito, não aos dados da rede de celulares.	Envolvidos e já familiarizados com a estrutura tecnológica usada para acesso e processamento dos dados. As condições dos dados, da ferramenta da Alpha e das interfaces dos parceiros puderam ser validadas pelos envolvidos.	O processo é transparente para usuários, havendo pouca exposição do modelo à opinião pública. Os envolvidos não veem conflitos éticos ou riscos aos usuários, cujos dados não serão transferidos para fora das bases da operadora. Entendem que negócio é uma forma de ampliar a oferta aos clientes que poderão se beneficiar das linhas de crédito.	Os custos da solução são pagos com folga pelos ganhos gerados na concessão de crédito e na venda de produtos e serviços da operadora.

Quadro 3.3 - Conteúdo – o ecossistema do caso da Alpha

Fonte: elaborado pela autora

Vale ressaltar que a Alpha não abandonou seu propósito de oferecer soluções que promovam a inclusão financeira, o que continua em seu horizonte. Enquanto trabalha nos negócios que conquistou, a Alpha planeja ações para incrementar sua atuação e se aproximar desse objetivo. Algumas dessas ações estão ligadas à obtenção de autorizações formais para o acesso aos dados e à parceria com empresas interessadas em oferecer crédito a grupos excluídos do sistema bancário e em utilizar novos modelos de operações de crédito.

3.7 Discussão

A discussão dos resultados desta pesquisa será dividida em três temas: a dinâmica do ecossistema de reúso de dados revelada pelo caso; conceitos e ferramentas que podem colaborar para a compreensão do reúso de dados; e a abordagem teórica da prática

operacionalizada pelo Multilevel Framework como alternativa para a pesquisa empírica de ecossistemas complexos.

A dinâmica do ecossistema de reúso de dados revelada pelo caso. O estudo atendeu a seus objetivos exploratórios, permitindo exercitar o modelo teórico proposto e iluminando questões interessantes para refletir sobre o caso e o cenário de reúso de dados que ele revela.

Em resumo, a assimetria de recursos (acesso aos dados) inviabilizou a implementação da solução da Alpha para avaliação de crédito com dados de telefonia celular para inclusão financeira. A empresa reformulou sua estratégia, oferecendo novos serviços que lhe trazem sustentabilidade financeira e a aproximam de atores que poderão lhe dar acesso a recursos relevantes no futuro. Além disso, articula-se também em outras frentes, como a busca por autorizações legais e parcerias.

Essa estratégia garante a sobrevivência da Alpha e de seu ecossistema, além de construir um caminho de desenvolvimento na direção de seus objetivos iniciais de inclusão financeira. Tal movimento sugere uma capacidade de resistência e reação dos ecossistemas frente às dificuldades que ameaçam a sua sobrevivência. Diante dos limites impostos pelos que detêm o controle sobre os recursos críticos, os agentes se articularam e realizaram manobras para criar e/ou acessar outros recursos e para criar novas vias de acesso aos recursos restritos. Essa dinâmica refletiu a complementaridade das epistemologias de posse e de prática do poder sugerida por Tello-Rozas, Pozzebon e Mailhot (2015).

Na visão de posse, o poder está ligado à autoridade formal e ao controle de acesso aos recursos. O poder é algo que os atores possuem e podem usar em sua interação com os outros, o que pode gerar restrições, distorções e dominações (Tello-Rozas, Pozzebon & Mailhot, 2015) – como ocorreu com as operadoras e bancos que negaram acesso aos dados para Alpha. Na visão da prática, o poder é algo relacional, que existe no contexto do qual emerge e no qual se reproduz e se transforma por meio das relações e ações que acontecem na prática (Tello-Rozas, Pozzebon & Mailhot, 2015) – como se vê nas manobras da Alpha para buscar outros recursos ou abrir caminhos alternativos aos recursos restritos.

O caso evidencia os impactos dos desequilíbrios gerados pelas assimetrias de poder que emergem da pluralidade do contexto e da complexidade da colaboração nesse ecossistema. Ao mesmo tempo, levanta a questão sobre as possibilidades que essa pluralidade abre aos agentes do ecossistema, por meio de suas micropáticas, buscarem caminhos alternativos para superar tais desequilíbrios. Esses espaços de manobras existiriam em um ambiente em que o

poder fosse centralizado (e não difuso) e os objetivos fossem únicos (e não múltiplos)? Nesse contexto, talvez houvesse menos desequilíbrios, mas a capacidade de reação diante dos que surgissem também poderia ser menor? Este estudo não tem elementos para responder essas questões, mas deixa a reflexão para futuras explorações.

Conceitos e ferramentas que podem colaborar para a compreensão do reúso de dados. Ao retratar as dimensões intrínseca, legal, tecnológica, societal e econômica na apresentação do reúso-de-dados-na-prática, este modelo ofereceu uma visão mais abrangente do reúso de dados, como clamam alguns estudiosos. A inclusão da dimensão intrínseca ao grupo de dimensões sugeridas por Custers e Bachnechler (2017) também se mostrou útil para destacar a dinâmica de relações, objetivos e interesses que mobilizam os atores diretamente envolvidos na construção e prática do reúso de dados.

A dimensão tecnológica permite apresentar a interação dos elementos materiais no ecossistema, como as abordagens sociomateriais têm sugerido (Orlikowski, 2007; Suchman, 2007). A ideia não é descrever a tecnologia, mas a “tecnologia trabalhando” (Orlikowski, 2007), suas capacidades e características relevantes para as práticas específicas, podendo influenciar e ser influenciadas por elas (Scott & Orlikowski, 2009). Implementar este conceito em pesquisas empíricas sem retornar à dualidade ou ao determinismo tecnológico de abordagens anteriores nos estudos de SI ainda é um desafio. Embora este estudo não represente uma solução ao tema, abre espaço para introduzir a dinâmica material na análise, um caminho que pode ser explorado por estudos que se aprofundem na reflexão de como usar a dimensão tecnológica para apresentar a dinâmica sociomaterial que permeia os ecossistemas na prática.

Por fim, o caso também demonstrou a importância do alinhamento entre todas as dimensões para viabilizar o reúso de dados. Na solução proposta inicialmente (avaliação de crédito com dados de telefonia para inclusão financeira), havia dúvidas nas dimensões legal e societal (havia base jurídica e ética? Qual seria a opinião pública?), gerando riscos e incertezas na dimensão econômica e intrínseca (quais seriam os reais benefícios e para quem? Poderia haver impacto para os negócios da operadora ou dos bancos? Quais as responsabilidades e riscos para ambos?). Na dimensão tecnológica, embora a Alpha tivesse confiança sobre a segurança e robustez de sua ferramenta, a falta de familiaridade da operadora de telefonia com o tipo de solução proposta e sua dificuldade em avaliar o funcionamento completo da ferramenta tornou a interação da operadora com a ferramenta nebulosa. Em contrapartida, na solução implementada após a adoção da nova estratégia, todas as dimensões estavam favoráveis. Assim, essa estrutura multidimensional mostrou-se adequada para promover uma

abordagem ampla, que pode contribuir para orientar a análise do reúso de dados para fins acadêmicos e práticos.

A abordagem da prática do Multilevel Framework como alternativa para a pesquisa empírica de ecossistemas complexos. O atual cenário tecnológico está criando contextos cada vez mais pluralistas e marcados pela colaboração complexa. Os estudos de SI precisam de modelos teóricos que ajudem a estudá-los. Este estudo corroborou a ideia de autores de que o uso de teorias da prática para explorar ecossistemas complexos é um caminho promissor.

A abordagem da prática considera a dinâmica do ecossistema, as relações mutuamente constitutivas entre seus elementos, a relevância de todos os componentes e a visão do todo e das partes de forma integrada (Feldman & Orlikowski, 2011). Ao retratar o processo e as práticas que dele emergem, a perspectiva da prática identificou as assimetrias de recursos entre os atores, questões relevantes para a compreensão da dinâmica dos ecossistemas, especialmente aqueles em que haja colaboração complexa (Feldman & Orlikowski, 2011; Tello-Rozas, Pozzebon & Mailhot, 2015, Vangen, 2016). Também demonstrou como os atores conseguiram lidar com tais assimetrias, criando caminhos alternativos para “sobreviverem” e buscarem seus objetivos, revelando a capacidade de reação do ecossistema.

O Multilevel Framework mostrou-se uma boa alternativa para a operacionalização da abordagem da prática nesse ecossistema complexo. O modelo possibilita incorporar diferentes níveis de análise (individual, grupal ou societal) (Pozzebon & Diniz, 2012), evidenciando os papéis dos atores, seus pressupostos, suas relações e participações no processo que dá origem à prática que, ao mesmo tempo, influencia esses atores e o processo. Permitiu, portanto, integrar a visão micro e macro do ecossistema em toda a sua complexidade e relações mutuamente constitutivas.

Além disso, a flexibilidade do modelo permitiu incorporar conceitos de reúso de dados que colaboraram com a leitura do caso, demonstrando sua capacidade de contribuir para pesquisas de múltiplos temas. Ele pode funcionar como uma meta-teoria capaz de articular outras teorias, oferecendo uma pluralidade que pode beneficiar a exploração de fenômenos mais complexos presentes em SI (Pozzebon, Mackrell & Nielsen, 2014). Essa pluralidade teórica, que pode ser também multiparadigmática, não é uma unanimidade no campo e, mesmo os pesquisadores favoráveis a ela, alertam para os riscos de fragmentação, incoerência e implausabilidade desse tipo de abordagem (Pozzebon, Mackrell & Nielsen, 2014). A utilização

de modelos como o Multilevel, que apoiem a operacionalização da articulação teórica, pode colaborar para o necessário enriquecimento do repertório acadêmico.

3.8 Conclusões, limites e oportunidades

Este trabalho pretende colaborar para o debate sobre o reúso de dados, uma prática promissora na era do BD, mas ainda pouco discutida no meio acadêmico. Ele propõe e exercita um modelo teórico para a análise empírica desse fenômeno, gerando três contribuições. A primeira refere-se às questões que surgiram a partir da dinâmica do ecossistema de reúso de dados, revelada pelo caso. A segunda é a proposta de conceitos e ferramentas que podem colaborar para a compreensão do reúso de dados, beneficiando a pesquisa e a prática desse fenômeno. A terceira é a abordagem teórica da prática, operacionalizada pelo Multilevel Framework, que foi adaptado para o estudo de um caso de reúso de dados, mas pode ser uma alternativa para a pesquisa empírica de ecossistemas complexos, envolvendo outras situações.

O caso explorado revela a importância das estratégias alternativas para superar assimetrias e tensões entre os elementos do ecossistema de reúso de dados. Tais alternativas demonstram o caráter criativo de ecossistemas e podem representar um caminho incremental para viabilizar inovações disruptivas. O caso da Alpha expôs a relevância das relações de poder, abrindo a discussão sobre suas abordagens, como posse e prática, ambas presentes em sua dinâmica. Ao refletir sobre um modelo que colabore para a abordagem holística do reúso de dados pelas lentes da prática, o estudo pode contribuir para o avanço desse fenômeno. Se acreditamos que o reúso de dados pode ser uma alternativa para alcançar os altos benefícios preconizados pelos entusiastas do BD, compreender a dinâmica de seus ecossistemas cada vez mais complexos pode ser de grande valia para a teoria e para a prática.

Os resultados aqui apresentados são limitados à experiência do caso investigado. Sua plausibilidade aumenta a probabilidade de que eles possam ser transferidos para outros contextos, o que é indicado como sugestão para futuros estudos. A solução que a Alpha implementou após a sua ampliação de serviços está em uma diferente categoria reúso de dados. Isso corrobora a sugestão de etapa anterior desta pesquisa sobre os diferentes potenciais de desafios entre as categorias. Um estudo de casos múltiplos abordando situações variadas da

matriz permitiria avançar a compreensão desse perfil de desafios e a discussão de alternativas para enfrentá-los, abrindo caminho para as inovações que eles oportunizam.

Outra sugestão para dar continuidade à exploração do modelo proposto seria o acompanhamento das próximas etapas do desenvolvimento do ecossistema da Alpha. A aplicação desse modelo em outros casos que tenham completado seu ciclo de desenvolvimento e envolvem outras áreas de negócios também poderia contribuir para o seu amadurecimento.

Referências Bibliográficas

Abbasi, A., Albrecht, C., Vance, A., & Hansen, J. (2012). Metafraud: A Meta-Learning Framework For Detecting Financial Fraud. *MIS Quarterly*, 36(4), 1293-A12.

Adner, R. (2006). Match Your Innovation Strategy To Your Innovation Ecosystem. *Harvard Business Review*, 84 (4), 98-107.

Adner, R. & Kapoor, R. (2010). Value Creation In Innovation Ecosystems: How The Structure Of Technological Interdependence Affects Firm Performance In New Technology Generation. *Strategic Management Journal*, 31, 306-333.

Aleksynska, M., & Cazes, S. (2016). Composite indicators of labour market regulations in a comparative perspective. *IZA Journal of Labor Economics*, 5(1), 3.

Altman, E. I (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy, *Journal of Finance*, 589-609.

Annanperä, E., Liukkunen, K., & Markkula, J. (2015). Innovation in evolving business ecosystem: A case study of information technology-based future health and exercise service. *International Journal of Innovation and Technology Management*, 12(04), 1550015.

Baesens, B., Bapna, R., Marsden, J. R., Vanthienen, J., & Zhao, J. L. (2016). *Transformational Issues of Big Data and Analytics in Networked Business*. *MIS quarterly*, 38(2), 629-631.

Berger, P. L., & Luckmann, T. (1966). *The social construction of reality: A treatise in the sociology of knowledge*. Anchor Books, Garden City, NY.

Beyerlein, M., Johnson, D., & Beyerlein, S. (Eds.). (2004). *Complex collaboration: Building the capabilities for working across boundaries*. Emerald Group Publishing Limited.

Beyers, R. J. (1964). The Microcosm Approach To Ecosystem Biology. *The American Biology Teacher*, 26(7), 491-498.

Bijker, W. E., & Law, J. (1992). *Shaping technology/building society: Studies in sociotechnical change*. MIT press.

Borgman, C. L. (2015). *Big data, little data, no data: Scholarship in the networked world*. MIT press.

Boyd, D. & Crawford, K. (2012) Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon. *Information, Communication & Society*, 15(5), 662-679.

Bruni, A. L., Murray, A., & Famá, R. (1998). Modelos brasileiros preditivos de risco de crédito: um estudo exploratório atual sobre as suas eficácias. *Periódico Tema*, 32, 148-167.

Chesbrough, H., Vanhaverbeke, W., & West, J. (Eds.). (2014). *New frontiers in open innovation*. Oup Oxford.

Clarke, R. (2016). Big data, big risks. *Information Systems Journal*, 26(1), 77-90.

Curty, R. G. (2015). *Beyond "Data Thrifting": An Investigation of Factors Influencing Research Data Reuse In The Social Sciences* (3713677). Available from ProQuest Dissertations & Theses Global. (1698463092).

Custers, B., & Bachlechner, D. (2017). *Advancing the EU Data Economy: Conditions for Realizing the Full Potential of Data Reuse*. *Information Polity*, 22(4), 291.

Custers, B., & Ursic, H. (2016). *Big Data and Data Reuse: A Taxonomy of Data Reuse for Balancing Big Data Benefits and Personal Data Protection*. *International Data Privacy Law*, 6(1), 4-15.

de Rezende Francisco, E., Aranha, F., Zambaldi, F., & Goldszmidt, R. (2007). Electricity Consumption as a Predictor of Household Income: a Spatial Statistics Approach. *In Advances in Geoinformatics* (pp. 267-282). Springer Berlin Heidelberg.

Denis, J. L., Langley, A., & Rouleau, L. (2007). Strategizing in pluralistic contexts: Rethinking theoretical frames. *Human relations*, 60(1), 179-215.

Diniz, E. H. (2007). Correspondentes Bancários E Microcrédito No Brasil: Tecnologia Bancária E Ampliação Dos Serviços Financeiros Para A População De Baixa Renda.

Faniel, I. M. & Jacobsen, T. E. (2010). Reusing scientific data: How earthquake engineering researchers assess the reusability of colleagues' data. *Computer Supported Cooperative Work (CSCW)*, 19(3), 355-375.

- Faraj, S. A., Kolker, E., Bevilacqua, L., & Özdemir, V. (2012). Collective Theranostics And Postgenomics Entrepreneurship: Rethinking Innovations As Knowledge Ecosystems Built By Complex Collaboration. *Expert Review Of Molecular Diagnostics*, 12(8), 787-790.
- Feldman, M. S., & Orlikowski, W. J. (2011). Theorizing practice and practicing theory. *Organization science*, 22(5), 1240-1253.
- Flick, U. (2004). *Uma Introdução à Pesquisa Qualitativa*. Porto Alegre, Bookman.
- Frosch, R.A. & Gallopoulos, N.E. (1989). *Strategies for manufacturing*. *Scientific American*, 261, 144-152.
- Giddens, A. (1984). *The constitution of society*. Cambridge: Polity.
- Iansiti, M. and Levien, R. (2004). Strategy As Ecology. *Harvard Business Review*, 82, 3: 68–78.
- Isenberg, D. (2010). *How to start an entrepreneurial revolution*. Harvard Business Review.
- Ketter, W., Peters, M., Collins, J., & Gupta, A. (2016). Competitive Benchmarking: An IS Research Approach To Address Wicked Problems With Big Data And Analytics. *MIS Quarterly*, 40(4), 1057.
- Langley, A. (1999). Strategies for theorizing from process data. *Academy of Management review*, 24(4), 691-710.
- Latour, B. (1992). Where Are the Missing Masses? The Sociology of a Few Mundane Artifacts. MIT Press, Cambridge, MA, 225-258.
- Luvizan, S., & Diniz, E. (2017). Big Data e o Uso Secundário de Dados: Desafios para a Qualidade de Dados e a Inovação. *Anais do Enanpad*, 2017.
- Mackenzie, A. (2015). The fintech revolution. *London Business School Review*, 26(3), 50-53.
- Martin, K. (2015). *Ethical issues in the big data industry*. *MIS Quarterly Executive*, 14(2), 67-85.
- Mervin, C. (1942). Financing Small Corporations. *Bureau of Economic Research*, New York.
- Miles, M. & Huberman, A. (1994). *Qualitative data analysis: An expanded sourcebook*. Sage.
- Moore, J. F. (1993). Predators and prey: a new ecology of competition. *Harvard Business Review*, 71(3), 75-83.

Moore, J. F. (1996). *The Death of Competition: Leadership and Strategy in the Age of Business Ecosystems*. Harper Business, New York.

Müller, O., Junglas, I., Brocke, J. V., & Debortoli, S. (2016). Utilizing big data analytics for information systems research: challenges, promises and guidelines. *European Journal of Information Systems*, 25(4), 289-302.

Nachira, F. (2002). Towards a network of digital business ecosystems fostering the local development. *European Commission Discussion Paper*, Bruxelles.

Newell, S., & Marabelli, M. (2015). Strategic opportunities (and challenges) of algorithmic decision-making: A call for action on the long-term societal effects of 'datification'. *The Journal of Strategic Information Systems*, 24(1), 3-14.

Orlikowski, W. J. (2000). Using technology and constituting structures: A practice lens for studying technology in organizations. *Organization science*, 11(4), 404-428.

Orlikowski, W. J. (2007). Sociomaterial practices: Exploring technology at work. *Organization studies*, 28(9), 1435-1448.

Orlikowski, W. J., & Scott, S. V. (2008). Sociomateriality: challenging the separation of technology, work and organization. *The academy of management annals*, 2(1), 433-474.

Park, H. & Leydesdorff, L. (2013) Decomposing Social And Semantic Networks In Emerging "Big Data" Research. *Journal of Informetrics*, 7(3), 756-765.

Pettigrew, A. M. (1985). Contextualist Research And The Study Of Organizational Change Processes. In E. Mumford, R. Hirschheim, G. Fitzgerald, & A. T. Wood-Harper (Eds.), *Research Methods in Information Systems* (pp. 53-72). New York: North Holland.

Pettigrew, A. M. (1990). Longitudinal field research on change: theory and practice. *Organization Science*, 1(3), 267-292.

Pettigrew, A. M. (1997). What is a processual analysis. *Scandinavian journal of management*, 13(1997), 4.

Pilinkiene, V., & Maciulis, P. (2014). Comparison of different ecosystem analogies: the main economic determinants and levels of impact. *Procedia-Social and Behavioral Sciences*, 156, 365-370.

Pinch, T. J., & Bijker, W. E. (1984). The social construction of facts and artefacts: Or how the sociology of science and the sociology of technology might benefit each other. *Social studies of science*, 14(3), 399-441.

Pozzebon, M. & De Freitas, H. (1998). Pela Aplicabilidade: Com Um Maior Rigor Científico-Dos Estudos De Caso Em Sistemas De Informação. *Revista de Administração Contemporânea*, 2(2), 143-170.

Pozzebon, M., & Diniz, E. H. (2012). Theorizing ICT and society in the Brazilian context: a multilevel, pluralistic and remixable framework. *BAR-Brazilian Administration Review*, 9(3), 287-307.

Pozzebon, M., Diniz, E. & Jayo, M. (2009). Adapting The Structurationist View Of Technology For Studies At The Community/Societal Levels. In Y. K. Dwivedi (Ed.), *Handbook of research on contemporary theoretical models in information systems* (pp. 18-33). New York: IGI Publishing.

Pozzebon, M., Mackrell, D., & Nielsen, S. (2014). Structuration bridging diffusion of innovations and gender relations theories: a case of paradigmatic pluralism in IS research. *Information Systems Journal*, 24(3), 229-248.

Pozzebon, M., Rodriguez, C. & Petrini, M. (2014). Dialogical Principles For Qualitative Inquiry: A Nonfoundational Path. *International Journal of Qualitative Methods*, 13, 293-317.

Rao, B., & Jimenez, B. (2011, July). A comparative analysis of digital innovation ecosystems. *In Technology Management in the Energy Smart World (PICMET)*, 2011 Proceedings of PICMET'11: (pp. 1-12). IEEE.

Samarajiva, R., & Lokanathan, S. (2016). *Using Behavioral Big Data For Public Purposes: Exploring Frontier Issues Of An Emerging Policy Arena*.

Scott, S. V., & Orlikowski, W. J. (2009). 'Getting the truth': exploring the material grounds of institutional dynamics in social media.

Siqueira, E., Diniz, E., & Albino, R. (2018). *Fintech Social: Definição, Categorização e Ilustrações Empíricas*.

Smith, R. F., & Winakor, A. H. (1935). *Changes In The Financial Structure Of Unsuccessful Industrial Corporations*. University of Illinois.

Stake, R. (1995). *The Art Of Case Study Research*. Sage.

Stake, R. E. (2008). Qualitative case studies. In N. K. Denzin & Y. S. Lincoln (Eds.), *Strategies of qualitative inquiry* (pp. 119-149). Thousand Oaks, CA, US: Sage Publications, Inc.

Stiglitz, J. E., & Weiss, A. (1981). Credit Rationing In Markets With Imperfect Information. *The American Economic Review*, 71(3), 393-410.

Suchman, L. (2007). *Human-machine reconfigurations: Plans and situated actions*. Cambridge University Press.

Tello-Rozas, S., Pozzebon, M., & Mailhot, C. (2015). Uncovering Micro-Practices and Pathways of Engagement That Scale Up Social-Driven Collaborations: A Practice View of Power. *Journal of Management Studies*, 52(8), 1064-1096.

Turner, M. A., Lee, A. S., Schnare, A., Varghese, R., & Walker, P. D. (2006). Give credit where credit is due: Increasing access to affordable mainstream credit using alternative data. *Chapel Hill, NC and Washington, DC: Political and Economic Research Council and The Brookings Institution*.

Vangen, Siv (2016). Developing Practice-Oriented Theory on Collaboration: A Paradox Lens. *Public Administration Review*, 77(2) pp. 263–272.

Walsham, G. (2002). Cross-cultural software production and use: a structurational analysis. *MIS quarterly*, 359-380.

Yoon, A. (2015). *Data reuse and users' trust judgments: Toward trusted data curation*. Available from ProQuest Central; ProQuest Dissertations & Theses Global.

Zammuto, R. F., Griffith, T. L., Majchrzak, A., Dougherty, D. J., & Faraj, S. (2007). Information technology and the changing fabric of organization. *Organization science*, 18(5), 749-762.

Zimmerman, A. S. (2003). *Data Sharing And Secondary Use Of Scientific Data: Experiences Of Ecologists*. Available from ProQuest Dissertations & Theses Global: Science & Technology; ProQuest Dissertations & Theses Global: Social Sciences.

Zimmerman, Ann. 2008. "New Knowledge from Old Data: The Role of Standards in the Sharing and Reuse of Ecological Data." *Science, Technology & Human Values* 631-652.

Zuboff, S. (2015). Big other: surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology*, 30(1), 75-89.

4 ARTIGO III – REÚSO DE DADOS: UMA JORNADA DE MUDANÇAS RUMO A NOVOS PARADIGMAS NO SETOR FINANCEIRO

4.1 Resumo

Este artigo pretende colaborar com a escassa literatura sobre o reuso de dados, apresentando um estudo de múltiplos casos que retratam esse fenômeno no setor financeiro. Seu objetivo é identificar quais são os desafios do reuso de dados e como os atores estão lidando com eles na prática. Para realizar essa missão, os casos selecionados correspondem a diferentes categorias, apresentando distintos níveis de desafios potenciais. O primeiro caso refere-se ao repropósito e à recontextualização de dados internos em um banco, ilustrando categorias de reuso de dados cujo patamar de desafios é denominado “direto”, uma vez que a organização tem acesso direto ao contexto de geração dos dados. O segundo traz o repropósito de dados externos públicos e privados em um *bureau* de crédito, refletindo tipos de reuso de dados que apresentam desafios de nível “intermedial”, já que a organização depende da interação com atores externos para equacionar questões emergentes. O terceiro concerne ao repropósito e à recontextualização de dados externos públicos e privados em três *fintechs* de crédito e cobrança, representando categorias de reuso de dados em que os desafios chegam ao seu potencial “extremo”, pois os atores envolvidos no processo podem pertencer a áreas diferentes e, portanto, estar mais distantes. A estrutura dessa pesquisa foi inspirada na abordagem de Stake (2006) para estudos de múltiplos casos como uma estratégia para explorar fenômenos complexos que ele denomina *quintain*. O conhecimento gerado em cada caso contribui para a compreensão do *quintain* como um todo. Para analisar os casos, o presente estudo utilizou o modelo conceitual Multilevel Framework, que foi adaptado para incorporar elementos que permitiram retratar o reuso de dados na prática. Essa leitura colaborou para identificar os desafios legais, tecnológicos, intrínsecos, sociais e econômicos que emergem desse processo. Ao refletir sobre tais desafios, foi possível identificar que uma parte significativa deles está associada a mudanças, sejam elas referentes à situação em questão ou ao contexto mais amplo em que o novo paradigma de reuso de dados, viabilizado pelo *big data* (BD), está se estabelecendo. Também identificou-se que os atores estão lidando com tais dificuldades por meio de uma dinâmica cíclica de ajuste/aprendizado, animada pela ação de indivíduos que mobilizam recursos e autoridades utilizando-se de abordagens de posse e de prática do poder.

