



Universidade Federal de Brasília - UnB
Faculdade de Economia, Administração, Contabilidade e Gestão de Políticas Públicas
Departamento de Gestão de Políticas Públicas

Frederico de Mendonça Ozanam

Dilemas éticos a partir do uso de Big Data para o aprimoramento da Gestão Pública no Brasil: Análise da produção de dados no Centro de Integração de Dados e Conhecimentos para Saúde

Brasília

2018

Frederico de Mendonça Ozanam

**Dilemas éticos a partir do uso de Big Data para o
aprimoramento da Gestão Pública no Brasil: Análise da produção de
dados no Centro de Integração de Dados e Conhecimentos para Saúde**

Monografia destinada à conclusão do curso de graduação em Gestão de Políticas Públicas da Universidade de Brasília como requisito parcial para a obtenção do título de bacharel.

Universidade de Brasília – UnB

Prof.^a Orientadora – Prof.^a Dr.^a Christiana Soares de Freitas

Brasília, DF

2018

dOZ99d de Mendonça Ozanam, Frederico
Dilemas éticos a partir do uso de Big Data para o
aprimoramento da Gestão Pública no Brasil: Análise da
produção de dados feita no Centro de Integração de Dados e
Conhecimentos para Saúde / Frederico de Mendonça Ozanam;
orientador Christiana Soares de Freitas; co-orientador
Fernanda Natasha Bravo Cruz. -- Brasília, 2018.
52 p.

Monografia (Graduação - Gestão de Políticas Públicas) --
Universidade de Brasília, 2018.

1. Big Data. 2. Dados Pessoais. 3. Centro de Integração
de Dados e Conhecimentos para a Saúde (Cidacs). 4. Dilemas
Éticos. 5. Teoria do Ator-Rede. I. Soares de Freitas,
Christiana , orient. II. Natasha Bravo Cruz, Fernanda , co
orient. III. Título.

Frederico de Mendonça Ozanam

**Dilemas éticos a partir do uso de Big Data para o
aprimoramento da Gestão Pública no Brasil: Análise da produção de
dados feita no Centro de Integração de Dados e Conhecimentos para Saúde**

A Comissão Examinadora, abaixo identificada, aprova o Trabalho de
Conclusão do Curso de Gestão de Políticas Públicas da Universidade de
Brasília do (a) aluno (a)

Frederico de Mendonça Ozanam

Prof^o Dra. Christiana Soares de Freitas

Professora Orientadora

Prof^a Dra Fernanda Natasha Bravo Cruz

Professora – Examinadora

Brasília, 29 de junho de 2018

*A todos aqueles que estiveram comigo nesta caminhada, meus pais, meus parentes,
minha namorada e meus amigos. Á minha mente, que por mais longa e difícil que fosse
a caminhada, conseguiu me fazer ficar mais forte.*

AGRADECIMENTOS

O presente trabalho é o fim de uma jornada, o fim de um sentimento que envolveu luta e suor como, mais ainda, dores e partidas, muitos ficaram para trás, o tempo foi fator determinante para perceber quem conseguiria chegar até o fim como quem também não, e para se perceber que o final é apenas um novo começo, uma nova batalha que envolve a disciplina, a vontade e a perseverança para mantê-la ainda acesa em nossos corações. O tempo mostrou que a conquista daqueles que conseguiram chegar ao fim não apaga os holofotes daqueles que não conseguiram chegar; a solidariedade talvez seja a maior riqueza do homem e sim, a competição existe como a compaixão também, logo, felizmente se observou a manutenção da vontade e da validade do interesse daqueles que não chegaram ao final como se era esperado, porém, mantiveram ainda a vontade em seus corações como também puderam ver a chama da oportunidade ainda acesa para o ambiente acadêmico.

Agradeço aos meus pais, Paulo e Yvone, aqueles que sempre me deram liberdade para pensar por conta própria, tiveram uma crença indubitável no meu valor e até onde poderia chegar; isso tudo permeado com muito carinho e compaixão, solidariedade que pude observar ser tão rara hoje em dia mas que cada vez mais ganha mais pessoas e se insere como um valor para a ação seja ela pública ou seja ela de um partícula, de um interessado. Meus pais se mantiveram sempre juntos e fizeram valer até mesmo minhas condições acima de qualquer viagem, qualquer vontade que tinham mas que, ao colocar na balança, sempre buscaram pender para o lado da conclusão de meu curso de nível superior, de minha formação acadêmica; serão sim recompensados de maneira justa por isso, não os falharei.

À minha querida namorada, Nayara, obrigado pelo tanto que você me fez bem, por todas as conversas, os encontros, foram momentos muito especiais para nós e que conseguimos guardar em nossas memórias. O apoio com trabalhos, com as dificuldades que tive na universidade, você sempre esteve de prontidão e interessada no “como resolver” acima de qualquer outra coisa; tua seriedade só me gerou mais confiança em ti e vontade de te ver mais feliz.

À professora Christiana, por acreditar em mim e confiar de que este trabalho poderia ser realizado. Sempre amiga, ouvinte atenta às minhas convicções desde à época das aulas e em que sempre tentava expor minhas dúvidas até às reuniões de monografia onde, com bastante leveza e transparência, tive a oportunidade de conhecer uma pessoa incrível que me

ajudou a entender aonde queria chegar neste trabalho e aonde poderíamos ir para que a sua confecção fosse a melhor possível.

À professora Fernanda, obrigado também por também acreditar em mim assim como a professora Christiana e, tendo em vista a matéria do projeto do trabalho final como a outra matéria que tive a felicidade de saber que me daria as aulas, pude conhecer uma pessoa também incrível que sempre encontrou um sorriso no rosto para poder acompanhar o seu dia a dia, sempre buscou tratar dos temas pertinentes à gestão de políticas públicas de uma maneira séria mas também feliz, mostrando como a leveza no tratamento que damos as coisas que nos cercam pode ser bastante oportuna ou até mesmo melhorar os nossos trabalhos em si, aprimorar o quão longe podemos chegar.

Por fim, quero agradecer a todos os meus amigos, aos meus pacientes colegas que longe ou perto nunca duvidaram de meu potencial e sempre buscaram me apoiar e me fazer levantar a cabeça quando mais precisei. Às vezes, quando somos perguntados como estamos, se está tudo bem, ou quando recebemos um gesto de carinho como um abraço, é o suficiente para simplesmente seguir em frente.

“Quando os próprios alunos estão criando, eles acabam aprendendo. E os professores serão o epicentro disto. Qualquer um que pensa diferente disto nunca teve um bom professor. Eu trocaria toda a minha tecnologia por uma tarde com Sócrates”

(Steve Jobs,2001)

RESUMO

Big Data e dados pessoais no Brasil são utilizados como ferramentas para gestão de políticas públicas, tendo como duas das principais instituições que tratam do tema da saúde pública a Fundação Oswaldo Cruz (Fiocruz) e o Centro de Integração de Dados e Conhecimentos para a Saúde (Cidacs) que buscam respectivamente ajudar na pesquisa e no desenvolvimento da tecnologia necessária para a coleta de informações e ao tratamento das mesmas para as políticas públicas da área da saúde. Big Data se insere como um conjunto de tecnologias novas que permitem um tratamento diferenciado em relação aos dados que são coletados por essas instituições. Esta pesquisa buscou analisar as implicações das formas de tratamento de dados e uso de Big Data no país, tanto por uma perspectiva ética e legal, analisando o direito à privacidade e à liberdade em face do tratamento que as instituições fazem sobre as informações disponíveis dos cidadãos, como por uma perspectiva social, utilizando a Teoria do Ator-Rede (ou Actor-Network Theory) que ajuda a interpretar as relações observadas, os discursos enunciados que representam pensamentos, opiniões e valores importantes acerca do tema.

Palavras-chave: Dados pessoais; Centro de Integração de Dados e Conhecimentos para a Saúde (Cidacs); dilemas éticos; Teoria do Ator-Rede; Big Data.