Os conceitos e as discussões substantivas apresentadas neste artigo contribuem para o avanço do reuso de dados na teoria e na prática.

Palavras-chave: big data; colaboração complexa; Multilevel Framework; reuso de dados.

4.2 Introdução

O reuso de dados não é um fenômeno novo, mas ganhou contornos diferenciados com as possibilidades tecnológicas e a disponibilidade de informações da era do BD (Abbasi, Sarker & Chiang, 2016; Baesens, Bapna, Marsden, Vanthienen & Zhao, 2016; Ketter, Peters, Colins & Gupta, 2016; Martin, 2015; Muller, Junglas, Brocke & Debortoli, 2016; Newell & Marabelli, 2015). Maior capacidade de armazenamento e processamento de dados, artefatos tecnológicos cada vez mais conectados, penetração das ferramentas móveis na rotina dos indivíduos, e digitalização de negócios e processos corporativos marcam esse novo cenário (Abbasi *et al.*, 2016; Baesens *et al.*, 2016; Ketter *et al.*, 2016; Martin, 2015; Muller *et al.*, 2016; Newell & Marabelli, 2015).

O reuso de dados é cercado de expectativas grandiosas em diversas áreas. Na saúde, espera-se que sua exploração inteligente permita avançar no diagnóstico e tratamento de doenças, melhorar a eficiência dos serviços de saúde e a gestão dos recursos na área (Grande, Mitra, Shah, Wan & Asch, 2014; Krumm *et al.*, 2014; Safran *et al.*, 2007). Na esfera pública, acredita-se que ele possa contribuir para ampliar a participação popular na gestão e na fiscalização, trazer mais inteligência aos serviços, aumentar a transparência e melhorar a administração dos recursos (Letouzé, 2012; Manyika *et al.*, 2011; Samarajiva & Lokanathan, 2016; Smolan, 2013). Na ciência, o reuso de dados aliado a capacidades tecnológicas e metodológicas promete acelerar a produção científica, abrir novas perspectivas de pesquisa, dinamizar e democratizar a ciência (Borgman *et al.*, 2015; Curty, 2015; Wallis, Rolando & Borgman, 2013). Nos negócios, ele é visto como uma força habilitadora capaz de elevar o patamar de processos existentes, além de viabilizar novos serviços e modelos de negócios (Aleksynska & Cazes, 2016; Baesens *et al.*, 2016; Martin, 2015).

As polêmicas que cercam o reuso de dados são proporcionais a tais expectativas de benefícios. Ameaças à democracia, utilização da informação de forma nociva à sociedade e a

livre concorrência têm preocupado estudiosos e especialistas (Newell & Marabelli, 2015; Samarajiva & Lokanathan, 2016; Zuboff, 2015). Esse debate vem sendo intensificado por escândalos de vazamento e mau uso de dados, como o episódio de Edward Snowden - que revelou práticas obscuras de captura e reuso de dados no governo americano - (Macaskill & Hern, 2018); as diversas ocorrências de vazamento de informações de usuários de gigantes de tecnologia, como o Yahoo e o *bureau* de crédito americano Equifax (Armerding, 2018); o nebuloso caso da Cambridge Analytica e a avalanche de procedimentos duvidosos do Facebook que vieram à tona a partir dele (Sherr, 2018). A repercussão de tais escândalos é irreversível e vem sendo associada ao “fim da lua-de-mel entre a indústria de tecnologia e os governos” (Sherr, 2018), marcado pela crescente onda de ações para aumentar a legislação e o controle dos produtos e serviços tecnológicos pelo mundo.

Embora essa vasta lista sobre privacidade, ética e direitos individuais e coletivos traga questões relevantes e urgentes, especialistas advertem que elas não são as únicas que merecem atenção na discussão do tema. É preciso avançar também no debate relacionado às implicações dessa prática na estrutura tecnológica e de dados, à construção de bases metodológicas adequadas aos novos modelos de uso de dados e à educação de profissionais e da sociedade, preparando-os para os novos paradigmas desse cenário (Baesens et al., 2016; Borgman et al., 2015; Curty, 2015; Grande et al., 2014; Custers & Ursic, 2016; Mundie, 2014; Samarajiva & Lokanathan, 2016; Zuboff, 2015). Pesquisadores alertam para a necessidade de estudos que abordem o reuso de dados de forma mais abrangente, incorporando suas dimensões legais, tecnológicas, societárias e econômicas (Borgman et al., 2015; Curty, 2015; Custers & Bachlechner, 2017; Newell & Marabelli, 2015; Samarajiva & Lokanathan, 2016; Zuboff, 2015).

Há alguns anos, a literatura acadêmica e os especialistas em BD vêm apontando questões ligadas ao reuso de dados (Boyd & Crawford, 2012). Ainda assim, poucos estudos se dedicaram a compreendê-lo em sua amplitude e profundidade (Curty, 2015; Custers & Ursic, 2016; Newell & Marabelli, 2015). O reuso dos dados, oriundos de fontes internas ou externas, para fins relacionados ou não aos da geração dessas informações, dá origem a um ecossistema complexo, no qual diversos atores interagem direta ou indiretamente por meio dos dados e das implicações de seu novo uso. A compreensão desse cenário requer examinar suas diversas perspectivas técnicas e sociais (Custers & Bachlechner, 2017; Newell & Marabelli, 2015; Samarajiva & Lokanathan, 2016; Zuboff, 2015). Também demanda abordagens capazes de revelar questões que emergem na prática, como requerem os estudos de fenômenos envolvendo colaboração complexa (Vangen, 2016).

Assim, este trabalho apresenta um estudo de múltiplos casos da área financeira, utilizando um modelo conceitual adaptado ao reuso de dados para analisá-lo pela perspectiva da prática. O objetivo é identificar os desafios que emergem do reuso de dados e propor um modelo para explicar como os atores lidam com tais desafios na prática.

O modelo conceitual utilizado é uma variação do Multilevel Framework (Pozzebon, Diniz & Jayo, 2009), que propõe a abordagem de fenômenos que envolvem a tecnologia pela perspectiva da construção social promovida pelos atores que interagem em um ambiente que, simultânea e continuamente, influencia e é influenciado por eles. Sua proposta original articula conceitos do contextualismo (Pettigrew, 1985, 1990), da Social Shapping of Technology (SST) (Bijker & Law, 1992; Pinch & Bijker, 1984; entre outros) e da tecnologia-na-prática (Orlikowski, 2000). O modelo foi adaptado para demonstrar o processo por meio da representação visual de seus principais eventos (Langley, 1999) e apresentar o reuso de dados na prática. Tal adaptação confirmou a flexibilidade do modelo e sua capacidade de incorporar conceitos específicos para potencializar sua aderência a temáticas particulares (Pozzebon & Diniz, 2012; Pozzebon, Mackrell & Nielsen, 2012).

A análise dos casos revelou que o reuso de dados emerge em um contexto complexo, marcado pela pluralidade e emergência (Denis, Langley & Rouleau, 2007; Faraj et al., 2012). Nele, atores oriundos de diferentes grupos sociais, movidos por diferentes pressupostos e objetivos, interagem em um processo que evolui por meio da dinâmica entre eles, bem como da dinâmica dos dados e suas condições de geração, captura e uso. A leitura dos casos, que correspondem a diferentes categorias de reuso de dados (repropósito ou recontextualização de dados internos ou externos públicos/privados), evidenciou desafios em todas as dimensões e permitiu identificar duas naturezas às quais eles se relacionam: mudanças e recursos. Ao observar os processos dos casos, foi possível verificar que os atores lidaram com essas dificuldades por meio de processos cíclicos de ajuste/aprendizado que foram estimulados pela ação mobilizadora de grupos que lideraram a articulação do ecossistema necessário para viabilizar o reuso dos dados. Tais grupos mobilizam recursos e autoridade, ou seja, utilizando abordagens de posse ou de prática de poder, como também se observa em outros casos envolvendo colaboração complexa (Tello-Rozas, Pozzebon & Mailhot, 2015).

O presente trabalho contribui para avançar o conhecimento sobre o reuso de dados, trazendo questões reais a serem endereçadas para a evolução desse fenômeno na teoria e na prática, além de conceitos que servem de alicerce para o desenvolvimento do conhecimento e abrem caminhos para novos estudos e debates sobre o tema. Ao incorporar as múltiplas

dimensões do reúso de dados no trabalho e propor uma leitura do caso por uma lente teórica da prática, este artigo atende ao chamado da literatura, que clama por pesquisas que abordem o tema sob perspectivas mais amplas. A operacionalização da abordagem teórica da prática por meio do Multilevel Framework adaptado também pode contribuir como referência para outros trabalhos que envolvam fenômenos de colaboração complexa, já que os pesquisadores ainda enfrentam dificuldades para implementar estudos empíricos que adotem tal visão nesses casos. Por fim, esta pesquisa também identifica, a partir da leitura dos casos práticos, correntes teóricas de sistemas de informação (SI) promissoras para investigar o tema, o que pode estimular o engajamento da área na exploração do reúso de dados, que muito pode se beneficiar do corpo de conhecimento desse grupo.

Na sequência, este artigo apresentará suas bases teóricas de reúso de dados e do modelo conceitual Multilevel Framework adaptado. Apresentará, então, sua metodologia inspirada na abordagem de múltiplos casos de Stake (2006). Posteriormente, irá expor os casos individualmente e a análise *cross-case*, discutindo os principais resultados e buscando dialogar com os objetivos do estudo. A conclusão, os limites e as sugestões para futuras pesquisas sobre o tema concluirão o trabalho.

4.3 Revisão teórica

4.3.1 Réúso de dados

Poucos estudos se dedicaram a determinar o conceito de reúso de dados, mas os trabalhos que fazem referência a ele, geralmente, definem-o como a utilização de dados para fins distintos daqueles para os quais foram gerados (Curty, 2015; Yoon, 2015).

Luvizan & Diniz (2017) propõem classificar o reúso de dados a partir da fonte dos dados e da finalidade de seu reúso. A fonte dos dados pode ser interna (da própria entidade que está promovendo o reúso dos dados), externa pública (dados disponíveis para acesso público) ou privada (dados pertencentes a entidades distintas da que está promovendo o seu reúso). Já a finalidade pode referir-se ou não aos propósitos originais dos dados. Seu reúso é chamado de repropósito, quando sua finalidade estiver relacionada à área de origem dos dados, ou de

recontextualização, quando sua finalidade concernir a uma área completamente distinta da original (Custers & Ursic, 2016).

O posicionamento do reúso de dados na matriz fonte x finalidade influencia seus potenciais desafios, já que o aumento da distância entre sua geração e reúso tende a dificultá-lo (Aleksynska & Cazes, 2016; Custer & Ursic, 2016; Yoon, 2015; Zimmerman, 2008). A figura 4.1 demonstra a variação do nível de desafios entre as categorias, que podem ser direto, intermedial e extremo.

		Fonte dos Dados			
		Dados Privados Internos	Dados Públicos	Dados Privados Externos	
Finalidade Reuso dos Dados	Repropósito	Repropósito de Dados Internos	Repropósito de Dados Públicos	Repropósito de Dados Privados Externos	Intermedial
	Recontextualização	Recontextualização de Dados Internos	Recontextualização de Dados Públicos	Recontextualização de Dados Privados Externos	Extremo

↓
Direto

Figura 4.1 - Matriz do esquema classificatório do reúso dos dados por fonte x finalidade

Fonte: Adaptado de Luvizan & Diniz, 2017

O repropósito ou a recontextualização de dados internos traz desafios, porém a entidade que atua nesse contexto tem acesso direto aos elementos envolvidos na geração e no reúso dos dados, potencializando sua capacidade de ação para superar questões emergentes (Luvizan & Diniz, 2017). No repropósito de dados externos públicos ou privados, o uso de dados pertencentes a outras entidades aumenta a quantidade de atores envolvidos e a complexidade da negociação necessária para equacionar as questões, que também podem ser mais numerosas e difíceis, tornando o nível de desafios “intermedial”. A recontextualização de dados privados externos públicos ou privados adiciona uma nova camada de complexidade à tratativa dos desafios devido à diversidade entre os atores, que podem ser de áreas distintas, regidas por interesses e pressupostos diferentes, chegando ao potencial extremo de desafios do reúso de dados.

As questões que desafiam o reúso de dados pertencem a inúmeras perspectivas. Assim, pesquisadores advertem sobre a importância das várias facetas do fenômeno para viabilizar e reforçar a relevância de estudos que utilizem abordagens capazes de capturar essa diversidade (Borgman et al., 2015; Custers & Ursic, 2016; Newell & Marabelli, 2015; Samarajiva & Lokanathan, 2016; Zuboff, 2015). Para colaborar com esse desenvolvimento, Custers e Bachlechner (2017) propõem 4 dimensões do reúso: legal (temas legais e

regulatórios), tecnológica (condições técnicas de ferramentas e dados), societal (padrões éticos e práticas socialmente aceitas) e econômica (viabilidade financeira). Uma etapa anterior desse estudo sugere acrescentar a dimensão “intrínseca” para abordar questões relacionadas aos grupos diretamente envolvidos no reúso e o processo pelo qual ele se articula para implementar e operar o reúso dos dados e os produtos/serviços viabilizados a partir dele. Todas as dimensões devem ser consideradas conjuntamente, pois, na prática, agem de forma interligada, influenciando-se mutuamente (Custers & Bachlechner, 2017). A figura 4.2 sumariza as dimensões propostas e suas temáticas principais.

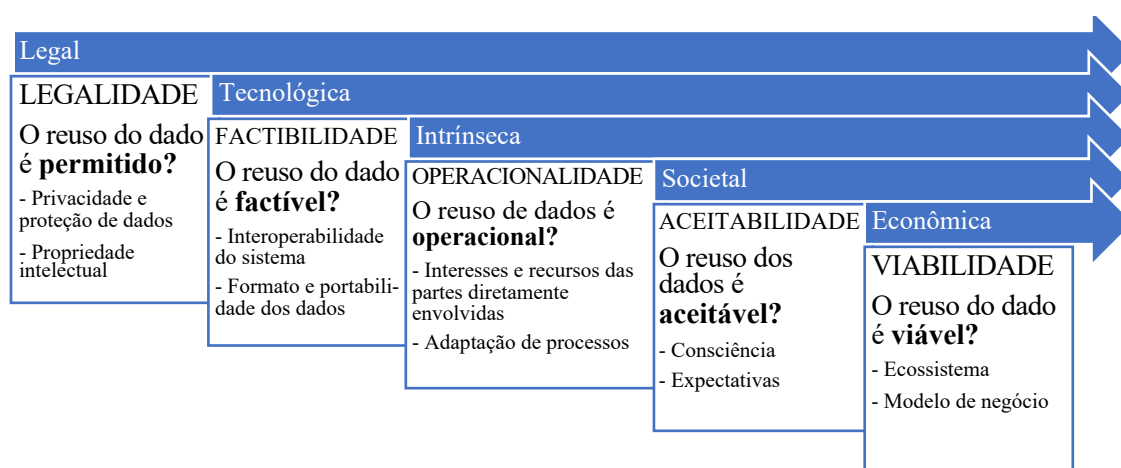


Figura 4.2 - Dimensões do reúso de dados que implicam em seus desafios
Fonte: adaptado de Custers e Bachlechner, 2017.

O reúso de dados também pode ser considerado um contexto pluralista, já que implica atividades baseadas em conhecimento que reúnem atores com diferentes objetivos e sem uma estrutura centralizada de poder (Denis, Langley & Rouleau, 2007). A interação entre indivíduos oriundos de diferentes grupos torna a colaboração mais complexa (Beyerlein, Johnson & Beyerlein, 2004; Faraj et al., 2012). A compreensão da dinâmica de ecossistemas pluralistas pode se beneficiar de abordagens teóricas da prática, capazes de apreender o processo que dá origem à colaboração complexa de seus agentes.

Lentes teóricas que focalizam os processos e a prática têm sido adotadas nos estudos de SI para compreender a dinâmica emergente do desenvolvimento e uso das tecnologias (Feldman & Orlikowski, 2011; Langley, 1999; Pettigrew, 1990; Pozzebon & Diniz, 2012). É também o caminho trilhado por pesquisadores que investigam situações de colaboração complexa, como a colaboração social (Vangen, 2016). Embora seu potencial seja reconhecido no meio acadêmico em diversas áreas, a operacionalização de pesquisas empíricas com tal abordagem ainda é um desafio que dificulta sua disseminação (Vangen, 2016). Um dos poucos

exemplos de modelos conceituais que dão suporte à exploração de casos empíricos pelas lentes da prática é o Multilevel Framework (Pozzebon, Diniz & Jayo, 2009), cuja adaptação resultou no modelo conceitual adotado neste trabalho.

4.3.2 Multilevel Framework adaptado

Criado em 2009, o Multilevel Framework considera que o fenômeno tecnológico é socialmente construído pela ação contínua de seus agentes, que o influenciam ao mesmo tempo que são influenciados por ele (Pozzebon & Diniz, 2012). O modelo utiliza os conceitos básicos do contextualismo (Pettigrew, 1985, 1990) para combinar componentes do SST (Bijker & Law, 1992; Pinch & Bijker, 1984; entre outros) e da tecnologia-na-prática (Orlikowski, 2000). Ele tem se mostrado flexível, podendo ser adaptado para incorporar conceitos mais específicos, o que amplifica seu poder de análise em diversos casos empíricos (Pozzebon & Diniz, 2012; Pozzebon, Mackrell & Nielsen, 2014). O presente estudo adotará a variação proposta na fase anterior dessa pesquisa, que deu origem ao modelo denominado Ecossistema do Reúso de Dados na Prática (figura 4.3), cujos componentes serão descritos na sequência.

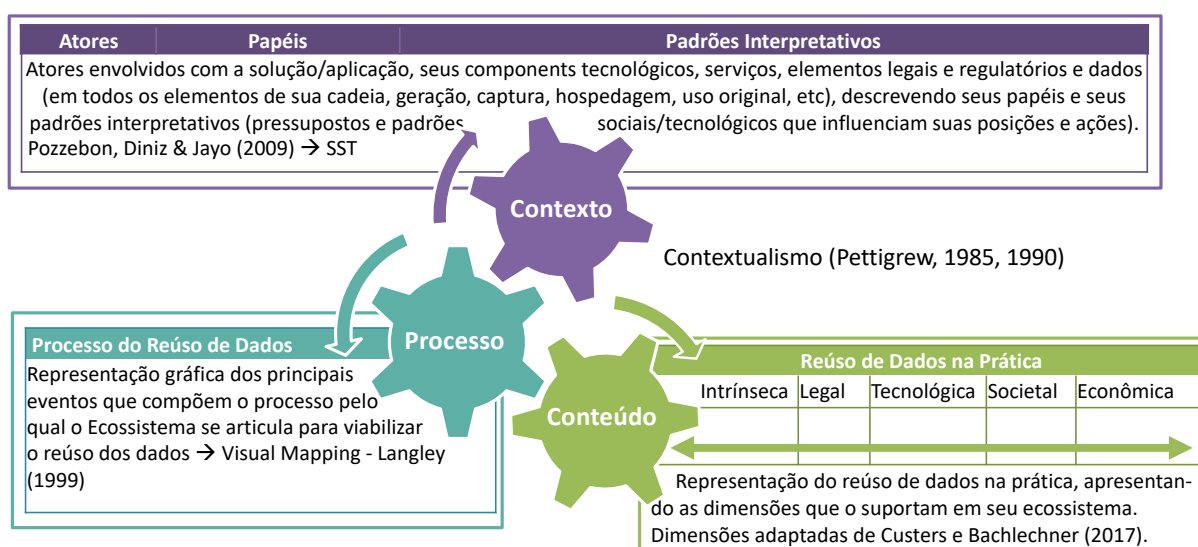


Figura 4.3 - Ecossistema do reúso de dados pelas lentes da prática

Fonte: adaptado do modelo Multilevel Framework de Pozzebon, Diniz e Jayo (2009)

Contexto – como no modelo original, o contexto é composto pelos grupos sociais envolvidos, seus papéis e pressupostos de ação, elementos apreendidos da SST (Pozzebon, Diniz & Jayo, 2009). Ele pode capturar todos os atores envolvidos no desenvolvimento e na

realização do reúso de dados, seus componentes tecnológicos, dados, elementos legais e regulatórios ou quaisquer outros que influenciem esse ambiente.

Processo – representação gráfica dos principais eventos organizados de forma lógica, capturando sua cronologia e relações em um mapa visual (Langley, 1999). Tal apresentação permite que o contorno do processo do qual emerge o reúso de dados seja delineado. A noção de que esse processo é formado por meio das negociações entre os atores, que constava no modelo original, não se perdeu, pois elas podem aparecer no mapa visual.

Conteúdo – o conteúdo que emerge do processo é representado pelo quadro reúso-de-dados-na-prática. Ele demonstra o reúso de dados em prática do momento retratado no estudo, destacando a situação corrente de suas dimensões intrínseca (proposta pela autora), legal, tecnológica, societal e econômica (Custers & Bachlechner, 2017).

Este modelo adaptado já foi exercitado em um caso exploratório na etapa anterior da pesquisa. Mostrou-se útil no direcionamento da abordagem do campo e na organização dos dados coletados, contribuindo para a análise dos dados e a compreensão da dinâmica prática do caso. Tal colaboração é relevante neste estudo, que tem a ambição de adotar o enfoque da prática em múltiplos casos. A metodologia adotada será apresentada a seguir.

4.4 Metodologia

A metodologia da pesquisa foi inspirada na abordagem de estudos de casos múltiplos de Stake (2006). Ela é indicada pelo autor como alternativa para investigar um *quintain*. Stake (2006) propõe a denominação *quintain* para ressaltar a abrangência e a complexidade envolvidas no objeto, fenômeno ou na condição a ser estudada. Em um estudo de casos múltiplos, o *quintain* é a arena ampla que o pesquisador busca compreender por meio de casos que contribuirão para a exploração de questões ou dimensões específicas. O conhecimento gerado em cada caso ajuda a entender o todo, ou seja, constrói a compreensão do *quintain* (Stake, 2006). Tal noção norteia o desenho da pesquisa e impacta suas escolhas teóricas, temáticas e metodológicas. Neste trabalho, o *quintain* é o reúso de dados. As questões a serem exploradas são o contexto em que se dá o reúso, os desafios que dele emergem e como os atores envolvidos os enfrentam.

Para facilitar o diálogo entre os casos, decidiu-se focar os estudos em uma única área: a financeira, que foi escolhida por seu histórico de intenso uso de tecnologia e de informação, a adoção do BD e as expectativas sobre os benefícios que esse novo paradigma de dados desperta na área (Diniz, Luvizan, Hino & Ferreira, 2018; Manyika *et al.*, 2011), fomentando a exploração de dados e gerando aplicações de vários tipos para seu reuso. A possibilidade de analisar situações de diferentes categorias de reuso de dados ajuda a compreendê-lo como um todo. A fronteira do *quintain* nesta pesquisa passou a ser, portanto, o reuso de dados na área financeira. A estrutura geral do estudo, suas etapas e escolhas serão descritas a seguir.

4.4.1 Design da pesquisa

O atual estudo qualitativo interpretativo/construtivista foi operacionalizado por meio de um estudo de casos múltiplos, que buscou compreender os desafios que emergem do reuso de dados e o processo pelo qual os atores lidam com eles na prática. A perspectiva qualitativa permite compreender os casos a partir de suas atividades no contexto e na situação particular (Flick, 2004; Langley, 1999; Stake, 2006). Tal compreensão é fundamental para a abordagem interpretativa/construtivista que pressupõe que o mundo e as pessoas, a realidade e seus agentes, constituem-se mutuamente pela experiência viva e contínua de suas interações (Giddens, 1984; Guba & Lincoln, 1994; Pozzebon, Rodriguez & Petrini, 2014; Sandberg, 2005; Stake, 2008, Walsham, 1995). Essa visão é coerente com o posicionamento epistemológico da pesquisa e seus objetivos.

Tal concepção também converge com o modelo conceitual adotado no trabalho, baseado em uma perspectiva contextual que permite retratar o fenômeno do estudo na prática. A literatura apresenta diferentes conceitos de processo (Van de Ven, 1992). Neste trabalho, o processo é definido como uma sequência cronológica de eventos que descreve a evolução “das coisas” (Langley, 1999; Pettigrew, 1997; Van de Ven, 1992). Tal conceito observa o processo em ação (Pettigrew, 1997), sendo apropriado para as abordagens teóricas da prática, que também buscam analisar como a vida social emerge da ação contínua dos atores (Denis, Langley & Rouleau, 2007; Feldman & Orlikowski, 2011). Assim, além de articular elementos do contextualismo (Pettigrew, 1985, 1990) e do SST, o modelo teórico adotado incorpora a

visão processual para explorar o reuso de dados que emerge da prática, sendo, portanto, coerente com o posicionamento metodológico e as questões da pesquisa.

Um desafio importante nesse desenho de pesquisa é a definição de critérios de qualidade que sejam capazes de orientar o rigor em sua execução e estabelecer parâmetros que a legitimem. Critérios de qualidade para a pesquisa qualitativa têm sido alvo de debate na academia há anos (Gioia, Corley & Hamilton, 2013; Guba & Lincoln, 1994; Pratt, 2009). Pesquisadores que adotam tal perspectiva, considerada mais subjetivista, são constantemente desafiados a justificar suas escolhas e submeter seus trabalhos a critérios, muitas vezes, positivistas/neopositivistas (Cunliffe, 2010; Pozzebon, Rodriguez & Petrini, 2014). Lincoln & Guba (1985) propõem quatro critérios de confiabilidade como alternativa para estudos construtivistas: credibilidade (os dados são plausíveis?); transferabilidade (os resultados podem ser estendidos a outros contextos?); dependabilidade (os resultados podem se repetir?); confirmabilidade (os dados podem confirmar os resultados?) (Lincoln & Guba, 1985). Tais critérios iluminaram a produção do artigo, que busca detalhar as etapas da pesquisa e suas escolhas teóricas e metodológicas, assim como relatar os caminhos e reflexões feitas ao longo do estudo para permitir que os leitores visualizem esse percurso. A citação de trechos originais da fala dos entrevistados e a incorporação de anexos com mais informações da pesquisa também busca oferecer ao leitor mais elementos para engajá-lo nas reflexões propostas.

4.4.2 Seleção dos casos

O estudo de múltiplos casos foi escolhido por sua capacidade de combinar o conhecimento de diversas situações para a compreensão de algo complexo, o *quintain*. Nesse tipo de estudo, cada caso dedica-se a compreender questões específicas e são selecionados por sua potencial contribuição para o entendimento do todo (Stake, 2006). Para orientar a seleção dos casos, foi realizada uma etapa anterior de pesquisa, propondo um esquema classificatório do reuso de dados. Os casos selecionados situam-se em diferentes categorias desse esquema, trazendo exemplos de aplicações com distintos potenciais de desafios. O quadro 4.1 representa a distribuição dos casos na tabela classificatória do reuso de dados.