LISTA DE ILUSTRAÇÕES

Figura 1 - Organograma Institucional do Ministério da Saúde	28
Figura 2 - Lista de Colaboradores e Apoiadores do Cidacs	36
Figura 3 - Desenho de Rede de Atores Institucionais correlacionadas ao Cidacs.....	36
Figura 4 - Lista de palestrantes do evento “ Abordagem Regulatória para o tratamento de dados pessoais”	39
Figura 5 - Desenho dos links entre bases de dados de diferentes programas sociais	44

LISTA DE ABREVIATURAS E SIGLAS

TAR – Teoria do Ator-Rede

ANT – Actor Network Theory

FIOCRUZ – Fundação Oswaldo Cruz

CIDACS – Centro de Integração de Dados e Conhecimentos para Saúde

FUB – Fundação Universidade de Brasília

FGV – Fundação Getúlio Vargas

IPEA – Instituto de Pesquisa Econômica Aplicada

EV – Entidades Vinculadas

HEMOBRAS- Empresa Brasileira de Hemoderivados e Biotecnologia

ANVISA – Agência Nacional de Vigilância Sanitária

FUNASA – Fundação Nacional de Saúde

ANS – Agência Nacional de Saúde Suplementar

MCI – Marco Civil da Internet

DNI – Documento Nacional de Identificação

Sumário

1. INTRODUÇÃO	13
1.1 Problematização, Justificativa e Objeto de Análise.....	15
2. ABORDAGENS TEÓRICO-METODOLÓGICAS	21
2.1 Big Data na área de Saúde	24
2.2 A Teoria do Ator-Rede (ANT- Actor Network Theory)	25
3. METODOLOGIA	27
4. ANÁLISE DE DADOS	29
4.1 A produção de big data no CIDACS : A “Plataforma de Dados” e a Produção dos dados.....	30
4.2 A Curadoria	31
4.3 A privacidade e a ética quanto ao tratamento dos dados.....	33
4.4 Links entre actantes	35
4.5 Cartografia de Controvérsias: o comum e o diferente entre os atores para as políticas públicas.....	38
4.7 Percepção dos pesquisadores sobre o uso de big data.....	44
5. CONSIDERAÇÕES FINAIS	48
REFERÊNCIAS	50

1. INTRODUÇÃO

Vivemos em uma era de inovação tecnológica ou de destruição tecnológica? Ou até mesmo ambas? Tem sido comum a evidência de que cada vez mais “o novo” tem sido atingido, buscado, até mesmo perseverado com tal veemência e velocidade que o estudo de novos produtos, tendências ou até mesmo dos anseios da sociedade em determinada época acaba por se tornar um trabalho muito mais árduo do que a tempos mais remotos, principalmente porque atingir uma posição de liderança com inovação não tem sido fácil e conservar tal posição tem sido ainda mais difícil (DAVILA e EPSTEIN, 2007)¹. Quantos *smartphones* entram no mercado propondo serem “um produto como nenhum outro jamais visto antes” para que em menos tempo do que o previsto, até mesmo pela própria fabricante, a demanda pela nova versão surja acalorada e impactando tomadas de decisão? Quantos hardwares e softwares são criados praticamente todos os dias ao redor do mundo? Até pouco tempo atrás, falava-se pouco sobre tais tipos de produtos e os mesmos eram vinculados a grupos específicos de desenvolvedores.

Sendo assim, a análise por parte das ciências humanas quanto a tal fenômeno social se faz necessária para que possamos entender o que está acontecendo, a velocidade criativa ou destrutiva em relação a aquilo que temos e que podemos vir a ter e se, dadas tais circunstâncias, ainda é possível ou não alguma escolha.

Schumpeter muito antes de *smartphones* e *tablets* já discutia o processo que denominou como “Destruição Criativa”. O dinamismo dos tempos modernos se assemelha e muito ao dinamismo das primeiras revoluções, o surgimento da energia a vapor quando a data da Primeira Revolução Industrial pode ser tão importante quanto o surgimento do primeiro computador pessoal/privado que, mais tarde, serviu para verdadeiras mudanças severas no modo de se negociar, comercializar e até mesmo produzir. A tais mudanças, advindas do surgimento de novos produtos ao mercado, mudanças estas que impactam o comportamento tanto da produção quanto dos consumidores, Schumpeter descreveu como um processo cíclico denominado “destruição criativa”.