		Fonte de Dados		
		Interno	Público	Privado Externo
Finalidade Reuso dos Dados	Repropósito	Caso Banco Nível dos desafios: Direto	Caso Bureau Nível dos Desafios: Intermedial	
	Recontextualização		Caso Fintechs Nível dos Desafios: Extremo	

Quadro 4.1 - Casos selecionados por tipo de desafio do reúso de dados

Fonte: elaborado pela autora

A seguir, apresenta-se os casos selecionados (identificados por codinomes para preservar a identidade das empresas), assim como a justificativa da escolha e o papel de cada situação no estudo:

- Caso Banco Ômega – foi selecionado para a investigação do repropósito e da recontextualização de dados internos. Conversas preliminares com gestores da área de dados do banco confirmaram a existência dessa prática de forma estruturada na organização, o que permitiu encontrar pessoas nas áreas de tecnologia, dados e negócios, as quais participaram de casos práticos. O repropósito e a recontextualização de dados internos foram associados ao nível de desafio denominado “direto”, que é o menos crítico dos três níveis estimados, conforme destacado em verde no quadro 4.1. O caso do banco também oferece a perspectiva do reúso de dados em uma grande corporação.

- Caso Bureau Sigma – foi escolhido por sua tradição em atividades que envolvem o reúso de dados. Tais atividades fazem parte de seu *core business* e, portanto, o *bureau* tem uma vasta experiência nessa área, sendo capaz de reportar não apenas os casos recentes do contexto do BD, mas também anteriores a ele. Essa perspectiva pode ser interessante para a compreensão da evolução do reúso dos dados ao longo do tempo e a identificação de características específicas do cenário de dados e tecnologia atuais. Em conversas preliminares com a diretoria do *bureau* foi possível identificar casos de repropósito de dados externos públicos e privados que correspondem ao potencial de desafios “intermedial”, destacado em amarelo no quadro 4.1.

- Caso *fintechs* – acompanhando as atividades das *fintechs* pela mídia especializada e por eventos da área, percebeu-se que essas organizações têm proposto usos mais inovadores dos dados, sendo caracterizados como recontextualização de dados externos públicos ou privados, que representa o potencial de desafios extremo, destacado em vermelho no quadro 4.1. Assim, o reúso de dados em uma *fintech* parecia completar adequadamente o rol de casos deste estudo. Ao iniciar o contato com esse tipo de empresa, porém, identificou-se que, em função de sua estrutura, geralmente nova e pequena, a quantidade de entrevistas seria muito limitada para compor um caso, o que ficaria desbalanceado em relação aos demais casos do

estudo. Para resolver a questão, decidiu-se elaborar o caso a partir da investigação de três *fintechs*. Para viabilizar tal composição, foram escolhidas empresas que atuam na mesma área: crédito e/ou cobrança. O estudo dessas organizações contribui também para a exploração do reúso de dados em meio a novos modelos de negócios de empresas jovens, denominadas *startups*.

O apêndice A apresenta o desenho completo da pesquisa, lembrando que este artigo refere-se apenas ao estudo de casos múltiplos. A figura também demonstra as questões específicas de cada caso.

4.4.3 Coleta dados

Os dados foram coletados por meio de entrevistas semi-estruturadas e realizadas pela autora entre 24/01/2018 e 07/12/2018. No fim, somaram-se 22 entrevistas (sete no Banco Ômega, nove no Bureau Sigma e seis nas *fintechs*) com duração média de 1h21min, totalizando 29h45min. O local e a duração das entrevistas, bem como a função e o codinome dos entrevistados são descritos na tabela 1. A seleção dos entrevistados foi baseada em conversas preliminares com interlocutores-chave nas organizações que colaboraram para identificar as pessoas com participação relevante em projetos e práticas de reúso de dados em cada empresa. Executivos e especialistas das áreas técnicas e de negócios envolvidos diretamente nesse campo compuseram o rol de entrevistados. No caso do Bureau Sigma, também foram incluídos três clientes externos para captar os efeitos práticos no reúso de dados que o serviço viabiliza. As entrevistas foram gravadas pela autora e transcritas por um profissional contratado. Os arquivos com as transcrições foram importados no Atlas.ti, software utilizado para análise dos dados.

Data	Caso	Local Entrevista	Duração	Função do Entrevistado	Codinome
24/01/2018	Banco Ômega	Banco Ômega - filial 1	44min	Gerente de uma das áreas de Diretoria de Dados	OG1
24/01/2018	Banco Ômega	Banco Ômega - filial 1	33min	Gerente de área de negócio	OG2
24/01/2018	Banco Ômega	Banco Ômega - filial 1	58min	Gerente de uma das áreas de Diretoria de Dados	OG3
03/06/2018	Banco Ômega	Banco Ômega - filial 2	1h	Gerente de uma área Técnica	OG4
03/06/2018	Banco Ômega	Banco Ômega - filial 2	1h 13min	Supervisor de uma área Técnica	OS1
21/06/2018	Banco Ômega	Banco Ômega - filial 1	43min	Analista Senior - especialista matemático - Área de Negócio	OA1
12/07/2018	Banco Ômega	Banco Ômega - filial 1	53 min	Analista Senior - especialista matemático - Área de Negócio	OA2
Total Caso Banco Ômega			6h4min	7 Entrevistas	
15/02/2018	Bureau Sigma	Bureau Sigma - filial 1	2h 09min	Diretor	SD1
15/02/2018	Bureau Sigma	Bureau Sigma - filial 1	1h	Funcionário da operação - atendimento cliente	SO1
15/02/2018	Bureau Sigma	Bureau Sigma - filial 1	3h19min	Diretor	SD2
23/02/2018	Bureau Sigma	Telefone	46min	Cliente	SC1
27/02/2018	Bureau Sigma	Telefone	1h21m	Cliente	SC2
28/02/2018	Bureau Sigma	Bureau Sigma - filial 2	1h 28min	Gestor TI	SG1
28/02/2018	Bureau Sigma	Bureau Sigma - filial 2	1h 57min	Gestor Projetos/Produto	SG2
28/02/2018	Bureau Sigma	Bureau Sigma - filial 2	2h 10min	Diretor	SD3
03/08/2018	Bureau Sigma	Telefone	43min	Cliente	SC3
Total Caso Bureau Sigma			14h53min	9 Entrevistas	
10/09/2018	Fintech Delta	Telefone	1h02min	Diretor Geral	FD3
23/11/2018	Fintech Delta	Fintech Delta - filial 1	2h15min	Diretor de Tecnologia	FD6
21/11/2018	Fintech Gama	Fintech Gama - filial 1	1h20min	Diretor Geral e de Operações	FD4 e FD5
07/12/2018	Fintech Gama	Fintech Gama - filial 1	1h18min	Diretor de Tecnologia	FD7
11/07/2018	Fintech Zeta	Escritório da autora	1h 29min	Diretor Desenvolvimento Negócios	FD1
06/08/2018	Fintech Zeta	Café em SP	1h 24min	Vice-Presidente Produto	FD2
Total Caso Fintechs			8h48min	6 entrevistas	
Total Geral			29h45min	22 Entrevistas	

Quadro 4.2 - Quadro de entrevistas dos casos

Fonte: elaborado pela autora.

Além das entrevistas, foram coletados outros documentos públicos pesquisados pela autora nos sites das empresas envolvidas, entidades do setor financeiro, dos órgãos públicos e portais de notícias ou fornecidos pelos entrevistados (por exemplo, casos da organização que foram publicados e materiais de divulgação). Essas informações não foram carregadas no Atlas.ti para compor a base de dados a ser codificada na pesquisa, mas serviram como base de apoio para enriquecer o entendimento de questões que surgiram nas entrevistas (especialmente sobre a história e os serviços das empresas). Elas também foram alternativas para comparar as perspectivas dos entrevistados e confirmar a interpretação do pesquisador, o que pode ser considerado um tipo de triangulação que contribui para a fidelidade dos relatos e a capacidade do texto transmitir os dados coletados (Creswell, 2007; Stake, 2006).

O protocolo das entrevistas (apêndice B) reuniu um conjunto de questões a serem abordadas para estimular os entrevistados a relatarem suas experiências de forma narrativa, contando histórias. O protocolo não foi usado como um guia sistemático rígido, mas uma referência que foi se ajustando ao longo da experiência em campo (Gioia et al., 2013). Como a pesquisa busca compreender o processo no qual o reúso de dados emerge, é importante que as informações coletadas possam expressar eventos significativos, suas relações e sua sequência

cronológica (Pettigrew 1990, Langley 1999). Privilegiar os relatos mais livres e as histórias dos entrevistados mostrou-se uma estratégia assertiva, permitindo a reconstituição do processo. Também limitou a interferência do entrevistador sobre os dados, já que as perguntas podem influenciar e/ou trazer vieses ao estudo (Creswell, 2007). Ao pedir para que os entrevistados contassem suas histórias, e não perguntar sobre questões específicas, permitiu a eles selecionar e abordar os pontos na sequência e dimensão que eles mesmos escolheram. Nessa estratégia, a entrevistadora fez indagações pontuais para elucidar questões que emergiam da própria história do participante, propondo algum novo tema para mudar de assunto quando a história já estava esgotada (o entrevistado está repetindo ou agregando detalhes e percepções muito semelhantes) ou quando o próprio entrevistado conclui o relato.

4.4.4 Análise dos dados

Os documentos foram importados no software Atlas.ti em três projetos, um para cada caso, já que, em primeiro momento, a análise dos casos foi feita individualmente, como recomenda Stake (2006). A codificação dos dados não partiu de categorias prévias, mas foi construída a partir das revelações oferecidas pelos próprios dados. O processo de análise foi inspirado na proposta de Strauss e Corbin (1998), que sugere três fases de codificação (aberta, axial e seletiva), as quais foram adaptadas às necessidades da pesquisa. Na codificação aberta, foi realizada a microanálise dos dados e optou-se pela criação de códigos mais descritivos, mantendo marcas de eventos e atores que ajudaram a preservar as vozes dos entrevistados e identificar a cronologia e as relações das questões abordadas, dando origem a um grande volume de códigos. Na etapa axial, os códigos foram agrupados em temas que refletiam atores, eventos, estratégias ou percepções, mantendo elementos que pudessem expressar as relações entre eles. A codificação seletiva buscou identificar os elementos mais relevantes para o caso e a história que eles contavam, construindo um mapa visual para representá-la. Além disso, foram adicionadas duas etapas para transpor os dados para o modelo teórico e para a análise *cross-case*. A tabela 2 apresenta o produto dessas etapas, que são detalhadas no apêndice C.

Etapa	Banco Ômega	Fintechs	Bureau Sigma
1 - Codificação aberta	464 códigos	633 códigos	340 códigos
2 - Codificação axial	91 agrupamentos	32 agrupamentos	45 agrupamentos
3 - Codificação seletiva	Mapa do caso	Mapa do caso	Mapa do caso
4 - Releitura lente conceitual	Multilevel Framework do caso	Multilevel Framework do caso	Multilevel Framework do caso
5 - Análise <i>cross-case</i>	Relatório unificado dos resultados dos casos		

Quadro 4.3 - Produtos da análise dos dados por etapa

Fonte: elaborado pela autora

4.5 Resultado dos casos

Para demonstrar o que foi capturado, será apresentado um resumo da história de cada caso e, em seguida, seus dados sumarizados no contexto, processo e conteúdo do Multilevel Framework.

4.5.1 O caso do Banco Ômega

O Banco Ômega é um dos maiores bancos do Brasil. Seu crescimento combinou a expansão orgânica da operação e a compra de outros bancos. É reconhecido pelo seu constante investimento em tecnologia e suas equipes buscam adotar as melhores práticas e ferramentas de TI do mercado. A operação dos bancos no Brasil é altamente regulada e tem que lidar com a complexidade tributária e a instabilidade econômica do país. A TI sempre foi um elemento fundamental na estratégia do Ômega para atender às demandas desse cenário.

Além da informatização de todo o fluxo de atendimento, o banco conta com diversos sistemas especializados para a gestão de suas áreas e produtos. A equipe de TI também é estruturada por sistemas e áreas de negócio, visando a especialização de seus times e a eficiência do atendimento. Essa estrutura também é alinhada com a organização hierarquizada de todo o banco e com suas políticas de governança de dados e gestão de riscos, que são orientadas à segregação de acessos e proteção dos dados.

Nos últimos anos, as tecnologias utilizadas pelo banco, ainda que fossem as mais avançadas em seus segmentos, foram se tornando limitadas para atender às novas demandas de manejar volumes crescentes de dados e oferecer ferramentas de inteligência e agilidade para a operação. Em meados de 2012, o banco começou a investir em novos padrões tecnológicos e iniciou a implementação de ferramentas associadas ao BD. Também começou a evoluir a ideia do reúso de dados e da importância de fomentá-lo, especialmente para resolver problemas de eficiência e qualidade de dados que, muitas vezes, eram duplicados entre as áreas, aumentando o custo de manutenção dessas informações e inviabilizando sua integridade.

Por meio da adoção de tais ferramentas, diversos projetos e eventos foram trazendo exemplos concretos do potencial desses modelos e do reúso de dados, mas sua adoção ainda acontecia de forma tímida, localizada em áreas e questões específicas, não sendo uma prática disseminada por todo o banco. Por volta de 2016, o banco incluiu em sua estratégia corporativa a exploração de dados (*data discovery*) para geração de valor para o negócio e o cliente, dando início à sua jornada de dados.

Além do direcionamento da alta gestão do Ômega para todas as áreas, também foram realizados grandes investimentos em tecnologia e pessoas, intensificando a formação e trazendo novos profissionais do mercado. Além disso, a diretoria de dados, com seu novo diretor, passou por transformações: foi elevada na hierarquia do banco, mudou o perfil e aumentou sua equipe, assumindo a missão de promover o reúso de dados em todas as áreas.

A expansão do reúso de dados gera implicações e desafios que retroalimentam as ações corporativas e da diretoria de dados, buscando solucionar as questões emergentes para seguir rumo ao uso qualificado de dados para trazer simplificação, agilidade e inteligência aos negócios. O apêndice D descreve o contexto, o processo e o conteúdo que compõe o Multilevel Framework do caso.

4.5.2 O caso do Bureau Sigma

O bureau Sigma é um dos *quatro bureaux* de crédito autorizados do Brasil. Ele nasceu da visão de experientes executivos do setor bancário, que acreditavam que a tecnologia poderia agilizar os serviços financeiros tradicionais e viabilizar novos modelos de operações. No início, enfrentou dificuldades devido aos limites tecnológicos da época e à falta de

maturidade do mercado para aderir ao modelo inovador que propôs. Mas foi se adaptando à demanda do mercado, mantendo-se na proposta de oferecer informações confiáveis para validação cadastral e decisões de crédito de instituições do setor financeiro e para a operação financeira de empresas de diversos setores, em especial, do varejo.

Foi pioneira no lançamento de uma ferramenta de *score* de crédito, que se tornou um componente importante de sua receita. Foi seguida pela concorrência, mas tem o diferencial da flexibilidade, oferecendo muitas opções de customização ao cliente. Depois desse lançamento, realizou melhorias em suas ferramentas e serviços ao longo do tempo, atendendo às demandas dos clientes, adequando-se aos avanços das tecnologias e da oferta de dados. Nenhuma grande mudança ou novo produto foi criado.

Por volta de 2016, o conselho de direção da Sigma identificou a necessidade de mudanças em sua gestão e operação. Embora a empresa estivesse saudável e tivesse uma participação de mercado estável, percebeu uma tendência de transformação em sua área que ameaçava o futuro de seu modelo atual. Após a posse de um novo presidente, iniciou-se uma grande reestruturação, visando garantir a sustentabilidade do negócio no longo prazo, estimular a inovação, além de trazer agilidade e inteligência de dados para o centro da operação.

Entre as principais mudanças, houve uma revisão da estrutura da empresa, rompendo com o modelo pesado e hierarquizado inspirado nas organizações financeiras tradicionais. A infraestrutura de tecnologia foi modernizada com o apoio de parcerias externas para a adoção de práticas mais ágeis, econômicas e tecnicamente superiores. Um novo esquema de governança tecnológica e de dados foi implementado, visando agilizar os processos e trazer conhecimentos multidisciplinares para as decisões sobre o desenvolvimento de seus produtos e oferta/uso de dados.

Essa nova estrutura é fundamental para o ambiente de mudanças em que a Sigma está inserida. As transformações da tecnologia, das leis, da oferta e do mercado de dados vêm mudando os padrões de acesso e captura de informações que ela utiliza ou pode utilizar. Também vêm mudando as expectativas e demandas de seus clientes. A Sigma precisa ser ágil para avaliar novas demandas e condições, assim como para identificar as oportunidades desse cenário no sentido de inovar e se diferenciar da concorrência, que agora inclui as agências de informação. Tais agências sempre existiram, mas suas possibilidades de atuação expandiram-se muito no cenário de dados atual, embora não tenham os mesmos acessos às bases oficiais do governo que continuam restritas a quatro *bureaus* de crédito autorizados.

Por meio do uso de dados, a Sigma pauta sua estratégia no desenvolvimento de inteligência para seus clientes e em sua própria operação. Seu objetivo é oferecer opções prontas e acessíveis para determinadas aplicações e clientes, além de viabilizar soluções diferenciadas que permitam aumentar o nível de automação e inteligência para consumidores dispostos a explorar as possibilidades da nova era de dados.

Ao longo da mudança e da operação em sua nova estrutura, mais orientada ao uso inteligente de dados, a Sigma vem percebendo uma série de desafios que demandam novas ações e ajustes da gestão e dos modelos de trabalho. O apêndice E apresenta o contexto, processo e conteúdo retratados pelo Multilevel Framework.

4.5.3 O caso das *fintechs*

O caso das *fintechs* foi composto a partir dos minicase de três empresas (Zeta, Delta e Gama) que serão descritos a seguir. As entrevistas colhidas permitiram compor o caso das *fintechs* de crédito. O apêndice F apresenta o contexto, processo e conteúdo do Multilevel Framework.

4.5.3.1 Minicase da *fintech* Zeta

A Zeta é uma *startup* do Vale do Silício, nos Estados Unidos da América (EUA), que está iniciando sua operação no Brasil. Tem profundas raízes na área de tecnologia, de onde vem seus fundadores e capital (uma de suas acionistas é uma grande fabricante de aparelhos eletrônicos e celulares). Possui grande expertise em BD, lógica de jogos (*game mechanics*) e experiência de clientes no uso de aplicativos. Seu propósito é aplicar esse conhecimento em uma oferta de serviço que promova a inclusão financeira, podendo promover, em alguns casos, a inclusão digital para viabilizar o acesso a serviços financeiros mais modernos.

Ela propõe um novo olhar no qual todas as pessoas sejam vistas como aptas a obter crédito e construir sua “pontuação” a partir de seu comportamento nesse ambiente. Assim, todos terão a oportunidade de construir um perfil positivo de crédito, independentemente de seus

históricos habilitadores (renda comprovada, propriedades, etc.) ou dificultadores (nome negativado, por exemplo) de crédito nos modelos tradicionais.

Os principais desafios do negócio da Zeta eram chegar às pessoas (penetração) e disponibilizar empréstimos para uma grande massa, o que requer capital e um modelo regulamentado de operação financeira. Apostou na parceria com operadoras de celular para resolver tais questões e formulou um modelo de negócios que viabiliza realizar seus propósitos de inclusão, gerando também valor aos parceiros.

Atualmente, as operadoras oferecem crédito de celular para pagamento futuro aos clientes de planos pré-pagos quando seu crédito chega ao fim. Tal ação inclui juros muito altos, pois é encarada pelas operadoras como um empréstimo, um serviço financeiro de risco. A proposta da Zeta é assumir a gestão desse processo, usando suas ferramentas de inteligência para oferecer pequenos valores a todos os clientes, indistintamente, e compor o perfil deles a partir de seu comportamento diante da oferta de crédito e das demais interações que passam a ter com a ferramenta, que vai capturando seus dados durante a navegação. À medida que esse perfil é criado, o cliente pode ter acesso a valores maiores de crédito, podendo chegar a financiar aparelhos celulares a taxas mais baixas, o que viabiliza a consumidores de baixa renda ter acesso a *smartphones* com internet (inclusão digital). Dessa forma, tais clientes poderão baixar os aplicativos da Zeta e de outras organizações que promovem serviços financeiros alternativos para a inclusão financeira.

A Zeta possui operações em diversos países e com muitas operadoras, havendo diferentes modelos de negócios para se adaptar às legislações e culturas locais, assim como aos requerimentos de cada parceiro. Em alguns casos, a Zeta assume todo o risco, pagando à operadora pelos créditos concedidos aos clientes. Em outros, a operadora compartilha o risco. A proposta é que as operadoras sempre ganhem com a fidelização e o aumento do consumo do usuário, sendo superior ao ganho que ela teria na forma de juros sobre os créditos para pagamento futuro. Ao passo que a operação funciona e ela expande sua atuação, a Zeta identifica desafios e oportunidades, refletindo sobre essas questões e buscando soluções que podem implicar em ajustes no serviço ou dar origem a mudanças mais profundas em sua solução, modelo de negócios e parcerias.

4.5.3.2 Minicase da *fintech* Delta

A Delta é precursora do modelo de cobrança digital estabelecido no mercado brasileiro. Seus fundadores, ex-executivos de grandes bancos, conheciam bem esse ambiente e acreditavam que o processo de cobrança e negociação de dívidas deveria ser modernizado pelas novas tecnologias, tornando o processo de negociação mais simples, menos traumático e mais conveniente aos devedores.

O mercado já operava com cobradoras físicas, que recebiam as carteiras dos credores, iniciavam a cobrança aos devedores e ofereciam propostas de negociação. Para aderir ao novo modelo, é preciso que os credores estejam dispostos a realizar algumas mudanças, especialmente se quiserem acessar todo o potencial de inteligência e automação que a cobrança digital pode oferecer.

O modelo de cobrança digital da Delta é baseado em uma plataforma que reúne as informações atualizadas das dívidas e condições de negociação disponíveis para os devedores visualizarem claramente seus débitos e possibilidades de quitação, enviar dúvidas e propostas de pagamento, ou solicitar atendimento personalizado. Assim, o cliente poderá realizar todo esse processo no local e horário de sua conveniência, sem passar pelo constrangimento da abordagem de cobrança. A ferramenta também oferece funcionalidades para ajudar a manter os acordos, como lembretes das datas de pagamento das parcelas e da necessidade de poupar ao longo do mês para honrar seu compromisso.

Para viabilizar essa operação, é necessário receber os dados dos credores sobre as dívidas e suas condições, e ter uma interação constante com eles para enviar as propostas recebidas dos devedores e receber suas respostas, assim como a atualização dos pagamentos já realizados. Também é importante que a plataforma capture dados de outras fontes para fomentar seus modelos de inteligência, que ajudam a localizar os clientes (corrigindo dados de contato, por exemplo) e trazê-los para a plataforma de negociação, especialmente em momentos mais propícios para negociação (término do pagamento de outra dívida, entrada de renda extra, etc).

A proposta de valor da Delta oferece um novo olhar sobre a cobrança, transformando-a em um novo modelo de relacionamento do credor com seu cliente. Para que esse valor se realize, no entanto, é preciso que os credores compreendam que não se trata de reproduzir o modelo da cobrança tradicional. Operar nesse novo cenário requer que os sistemas

dos credores estejam aptos às integrações necessárias com a plataforma e que tenham capacidade de resposta compatíveis com as demandas de uma operação on-line. Também exige que sua visão de cobrança seja modificada para que a abordagem do cliente seja feita com inteligência de dados, respeitando os momentos em que ele reúne mais condições de pagamento, usando os canais de contato adequados e uma oferta de serviços atrativa ao cliente.

A Delta está atenta às oportunidades e aos desafios que surgem ao longo de suas operações. Ela busca constantemente endereçá-los, ajustando seus produtos e serviços ou criando novas soluções, parcerias e modelos de negócios, se necessário ou oportuno. Um exemplo disso foi seu recente reposicionamento para ser não apenas uma plataforma que ajuda o cliente a quitar suas dívidas, mas também a se manter sem dívidas. Já possui algumas iniciativas para expandir os serviços nesse sentido, como uma plataforma complementar que indica caminhos para que o usuário possa gerar uma renda adicional. O acesso a essa plataforma é aberto, podendo atender não apenas os clientes devedores da plataforma de cobrança, mas qualquer pessoa que a acesse pela internet. As informações dessa aplicação também compõem a base de dados para a inteligência de cobrança da Delta.

4.5.3.3 Minicase da *fintech* Gama

A Gama nasceu do encontro de ex-executivos de grandes instituições financeiras do mercado de crédito, que tinham em comum a paixão pela tecnologia e a percepção que esse mercado poderia se beneficiar dela para melhorar seus serviços. Inspirados por novos modelos de crédito que surgiam em outros países, trabalharam para elaborar um tipo de serviço que trouxesse a conveniência do mundo digital para esse processo, até então penoso para quem solicitava um empréstimo.

Já havia algumas iniciativas no Brasil, mas sua plataforma foi pioneira e propôs o modelo que foi adotado pelo mercado de crédito on-line desde então. Como no caso da cobrança digital, sua plataforma também requer a integração com os sistemas dos credores. Nesse caso, eles devem receber as solicitações de empréstimo dos clientes (usuários da plataforma), responder com suas ofertas de crédito, receber e responder contrapropostas dos clientes. A plataforma oferece aos credores a possibilidade de trabalhar com a pontuação de crédito que ela calcula ou com a pontuação definida pelo sistema do credor. Eles definem sua política na

plataforma para que ela apenas direcione as solicitações de crédito que atendam aos seus critérios.

Inicialmente, os credores resistiram ao modelo, pois se incomodavam com a oferta de múltiplas propostas de crédito aos clientes, que poderiam facilmente compará-las. A Gama teve o trabalho de educá-los, mostrando que essa comparação já acontecia no mundo físico e que o modelo oferecia mais oportunidades do que riscos à competitividade dos participantes. Após o convencimento dos primeiros credores, a adesão dos demais avançou com facilidade e hoje a Gama tem grandes financeiras e bancos operando em sua plataforma.

Além da conveniência para todos os envolvidos, a Gama aposta em um modelo de inteligência de dados para que o usuário consiga obter a melhor oferta de crédito possível e o credor receba as demandas dos clientes mais apropriados ao perfil que lhe interessa.

Embora ainda capte dados de fontes variadas (credor, *bureaus* de informação, etc.), a Gama tem solicitado cada vez mais dados para o próprio cliente. Em seus testes de validação, verificou-se, por exemplo, que a melhor fonte de dados de contato é o formulário preenchido na solicitação do empréstimo. A *startup* também foi pioneira em solicitar ao cliente informações sensíveis, como perguntar se ele está negativado. Para a surpresa de muitos, o usuário as fornece com transparência. Além disso, a Gama solicita acesso aos dados do extrato bancário digital do cliente, o que também é concedido pela maioria. A empresa acredita que o cliente está interessado em obter a melhor oferta de crédito possível e está disposto a compartilhar informações que possam contribuir para isso, desde que confie na plataforma. Por essa razão, a Gama toma muito cuidado na interação com o cliente e trata com cautela o uso de dados de redes sociais para evitar a quebra de confiança do usuário em relação à plataforma.

Esses são alguns exemplos da evolução que vem ocorrendo na Gama, que se mantém atenta às demandas e oportunidades, buscando novas soluções ou parcerias para atendê-las.

4.6 Análise *cross-case*

Ao reportar a análise *cross-case*, diversos pesquisadores enfatizam as similaridades encontradas nos casos. No entanto, Stake (2006) destaca a importância de discutir suas diferenças, lembrando dos distintos papéis de cada caso na compreensão do *quintain* da

pesquisa. Esse relato da análise *cross-case* apresentará similaridades e diferenças dos casos, buscando refletir sobre o que elas revelam acerca do reúso de dados.

4.6.1 O contexto

O resultado dos casos demonstrou que o reúso de dados está inserido em um contexto pluralista, assim como o caso exploratório da etapa anterior já havia indicado. Um contexto pluralista é marcado por atividades baseadas em conhecimento, múltiplos objetivos e poder difuso (Denis, Langley & Rouleau, 2007). O reúso de dados pode ser visto como uma atividade baseada em conhecimento, necessário em suas diversas etapas: da captura dos dados até o desenvolvimento de ferramentas/métodos para reutilizá-los para novos fins. Ele implica múltiplos atores, diretamente envolvidos no processo ou agindo no ambiente que o influencia. Tais atores podem ser de áreas distintas e possuir diferentes pressupostos tecnológicos e objetivos em relação ao reúso. Nos casos da área financeira abordados neste estudo, os grupos envolvidos são geradores, depositários (que “guardam”) ou consumidores (que reusam) dos dados; provedores das ferramentas e serviços que reusam os dados (banco, *bureau*, *fintechs*); fornecedores de tecnologias; legisladores e clientes dos serviços viabilizados pelo reúso (clientes do banco e do *bureau*, credores, devedores e tomadores de crédito); além da sociedade; dos mercados de dados e do mercado financeiro, no qual se dá o reúso. Nesse contexto de tantos personagens, não há um núcleo de poder centralizado que detenha toda a autoridade e o controle sobre os recursos, sendo, portanto, um ambiente de poder difuso entre os tantos atores que dele participam.

Além da pluralidade de áreas às quais pertencem os atores que participam do reúso de dados, tais atores podem operar em parcerias (como a das *fintechs* Delta e Gama com empresas credoras), alianças formais (como no caso da *fintech* Zeta, cujo acionista é uma fabricante de celulares) ou operações conjuntas pontuais (como no exemplo do Bureau Sigma com empresas das quais compra dados eventualmente). A colaboração entre os atores no contexto do reúso de dados pode, portanto, ser denominada complexa, já que cruzará fronteiras entre áreas, negócios, grupos formais e culturas (Faraj et al., 2012).

Ademais dessa complexidade inerente ao reúso, o ambiente tecnológico, legal e societal no qual ele está inserido passa por transformações que impactam o mercado de dados

e as possibilidades de serviços financeiros, movimentando ainda mais seu contexto. Dessa forma, as empresas do setor financeiro percebem a necessidade de se adaptar ao novo cenário para sobreviver, sendo forçadas a rever seus padrões. Ao mesmo tempo que aumenta seus desafios, essa condição também traz oportunidades ao mercado, pois permite a operação de novos modelos de serviços e novos entrantes.

4.6.2 O processo e o conteúdo que dele emerge

Em todas as empresas participantes do estudo, o reuso de dados foi descrito em meio a um processo de mudança mais amplo e motivado por objetivos do negócio. Nos casos do Bureau Sigma e do Banco Ômega, que são mais antigos, ele já acontecia, mas as empresas realizaram uma grande mudança de estrutura e estratégia, apostando no reuso de dados como uma ferramenta potencializadora de inteligência e novas oportunidades de negócio. Já nas *fintechs* Zeta, Delta e Gama, que são mais jovens, ele faz parte da lógica do negócio que deu origem às companhias. Em todos os casos, a implementação do reuso de dados tem gerado desafios que retroalimentam ações de mudança, ajustando processos, estruturas ou modelos de negócios às condições e oportunidades emergentes, em uma dinâmica cíclica contínua. A fala do entrevistado abaixo, referindo-se a uma “jornada”, ilustra tal dinâmica:

Então, acho que a mudança cultural tá passando por a liderança dar mais abertura para as pessoas, e a gente tentar trazer alguns perfis também de pessoas que têm essa pegada de viver a tecnologia e de instigar a mudança e a transformação. Então, eu acho que tá numa jornada... OG2 do Banco Ômega.

Os três casos relatam questões ligadas às cinco dimensões consideradas no modelo (legal, tecnológica, intrínseca, societal e econômica), corroborando a relevância destas para viabilizar o reuso de dados. Ao classificar as menções dos entrevistados sobre os desafios encontrados na prática, pode-se observar o foco de cada caso nessas dimensões (gráfico 4.1). É importante lembrar que tais dimensões dialogam e se influenciam mutuamente, havendo questões que poderiam impactar em múltiplas delas. Essa associação das questões citadas pelos entrevistados com as dimensões não pretende ser uma definição absoluta, mas apenas indicar focos de preocupação que emergem dos casos e refletir sobre eles.

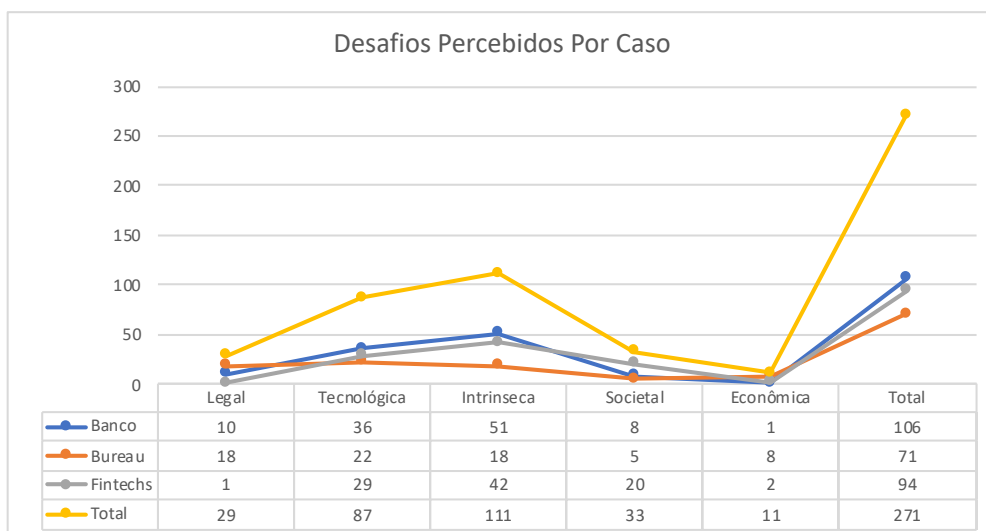


Gráfico 4.1 - Desafios do reúso de dados percebidos por caso

Fonte: elaborado pela autora

No geral, a dimensão intrínseca destaca-se entre os desafios percebidos em todos os casos, o que corrobora a inclusão da dimensão proposta em etapa anterior dessa pesquisa. A dimensão tecnológica também se sobressai em todos os casos, indicando os desafios práticos para a implementação das tecnologias disponíveis e para o reúso dos dados existentes. Já a dimensão econômica é pouco lembrada pelos entrevistados, sugerindo que, embora haja necessidade de investimentos, a sustentabilidade financeira do reúso não está entre as questões mais críticas.

Ainda que haja certa coerência na distribuição das questões entre as dimensões dos casos, nota-se algumas particularidades em cada um. O caso do Bureau Sigma retrata o repropósito de dados externos públicos e privados e destaca a preocupação com as questões econômicas e legais, o que é coerente com seu cenário atual. Mudanças na política de acesso a bases oficiais do governo vêm aumentando o custo desses dados, que são o diferencial de seu negócio, como relata um dos diretores da Sigma: “[...]eles concederam agora uma nova porta de entrada pra quem é *bureau* oficial, que é link direto [...] só que é uma informação que se tornou muito cara [...]” – SD3 da Sigma.

Novos concorrentes chegam ao mercado e a pressão por redução de preço aumenta entre os clientes, especialmente aqueles que buscam por serviços padrão com menor nível de inteligência:

Vou te fornecer uma base de pessoas que são perfil do seu produto. Só que isso vai custar dez vezes mais. Então, não quero... Ele prefere fazer 30 ligações para fazer uma venda, gastar o dobro do preço do que pagar metade e ter uma ligação certa. Não consegue fazer essa conta. SD2 da Sigma.

As leis ainda estão se delineando, não oferecendo regras claras, o que permite atuações de concorrentes que se considera pouco éticas e que “poluem” o mercado de dados.

O caso das *fintechs* destaca-se pela maior ênfase às questões societárias, o que também é coerente com seu cenário. O reúso de dados das *fintechs* envolve o repropósito e a recontextualização de dados externos públicos ou privados, ou seja, ele necessita de informações geradas pela sociedade em situações diversas. O comportamento da sociedade na criação e concessão de acesso aos dados impacta diretamente na viabilidade desse reúso, do qual dependem seus negócios. Além disso, o reúso de dados nas *fintechs* faz parte de um novo modelo de serviços que necessita da aceitação pública para se consolidar e desenvolver.

A fonte de dados para nós não importa, desde que possamos ter acesso... Acho que o principal desafio que temos é explicar aos consumidores de forma que eles se disponham a nos dar acesso a seus dados, porque eles entendem o que significa emprestar assumindo riscos... Obtendo, em contrapartida, taxas de juros mais competitivas. FD2 da Zeta

Já no Banco Ômega, as preocupações estão fortemente concentradas nas dimensões intrínseca e tecnológica, o que também é coerente com o tipo de reúso estudado no caso: repropósito ou recontextualização de dados internos. A dimensão econômica é marcada pela convicção sobre o valor do reúso de dados e sua capacidade de compensar todos os investimentos, protegendo sua participação de mercado da concorrência e gerando novos negócios e receitas, como ilustra a fala de um dos gerentes da área de dados do Ômega:

O ponto que a gente quer chegar aqui é um pouco além, porque hoje eu já sei que mais dado vai melhorar meus negócios atuais, e a gente já tem feito isso... A gente já sabe que a maneira de ganhar dinheiro que os bancos têm há anos, ele vai ter que sofrer pelas inovações. E a gente acredita que dados é o coração dessa nova fase. E como que eu penso novos produtos e novas receitas pro banco a partir de dados... A gente quer melhorar muito o nosso atendimento, ser centrado no cliente, oferecer sempre o melhor produto pro cliente... Porque, hoje, o cliente tem opção, o cliente tem bons atendimentos fora daqui... E talvez a forma de remuneração seja outra, com informação eu consigo ter modelos melhores, eu consigo ter mais inteligência, tomar melhores decisões. A gente precisa traduzir essas informações em dinheiro e mudar a maneira como a gente faz os negócios. É um desafio nosso para uma próxima fase. – OG1 do Ômega

4.6.3 Os desafios do reúso de dados

Ao refletir sobre a natureza dos desafios mencionados, percebe-se que uma parte significativa deles refere-se a algum tipo de mudança que o reúso de dados implica: na tecnologia, na fonte de dados, no processo em que o reúso está inserido, no comportamento das pessoas sobre suas informações, na forma de trabalhar com dados, etc. Um dos gestores da Sigma ilustra esse ambiente de mudanças:

Então, as tecnologias de produtos nossos foram acompanhando essa mudança. Tecnologias que a gente nunca imaginou que ia estar utilizando pra dados como performance, indexação... A gente nunca tinha pensado cinco anos atrás; a gente utiliza hoje. SG2 da Sigma

O reúso de dados já existia, mas é algo novo na roupagem atual do BD. Portanto, as mudanças relatadas neste momento não estão ligadas apenas ao caso em questão, mas fazem parte de um pano de fundo sobre o qual uma nova forma de usar dados está se delineando. É possível, portanto, questionar se, em um momento futuro, quando essas bases já estiverem consolidadas, tais desafios diminuirão.

O gráfico 4.2 demonstra a forte presença de questões ligadas a mudanças em todas as dimensões do reúso de dados. Aqui também não se pretende uma classificação absoluta ou uma análise quantitativa, mas apenas ilustrar como o cenário atual de transformações parece impactar os desafios desse âmbito.

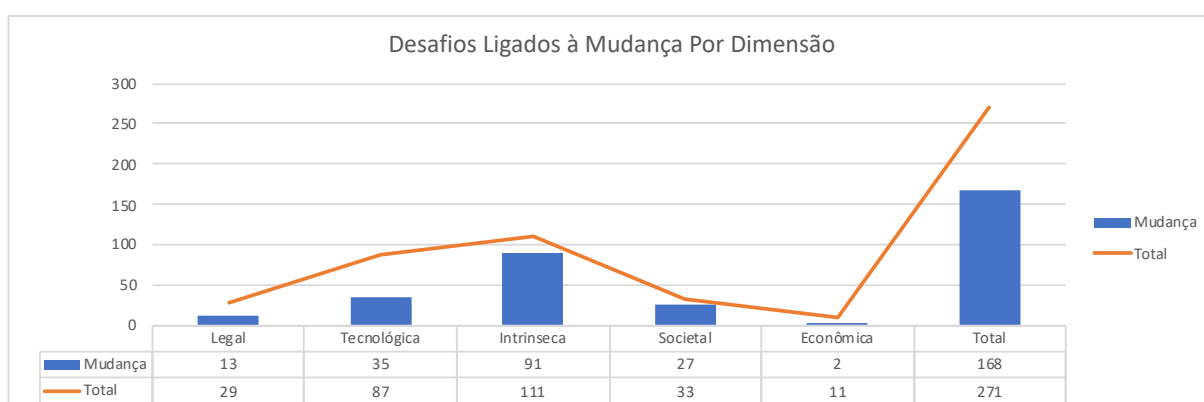


Gráfico 4.2 - Desafios ligados à mudança por dimensão
Fonte: elaborado pela autora

O apêndice G traz um resumo de todos os desafios listados nos casos por dimensão. A análise dos desafios mencionados apontou para seis temáticas:

- a) tecnologia;
- b) acesso aos dados;

- c) confiança para ter acesso e legitimar o uso dos dados;
- d) mudança de paradigma da qualidade e curadoria de dados;
- e) ajuste de processos;
- f) falta de profissionais capacitados.

A seguir, uma reflexão sobre essas temáticas:

a) a tecnologia merece lugar de destaque em todos os casos. É opinião unânime de que as técnicas e ferramentas de dados atingiram altíssimo padrão e que ainda há grande potencial inexplorado na tecnologia existente. Também reconhecem a contribuição das ferramentas abertas e dos novos modelos de serviço (*open source, cloud, SAAS, etc.*) para viabilizar o acesso à tecnologia de ponta e prover a escalabilidade e a dinâmica necessárias no cenário de dados atual, como aponta o supervisor de uma das áreas técnicas do Ômega:

E aí ultimamente tem crescido muito a criação de ferramentas *open source*. Então, em algum lugar alguém que já teve esse problema e já resolveu de uma forma... É muito mais simples do que era no passado... Mais aberta, atualização cada vez mais constante, muita gente olhando pras mesmas coisas, que aí gera mais documentação sobre essas tecnologias. Então, nesse lado eu não vejo que seja um gargalo. – OS1 do Ômega

Todos reconhecem, porém, que a implementação dessas tecnologias ainda não está 100% resolvida na prática, como exemplifica um dos diretores da Delta:

Trabalhar na nuvem elimina uma série de dificuldades, mas traz algumas novas dificuldades que antes não existiam e que agora têm que ser percebidas, e que não tem um roteirinho[...]. FD6 da Delta

Algumas dificuldades mencionadas nos casos foram: incompatibilidades e problemas de integração com sistemas legados, falta de padrões e boas práticas para implantação de ferramentas novas ainda pouco utilizadas, procedimentos de governança e controle inadequados às novas estruturas tecnológicas, falta de conhecimento sobre como implementar os controles adequados às novas estruturas;

b) o acesso aos dados apareceu de forma distinta entre os casos. O caso do Banco Ômega retrata reúso de dados internos e, embora a gestão do acesso às informações esteja sob sua tutela, sua adaptação ao novo cenário de estímulo à exploração de dados é um dos desafios de sua jornada. Seus modelos de controle de acesso, baseados na segregação para proteção dos dados, precisam ser revistos para permitir a exploração das informações sem perder a segurança, sendo esse um desafio muito ligado à dimensão intrínseca. Um dos gerentes da área de dados reflete sobre o apoio da Diretoria de dados nessa governança para ajudar a enfrentar essa problemática:

A gente vai mais a fundo em cada caso e a gente consegue aprovar alguns usos que não têm risco pro banco, ou risco baixíssimo, que não tem nenhum tipo

de inconsistência jurídica, inconsistência legal ou inconsistência ética... A gente consegue aprovar muito mais iniciativas que, por desconhecimento ou por falta de uma área que analisa profundamente os riscos em cada caso, a gente simplesmente rejeitava, negava e não fazia. OG1 do Ômega

A equipe técnica do banco também destaca a mudança de paradigma necessária na governança dos dados. O uso de uma informação pode ser apropriado para uma finalidade e ilegal/anti-ético/inadequado para outra. A exploração dos dados perseguida pelo banco envolve o uso de outras ferramentas além dos sistemas nos quais essas informações foram geradas. Após a captura dos dados em diversas aplicações, é muito difícil controlar que sejam usados apenas para os fins que foram autorizados na liberação de acesso, o que ainda é um desafio em discussão nas equipes técnicas.

Já no caso do Bureau Sigma, o processo de captura de informações é mais crítico. A obtenção de dados de qualidade, mesmo nas bases oficiais do governo, tem se tornado mais trabalhosa e cara, uma vez que fontes gratuitas têm se tornado pagas (monetização de dados). Em um mercado em que o cliente pressiona fortemente por preços mais baixos, a busca constante por fontes alternativas que mantenham a qualidade dos dados e reduzam os custos de captura passou a ser uma atividade crítica para o negócio, sendo esse um desafio que pode impactar na dimensão econômica. Um dos diretores do Sigma reflete sobre esse dilema:

[...] a dificuldade que a gente está vendo hoje é: como pegar essas informações que são públicas, que deveriam ser públicas, mas não são mais? Por que está todo mundo fechando as portas [...]. SD2 da Sigma

Por fim, as *fintechs* possuem modelos de operação construídos com base em acesso a dados que já se mostraram viáveis quando a operação foi desenhada. Também se mostram constantemente atentas a criar novas formas de capturar dados, ainda que isso impacte seu modelo de negócio (por exemplo, parceria com novas plataformas), como relata um dos diretores da Gama:

Nossos dados vêm do consumidor ou eles vêm de informações que a gente busca desse consumidor no mercado. Então, eu posso eventualmente buscar uma informação desse consumidor num determinado, isso a ideia, em determinados sites de compras... E com tecnologia você consegue trazer outras informações dele próprio. FD4 da Gama

c) a importância da confiança para ter acesso e legitimidade para usar os dados mostrou-se uma preocupação em todos os casos. Eles partilham a percepção de que as pessoas estão dispostas a fornecer seus dados se receberem benefícios em troca, corroborando a literatura (Custers & Ursic, 2016). Por isso, todos se preocupam em se manter dentro da legalidade e alinhados à conduta reconhecida pela ética da sociedade. A postura que as empresas adotam em relação às dimensões legais e societárias, no entanto, é um pouco diferente. O banco assume uma

postura ativa, buscando caminhos lícitos para influenciar o mercado de dados, referindo-se à formulação de leis (papel articulador junto a entidades formais) e de práticas entre empresas e indivíduos (papel educativo), como comenta um dos gerentes da área de dados do Ômega:

E a nossa postura como banco, como influenciador da sociedade, a gente tá construindo junto. Então o nosso jurídico junto com nosso jurídico de dados tá em discussões com o Bacen, Febraban, que a gente tá ajudando a escrever isso, ajudando a regular essa história. Então, tem coisas que não estão prontas ainda, que a gente tá sendo influenciador aqui pra fazer. OG1 do Ômega

Já o Bureau Sigma, embora questione muito e demonstre grande preocupação com os impactos das mudanças em curso, assume uma posição mais passiva. Reclama daquilo que o incomoda ou prejudica, mas não mencionou ações concretas sobre tais incômodos.

Por fim, as *fintechs* abordam o tema com mais naturalidade, dando menos ênfase a ações na dimensão legal e societal e concentrando-se mais nas dimensões tecnológicas e intrínsecas. Elas reconhecem a importância dos avanços legais e acreditam que, embora eles possam impactar o mercado de dados e todos os serviços que dele derivam, cabe às empresas ajustar seus modelos de negócio ao que é permitido, encontrando soluções na própria tecnologia para atender tais padrões. Um dos diretores da Gama, por exemplo, descreve a alternativa criada pela empresa para obter dados da conta corrente dos solicitantes de crédito, como alternativa para a falta de legislação para o compartilhamento dos dados bancários:

O acesso à conta corrente, como os bancos brasileiros não tem as APIs deles, é feito através de *crawling* dos sites internet banking [...] Com o consentimento do cliente, *opt in*, todos os esclarecimentos, com todas as preocupações de segurança que você pode ter. FD7 da Gama

Também acreditam no avanço natural das práticas e padrões éticos socialmente aceitos, no amadurecimento do mercado e numa interferência limitada do estado na regulamentação de dados, suficiente para proteger os cidadãos e interesses coletivos básicos, sem ser excessivamente prescritivo e “engessar” o mercado com definições operacionalmente inviáveis, que limitem o desenvolvimento e a inovação. Eles chamam a atenção para a influência dos incidentes do Facebook e de outros grandes casos de vazamento de dados sobre as regras que vêm sendo criadas. Lembram que reúso de dados não diz respeito apenas às informações provenientes das redes sociais e nem às empresas gigantes que as operam. Alertam para os riscos de se estabelecer regras muito detalhadas no momento de transição pelo qual estamos passando. Acreditam que os cidadãos e as empresas estão experimentando coisas e aprendendo para, então, poderem formar suas opiniões e se posicionarem sobre normas e leis;

d) a mudança de paradigma da qualidade e da curadoria de dados é destacada de alguma forma em todos os casos. Qualidade dos dados sempre foi uma questão crítica para seu uso e o

desafio relatado aqui é a necessidade de mudar o paradigma de seus instrumentos. Os critérios para controle de qualidade de dados devem ser orientados pela finalidade de seu uso. A qualidade do dado que atende a um propósito pode ser insuficiente para outro. Ao fomentar a exploração de informações, torna-se muito mais difícil definir critérios de qualidade de dados, já que nem todo uso é conhecido na geração destes. Além disso, os dados criados por uma área podem não satisfazer o critério de qualidade para o uso de outra, criando novas demandas para as áreas geradoras. A qualidade da documentação e da curadoria dos dados em um ambiente em que áreas distintas deverão ser capazes de identificá-los e explorá-los também foi apontada entre os novos desafios, como expõe um supervisor da área técnica do Ômega:

No final você só vai conseguir garantir um reúso do dado se você conseguir gerar coisas de qualidade, seja no nível um, que é a documentação, seja no nível dois, que é o cara estar gerando a informação da melhor forma possível. Mas é que esse nível dois vai depender muito... talvez mais da área de negócios – que é dona desse dado –, mas também do consumidor (do dado), porque o que ele espera de qualidade de um lado pode ser diferente do cara de baixo (referindo-se ao gerador do dado. OS1 do Ômega

e) a necessidade de ajustar processos também apareceu em todos casos, porém com diferentes nuances. No caso do Banco Ômega, que abordou o reúso de dados interno, a demanda de mudanças é concentrada nos processos do próprio banco e diz respeito à governança, distribuição de responsabilidades entre áreas (especialmente consumidoras e geradoras de dados), atualização de dados e fluxos de comunicação das diferentes equipes de TI, por exemplo. Como parte dessa discussão, um dos gerentes de uma área técnica do banco reflete:

Tem outros times que faziam de forma mais centralizada, e aí gera um pouco desse desconforto, dessa resistência, pelo trabalho adicional que isso pode gerar, pela responsabilização por algo que o cara não tinha ciência... E, pra negócios, a mesma coisa: o time de modelagem, risco de crédito, tem uma plataforma mega robusta que eles cuidam, pagam a conta dela [...] Então, tem um lado de custeio interno das áreas que acho que também pode prejudicar um pouco esse uso disseminado e aberto dos dados. OG4 do Ômega

Já no caso do Bureau Sigma, surge a necessidade de incluir mais pessoas no processo de decisão sobre o uso de dados e a tecnologia, como forma de validar a adequação do uso e capturar as oportunidades do mercado. O caso também reporta a mudança de processos nos clientes, como relembra o diretor da seguradora que implementou um modelo de inteligência em parceria com a Sigma para verificação de fraudes:

Ali houve um desgaste muito grande. Em muitas situações, eu interferia no processo deles porque, imagina um esquema de produção, eu enfiava a mão na esteira, pegava um determinado produto e falava assim: ‘Dá aqui que esse aqui eu quero olhar’. Então, acabei sendo uma barreira no processo normal da companhia. SC2, cliente da Sigma

O caso das *fintechs* destaca a necessidade de mudança nos processos das empresas parceiras (credores) para que elas possam usufruir de todos os benefícios do modelo digital. Para as *fintechs*, ainda que suas plataformas estejam preparadas para tornarem o atendimento do cliente digital e trazer mais inteligência ao negócio, sua capacidade de fazê-lo depende dos processos dos credores. Exigências de documentos físicos e assinaturas, por exemplo, reduzem os benefícios de um credor expandir as fronteiras de atendimento para além das localidades em que já atua com lojas físicas. Processos de atualização de dados morosos ou falta de integração entre sistemas retarda a chegada de informações na plataforma, perdendo a janela de oportunidade para abordar o cliente. Ainda que as causas e situações reportadas pelas *fintechs* sejam distintas, é consenso entre elas que o reuso de dados pode demandar mudanças nos processos das empresas, seja para viabilizá-lo ou para permitir que o valor do novo racional de negócio se realize em sua plenitude, como ilustra um dos diretores da Delta:

No espaço digital tem credor passando a mesma carteira pra mais de um (cobrador) [...] é um *mindset* antigo pra uma abordagem nova [...] Credores que não conseguiram entender ainda que a relação com o devedor no ambiente digital é uma relação diferente. E que é nela que existe a possibilidade de criar uma relação de longo prazo com aqueles que não podem pagar agora a dívida. Essa relação tá de alguma maneira sendo mal resolvida e o pessoal perdendo com isso. FD3 da Delta

f) a falta de profissionais capacitados também está presente em todos os casos. A necessidade de pessoas preparadas para trabalhar com as novas ferramentas tecnológicas e os métodos matemáticos e estatísticos da era do BD foi reportada em todos os casos. Vale destacar que tal preocupação também apareceu na entrevista do executivo americano, referindo-se à escassez dessa mão-de-obra não apenas no Brasil, mas também em outros países em que atuam, inclusive nos EUA, sede de sua operação. As empresas relatam que os profissionais existem, mas em quantidade menor do que a demanda do mercado, o que desafia sua contratação e retenção. Todas adotaram a estratégia de desenvolvimento interno para mitigar essa escassez, como ilustra um gerente de área técnica do Banco Ômega:

A gente tá querendo formar cientistas de dados no banco, que eles consigam fazer um uso cada vez mais avançado dos dados pra resolver os problemas de negócios. Isso acho que é uma jornada de médio prazo. OG4 do Ômega

As empresas criticam o conteúdo da educação formal, que julgam não estar adaptado totalmente ao novo cenário de dados na área de TI e em outros segmentos que também irão se transformar pela utilização de informações. Alguns acreditam que faltam conteúdos conceituais e analíticos nos cursos, como sugere um dos diretores da Delta:

Eu falo que programador brasileiro aprende a programar do jeito que uma pessoa que não é programadora aprende a falar inglês. Ele aprende pela

sintaxe, pelo uso prático. Ele não aprende a lógica, o conceito, o fundamento. Como é que eu falo ‘por favor, me dê um café?’ Aí a pessoa vai lá e ensina a frase. FD6 da Delta

O executivo americano menciona que nos EUA os conteúdos acadêmicos estão adaptados, porém não dão conta de formar as pessoas na quantidade e no tempo que as empresas precisam. Ele acredita que essa falta é transitória e que sua solução será acelerada pela atração de profissionais para as carreiras de dados em função dos altos salários que a área tem oferecido. Já os profissionais brasileiros apresentam mais preocupações, chegando a questionar a influência da qualidade da educação básica no desenvolvimento das habilidades requeridas pelas ciências de dados contemporâneas:

É reflexo da realidade brasileira. Você tem, em geral, o ensino básico médio fundamental ruim, o que faz com que as pessoas não consigam entrar em boas faculdades. Então, enquanto mudanças estruturais não acontecerem na sociedade, você vai continuar convivendo com essa realidade. Não é no meio do processo que você vai resolver, não é no fim, é no início. FD7 da Gama

4.7 Discussão dos resultados

Após a análise *cross-case*, Stake (2006) recomenda que o pesquisador utilize os resultados dos casos, suas disparidades e uniformidades, para refletir sobre o *quintain* da pesquisa. O resultado dessa reflexão é apresentado a seguir. Ela foi organizada sob três temáticas que dialogam com as questões deste estudo: os potenciais desafios do reuso de dados e como os atores lidam com tais desafios na prática.