Ferrarezi e Tomachesk (2010) já propunham falar de inovações com um maior cuidado; não se pode pré-determinar mudanças e inovações como processos simples e com passos e séries de etapas que podem ser planejadas antes de ocorrerem. A inovação vai

¹ DAVILA. Tony, EPSTEIN. Marc. SHELTON, Robert. “ As Regras da Inovação – Cap.1 Conduzindo o sucesso: a forma como você inova define o que você inova” (pág. 21-23).

moldando um novo sistema, ainda que não percebamos isso de imediato mas apenas com o passar do tempo, o que remete também às ondas de destruição criativa já que, por mais vertiginosa que possa parecer cada ponto mais íngreme de cada onda, a ideia é de que as mudanças que impactem o desenvolvimento com inovações significativas venham atreladas também de passos que não foram pré-determinados e que acabam alavancando processos contínuos resultantes de tais mudanças.

Pensando nos dias atuais e em tais processos, temos como uma das últimas ondas do quadro as redes digitais, os softwares e as novas mídias; todas essas são mudanças que incrementam a concepção que temos do que vemos como sociedade, do que queremos como indivíduos. A informação por si só acabou se tornando uma das mais drásticas mudanças que vemos ocorrer com o tempo, a forma pela qual está posta atualmente, levando em consideração a forma que tínhamos há tempos atrás, mudou para uma forma completamente diferente. Hoje, por exemplo, fala-se do controle da informação por mecanismos virtuais. O conhecimento da tecnologia, então, faz-se fundamental para que se acompanhe tais mudanças. Para se ter uma idéia, segundo Gantz e Reinsel (2011), se tomarmos o período entre 2006 e 2011 e o analisarmos, veremos que a quantidade de informação (sejam elas em textos, vídeos, fotos, filmes, documentos dos mais variados) é tamanha que já ultrapassou o trilhão de *gigabytes*. Já se usa a nomenclatura de *zettabytes* para se tratar do assunto, 1 trilhão de *gigabytes* equivale a 1 zettabyte, e isso, como preceituavam os autores, ainda é acompanhado de um aumento desse valor em dobro a cada dois anos. Então, já naquela época se atentava para a ideia de que os 1.8 trilhões de *gigabytes* que se observava em praticamente 500 quadrilhões de arquivos, é uma quantidade de *bits* que corresponderia a praticamente a quantidade de estrelas no nosso universo físico, a quantidade de estrelas que vemos e que não vemos no céu.

O conhecimento que vem a nosso interesse aqui, neste trabalho, e que trata dessa quantidade imensa de dados e arquivos disponibilizados na internet, é o do Big Data, que hoje funciona como um instrumento a serviço tanto de entidades governamentais como de entes privados atendendo a interesses importantes não só destes como, muitas vezes também, da sociedade.

Big data significa uma grande quantidade de dados e de bases de dados que são armazenadas e organizadas para fins diversos. Apesar de o nome em si conseguir explicar bastante à primeira vista (o que de fato é bastante útil), ainda assim carece de mais informações sobre o quão inovador é este novo tipo de armazenamento que, frisamos aqui,

apesar do nome, não se refere apenas ao armazenamento em si mas também às características que o envolvem, como variedade, velocidade, volume e, principalmente, valor.

Cada vez mais as empresas famosas vêm se utilizando de uma ferramenta comum e que se ancora em Big Data em vários momentos: o mapeamento de consumidores (DUARTE, 2013). O mapeamento consiste em se observar certos padrões de consumo, hábitos de um público alvo determinado e traduzir a identificação em uma ferramenta objetiva não só para propagandas e marketing como também para a própria produção do conteúdo em si, uma inovação recente que consiste no uso de uma base de dados previamente construída para garantir a execução de uma tarefa que tem como base de sua ação a seguinte pergunta: “O que você quer?”.