4.7.1 Potenciais desafios do reuso de dados

Os desafios identificados nos casos analisados foram posicionados nas dimensões legais, tecnológicas, intrínsecas, sociais e econômicas. Tais dimensões foram baseadas nas perspectivas propostas por Custers e Bachlechner (2017), às quais se somaram a perspectiva intrínseca sugerida em uma fase anterior da pesquisa. Todas as questões apontadas pelos casos puderam ser associadas a alguma dessas dimensões, assim como todas as dimensões foram associadas a alguma questão. Tal resultado sugere que todas as dimensões são relevantes e

suficientes para retratar os desafios do reuso de dados de forma abrangente. Essa visão pode apoiar pesquisas que o discutam sob uma ótica mais ampla, o que a literatura vem demandando. Também alerta para a multidisciplinaridade do reuso, que deve ser observada na teoria e nas iniciativas práticas.

As temáticas principais dos desafios mencionados nos casos são a tecnologia, o acesso a dados, a confiança para ter acesso e legitimar o uso dos dados, a mudança de paradigma da qualidade e curadoria de dados, o ajuste de processos e a falta de profissionais capacitados. Uma parte significativa das questões sobre elas diz respeito às mudanças, as quais são associadas à adoção/implementação do reuso de dados e, portanto, tendem a ser situacionais, desaparecendo ou se transformando à medida que o reuso se consolida na prática. Entre tais questões, observa-se algumas relacionadas à mudança específica, ou seja, referentes à implementação do reuso de dados em questão, enquanto outras referem-se à mudança do cenário tecnológico e de práticas que caracteriza o novo paradigma desse processo que está sendo viabilizado pelo BD. Modelos teóricos que investigam mudanças, adoção de tecnologias e ecossistemas de inovação são alguns exemplos de abordagens que poderiam contribuir para as pesquisas nesse contexto de transformação.

Embora haja nuances entre os desafios identificados nas diferentes categorias do reuso de dados, as temáticas mostraram-se predominantemente recorrentes entre elas. No entanto, a complexidade das questões pode variar de um tipo de reuso para outro. Por exemplo, a necessidade de mudança de processos apareceu em todas as categorias, mas quando o reuso é interno trata-se de processos da própria empresa que está promovendo-o (como no caso do Banco Ômega). Já no repropósito de dados externos, pode envolver outras entidades, fora do controle da empresa, que estão promovendo o reuso de dados (como no caso do Bureau Sigma). Se o reuso for uma recontextualização de dados externos, além de envolver entidades externas, estas podem ser ainda mais distantes da realidade daquelas que promovem o reuso (como no caso das *fintechs*), o que torna a mudança ainda mais desafiadora. Tal resultado corrobora os estudos sobre o reuso de dados, que afirmam que seus desafios se intensificam em função da distância entre a geração e seu reuso (Custers & Ursic, 2016; Zimmerman, 2008). Também contribui com os diferentes níveis de potenciais desafios propostos em etapa anterior da pesquisa (direto, intermedial e extremo).

Vale observar também que o mesmo tema pode impactar diferentes dimensões em cada caso, como ilustrado no tema “acesso a dados”, que no Banco Ômega trouxe mais desafios intrínsecos; no Bureau Sigma despertou questões econômicas; e nas *fintechs* foi endereçado de

forma que pode trazer impactos em todas as dimensões. Tal exemplo demonstra a inviabilidade de generalizações sobre os desafios do reúso de dados em função de seus tipos ou dimensões. A proposta deste trabalho ao apresentar tal segmentações é apenas encontrar caminhos para organizar uma discussão ampla, identificando pontos de diálogo entre as situações e questões. Não há aqui a intenção de determinar medidas ou relações causais de aplicação geral.

4.7.2 Como os atores estão lidando com os desafios que emergem do reúso de dados

Duas dinâmicas principais marcaram a forma como os atores lidaram com os desafios do reúso de dados nos casos reportados. Elas foram denominadas de ciclos de ajuste/aprendizado e mobilização na figura 4.4.

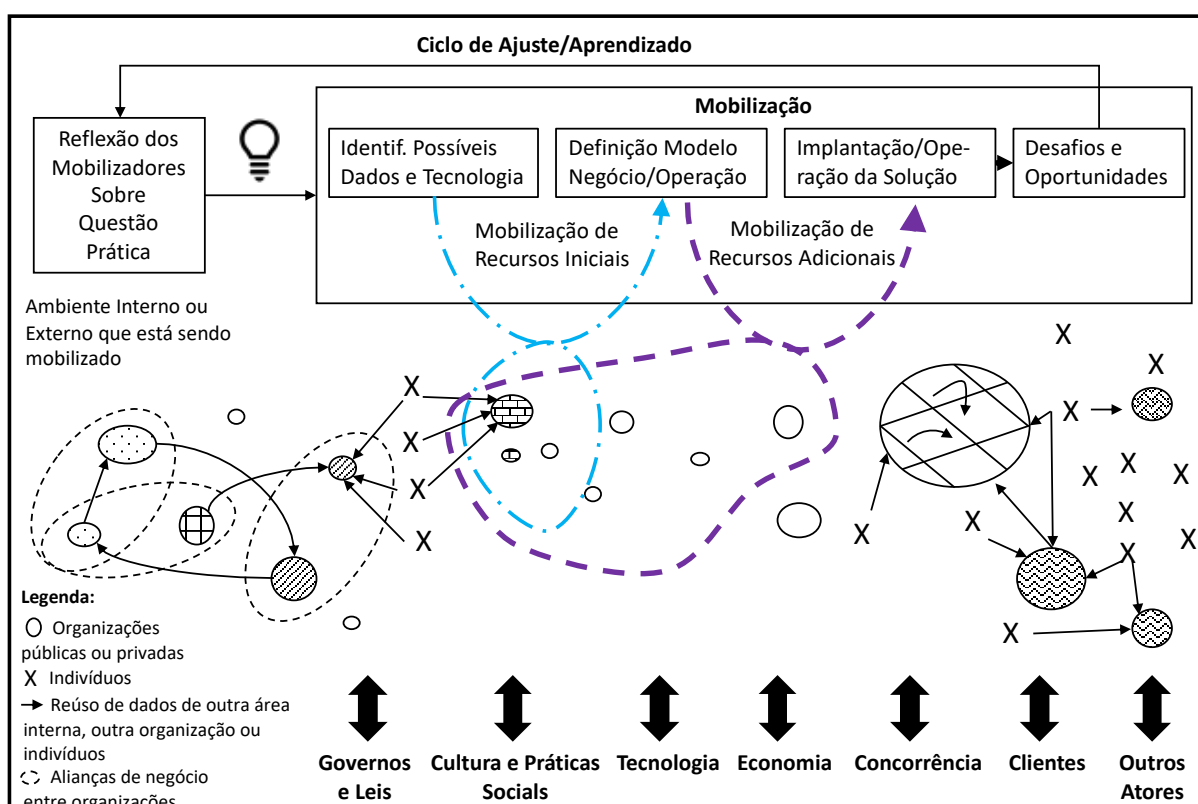


Figura 4.4 - Dinâmica do reúso de dados para lidar com os desafios de seu ecossistema

Fonte: elaborada pela autora

Os ciclos de ajuste/aprendizado aparecem no caso do banco e do *bureau* a partir de uma iniciativa de mudança estratégica corporativa. Já nas *fintechs*, os ciclos se deflagram pela própria criação das empresas. Tais ciclos têm em comum a identificação dos desafios e das

oportunidades do reúso de dados na prática, retroalimentando as ações em um processo contínuo de ajuste, que visa superar os desafios e consolidar o reúso e os negócios/processos que ele viabiliza. Essa dinâmica cíclica e contínua é coerente com a que é observada em outros fenômenos considerados emergentes, como o “strategyzing” (Denis, Langley & Rouleau, 2007, Jarzabkowski et al., 2007) e o “learning” (Argyris & Schon, 1997). Os estudos desses diferentes fenômenos compartilham a noção de que eles se constituem na prática, em um processo constante que ocorre em um contexto social, influenciando e sendo influenciado por ele. Compreender a dinâmica cíclica e contínua do reúso de dados o posiciona nesse grupo, permitindo que se beneficie do corpo de conhecimento deste. Isso também confirma a necessidade de utilizar perspectivas teóricas da prática para investigá-lo.

A dinâmica identificada nos casos também sugere a importância dos mobilizadores para articular o ecossistema necessário para o reúso de dados. Em todos os casos, um grupo se articulou para mobilizar os recursos e as interações necessárias para colocá-lo em prática. Esse grupo atuou catalisando recursos, mediando atores e buscando reunir as condições adequadas para o reúso. Esse processo está em linha com a literatura de ecossistemas de inovação que, assim como os ecossistemas de reúso de dados, caracterizam-se pela diversidade de interesses e recursos de seus participantes (Annanpera, Liukkunen & Markkula, 2015) e pela distância cognitiva entre eles (Dooley & O'sullivan, 2007). Nesse contexto, a figura de alguma liderança torna-se fundamental para superar as barreiras da rede e possibilitar a colaboração entre os atores (Annanpera, Liukkunen & Markkula, 2015). Essa liderança pode ser exercida por diversos participantes do ecossistema, em diferentes momentos e com papéis distintos (Annanpera, Liukkunen & Markkula, 2015; Gemünden et al., 2007). A função da liderança pode estar ligada aos recursos (“promotor do poder”), ao conhecimento técnico (“promotor da *expertise*”), ao conhecimento intraorganizacional (“promotor do processo”) ou à construção de alianças (promotor de relacionamentos) (Gemünden et al., 2007). A implementação do reúso de dados nos casos investigados revela a atuação de grupos articulando o acesso a informações e outros recursos, mobilizando seu conhecimento em tecnologia e no processo em questão e sua capacidade de fazer conexões e parcerias para promover a colaboração. Tal movimento pode acontecer no ambiente interno ou externo da organização, dependendo dos recursos que estiverem sendo mobilizados. No caso do banco, por exemplo, houve grande movimentação de recursos internos (pessoas, áreas, capacidades, dados internos, etc), enquanto nas *fintechs* a mobilização de recursos externos foi mais frequente (parcerias).

Também pode-se observar que a atuação dos grupos está ligada à movimentação de recursos e/ou de autoridade, elementos associados ao poder pela literatura (Hardy & Phillips, 1998). A forma como os mobilizadores atuaram nos casos revela duas abordagens de poder: como posse ou como prática. A abordagem da posse vê o poder como algo que se possui (Hardy & Phillips, 1998, Hardy, Philips & Lawrence, 1998), enquanto a abordagem da prática o considera algo relacional, que se constrói nas interações entre os atores (Marshall & Rollinson, 2004). Nos casos apresentados, observa-se a dinâmica da posse quando os atores lançam mão de seu acesso a recursos ou autoridade para viabilizar o reúso (como o banco e o *bureau* reorganizando suas estruturas internas e instrumentando grupos específicos para disseminar o reúso de dados pela empresa). Já a dinâmica da prática manifesta-se por meio de alternativas criadas pelos mobilizadores para ampliar seus recursos (como as novas formas de abordar os usuários para capturar seus dados e permitir a avaliação de risco sem depender de informações caras dos *bureaus* de crédito, que inviabilizariam a operação de uma *startup*).

Essa mesma interação das dinâmicas de posse e prática do poder foi verificada por Tello-Rozas, Pozzebon & Mailhot (2015) em sua pesquisa sobre colaboração social. As autoras destacam que as novas formas complexas de colaboração social requerem um novo tipo de mobilizador, que lide com os recursos e a autoridade de maneira diferente da que é retratada na literatura tradicional. Elas sugerem que esse novo perfil tenha “habilidade de gerenciar a interação entre recursos e autoridade, por exemplo, entre compartilhamento/cocriação de recursos abundantes e a descentralização/delegação coordenada de autoridade informal” (Tello-Rozas, Pozzebon & Mailhot, 2015, p. 1088). Os casos estudados sugerem que o ecossistema complexo do reúso de dados também pode demandar esse novo perfil mobilizador para articular a colaboração complexa que essa prática requer.

Por fim, o processo de mobilização pode dar origem a um ecossistema plural, emergente e dinâmico. Plural pela diversidade de atores e grupos que podem estar direta ou indiretamente envolvidos no reúso de dados (geradores, consumidores e intermediários dos dados, usuários dos serviços/ferramentas viabilizadas pelo reúso, etc). Emergente pois se forma a partir das demandas necessárias para viabilizar o reúso na prática. Dinâmico porque vai evoluindo ao longo do tempo para superar desafios e/ou realizar as oportunidades que surgem em sua prática. Essa evolução pode ser motivada pela dinâmica de processos, relações, interesses e interações entre os atores envolvidos no reúso dos dados, como também das bases legais, éticas e econômicas que o sustentam. Esse contexto confirma a relevância de analisar o reúso de dados por lentes teóricas da prática, que permitam capturar tal emergência e dinâmica,

sem as quais não se pode compreender sua essência. Além disso, a evolução desse ecossistema também é afetada pela dinâmica dos próprios dados (mudança do conteúdo/perfil dos dados, do que ele revela, etc.) e das suas condições de geração, acesso e uso, sugerindo potencial de abordagens sociomateriais para a investigação de ambientes de reúso de dados.

4.8 Conclusão

Este artigo apresenta o estudo de três casos de reúso de dados no setor financeiro, analisando-os pelas lentes do Multilevel Framework, que foi adaptado para ressaltar as dimensões legais, tecnológicas, intrínsecas, sociais e econômicas desse fenômeno na prática. O objetivo do estudo é identificar os desafios que emergem do reúso dos dados e propor um modelo para explicar o processo pelo qual os atores lidam com tais dificuldades. A análise dos casos corroborou a diferença dos níveis de desafios envolvidos em cada categoria de reúso. Esses diferentes níveis, denominados direto, intermedial e extremo, estão relacionados à distância entre a geração e o reúso dos dados. Os casos também revelaram que a tratativa de desafios relacionados a temas semelhantes desencadeou efeitos em dimensões diferentes do reúso, reforçando a importância de compreender o processo pelo qual os atores lidam com os desafios que dele emergem na prática. A pesquisa propõe, por fim, que tal processo é caracterizado por uma dinâmica cíclica e contínua de ajuste/aprendizado e pela ação mobilizadora de grupos (Annanpera, Liukkunen & Markkula, 2015; Gemunden et al., 2007) que interagem com os recursos e a autoridade do ecossistema por meio de abordagens de posse (Hardy & Phillips, 1998, Hard, Philips & Lawrence, 1998) e de prática (Marshall & Rollinson, 2004) de poder.

O artigo oferece algumas contribuições para a teoria e a prática do reúso de dados. Primeiro, atende o chamado da escassa literatura sobre o tema, que clama por estudos que adotem perspectivas mais amplas para abarcar os diversos aspectos envolvidos nesse fenômeno no atual cenário do BD (Borgman et al., 2015; Boyd & Crawford, 2012; Curty, 2015; Custers & Bachlechner, 2017; Newell & Marabelli, 2015; Samarajiva & Lokanathan, 2016; Zuboff, 2015). Ele dialoga com essa missão ao caracterizar o reúso de dados como um fenômeno que ocorre em um ecossistema plural, emergente e dinâmico, explorando suas dimensões legais, tecnológicas, intrínsecas, sociais e econômicas. As discussões e os conceitos apresentados

neste trabalho expandem o conhecimento sobre os diversos aspectos do reúso de dados e identificam questões que podem contribuir para o desenvolvimento acadêmico e prático da temática.

Como segunda contribuição, este estudo identifica pontos de diálogo entre questões relevantes do reúso de dados e correntes teóricas que podem contribuir para sua investigação, como a colaboração complexa (Faraj *et al.*, 2012), a sociomaterialidade (Orlikowski, 2007; Orlikowski & Scott, 2008) e as teorias da prática (Denis, Langley & Rouleau, 2007; Feldman & Orlikowski, 2011). Tal perspectiva pode iluminar caminhos teóricos para novos estudos sobre o reúso, especialmente em SI, onde é pouco explorado. A participação de SI na pesquisa sobre o assunto é fundamental, já que o corpo de conhecimento da área pode contribuir para a construção de pesquisas capazes de implementar a abordagem ampla que esse complexo fenômeno demanda. Por outro lado, o reúso de dados já é uma realidade e, como diversas passagens dos casos demonstram, sua realização implica a revisão de várias práticas relacionadas à TI. Portanto, não se atentar ao novo paradigma de uso de dados nos estudos de SI pode tornar seu conhecimento sobre tais práticas obsoleto.

Por fim, ao utilizar uma adaptação do Multilevel Framework (Pozzebon, Diniz & jayo, 2009) para operacionalizar a perspectiva da prática no estudo do reúso de dados, esse trabalho pode contribuir para a evolução de trabalhos empíricos que adotem a abordagem da prática para investigar fenômenos complexos. Diversos pesquisadores vêm destacando a dificuldade de operacionalizar lentes conceituais da prática em pesquisas empíricas, especialmente em contextos que envolvam a colaboração complexa (Vangen, 2016). O presente estudo demonstrou que a utilização do Multilevel Framework como uma lente integradora de múltiplos casos pode ser uma boa alternativa para a exploração de objetos complexos, como o *quintain* abordado nesse trabalho.

4.9 Limites e sugestões para futuros estudos

A estrutura da pesquisa não permite que seus resultados sejam generalizados, porém é plausível que seus achados possam ser transferidos para outros casos, o que poderia ser verificado por novos estudos empíricos. A replicação deste estudo em outros setores também

seria uma boa alternativa para verificar se o cenário encontrado na área financeira coincide com o cenário de outras áreas, afirmação que a presente pesquisa não tem condições de fazer.

Para futuros estudos, sugere-se explorar empiricamente os diversos desafios do reúso de dados identificados nos casos. Refletir sobre eles pode dar origem a diversos temas de pesquisa como: modelos de segurança e gestão de acesso orientados pela finalidade do uso dos dados; gestão e critérios da qualidade de dados para fins não previstos; curadoria de dados internos e externos; impactos do reúso de dados nos processos e estruturas organizacionais; inovações baseadas em reúso de dados; impactos sociais do reúso de dados; dinâmicas de poder no ecossistema de reúso de dados; o papel da confiança no reúso de dados; impacto dos avanços legais no mercado de dados; entre outros. Além disso, indica-se para outros trabalhos aplicar as perspectivas teóricas que dialogaram com o reúso de dados neste estudo, como a sociomaterialidade e teorias da prática e da colaboração complexa. Essa pode ser uma via promissora para expandir o conhecimento sobre o tema e para a evolução das próprias teorias, que se beneficiarão do aprendizado que pode emergir da exploração empírica de um fenômeno emergente e complexo como o reúso de dados na era do BD. A dinâmica mobilizadora revelada no estudo também poderia ser investigada usando novas abordagens, como a Teoria do Ator-Rede (ANT). Um elemento fundamental da ANT é a translação, processo interativo entre os atores, que pode ser dividido em 4 momentos: problematização, interesse, engajamento e mobilização (Denis, Langley e Rouleau, 2007). A identificação desses momentos, seus atores e processos em casos empíricos de reuso de dados pode colaborar para expandir a compreensão dessa dinâmica na prática.

Referências bibliográficas

Abbasi, A., Albrecht, C., Vance, A., & Hansen, J. (2012). Metafraud: A Meta-Learning Framework For Detecting Financial Fraud. *MIS Quarterly*, 36(4), 1293-A12.

Aleksynska, M., & Cazes, S. (2016). Composite indicators of labour market regulations in a comparative perspective. *IZA Journal of Labor Economics*, 5(1), 3.

Argyris, C., & Schön, D. A. (1997). Organizational learning: A theory of action perspective. *Reis: Revista Española de Investigaciones Sociológicas*, (77/78), 345-348.

Armerding, T. (2018). The 18 biggest data breaches of the 21st century. CSO Oline. In: <https://www.csoonline.com/article/2130877/data-breach/the-biggest-data-breaches-of-the-21st-century.html>

Annanperä, E., Liukkunen, K., & Markkula, J. (2015). Innovation in evolving business ecosystem: A case study of information technology-based future health and exercise service. *International Journal of Innovation and Technology Management*, 12(04).

Baesens, B., Bapna, R., Marsden, J. R., Vanthienen, J., & Zhao, J. L. (2016). *Transformational Issues of Big Data and Analytics in Networked Business*. MIS quarterly, 38(2), 629-631.

Beyerlein, M., Johnson, D., & Beyerlein, S. (Eds.). (2004). Complex collaboration: Building the capabilities for working across boundaries. Emerald Group Publishing Limited.

Bijker, W. E., & Law, J. (1992). *Shaping technology/building society: Studies in sociotechnical change*. MIT press.

Borgman, C. L., Darch, P. T., Sands, A. E., Pasquetto, I. V., Golshan, M. S., Wallis, J. C., & Traweek, S. (2015). Knowledge infrastructures in science: Data, diversity, and digital libraries. *International Journal on Digital*.

Boyd, D. & Crawford, K. (2012) Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon. *Information, Communication & Society*, 15(5), 662-679.

Creswell, J. W. (2007). *Projeto de pesquisa métodos qualitativo, quantitativo e misto*. Porto Alegre: Artmed.

Curty, R. G. (2015). *Beyond "Data Thrifting": An Investigation of Factors Influencing Research Data Reuse In The Social Sciences*. Available from ProQuest Dissertations & Theses Global.

Cunliffe, A. L. (2011). Crafting qualitative research: Morgan and Smircich 30 years on. *Organizational Research Methods*, 14(4), 647-673.

Custers, B., & Bachlechner, D. (2017). *Advancing the EU Data Economy: Conditions for Realizing the Full Potential of Data Reuse*. Information Polity, 22(4), 291.

Custers, B., & Ursic, H. (2016). *Big Data and Data Reuse: A Taxonomy of Data Reuse for Balancing Big Data Benefits and Personal Data Protection*. *International Data Privacy Law*, 6(1), 4-15.

Denis, J. L., Langley, A., & Rouleau, L. (2007). Strategizing in pluralistic contexts: Rethinking theoretical frames. *Human relations*, 60(1), 179-215.

- Diniz, E. H., Luvizan, S. S., Hino, M. C., & Ferreira, P. C. (2018). Unveiling the Big Data Adoption in Banks: Strategizing the Implementation of a New Technology. *In Digital Technology and Organizational Change* (pp. 149-162). Springer, Cham.
- Dooley, L., & O'SULLIVAN, D. A. V. I. D. (2007). Managing within distributed innovation networks. *International Journal of Innovation Management*, 11(03), 397-416.
- Faraj, S. A., Kolker, E., Bevilacqua, L., & Özdemir, V. (2012). Collective Theranostics And Postgenomics Entrepreneurship: Rethinking Innovations As Knowledge Ecosystems Built By Complex Collaboration. *Expert Review Of Molecular Diagnostics*, 12(8), 787-790.
- Feldman, M. S., & Orlikowski, W. J. (2011). Theorizing practice and practicing theory. *Organization science*, 22(5), 1240-1253.
- Flick, U. (2004). *Uma Introdução à Pesquisa Qualitativa*. Porto Alegre, Bookman.
- Gemünden, H. G., Salomo, S., & Hölzle, K. (2007). Role models for radical innovations in times of open innovation. *Creativity and innovation management*, 16(4), 408-421.
- Giddens, A. (1984). *The constitution of society*. Cambridge: Polity.
- Gioia, D. A., Corley, K. G., & Hamilton, A. L. (2013). Seeking qualitative rigor in inductive research: Notes on the Gioia methodology. *Organizational research methods*, 16(1), 15-31.
- Grande, D., Mitra, N., Shah, A., Wan, F., & Asch, D. A. (2014). *The Importance of Purpose: Moving Beyond Consent in the Societal Use of Personal Health Information*. *Annals of internal medicine*, 161(12), 855-862.
- Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. *Handbook of qualitative research*, 2(163-194), 105.
- Hardy, C., & Phillips, N. (1998). Strategies of engagement: Lessons from the critical examination of collaboration and conflict in an interorganizational domain. *Organization science*, 9(2), 217-230.
- Hardy, C., Phillips, N., & Lawrence, T. (1998). Distinguishing trust and power in interorganizational relations: Forms and facades of trust. *Trust within and between organizations*, 64-87.
- Jarzabkowski, P., Balogun, J., & Seidl, D. (2007). Strategizing: The challenges of a practice perspective. *Human relations*, 60(1), 5-27.
- Ketter, W., Peters, M., Collins, J., & Gupta, A. (2016). Competitive Benchmarking: An IS Research Approach To Address Wicked Problems With Big Data And Analytics. *MIS Quarterly*, 40(4), 1057.