Big Data vem sendo utilizado também no setor público. O Ministério da Justiça, por exemplo, possui o *IBM Watson Explorer*, que serve para a mineração de dados em larga escala e em grandes tamanhos, terabytes², para sua posterior conversão em gigabytes, permitindo, assim, uma busca por “palavras-chave” que facilita a ação do órgão para o combate à lavagem de dinheiro e a outros atos ilícitos.³

A proposta deste trabalho tem como escopo a análise de Big Data na área de saúde no setor público e, mais precisamente, a análise do uso de Big Data para o incremento e aprimoramento de políticas públicas nacionais para a saúde, tornando-as, por vezes, mais precisas e eficientes.

1.1 Problematização, Justificativa e Objeto de Análise

Quando analisamos a imensa quantidade de dados que estão à disposição dos diversos órgãos governamentais ou das diversas instituições de capital privado que existem ao redor do mundo, podemos parar para nos perguntar: para quê? Por que há um interesse tão grande nos gostos, nas vontades de cada um e o quão relevantes são essas informações para a determinação dos objetivos ou até mesmo do cotidiano dessas instituições? Seria ético?

Em abril de 2018, Mark Zuckerberg, atual presidente e fundador da rede social Facebook, passou por uma sabatina no Congresso Americano dos Estados Unidos. Foi indagado em relação ao uso indevido de dados pessoais de milhões de usuários de sua

² Nomenclatura para tamanho de dados contidos em arquivo ou programa de computador, 1 terabyte equivale a 1000 gigabytes.

³ Corporação IBM. “Ministério da Justiça do Brasil: REDE-LAB do Brasil identifica ativos ilícitos com a ajuda do IBM Watson Explorer” Disponível em: <<https://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=IMC14890BRPT>> Acesso em: 8 set.2017.

plataforma. Os principais veículos de informação apontam que se chegaria ao número de 87 milhões de usuários, sendo 443 mil brasileiros. A rede social teria, segundo aqueles que acusam Zuckerberg, influenciado nas últimas eleições presidenciais norte-americanas (as quais elegeram o então presidente dos Estados Unidos, Donald Trump) ao ceder à empresa *Cambridge Analytica* dados pessoais de milhões de usuários, o que teria não só influenciado no resultado das eleições presidenciais como também demonstrado um atentado sério e gravíssimo ao direito de privacidade e à proteção de dados pessoais de indivíduos de inúmeros países, incluindo o Brasil.⁴

Em *Digital Rights in Latin America and the Caribbean* (Direitos Digitais na América Latina e no Caribe), as autoras afirmam a necessidade da garantia de direitos humanos na internet (Freitas *et al.*, 2018). No capítulo intitulado “*Personal Data Protection and Right to Privacy*” (Proteção de Dados Pessoais e Direito à Privacidade) é demonstrado como vários países da América Latina já possuem leis para proteção dos dados pessoais dos seus cidadãos; mas o Brasil ainda não possui uma lei específica para tanto. Um dos problemas atuais é a extração de dados pessoais por grandes corporações sem o prévio consentimento de seus proprietários. Nesse sentido, Freitas *et al* afirmam:

“[...]A delicate issue nowadays is personal data extraction by corporations without their owner’s consent. The most recent scandal involving Facebook and Cambridge Analytica revealed the citizen’s unawareness and vulnerability. Using apparently innocent tests to check personalities the company extracted citizen’s personal data without their consent to influence elections in Great Britain and in the United States” (FREITAS *et al*, p.11, 2018).

Castells (2017), ao descrever e analisar a sociedade em rede, ressalta sua estrutura social construída em torno de redes ativadas por tecnologias de comunicação e de informação (TICs), processadas digitalmente e baseadas na microeletrônica; seriam estruturas formando novos arranjos sociais, arranjos em que as relações de consumo, reprodução, experiência e poder seriam assim expressas segundo códigos que se organizam de acordo com cada cultura. A capacidade que essas redes possuem de ferir a autodeterminação política das pessoas e de inserir “cultura” dentro das comunidades e grupos que são formados dentro delas é muito grande.