- Krumm, R., Semjonow, A., Tio, J., Duhme, H., Bürkle, T., Haier, J., ... & Breil, B. (2014). *The need for harmonized structured documentation and chances of secondary use—Results of a systematic analysis with automated form comparison for prostate and breast cancer*. *Journal of biomedical informatics*, 51, 86-99.
- Langley, A. (1999). Strategies for theorizing from process data. *Academy of Management review*, 24(4), 691-710.
- Letouzé, E. (2012). *Big Data for Development: Challenges & Opportunities*. UN Global Pulse.
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry* (Vol. 75). Sage.
- Luvizan, S., & Diniz, E. (2017). Big Data e o Uso Secundário de Dados: Desafios para a Qualidade de Dados e a Inovação. *Anais do Enanpad*, 2017.
- MacAskill, E. & Hern, A. (2018) Edward Snowden: 'The people are still powerless, but now they're aware'. The Guardian. In: <https://www.theguardian.com/us-news/2018/jun/04/edward-snowden-people-still-powerless-but-aware>
- Manyika, J *et al.* (2011). *Big Data: The Next Frontier For Innovation, Competition, And Productivity*. McKinsey Global Institute.
- Marshall, N., & Rollinson, J. (2004). Maybe Bacon Had a Point: The Politics of Interpretation in Collective Sensemaking 1. *British Journal of Management*, 15(S1), 71-86.
- Martin, K. (2015). *Ethical issues in the big data industry*. *MIS Quarterly Executive*, 14(2), 67-85.
- Müller, O., Junglas, I., Brocke, J. V., & Debortoli, S. (2016). Utilizing big data analytics for information systems research: challenges, promises and guidelines. *European Journal of Information Systems*, 25(4), 289-302.
- Mundie, C. (2014). Privacy Pragmatism: Focus on Data Use, Not Data Collection. *Foreign Affairs*. 93(2). Disponível em: <http://rational.com/privacy-pragmatism-focus-on-data-use-not-data-collection/>
- Newell, S., & Marabelli, M. (2015). *Strategic opportunities (and challenges) of algorithmic decision-making: A call for action on the long-term societal effects of 'datification'*. *The Journal of Strategic Information Systems*, 24(1), 3-14.
- Orlikowski, W. J. (2000). Using technology and constituting structures: A practice lens for studying technology in organizations. *Organization science*, 11(4), 404-428.
- Orlikowski, W. J. (2007). Sociomaterial practices: Exploring technology at work. *Organization studies*, 28(9), 1435-1448.

Orlikowski, W. J., & Scott, S. V. (2008). Sociomateriality: challenging the separation of technology, work and organization. *The academy of management annals*, 2(1), 433-474.

Pettigrew, A. M. (1985). Contextualist Research And The Study Of Organizational Change Processes. In E. Mumford, R. Hirschheim, G. Fitzgerald, & A. T. Wood-Harper (Eds.), *Research Methods in Information Systems* (pp. 53-72). New York: North Holland.

Pettigrew, A. M. (1990). Longitudinal field research on change: theory and practice. *Organization Science*, 1(3), 267-292.

Pettigrew, A. M. (1997). What Is A Processual Analysis. *Scandinavian Journal of Management*, 13(1997), 4.

Pinch, T. J., & Bijker, W. E. (1984). The social construction of facts and artefacts: Or how the sociology of science and the sociology of technology might benefit each other. *Social studies of science*, 14(3), 399-441.

Pozzebon, M., & Diniz, E. H. (2012). Theorizing ICT and society in the Brazilian context: a multilevel, pluralistic and remixable framework. *BAR-Brazilian Administration Review*, 9(3), 287-307.

Pozzebon, M., Diniz, E. & Jayo, M. (2009). Adapting The Structurationist View Of Technology For Studies At The Community/Societal Levels. In Y. K. Dwivedi (Ed.), *Handbook of research on contemporary theoretical models in information systems* (pp. 18-33). New York: IGI Publishing.

Pozzebon, M., Mackrell, D., & Nielsen, S. (2014). Structuration Bridging Diffusion Of Innovations And Gender Relations Theories: A Case Of Paradigmatic Pluralism In IS Research. *Information Systems Journal*, 24(3), 229-248.

Pozzebon, M., Rodriguez, C. & Petrini, M. (2014). Dialogical Principles For Qualitative Inquiry: A Nonfoundational Path. *International Journal of Qualitative Methods*, 13, 293-317.

Pratt, M. G. (2009). From the editors: For the lack of a boilerplate: Tips on writing up (and reviewing) qualitative research.

Safran, C., Bloomrosen, M., Hammond, W. E., Labkoff, S., Markel-Fox, S., Tang, P. C., & Detmer, D. E. (2007). Toward a national framework for the secondary use of health data: an American Medical Informatics Association White Paper. *Journal of the American Medical Informatics Association*, 14(1), 1-9.

Samarajiva, R., & Lokanathan, S. (2016). *Using Behavioral Big Data For Public Purposes: Exploring Frontier Issues Of An Emerging Policy Arena*.

Sandberg, J. (2005). How do we justify knowledge produced within interpretive approaches? *Organizational research methods*, 8(1), 41-68.

Sherr, I. (2018). Facebook, Cambridge Analytica and data mining: What you need to know. CBNET. In: <https://www.cnet.com/news/facebook-cambridge-analytica-data-mining-and-trump-what-you-need-to-know/>

Smolan, R. (2013). The human face of big data.

Stake, R. E. (2006). *Multiple Case Study Analysis*. Guilford Press.

Stake, R. E. (2008). Qualitative case studies. In N. K. Denzin & Y. S. Lincoln (Eds.), *Strategies of qualitative inquiry* (pp. 119-149). Thousand Oaks, CA, US: Sage Publications, Inc.

Strauss, A., & Corbin, J. (1998). *Basics of Qualitative Research: Procedures And Techniques For Developing Grounded Theory*. Sage Publications.

Tello-Rozas, S., Pozzebon, M., & Mailhot, C. (2015). Uncovering Micro-Practices and Pathways of Engagement That Scale Up Social-Driven Collaborations: A Practice View of Power. *Journal of Management Studies*, 52(8), 1064-1096.

Van de Ven, A. H. (1992). Suggestions for studying strategy process: A research note. *Strategic management journal*, 13(S1), 169-188.

Vangen, Siv (2016). Developing Practice-Oriented Theory on Collaboration: A Paradox Lens. *Public Administration Review*, 77(2) pp. 263–272.

Wallis, Jillian C., Elizabeth Rolando, and Christine L. Borgman. 2013. "If We Share Data, Will Anyone Use Them? Data Sharing and Reuse in the Long Tail of Science and Technology." *PLoS One* 1-17.

Walsham, G. (1995). Interpretive case studies in IS research: nature and method. *European Journal of information systems*, 4(2), 74-81.

Yoon, A. (2015). *Data reuse and users' trust judgments: Toward trusted data curation*. Available from ProQuest Central; ProQuest Dissertations & Theses Global.

Zimmerman, Ann. 2008. "New Knowledge from Old Data: The Role of Standards in the Sharing and Reuse of Ecological Data." *Science, Technology & Human Values* 631-652.

Zuboff, S. (2015). Big other: surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology*, 30(1), 75-89.

Apêndices

APÊNDICE A – Diagrama da estrutura geral da pesquisa

O QUINTAIN: REUSO DE DADOS NO SETOR FINANCEIRO

Objetivo Principal da Pesquisa para o Quintain:

Caracterizar o reuso de dados no setor financeiro, identificar seus potenciais desafios e propor um modelo para explicar o processo pelo qual os atores envolvidos lidam com tais desafios.



Fonte: elaborado pela autora

Artigo I – Caracterizar o Reuso de Dados

Rev. Bibliográfica: definição conceitos e posicionamento literatura existente

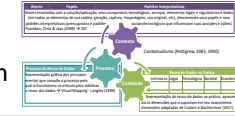


Meta-análise: identificação características, proposta esquema classificatório e níveis de potenciais desafios

		Fonte dos Dados		
		Dados Internos	Dados Públicos	Dados Privados Externos
Finalidade do Reuso de Dados	Repropósito	●	●	●
	Recontextualização	●	●	●

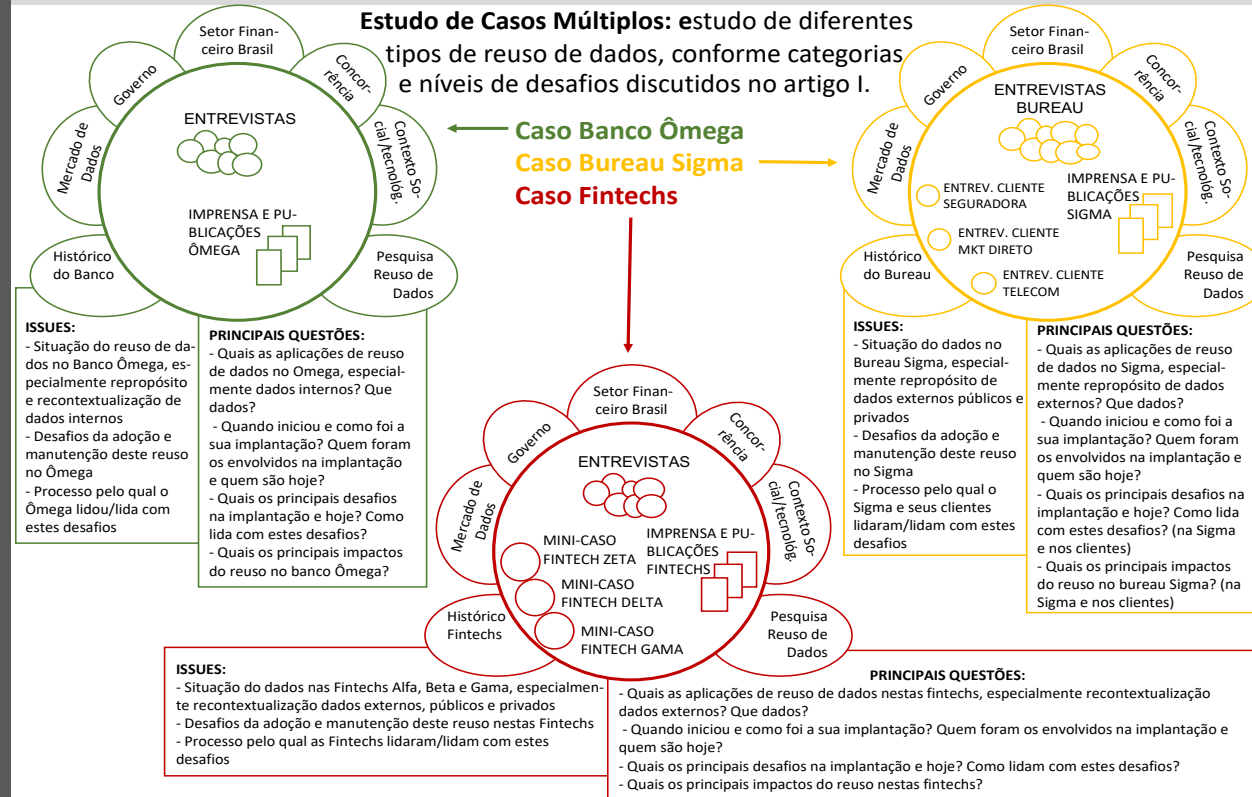
Artigo II – Modelo p/ caso empírico

Proposta de modelo conceitual: apresentação de um modelo teórico para investigar casos empíricos de reuso de dados, exercitando-o em um caso exploratório.



Artigo III – Identificar desafios e explicar como atores lidam com eles na prática

Estudo de Casos Múltiplos: estudo de diferentes tipos de reuso de dados, conforme categorias e níveis de desafios discutidos no artigo I.



APÊNDICE B - Protocolo das entrevistas – estudo de múltiplos casos de reúso de dados

1. Estrutura geral

Entrevistadora: Simone da Silva Luvizan

Orientador: Eduardo Henrique Diniz

Entidades da pesquisa de campo: Banco Ômega, Bureau Sigma, *fintechs* Zeta, Delta e Gama.

Objetivo do estudo: identificar os desafios que emergem do reúso de dados e explicar como os atores lidam com tais desafios na prática.

Método de coleta: entrevistas semiestruturadas presenciais (em casos de exceção por telefone).

Método de registro: gravação em áudio para posterior transcrição e anotações manuais feitas pelo entrevistador durante entrevista.

Seleção dos entrevistados: feita a partir de entrevista(s) exploratória(s) com interlocutor(es), que ocupe(m) posição(ões)-chave e que tenha(m) conhecimento sobre casos concretos de reúso de dados na empresa, assim como o trânsito necessário para viabilizar o acesso a atores envolvidos nessas experiências na área de TI e de negócio.

2. Temas de referência para as entrevistas

As entrevistas partirão de temas que servirão como referência do conteúdo a ser explorado, podendo incorporar questões emergentes. Os pontos não precisam ser seguidos sequencialmente como um roteiro, priorizando estimular o entrevistado a contar suas histórias livremente.

Abertura das entrevistas

Todas as entrevistas devem iniciar fornecendo as seguintes informações ao entrevistado:

- Apresentação do entrevistador e da pesquisa;
- Solicitação de permissão para gravação da entrevista;
- Esclarecimentos sobre a confidencialidade das respostas individuais e do nome da empresa no caso;
- Esquema da entrevista – será conduzida como uma conversa e, além dos itens propostos pela entrevistadora, o interlocutor pode falar sobre outras questões que julgue

interessantes ou relevantes para o caso (valorizar o conhecimento do entrevistado sobre o caso e a importância de captar a visão dele).

Conteúdo – temas a explorar na entrevista

Posicionamento do entrevistado

1. Apresentação do entrevistado (nome, formação, função e área de atuação).
2. Quanto tempo está na empresa, sua história na organização (áreas/posições anteriores, projetos de que participou, etc).

História do reúso de dados na empresa

3. Existe ou não, desde quando, em que áreas, em quais suas aplicações.
4. Como surgiu a ideia e como começou.
5. Quem eram os envolvidos e como foi o processo de início dessa prática.

Estrutura para o reúso de dados

6. Tecnologia utilizada.
7. Dados utilizados.
8. Pessoas envolvidas e seus perfis/papéis.

Processo de implantação/adoção de soluções que fazem reúso de dados

9. História dos projetos de reúso de dados.
10. Ferramentas e dados envolvidos.
11. Desafios da implantação e do uso efetivo.
12. Como lidou com tais desafios.
13. Aceitação/adoção das soluções.
14. Resultados esperados e realizados pelos casos.

Perspectivas do reúso de dados na empresa

15. Impactos dos casos de reúso de dados implementados e expectativas para futuro.
16. Desafios e oportunidades para a expansão do reúso de dados na empresa.
17. Como pretende explorar tais desafios e oportunidades.

Fechamento das entrevistas

Concluir as entrevistas com as seguintes colocações:

- Verificação de questões adicionais – há algo que ele julgue relevante e não foi abordado;
- Agradecimento pela participação;

- Abertura para contato futuro caso o entrevistado queira agregar alguma informação ou tenha alguma questão sobre a pesquisa.

APÊNDICE C – Processo de análise dos dados

Etapa 1 – Codificação aberta: foi realizada a microanálise dos dados, buscando associar códigos a trechos das entrevistas (palavras, parte de frases, frases inteiras ou conjunto de frases). No modelo de Strauss e Corbin (1998), nesta etapa busca-se gerar propriedades e categorias por meio de comparações teóricas e entre os códigos. Este estudo optou por gerar, nesta fase, códigos mais descritivos, mantendo marcas de fatos, momentos e atores que ajudem a identificar a sequência cronológica e as relações eventos/questões relatadas nas entrevistas (Creswell, 2007; Flick, 2004). Assim, embora tenham surgido alguns códigos mais interpretativos e que permitiram uma reflexão cíclica mais rica, aglutinando dados, a maioria dos códigos descrevem situações particulares, não podendo ser associados a outros para preservar as vozes dos entrevistados e a sua capacidade de “contar a história do caso”. Alguns pesquisadores sugerem que a primeira etapa da codificação pode ser uma redução dos dados originais (Miles & Huberman, 1994) e que deve manter as vozes informantes e os momentos trazidos pelas perspectivas das pessoas e pelo tempo (Gioia et al., 2013). Essa estratégia deixou as primeiras etapas de codificação bem trabalhosas, pois deu origem a uma grande quantidade de códigos. No entanto, foi extremamente útil para enriquecer a análise das fases seguintes. “Você tem que se perder antes de se encontrar” (Gioia et al., 2013 pg. 20).

Etapa 2 – Codificação axial: Strauss e Corbin (1990) sugerem que esta etapa produza uma estruturação, que chamam de “paradigma de codificação”, revelando os fenômenos, as condições causais, os contextos, as condições intermediárias, estratégias de ação/interação e consequências. O importante é que essa fase revele relações entre os códigos que sejam suportadas pelos dados. Nesta etapa, foi realizado o agrupamento dos códigos em temas (podendo cada código pertencer a mais de um tema), que refletiam atores, eventos, estratégias, percepções, etc. Após realizar esse agrupamento no Atlas.ti, os dados foram exportados para uma planilha eletrônica que facilitou uma nova leitura dos grupos e seus conteúdos, permitindo analisar melhor cada grupo, reconhecer suas relações ou “subcategorias”. Nesse processo analítico, o pensamento move-se entre o indutivo (gerando conceitos e relações a partir dos dados) e o dedutivo (testando os conceitos e relações, contrapondo-os aos dados) (Flick, 2004; Godoy, Bandeira-De-Mello & Silva, 2006; Strauss & Corbin, 1998). As anotações e comentários gerados nesta fase foram muito úteis na etapa seguinte.

Etapa 3 – Codificação seletiva: esta etapa dá continuidade à anterior, buscando agora as categorias e relações mais relevantes para contar a história do caso, que passa a ter um conceito

central com o qual as demais categorias se relacionam (Strauss & Corbin, 1990). Durante este processo, há nova interação com os elementos das etapas anteriores e com os dados, buscando completar a compreensão ou enriquecer as categorias selecionadas (Flick, 2004; Godoy, Bandeira-De-Mello & Silva, 2006). As categorias destacadas, assim como suas relações e conteúdos principais, foram representadas por um mapa visual (Langley, 1999). O diagrama de cada caso ajudou a visualizar sua história, contando a dinâmica do reuso de dados nas diferentes situações.

Etapa 4 – releitura por lente conceitual: os resultados de cada caso foram relidos e sumarizados no Multilevel Framework adaptado. Os elementos do mapa visual permitiram preencher quase todo o Multilevel Framework dos casos, havendo necessidade de recorrer aos dados apenas para a confirmação de detalhes ou questões muito específicas. Esta etapa permitiu que os dados analisados sem uma lente teórica a priori fossem organizados em um modelo conceitual único, contribuindo para a análise cruzada dos casos na próxima fase.

Etapa 5 – Análise *cross-case* – os procedimentos desta etapa foram inspirados em Stake (2006). Ele propõe que os relatórios dos casos individuais sejam relidos e que seus resultados sejam analisados na perspectiva do *quintain*, ou seja, das questões centrais da pesquisa. Dessa forma, mais do que um comparativo entre os casos, a análise *cross-case* busca construir uma compreensão sobre o *quintain* que tenha evidências nos casos, suas uniformidades ou disparidades (Stake, 2006). O autor sugere que os temas centrais investigados na pesquisa sejam destacados e que o pesquisador busque a contribuição de cada caso para esse tema. Também indica que se criem tabelas complementares que possam trazer registros das contribuições dos casos para esses temas, facilitando visualizar toda a informação colhida sobre cada assunto. Nesta fase, foram identificados dois temas centrais: 1) desafios que emergem do reuso de dados e 2) como os atores lidam com essas dificuldades. Para a análise do tema 1, os desafios identificados nos três casos foram levados para uma planilha, analisados e categorizados, oferecendo uma perspectiva mais ampla dos desafios do reuso de dados como um todo. Para analisar o tema 2, foi feita uma leitura conjunta do contexto e processo dos três casos apresentados no Multilevel Framework, gerando uma análise única para essa dinâmica, dando origem às reflexões apresentadas nos resultados e discussões do caso.

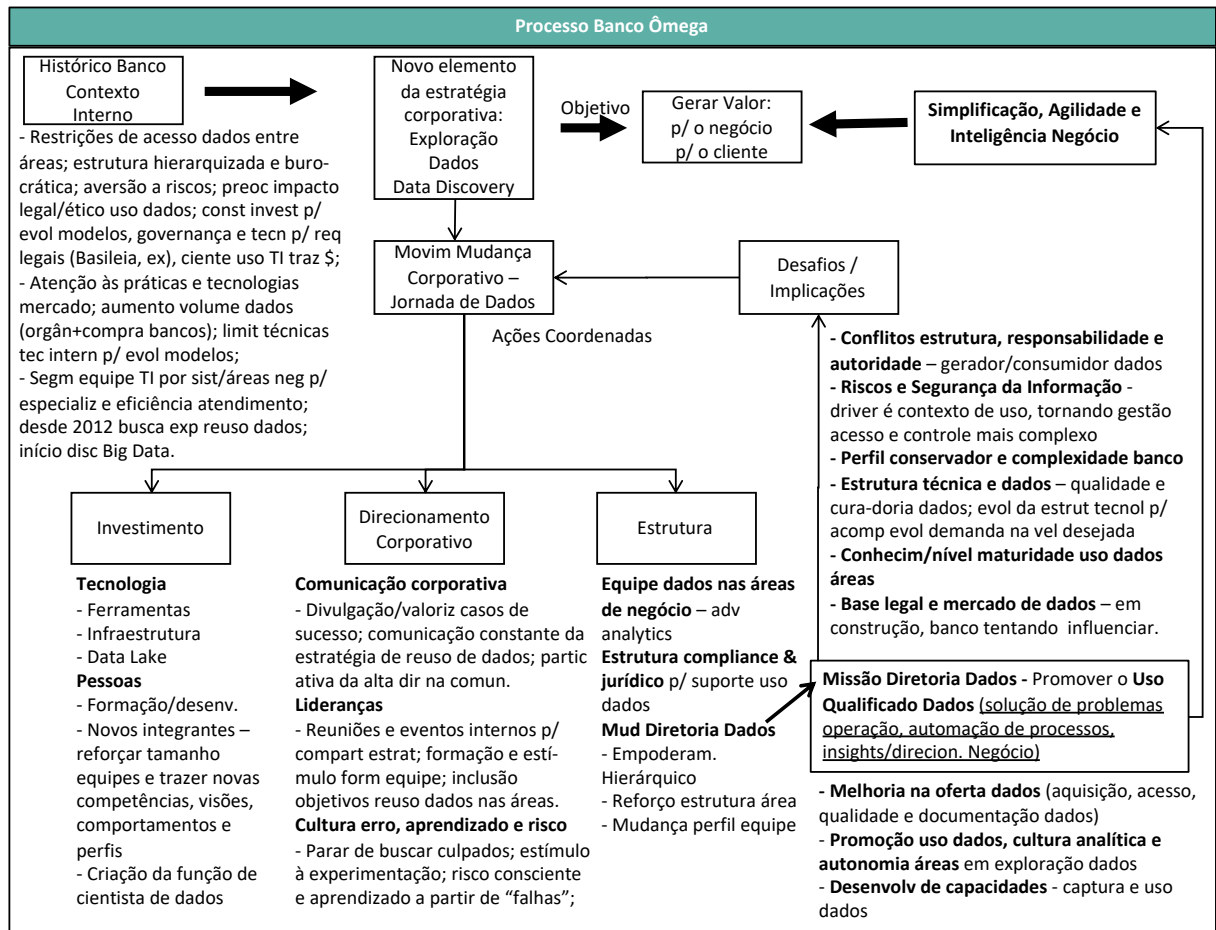
APÊNDICE D – Multilevel Framework do caso Banco Ômega

Contexto do caso Ômega

Contexto Banco Ômega		
Atores	Papéis	Padrões interpretativos
Setor financeiro	Oferta de serviços financeiros.	<ul style="list-style-type: none"> - Esgotamento do modelo bancário tradicional; - Setor de alto risco e conservador no uso de dados, atuando para influenciar o mercado de dados e sua legislação; - Possui associações, como a Febraban, para articular bancos e representá-los em discussões relevantes para a área; - Novos entrantes com diferentes perfis operação-<i>fintechs</i>.
Concorrentes	Oferta de serviços financeiros.	<ul style="list-style-type: none"> - Concorrência de alto nível e muito competitiva; - Novos entrantes com diferentes perfis operação-<i>fintechs</i>.
Governo	Responsável pelas regulamentações do setor financeiro e uso de dados, e pelas bases de dados oficiais empresas/cidadãos.	<ul style="list-style-type: none"> - Forte regulação do setor financeiro no Brasil, demanda altos padrões de governança e transparência dos bancos; - Responsável pelo desenvolvimento de bases legais para o uso de dados que estão em curso, tendo de conciliar os interesses dos diversos setores da sociedade; - Usuário conservador de tecnologia, demanda serviços de dados, governo requer novos modelos e tecnologias.
Sociedade	Cidadãos, consumidores dos serviços bancários e/ou mão-de-obra do banco.	<ul style="list-style-type: none"> - Evolução do perfil das pessoas enquanto profissionais (funcionários ou potenciais funcionários) – mais dinâmico e ávido por desenvolvimento constante, valoriza o ambiente trabalho que ofereça condições para isso; - Evolução do perfil das pessoas enquanto clientes ou potenciais clientes – mais críticos em relação ao serviço e abertos a compartilhar dados se tiver benefício.
Mercado de dados	Comercialização de dados e serviços baseados em dados.	<ul style="list-style-type: none"> - Em desenvolvimento e pouco organizado; - Muitas empresas não estão preparadas para “produzir”/“monetizar” seus dados.
Mercado de tecnologia	Desenvolvimento e fornecimento de infraestrutura, ferramentas e aplicações de tecnologia/dados.	<ul style="list-style-type: none"> - Evolução tecnológica e da geração de dados nas últimas décadas; - Visão de dados como uma disciplina própria; - Aumento da oferta de tecnologias de dados.
Equipe interna: área de TI	Desenvolvimento, implantação e sustentação da TI do banco.	<ul style="list-style-type: none"> - Equipe grande e bem segmentada por tecnologias e áreas de negócio; diversos profissionais de carreira no banco; - Focados na eficiência dos serviços de sua área de atuação, atentos aos padrões de governança e segurança do banco.
Equipe interna: área de dados	Desenvolvimento de padrões, práticas, técnicas e ferramentas para uso de dados	<ul style="list-style-type: none"> - Equipe híbrida – profissionais de carreira do banco e novos, idades e perfis também variados; - Focados na missão de disseminar a exploração de dados pelo banco e criar padrões de governança, metodologia, tecnologia e conhecimento para o uso qualificado de dados nas áreas.
Equipe interna: áreas de negócio	Funções diversas de negócio no banco.	<ul style="list-style-type: none"> - Focados na eficiência e resultados de suas áreas; - Cultura de segurança de informação e aversão a risco fortemente estabelecida nas áreas; - Diferentes níveis de conhecimento e utilização de ferramentas para o uso de dados entre as áreas; - Objetivos referentes ao uso de dados ou objetivos de negócio que o demandam foram incorporados às metas da área.

Fonte: elaborado pela autora

Processo do caso Ômega



Fonte: elaborado pela autora

Conteúdo – caso Banco Ômega

Conteúdo Banco Ômega				
Reúso dos dados na prática: recontextualização e repropósito de dados internos Jornada de dados para promover o reúso de dados das bases internas (disponibilizados pelos clientes ou capturados/produzidos ao longo dos serviços e processos internos das diferentes áreas) para automação/simplificação de processos e inteligência de negócios no banco.				
Legal	Tecnológica	Intrínseca	Societal	Econômica
Leis sempre influenciaram a tecnologia do setor e passa por mudanças em: a) serviços financeiros (pode ampliar a possibilidade de atuação do banco e abrir mercado para as <i>fintechs</i>); b) proteção e uso de dados (pode influenciar o mercado de dados em fase de formação).	Estruturas tradicionais chegam ao limite da capacidade e restringem o reúso de dados. Evolução tecnológica possibilitou expandir o reúso de dados e melhorar os serviços e a eficiência do banco. Estrutura para experimentação e uso de novas ferramentas disponíveis no banco. Adoção dessas tecnologias não é <i>plug-and-play</i> e demandam formação de equipes de TI e demais áreas, adaptação aos padrões de governança, compatibilização de processos e ferramentas legadas, revisão da infraestrutura, etc. Algumas ferramentas ainda não têm casos práticos em ambientes complexos como o banco.	Cultura de segurança e estrutura de TI (tecnologia e pessoas) orientadas à proteção dos dados, segmentando o acesso às áreas e ferramentas. A exploração de dados subverte essa lógica, demandando mudanças nas estruturas e nos processos. A área de dados é empoderada para assumir essa missão, capturar os desafios que emergem na prática e endereçar as discussões e mudanças necessárias, assim como prover ferramentas e formação para as áreas. Grande investimento em TI e em pessoas e suporte da alta direção.	Mudanças de comportamento em curso – grupos estão mais conectados e têm mais expectativas sobre serviços digitais. Pessoas estão mais abertas a compartilhar dados e usar tecnologias se tiverem benefícios. Valorizam mais os ambientes de trabalho desafiadores, dinâmicos, sendo que o tamanho da empresa não é suficiente para atrair e reter os melhores talentos.	Mudanças tecnológicas requerem altos investimentos, mas trazem grande retorno para o banco. Limites técnicos dos modelos atuais para lidar com o crescimento do volume de dados requer atualizações para a continuidade da operação. Esgotamento do modelo de serviço tradicional do banco é iminente e a reformulação de processos/serviços é fundamental para a sustentabilidade do negócio no longo prazo.

Fonte: elaborado pela autora

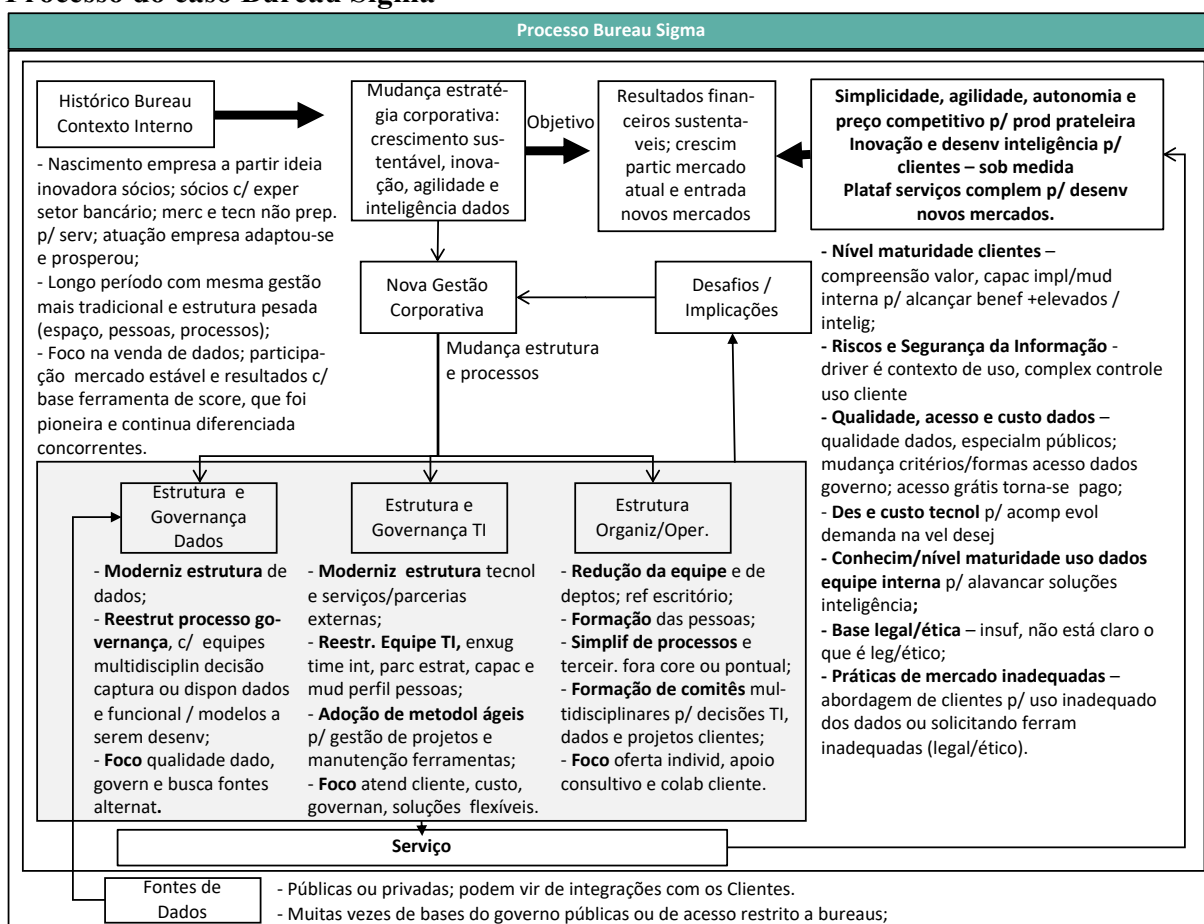
APÊNDICE E – Multilevel Framework do caso Bureau Sigma

Contexto do caso Bureau Sigma

Contexto Bureau Sigma		
Atores	Papéis	Padrões interpretativos
Cliente	Compradores de dados e/ou serviços de inteligência baseada em dados.	<ul style="list-style-type: none"> - Sensível a preço e atento à concorrência; - Demanda agilidade e ofertas variadas (<i>commodities</i> e soluções específicas); demanda predição; - Diferentes níveis de conhecimento/maturidade e novas tecnologias/dados: alguns maiores têm grande conhecimento, altos padrões de exigência e equipes/processos qualificados para desenvolver e suportar os projetos; a maioria dos pequenos e médios, e alguns grandes, tem pouco conhecimento (às vezes repete o que ouve e acha que conhece), pouca estrutura e verba para desenvolver projetos específicos, baixa percepção de valor e temor risco/resistência a mudança.
Concorrentes	Ofertas de dados e de serviços de inteligência baseada em dados.	<ul style="list-style-type: none"> - Mercado de <i>bureau</i> de crédito restrito a quatro empresas no Brasil, mas altamente concentrado; - Concorrência muito competitiva; - Novos entrantes – <i>bureaus</i> de informação e outros tipos de serviço, limitados no que se refere a informações de bases do governo só disponíveis para <i>bureaus</i> de crédito oficiais.
Governo	Responsável pelas regulamentações do setor financeiro e uso de dados, e pelas bases de dados oficiais empresas/cidadãos.	<ul style="list-style-type: none"> - Lento na atualização de leis para atender a novas demandas de dados na sociedade; abordagem legal de crédito protege fraudadores e não beneficia bons pagadores; - Começando a transformar a disponibilização de suas bases de dados em negócio; - Estrutura interna de processos e tecnologia não adequada para as demandas atuais e se ajustando aos poucos.
Sociedade	Conjunto da sociedade (por exemplo: cidadãos, empresas, entidades não-governamentais, associações, órgãos e agentes públicos).	<ul style="list-style-type: none"> - Evolução de conhecimento e adoção de tecnologia, mas conhecimento em geral ainda é muito superficial; - Pouco cuidado com suas informações digitais (redes sociais, por exemplo).
Mercado de dados	Comercialização de dados e serviços baseados em dados.	<ul style="list-style-type: none"> - Em desenvolvimento e pouco organizado; - Muitas empresas estão fazendo mau uso da tecnologia e dos dados, criando um mercado negro de dados.
Mercado de tecnologia	Desenvolvimento e fornecimento de infraestrutura, ferramentas e aplicações de tecnologia e dados.	<ul style="list-style-type: none"> - Evolução e disseminação da tecnologia e da geração de dados nas últimas décadas; - Aumento da oferta e redução do custo da tecnologia.

Fonte: elaborado pela autora

Processo do caso Bureau Sigma



Fonte: elaborado pela autora

Conteúdo – caso Bureau Sigma

Conteúdo Bureau Sigma				
Reúso dos dados na prática: repropósito de dados externos públicos e privados				
Reúso de dados cedidos pelos clientes ou capturados de fontes públicas/privadas, compondo sua base para venda de dados, serviços de prateleira mais diferenciados ou serviços de inteligência customizados.				
Legal	Tecnológica	Intrínseca	Societal	Econômica
Atividade do bureau de crédito regulamentada e restrita a quatro empresas. Falta de leis claras sobre o uso de dados permite práticas desleais que contaminam o mercado. Falta cadastro positivo e modernização das leis de proteção de crédito (fraudes e maus pagadores reincidentes).	Tecnologia disponível e mais acessível financeiramente. Porém, requer profissionais capazes de transformá-la em aplicações inteligentes; falta entendimento das pessoas nas áreas sobre novos modelos que a tecnologia e o uso dos dados viabilizam. Validação e preparação de dados ainda não é trivial.	Atuação dos <i>bureaus</i> de crédito já estabelecida no mercado. Principais clientes - setor financeiro ou serviços financeiros internos de empresas diversas. Perfil de finanças é conservador, sensível a riscos e a custos, oferecendo resistência a mudanças. Ao mesmo tempo, a área quer os benefícios da 'nova era de dados', especialmente na redução de custos ref consulta dados. Na oferta de cadastro qualificado/segmentado, clientes também acreditam que disponibilidade dados internet e novas tecnologias simplificaram o serviço, que deveria 'baratear'. Agências de informação usando recursos duvidosos e oferecendo dados má qualidade a preços baixos, conturbando o mercado.	Figura dos <i>bureaus</i> de crédito, conhecida pela sociedade e associada ao cadastro negativo de crédito, não havendo grande polêmica sobre sua atuação. Comportamento sociedade referente aos seus dados e opinião pública sobre a ética no uso de dados influencia a definição de leis e práticas no mercado dados.	Remuneração baseada no consumo de dados. Pressão para a redução de preços e o aumento da concorrência no mercado de dados demandam eficiência interna. "Produtos de prateleira" com custos menores e "produtos customizados" com mais inteligência e adaptados à necessidade do cliente, com custos maiores e implantação mais complexa para clientes que tenham recursos e reconheçam o valor para seus negócios.

Fonte: elaborado pela autora

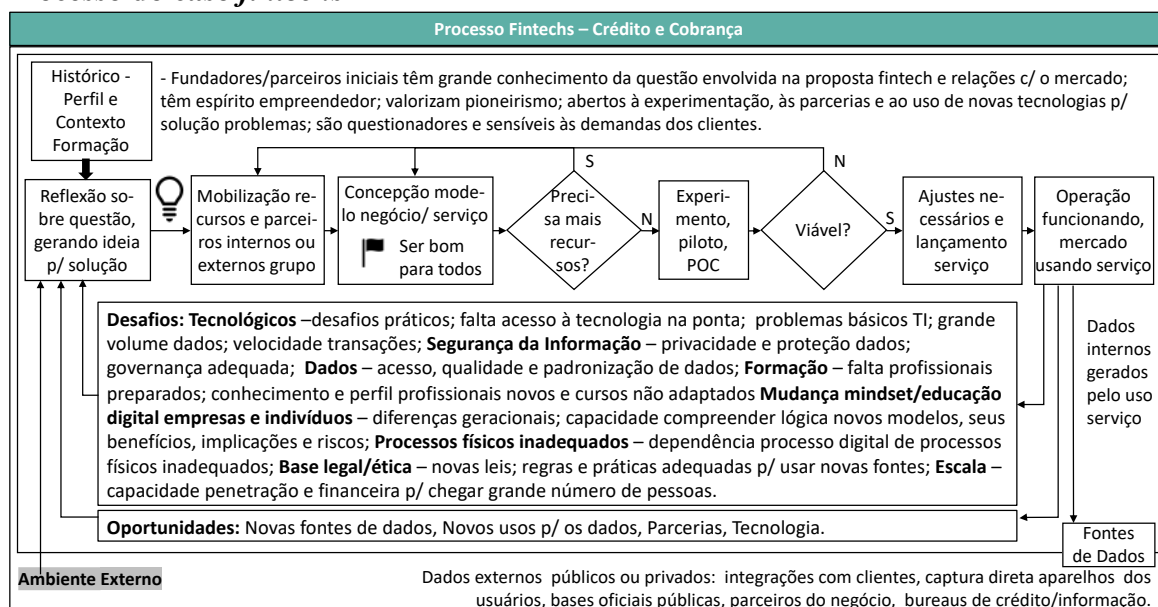
APÊNDICE F – Multilevel Framework do caso *fintechs*

Contexto do caso *fintechs*

Contexto		
Atores	Papéis	Padrões interpretativos
Bancos	Contratantes de serviços de <i>fintechs</i> .	- Conservadores; avessos a risco; operam sob forte regulamentação; são focados em negócios tradicionais; gostam de TI, mas não são ágeis em inovações pelo risco; não têm boas ofertas de crédito para classes mais baixas.
Concorrentes	Oferta de serviços de crédito e cobrança.	- Concorrentes são outras <i>fintechs</i> e instituições financeiras tradicionais (bancos, cobradoras, financeiras, etc). - Após o modelo pioneiro funcionar, novos entrantes seguem seus passos; disponibilidade de dados e tecnologias facilita entrada; - Concorrência cada vez mais competitiva – com instituições financeiras tradicionais e entre as próprias <i>fintechs</i> (novos modelos ou cópia do pioneiro).
Credores	Utilização da <i>fintech</i> como canal para oferecer suas propostas de crédito ou negociação de dívidas.	- Bancos, financeiras tradicionais, investidores e novos operadores de concessão de crédito; empresas de setores diversos credores de dívidas a serem negociadas. Houve resistência inicial, mas percebeu-se que há riscos em plataforma similares ao modelo tradicional, que há potencial de mais negócios e com resultados melhores no digital; nem sempre compreendem nova proposta e/ou estão preparados para acessar todo o valor do novo modelo de serviço.
Tomadores/devedores	Contratantes dos serviços de negociação de dívidas e crédito.	- Percebem conveniência do serviço das plataformas digitais; - Alguns não tinham acesso a crédito nos canais tradicionais.
Mercado de crédito	Serviços de crédito (empréstimos, financiamentos, negociação de dívidas, etc).	- Mais dados disponíveis para melhorar a segmentação de clientes e avaliação de risco; há novos entrantes, mas sua participação no mercado ainda é pequena; crédito inacessível ou caro para classes mais baixas; barreiras para benefícios a <i>fintechs</i> - capacidade de parceiros de integrar seus sistemas com plataformas, além de processos físicos obsoletos/inadequados que se conectam com o processo digital da plataforma e exclusão digital.
Governo	Responsável pelas regulamentações do setor financeiro (incluindo <i>fintechs</i>) e uso de dados, e pelas bases de dados oficiais de empresas/cidadãos.	- Ainda não regulamentou o cadastro positivo e o <i>open banking</i> em discussão; tema “dados” sendo discutido por governos de vários países – a exemplo do Regulamento Geral sobre a Proteção de Dados (GDPR); leis caminham para bloquear uso de dados ao máximo; estão influenciados por problemas como o do Facebook em relação ao vazamento dados.
Sociedade	Conjunto da sociedade (por ex.: cidadãos, empresas, entidades não-governamentais, associações, órgãos e agentes públicos).	- Evolução da cultura digital; pessoas dispostas a compartilhar dados em troca de benefícios; aproximadamente 40% clientes pré-pago não têm <i>smartphone</i> /internet; população de baixa renda não consegue crédito ou paga muito caro por ele; inclusão digital poderia dar acesso a modelos alternativos crédito para esses grupos.
Mercado de dados	Comercialização de dados e serviços baseados em dados.	- O aumento da disponibilidade de dados reduziu a necessidade de <i>bureaus</i> de crédito tradicionais; complexidade captura dados de redes sociais e <i>e-commerce</i> /navegação e a internet abre espaço para <i>bureaus</i> de informação.
Mercado de tecnologia	Desenvolvimento e fornecimento de infra-estrutura, ferramentas e aplicações de tecnologia e dados.	- Evolução da oferta e redução do custo de tecnologia, especialmente infraestrutura (<i>cloud</i>) e ferramentas BD (por exemplo: ferramentas <i>open source</i> ou comercializadas como serviço da Google); aumento significativo disponibilidade de dados.

Fonte: elaborado pela autora

Processo do caso *fintechs*



Fonte: elaborado pela autora

Conteúdo – caso *fintechs*

Conteúdo				
Reúso dos dados na prática: recontextualização e repropósito de dados externos públicos e privados Reúso de dados disponibilizados pelos clientes ou capturados de fontes públicas, parceiros ou bureaus de crédito/informação para segmentar clientes, definir pontuação de crédito e prover inteligência e automação ao processo e à interação com os clientes.				
Legal	Tecnológica	Intrínseca	Societal	Econômica
Leis de proteção de dados em evolução, limitando algumas operações, mas também organizando o mercado e dando mais segurança para que clientes compartilhem seus dados em benefício próprio. Mudanças de regras de operações financeiras permitem a atuação novos entrantes.	Tecnologia disponível, porém falta profissionais preparados para utilizar todo o seu potencial e alavancar as ferramentas das <i>fintechs</i> . Ainda convivem com problemas triviais de tecnologia, especialmente na interação com empresas credoras com tecnologias e processos obsoletos, o que limita alguns serviços. Entre tomadores/devedores ainda há falta de acesso à tecnologia (<i>smartphone</i> e <i>internet</i>), dificultando a penetração das <i>fintechs</i> .	Empresas credoras superaram resistências iniciais e aderiram aos modelos digitais de crédito e cobrança, embora nem todas compreendam o novo paradigma e estejam adaptadas a ele (processos internos). Tomadores e devedores aderiram rapidamente ao modelo, pois percebem a conveniência do serviço. Alguns não conseguiam obter crédito e negociar nos modelos tradicionais. Aceitam fornecer dados para melhorar sua negociação e/ou autorizam a captura de dados pelos aplicativos, facilitando o acesso a dados das <i>fintechs</i> .	Nascidas em ambientes tecnológicos, compreendem as práticas de uso de dados socialmente aceitas, prezando pela proteção dos dados, transparência do uso e pelo direito dos usuários obterem benefícios de seus próprios dados, o que contribui para conquistar confiança e apoio público.	Propostas de valor estruturadas para trazer ganhos a todos os envolvidos: melhores condições para tomadores/devedores, mais negócios e menores custos aos credores, mantendo sua liberdade de fixar taxas e níveis de risco admitidos em sua operação. Remuneração das <i>fintechs</i> normalmente baseada nas taxas de conversão. Operam com estruturas enxutas, muito informatizadas, capazes de absorver o volume necessário de operações para atingir os resultados financeiros esperados.

Fonte: elaborado pela autora

APÊNDICE G – Resumo dos desafios citados nos casos por dimensão

Caso	Dimensão	Desafios
Bureau	Econômica	Monetização dos dados para comercializar todas as informações e cruzamentos possíveis. Retração econômica do mercado e o foco dos clientes no preço da solução como seu primordial valor percebido. Alto custo para captura massiva de dados das bases oficiais do governo. <i>Timing</i> da informação para monetizar o negócio.
Fintechs	Econômica	Dificuldade de contratar bons profissionais de desenvolvimento e <i>analytics</i> no valor que possam pagar. Desafio – cenário econômico instável e condições para empreender no Brasil.
Banco	Econômica	Demanda muitos recursos para a exploração de grandes massas de dados e para a confirmação do valor dos modelos.
Banco	Intrínseca	Jornada de dados – promoção da exploração de dados e de seu uso qualificado para automatizar processos, resolver problemas e gerar inteligência, aproveitando dados entre áreas e processos ou capturando dados externos. Visão das demandas, da estrutura e cultura de trabalho em silos. Ferramentas e equipes de TI específicas para cada negócio, trazendo agilidade e especialização no suporte à área, mas também conflitos para a jornada de dados do banco. Cultura de qualidade de dados considerando novos usos, expandindo as ilhas de qualidade do banco para que os dados sejam cada vez mais controlados, monitorados e governados. Geração de valor para o cliente e para o negócio. Mudança da cultura do erro e estímulo à experimentação nas áreas. Aumento da responsabilidade das pessoas à medida que têm mais acesso aos dados. Diferentes níveis de maturidade das áreas para lidar com dados. Resistência a mudança, conservadorismo e conflitos de interesse gerados pelo reúso de dados em diversas situações entre geradores e consumidores dos dados. Pessoas capacitadas na área de dados para extrair o melhor da tecnologia e promover a jornada de dados do banco. Impacto do turnover na equipe de TI sobre as atualizações das interfaces com <i>data lake</i> . Geração de modelos tecnicamente bons e em tempo para atuar. Gestão do risco da exploração e do reúso de dados. Controle – o “dono” do dado pode controlar a concessão de acesso, mas não terá domínio sobre todo o seu reúso. Atualização constante das equipes, atração e retenção de talentos. <i>Timing</i> da atualização tecnológica para atender às demandas dinâmicas e para ser competitiva na contratação de profissionais.
Bureau	Intrínseca	Garantia da operação dentro dos padrões éticos diante das possibilidades dos dados existentes. Falta de compreensão do cliente sobre o valor de dados e modelos específicos para as suas necessidades e sobre os limites dos modelos <i>standard</i> . Dúvidas da equipe comercial e dos clientes sobre a qualidade dos dados por falta de conhecimento e comparações inadequadas com outras fontes. Resistência à mudança nos clientes por questões de estrutura e pelo medo do erro dos funcionários envolvidos. Aceitação da aplicação mais ampla de modelos preditivos em diversas áreas. Compartilhamento de dados dos clientes para a construção de modelos específicos. Alterações no acesso/bloqueio de dados das bases oficiais públicas. Dados não qualificados para seus fins, gerando resultados ruins na utilização. Risco de uso indevido de dados, mesmo que a segurança da informação esteja bem instrumentada.

Caso	Dimensão	Desafios
Fintechs	Intrínseca	Contratar pessoas capacitadas em desenvolvimento, <i>analytics</i> e <i>data science</i> e que tenham o perfil para atuar no ambiente rápido da <i>fintech</i> . Mercado para modelos de atuação inovadores. Mudança do <i>mindset</i> nas empresas para aderir aos novos serviços e compreender suas propostas de valor. Ferramentas e processos operacionais obsoletos e incompatíveis com as soluções digitais, inviabilizando ou reduzindo o valor desses modelos. Preparo e disposição dos clientes para usar a inteligência gerada. Diferenças geracionais e falta de educação digital das pessoas dificulta trazer os consumidores para a plataforma <i>fintech</i> . Adaptação das soluções das <i>fintechs</i> às diferenças dos parceiros e do mercado entre países. Objetivos de curto prazo dos gestores nos credores incompatíveis com a lógica de relacionamento “modelo digital de crédito/cobrança”. Medo das pessoas, empresas e dos governos em relação às <i>fintechs</i> . Sensibilidade dos credores para a importância da integração dos sistemas e a inteligência analítica na relação com devedores/tomadores e na conversão. Penetração – trazer consumidores para a plataforma ou pessoas baixarem e manterem o aplicativo no celular. Conciliação dos interesses da plataforma, dos credores e usuários (consumidores finais). Limites do serviço e da entrega de valor em função da estrutura e da colaboração dos credores.
Banco	Legal	Falta de regulamentações específicas. Impacto de novas leis sobre o reúso de dados no banco e sobre a visão e o risco dessa prática. Responsabilidade quando os dados são usados para novos fins. Processo de aquisição de dados para garantir que seu gerador está ciente e podem ser vendidos/cedidos. Termos de uso claros para os clientes. Riscos de vazamento ou uso indevido dos dados. Limites legais para o reúso de dados em situações críticas, como negociação de dívidas.
Bureau	Legal	Definições sobre uso indevido/discriminatório de dados e risco de algoritmos tendenciosos. Mudanças nos procedimentos e regras do governo, impactando a demanda de uso de dados das bases oficiais. Limites de acesso gratuito às bases oficiais públicas incompatíveis com o atendimento de requerimentos legais e com a realidade do mercado. Mercado negro de dados e uso de informações para golpes e atividades ilegais. Preocupação com a exposição e proteção dos indivíduos por meio dos dados.
Fintechs	Legal	Desafio – haverá muitas novas fontes de dados, a questão é a base legal para poder utilizá-las.
Banco	Societal	Há poucos profissionais capacitados e experientes na área de dados e eles estão sendo muito disputados no mercado. O perfil das pessoas mudou e continuará mudando, afetando suas expectativas em relação ao ambiente de trabalho, o que requer que as empresas também revejam seus modelos para atrair e reter os melhores talentos. Nem todas as empresas estão preparadas para “produzirem” seus dados. Validação do uso dos dados para mitigar os riscos de imagem (opinião pública) para o banco.
Bureau	Societal	Impacto do medo, das restrições legais e do comportamento social sobre o acesso a dados no futuro. Falta de atualização das pessoas sobre a evolução das coisas em geral, impactando seu preparo para o novo cenário de dados. Pôlemica pública sobre algoritmos tendenciosos e discriminatórios, necessidade de certo nível de informação para a compreensão do contexto de uso para avaliar cada situação. Interferência das práticas adotadas pelas pessoas sobre a qualidade e o comportamento dos dados que são gerados por elas em suas diversas interações com a tecnologia (uso do celular, e-mail, etc.).

Caso	Dimensão	Desafios
Fintechs	Societal	Agilidade na formação das pessoas para acompanhar a dinâmica dos conteúdos atuais. Desenvolvimento da capacidade das pessoas aprenderem constantemente. Questões microeconômicas, direitos individuais, responsabilidades sociais e diferenças culturais que afetam a discussão sobre os limites e práticas “adequadas” da tecnologia e do uso de dados. Mudanças de atitude de empresas para viabilizar a criação de padrões de dados. Capacidade de as pessoas compreenderem o funcionamento, os benefícios e riscos dos novos modelos de inteligência em suas diferentes aplicações. Má qualidade da formação básica e da grande massa de profissionais no Brasil. Viabilizar parcerias entre empresas e universidades para melhorar a formação dos profissionais de dados.
Banco	Tecnológica	Conversa entre Tableau e Hadoop ainda não funciona 100% no banco. Adaptação de processos das equipes para minimizar erros de interface e manter a integridade de <i>layout</i> e conteúdo (significados, codificações, frequência, etc.) entre as bases originais e o <i>data lake</i> . Novas formas de implementar padrões de qualidade de dados, sem conhecer seus requisitos, para garantir a qualidade das informações em utilizações ainda não previstas. Adaptação da governança para introdução das novas tecnologias no ambiente e para o controle do uso de dados usados em softwares menos controlados do que os sistemas tradicionais (legados). Agilizar a adoção de novas tecnologias no banco para reduzir a demora da atualização tecnológica em relação ao mercado. Implementação do “repositório do cliente”, consolidando alto volume de dados de diversos sistemas e formatos e garantindo a segurança e agilidade necessárias para o acesso on-line. Ferramentas para usuários explorarem dados com autonomia, garantir sua qualidade, viabilizar a integração de informações de diferentes formatos e origens, e para assegurar a integridade dos dados entre os diferentes repositórios. Níveis de qualidade e tempestividade diferentes de acordo com a utilização da informação, inclusive usos futuros não previstos. Maior risco de vazamento de dados devido ao aumento de seu acesso nas áreas. Ferramentas e processos para documentação dos dados. Demanda de conhecimento das informações e de negócio para analisar a coerência dos resultados e do período de experimentação para mitigar riscos operacionais e técnicos de novos modelos/ferramentas.
Bureau	Tecnológica	Limites da tecnologia para atender a demandas/expectativas de velocidade e processamento de dados. Qualidade e interpretação de informações. Alto risco de qualidade de dados não-estruturados. Garantia formal de disponibilidade da plataforma. Mudanças na política de acesso a dados em sites oficiais. Aumento da dificuldade de acesso e captura de informações de bases de dados oficiais. Evolução do cenário de dados tornando mais fácil validar e burlar. Falta de conhecimento, limitando o desenvolvimento de soluções e serviços que a tecnologia/dados permitiria.

Caso	Dimensão	Desafios
<i>Fintechs</i>	Tecnológica	<p>Calibração de modelos. Criação de estruturas para segurança e segregação de dados dos clientes. Sistemas dos clientes não preparados para conexões externas. Desafios práticos para a implementação de tecnologias disponíveis, como qualidade de dados, interações com parceiros e outras ferramentas, procedimentos para implantação, necessidade de novos conhecimentos e habilidades, adaptação de processos, capacidade de escolher componentes adequados e de lidar com os desafios específicos das novas tecnologias, como o <i>cloud</i>. Padronização de dados de grandes <i>players</i>, como operadoras de celulares. Clientes não adaptados ao <i>cloud</i>, não estando preparados para auditar estruturas nesse modelo de seus fornecedores. Limites da estrutura da operadora de celular para implementação de solução SSD. Falta de preparo das pessoas, colocando em dúvida o resultado de experimentações. Ambiente bastante dinâmico e com muitas novidades, não havendo conhecimento da estrutura e melhores práticas estabelecidas para diversas questões emergentes. Limitação da tecnologia na ponta: celulares dos clientes ruins, sem capacidade instalação de aplicativo, com pouco ou nenhum acesso à internet. Questões técnicas para capturas de dados em grandes volumes de redes sociais (bloqueio <i>crawler</i>) e para em massa envio de e-mails para endereços não validados (bloqueio <i>spam</i>). Garantia da privacidade. Dificuldade de atualizar e controlar a qualidade dos dados quando não está próximo de sua geração. Dificuldade de integração com clientes – sistemas legados obsoletos e/ou despreparados para integrações. Necessidade de agilidade nas transações e trabalhar com grande volume de dados no início da esteira operacional. Problemas básicos de TI, como indisponibilidade de sistemas no período noturno, incompatível com serviços on-line.</p>

Fonte: elaborado pela autora

5 CONCLUSÃO DA TESE

As três etapas da pesquisa foram articuladas em três artigos independentes e respondem, em conjunto, ao objetivo central da tese: **caracterizar o reúso de dados, identificar os desafios que dele emergem e propor um modelo para explicar como os atores lidam com tais desafios na prática.** Cada artigo aborda aspectos distintos desta ampla questão, conforme descrito abaixo no resumo do conteúdo e resultados da tese. Na sequência, um sumário das contribuições da pesquisa reflete sobre suas colaborações substantivas, metodológicas e teóricas. Para finalizar, apresenta-se um apanhado dos limites e sugestões para novos estudos que foram detalhados nos artigos.

5.1 Conteúdo e resultados da tese

O artigo I reflete sobre o conceito do reúso de dados e a importância desse fenômeno, justificando sua relevância na era do BD e evidenciando a escassez de estudos acadêmicos sobre o tema, especialmente no campo de SI. Os resultados da revisão bibliográfica dão suporte à discussão sobre a literatura existente e sobre alguns desafios do reúso de dados que poderiam ser investigados pelas pesquisas de SI, problematizando a pequena atuação da área nessa temática. Após a meta-análise de casos empíricos publicados em artigos sobre BD nos últimos cinco anos em periódicos de SI, o estudo propõe um modelo classificatório do reúso de dados. Ele é baseado nas dimensões de finalidade do reúso (repropósito ou recontextualização) e fonte do dado (interna ou externa pública/privada). A discussão dessas categorias sugere diferentes níveis de desafios potenciais em cada situação, corroborando a pertinência e a relevância dessa classificação para a investigação do reúso de dados. Para refletir a influência da distância entre a geração e o reúso dos dados sobre os desafios (Curty, 2015; Custers& Ursic, 2016; Zimmerman, 2008), o nível desses potenciais desafios nas categorias foi denominado “direto”, “intermedial” e “extremo”, como demonstra o quadro 5.1.

		Fonte dos Dados		
		Dados Internos	Dados Públicos	Dados Privados Externos
Finalidade do Reuso dos Dados	Repropósito	Direto Pode haver questões tecnológicas, sociais, legais e econômicas, porém a proximidade entre a geração e o reuso dos dados (ambos na mesma organização) pode facilitar seu encaminhamento.	Intermedial Pode haver questões tecnológicas, sociais, legais e econômicas, que tendem a ser mais complexas e seu encaminhamento pode depender de atores externos à organização.	
	Recontextualização		Extremo Pode haver questões tecnológicas, sociais, legais e econômicas, que tendem a ser mais complexas e seu encaminhamento pode depender de atores externos à organização e ao contexto do reuso, aumentando a distância destes atores, o que pode dificultar o processo de reuso.	

Quadro 5.1 - Nível de desafios potenciais do reuso de dados por categoria

Fonte: elaborado pela autora para o artigo I

O artigo II apresenta o modelo conceitual Multilevel Framework (Pozzebon, Diniz & Jayo, 2009) e propõe ajustá-lo no sentido de incorporar elementos que potencializem sua contribuição para investigar fenômenos que envolvam reuso de dados na prática. O modelo ajustado traz a representação do processo por um mapa visual (Langley, 1999). O conteúdo, que originalmente era demonstrado no modelo pela tecnologia-na-prática (Orlikowski, 2000), passou a refletir o reuso-de-dados-na-prática, evidenciando as dimensões legais, tecnológicas, intrínsecas, societárias e econômicas do reuso (adaptadas de Custers & Bachlechner, 2017). Para exercitar o modelo proposto, foi realizado um estudo de caso exploratório na Alpha. A figura 5.1 demonstra o modelo com as referidas adaptações. A empresa possui uma ferramenta de análise de risco de crédito a partir de perfis comportamentais baseados em dados da rede de celulares. Esse caso enquadra-se na categoria com potencial de desafios extremo, de acordo com a classificação proposta no artigo I: recontextualização de dados privados externos. A análise do caso demonstra que o reuso de dados acontece em um contexto pluralista (Denis, Langley & Rouleau, 2007) e marcado pela colaboração complexa (Faraj et al., 2012), dando origem a uma série de desafios que emergem na prática. Tal visão corrobora o uso de modelos teóricos que operacionalizem as teorias baseadas na prática para a investigação do reuso de dados, como faz o Multilevel Framework proposto na pesquisa.

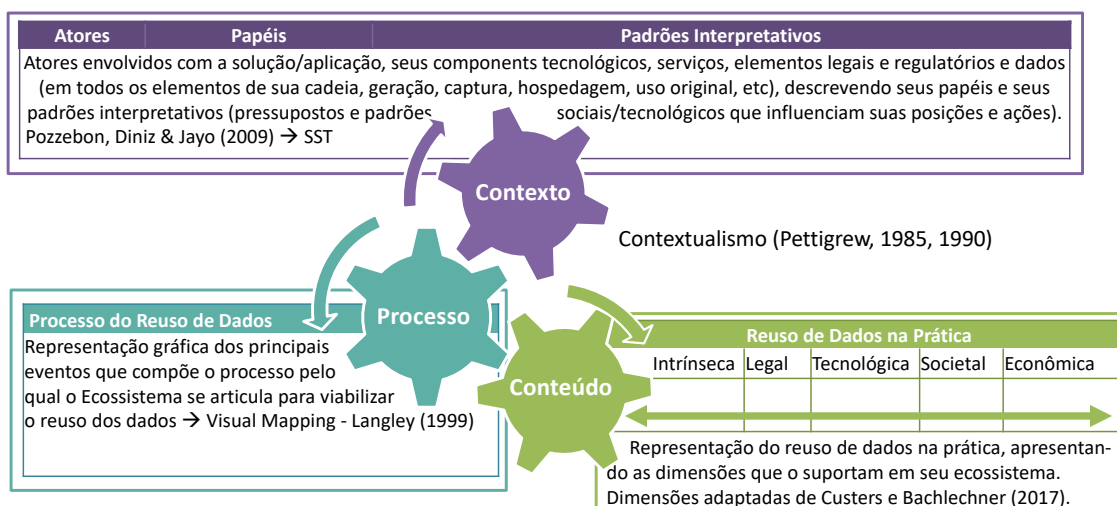


Figura 5.1 - Ecosistema do reuso de dados pelas lentes da prática

Fonte: adaptado do modelo Multilevel Framework (Pozzebon, Diniz & Jayo, 2009)

O artigo III apresenta um estudo de casos múltiplos. Os casos investigados representam diferentes categorias de reuso e ilustram os três níveis de desafios potenciais sugeridos no artigo I: direto, intermedial e extremo. Os casos são analisados individualmente, utilizando a lente teórica do Multilevel Framework adaptado, proposto no artigo II. Em seguida, uma análise *cross-case* apresenta as principais temáticas dos desafios identificados nos casos: tecnologia; acesso aos dados; confiança para ter acesso e legitimar o uso dos dados; mudança de paradigma da qualidade e curadoria de dados; ajuste de processos impactados pelo reuso dos dados; e falta de profissionais capacitados.

A reflexão sobre esses temas evidencia suas dimensões legais, tecnológicas, intrínsecas, sociais e econômicas. Analisando o processo dos casos, o artigo propõe um modelo para explicar como os atores estão lidando com tais desafios (figura 5.2). Ele revela uma dinâmica cíclica de aprendizado/ajustes que é promovida pelos grupos que mobilizam os recursos necessários para viabilizar o reuso de dados. A atuação dos mobilizadores alterna-se entre as dinâmicas de posse e de prática do poder, corroborando estudos que também identificaram a combinação dessas abordagens em outros casos envolvendo colaboração complexa (Tello-Rozas et al., 2015).

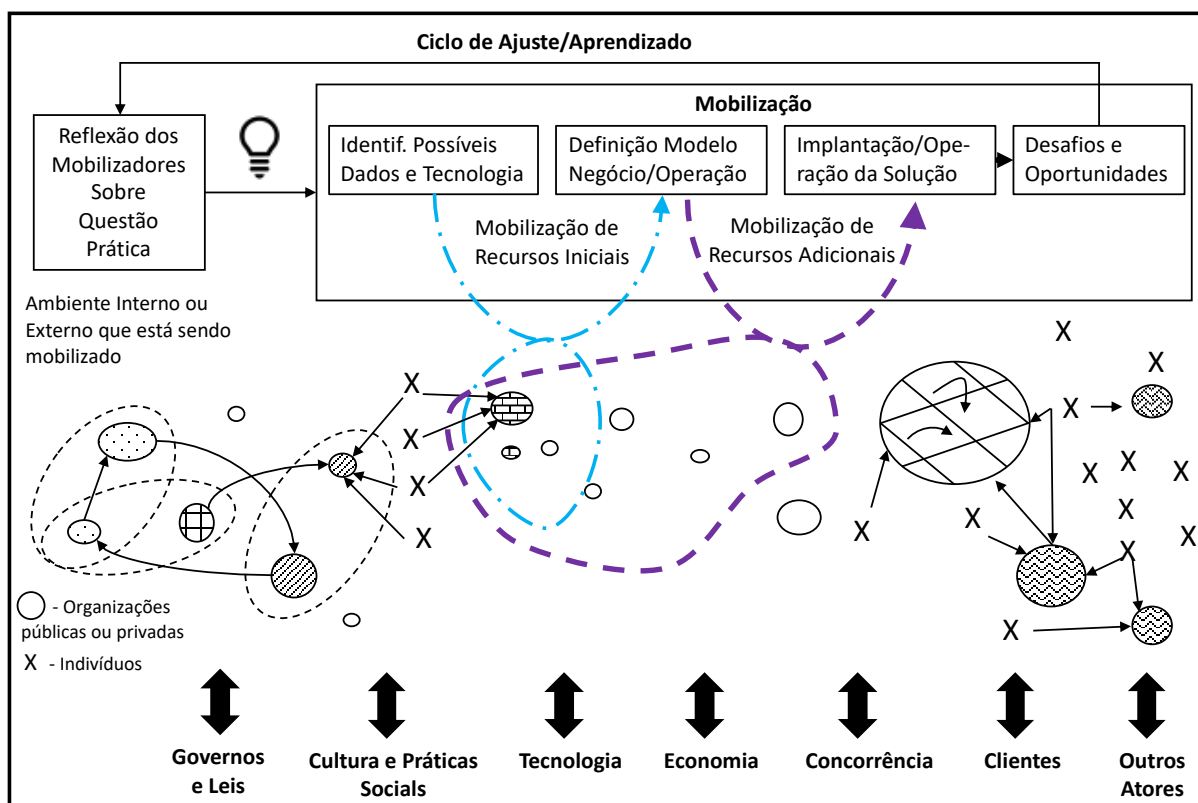


Figura 5.2 - Dinâmica do reúso de dados para lidar com os desafios de seu ecossistema
 Fonte: elaborado pela autora para o artigo III

5.2 Contribuições da Tese

Essa tese oferece contribuições para o reúso de dados, para a área financeira e para a pesquisa, conforme descrito a seguir.

5.2.1 Para o reúso de dados

A primeira contribuição da tese refere-se à proposta de definições e conceitos sobre o reúso de dados, colaborando para o desenvolvimento da pesquisa acadêmica do tema. Em sua discussão acerca da orientação conceitual para o avanço das pesquisas sobre o *strategyzing* (estratégia-na-prática), Jarzabkowski, Balogun & Seidl (2007) chamavam a atenção para a importância de um quadro coeso, definindo o que é estratégia, quem são os estrategistas, o que eles fazem, o que a análise dos estrategistas e suas ações explicam e como as teorias sociais e

organizacionais contribuem para a análise da estratégia-na-prática. Analogamente, o avanço das pesquisas sobre o reúso-de-dados-na-prática também pode se beneficiar da definição do que é reúso de dados, quem são os envolvidos nessa prática, o que fazem, o que a análise desses atores e suas ações explica, e como as teorias existentes contribuem para a compreensão do reúso de dados na prática. Embora o presente estudo não tenha esgotado tais questões, seu debate passou por todas elas, colaborando para a definição de conceitos e estimulando novas pesquisas que dialoguem com eles. Este trabalho ajuda a compreender o cenário do reúso de dados, os atores envolvidos e as questões que emergem em seu contexto, despertando a reflexão sobre seus impactos na prática e sobre os limites e as oportunidades de teorias existentes explorarem esse fenômeno.

O esquema classificatório do reúso de dados permite recortá-lo, o que pode ser útil para viabilizar a pesquisa desse fenômeno complexo. A definição das dimensões legais, tecnológicas, intrínsecas, sociais e econômicas (adaptadas de Custers & Bachlechner, 2017) oferecem uma referência que pode ajudar na produção de abordagens mais abrangentes do reúso de dados, como vem pedindo a literatura sobre o tema (Borgman, 2015; Custers & Ursic, 2016; Custers & Bachlechner, 2017, Newell & Marabelli, 2015; Samarajiva & Lokanathan, 2016; Zuboff, 2015). Por fim, ao posicionar o reúso de dados como um fenômeno que se passa em um contexto pluralista (Denis, Langley & Rouleau, 2007) e marcado pela colaboração complexa (Faraj *et al.*, 2012), o atual estudo estabelece o diálogo com pesquisas de outros fenômenos que também possuem essas características, sugerindo que o conhecimento sobre eles pode contribuir para o entendimento do reúso de dados

A terceira contribuição da tese para o reúso de dados refere-se ao campo de SI. A presente pesquisa estimula o engajamento dos pesquisadores de SI na exploração do reúso de dados, problematizando a escassa participação da área na discussão do tema ao evidenciar o impacto dessa prática sobre temáticas que fazem parte da agenda dos estudos de SI. Ao longo dos artigos também são apresentados inúmeros temas de SI a serem investigados. Além disso, o estudo chama a atenção para o potencial do perfil dos pesquisadores de SI e do corpo de conhecimento da área para a exploração do reúso de dados, especialmente em abordagens mais abrangentes que a literatura vem solicitando. A identificação do potencial de teorias como a sociomaterialidade e as teorias da prática e da colaboração complexa, por exemplo, completa esse estímulo ao inspirar caminhos teóricos para as pesquisas de SI sobre reúso de dados.

5.2.2 Para a área financeira

Esse trabalho também oferece uma contribuição substantiva. Ao apresentar casos de reúso de dados no setor financeiro, o estudo evidencia o momento de mudanças nessa área, marcado pelas atuais possibilidades tecnológicas, novos entrantes na concorrência, alterações na legislação do setor e novo perfil dos clientes e das suas demandas. Os casos ilustram como atores, tradicionais e recém-chegados, estão se articulando para sobreviver e prosperar nesse novo cenário, além de demonstrar o papel central que a tecnologia e os dados estão ocupando na arena da área financeira.

5.2.3 Para a pesquisa

A estrutura da pesquisa oferece uma referência metodológica para estudos sobre fenômenos complexos e emergentes como o reúso de dados. Nessa tese, para abordar bem o campo foram necessárias etapas anteriores para a definição de conceitos e de um esquema classificatório, criando parâmetros para a seleção dos múltiplos casos. Tais casos representavam situações distintas de reúso e seu conjunto ofereceu uma visão geral desse quintain.

A tese também demonstra a utilização do Multilevel Framework como uma alternativa para operacionalizar a abordagem teórica da prática no estudo empírico de um fenômeno de colaboração complexa. Assim como as pesquisas de SI vêm adotando a perspectiva da prática há alguns anos para a expansão da visão dos estudos envolvendo fenômenos tecnológicos, os estudos sobre colaboração complexa também destacam o potencial dessa abordagem teórica. Alguns autores alertam, porém, para a dificuldade de operacionalizar pesquisas empíricas sobre colaboração complexa que utilizem as lentes da prática. Ao utilizar o Multilevel Framework para analisar os casos de reúso de dados, a presente pesquisa demonstra como o modelo pode ser utilizado para esse fim. Ele foi adaptado para incorporar conceitos do reúso de dados pertinentes ao estudo. No entanto, o Multilevel Framework pode ser utilizado em sua forma original ou pode ser adaptado para focalizar elementos relevantes de outros fenômenos. Ele é, portanto, uma opção para operacionalizar estudos sobre diversas temáticas abordadas na perspectiva da prática, incluindo os fenômenos de colaboração complexa.

Por fim, esta tese também oferece uma contribuição teórica ao propor o esquema classificatório do reúso de dados e do modelo que explica o processo pelo qual os atores lidam com seus desafios na prática. Embora as teorias gerais da área de Administração sejam de extrema importância, é cada vez mais relevante a construção de teorias mais específicas, que sejam ferramentas para os pesquisadores diante da complexidade das questões emergentes do campo de SI (Majchrzak, Markus & Wareham, 2016). A definição de um esquema classificatório para o reúso de dados é uma contribuição teórica do tipo I, de acordo com a taxonomia das teorias de SI proposta por Gregor (2006). As teorias do tipo I visam definir “o quê”, restringindo-se à análise e à descrição de um fenômeno (Gregor, 2006). Embora não permitam estabelecer relações causais, previsões ou prescrições, esse tipo de teoria é importante para a compreensão de um fenômeno (Gregor, 2006). Já o modelo proposto para explicar a dinâmica dos atores na prática representa uma contribuição teórica do tipo II na taxonomia de Gregor (2006). As teorias do tipo II visam explicar o fenômeno e não permitem previsões (Gregor, 2006). Tanto o esquema classificatório do reúso quanto o modelo que explica sua dinâmica na prática estão concentrados na compreensão do reúso de dados e como os atores estão lidando com seus desafios. Teorias que colaborem para a compreensão do problema são fundamentais para o avanço do conhecimento relacionado a questões complexas e emergentes (Goes, 2015; Majchrzak, Markus & Wareham, 2016).

5.3 Limites e sugestões para estudos futuros

Os dados utilizados nesse estudo, assim como suas escolhas teóricas e metodológicas, foram adequados aos objetivos de cada etapa do estudo, porém não permitem generalizações. A plausibilidade de suas propostas, no entanto, sugere que eles podem ser transferidos para outros contextos, o que aumenta seu potencial de contribuições e poderia ser explorado por futuras pesquisas.

Essa tese propôs conceitos para caracterizar o reúso de dados e modelos que ajudam a refletir sobre sua dinâmica, como, por exemplo, o esquema classificatório (artigo 1), o reúso-de-dados-na-prática para retratar o conteúdo emergente no Multilevel Framework (artigo 2) e o modelo que representa a dinâmica para lidar com os desafios do reúso de dados (artigo 3).

Tais conceitos e modelos podem ser explorados em futuros estudos que confirmem sua lógica, verifiquem sua aderência e investiguem suas nuances em diferentes contextos empíricos.

Novos estudos empíricos poderiam verificar a distribuição dos casos entre as categorias de reúso de dados. Estudos de casos de outras áreas poderiam verificar como a dinâmica identificada nos casos da área financeira se repetem ou divergem de outros segmentos. Trabalhos empíricos poderiam se dedicar ao aprofundamento das questões específicas de cada dimensão do reúso de dados (legal, tecnológica, intrínseca, societal e econômica). Também poderiam explorar os desafios e os insights que emergiram nos casos, refletindo sobre como eles se relacionam com as dimensões do reúso de dados e/ou com as categorias de desafios do Big Data denominados 5 V's.

Ao longo dos artigos, foram mencionadas diversas temáticas da área de SI impactadas pelo reúso de dados, como a segurança, qualidade e curadoria de dados, os impactos do reúso de dados sobre a estrutura organizacional, impactos societais do reúso de dados e impactos do cenário legal emergente sobre o reúso de dados, por exemplo. Tais questões poderiam ser estudadas em próximos trabalhos da área, que também poderiam endereçar a sugestão de adotar as lentes teóricas da ANT e da sociomaterialidade.

Referências bibliográficas

Borgman, C. L. (2015). *Big data, little data, no data: Scholarship in the networked world*. MIT press.

Curty, R. G. (2015). *Beyond "Data Thrifting": An Investigation of Factors Influencing Research Data Reuse In The Social Sciences* (3713677). Available from ProQuest Dissertations & Theses Global. (1698463092).

Custers, B., & Bachlechner, D. (2017). *Advancing the EU Data Economy: Conditions for Realizing the Full Potential of Data Reuse*. Information Polity, 22(4), 291.

Custers, B., & Ursic, H. (2016). *Big Data and Data Reuse: A Taxonomy of Data Reuse for Balancing Big Data Benefits and Personal Data Protection*. International Data Privacy Law, 6(1), 4-15.

Denis, J. L., Langley, A., & Rouleau, L. (2007). *Strategizing in pluralistic contexts: Rethinking theoretical frames*. Human relations, 60(1), 179-215.

- Faraj, S. A., Kolker, E., Bevilacqua, L., & Özdemir, V. (2012). *Collective Theranostics And Postgenomics Entrepreneurship: Rethinking Innovations As Knowledge Ecosystems Built By Complex Collaboration*. *Expert Review Of Molecular Diagnostics*, 12(8), 787-790.
- Gregor, S. (2006). *The Nature of Theory in Information Systems*. *MIS quarterly*, 611-642.
- Jarzabkowski, P., Balogun, J., & Seidl, D. (2007). *Strategizing: The challenges of a practice perspective*. *Human relations*, 60(1), 5-27.
- Langley, A. (1999). *Strategies for theorizing from process data*. *Academy of Management review*, 24(4), 691-710.
- Majchrzak, A., Markus, M. L., & Wareham, J. (2016). *Designing for Digital Transformation: Lessons for Information Systems Research from the Study of ICT and Societal Challenges*. *MIS Quarterly*, 40(2), 267-277.
- Newell, S., & Marabelli, M. (2015). *Strategic opportunities (and challenges) of algorithmic decision-making: A call for action on the long-term societal effects of 'datification'*. *The Journal of Strategic Information Systems*, 24(1), 3-14.
- Orlikowski, W. J. (2000). *Using technology and constituting structures: A practice lens for studying technology in organizations*. *Organization science*, 11(4), 404-428.
- Pozzebon, M., Diniz, E., & Jayo, M. (2009). *Adapting the Structurationist View of Technology for Studies at the Community/Societal Levels*. *Handbook of research on contemporary theoretical models in information systems*, 18-33.
- Samarajiva, R., & Lokanathan, S. (2016). *Using Behavioral Big Data For Public Purposes: Exploring Frontier Issues Of An Emerging Policy Arena*.
- Tello-Rozas, S., Pozzebon, M., & Mailhot, C. (2015). *Uncovering Micro-Practices and Pathways of Engagement That Scale Up Social-Driven Collaborations: A Practice View of Power*. *Journal of Management Studies*, 52(8), 1064-1096.
- Zimmerman, Ann. 2008. *New Knowledge from Old Data: The Role of Standards in the Sharing and Reuse of Ecological Data*. *Science, Technology & Human Values* 631-652.
- Zuboff, S. (2015). *Big other: surveillance capitalism and the prospects of an information civilization*. *Journal of Information Technology*, 30(1), 75-89.