Fato é que a privacidade, como tínhamos há tempos atrás é posta em xeque. Não apenas o Facebook diz ser um espaço que protege os dados pessoais de seus usuários, mas praticamente todas as demais redes sociais promovem a ideia de que estamos nos utilizando de ferramentas que servem aos nossos interesses.

⁴ Portal de Notícias BBC Brasil. “Facebook admite uso indevido de dados de 87 milhões de usuários, 443 mil no Brasil” 2018. Disponível em: < <https://www.bbc.com/portuguese/geral-43646687>>.

A ideia dessas corporações é de que não é apenas a promoção de seus lucros e ativos financeiros o que importa, mas, também, a promoção de espaços cada vez mais democráticos de demonstrações de respeito à opinião alheia e ao diálogo. Entretanto, como usuários, sabemos que não é tão certa tal definição. A sabatina que aconteceu no Congresso Americano demonstra isso, o uso indevido não afeta apenas o Facebook como empresa e suas ações na Bolsa de Valores americanas; nos afeta socialmente também na medida em que vem aqui ao nosso interesse buscar o real "indevido" no uso das ferramentas que mineraram os dados de milhões de pessoas: há um dilema ético posto? As pessoas realmente se importam com sua privacidade e com o que está sendo feito com seus dados pessoais? Independente de se importarem ou não, teria importância para os governos?

Existem pessoas que possuem o mais variado número de opiniões em relação às mais diferentes temáticas possíveis; temos pessoas que torcem por determinado time de futebol e que podem se satisfazer ao entrar em determinado site de compra e venda de produtos esportivos e encontrarem camisas, tênis, roupas das mais variadas que tem relação com a equipe a qual torcem, entretanto, e quanto às opiniões políticas ou religiosas dessas mesmas pessoas? Será que as mesmas teriam tanto interesse assim no uso de seus dados para a oferta, pela mesma loja virtual, de produtos que tivessem a ver com suas opiniões políticas, religiosas ou até mesmo que tivessem a ver com a sua orientação sexual? Ou esses dados são até mesmo mais do que meramente dados pessoais, significando de fato dados "íntimos" que devem ser preservados de acordo com o direito à privacidade e à proteção de dados pessoais?

Tais indagações buscam amparo na ética e, ainda que haja reflexões em relação ao que seja ético ou não, aos reais limites que podemos colocar no uso de informações pessoais, o fato é que a possibilidade da coleta desses dados já existe e dificilmente poderá ser freada. Não é difícil termos pessoas que voluntariamente exprimem suas opiniões, seus votos, suas crenças e até mesmo exibem os parceiros com os quais se relacionam seja em uma ou até mais redes sociais; há uma verdadeira cultura que busca pela exposição, pela demonstração até mesmo, segundo alguns estudos, por uma auto validação, uma auto validação social para uma geração de valor de si mesmo em face do outro, o próprio termo "face" tem relação com a rede social Facebook, pois a mesma em sua origem surgiu com a ideia de promover um julgamento de estudantes em relação aos rostos de cada um, qual estudante seria mais bonito do que qual e assim se promovendo a competição por "quais eram os mais belos" da universidade em que seus fundadores estudaram.

A questão da privacidade e de seus usos, sejam eles por empresas como Facebook ou até mesmo entidades governamentais, acaba interferindo de maneira direta e indireta no

âmbito da ação pública e das políticas públicas que temos desenvolvido para salvaguardar a população com medidas que buscam sanar, muitas vezes, aquilo que falta em uma ação governamental, para mais diretamente tentar proteger, tentar solucionar em regiões e em casos mais específicos. Programas que tratam da fome, da miséria, da falta de acesso à energia ou a outros elementos que são essenciais e que deveriam ser acessíveis a qualquer ser humano, encontram também amparo no uso dos dados de brasileiros.

Tendo em vista a importância e o reflexo que tal tecnologia pode vir a trazer na Gestão Pública como um todo, surge a seguinte pergunta norteadora deste trabalho: *como Big Data vem sendo utilizado para o processo de formulação de políticas e ações na área de saúde?*

O objetivo, neste trabalho, será a análise de processos sociotécnicos que envolvem o uso de Big Data pela Fundação Oswaldo Cruz (FIOCRUZ), mais especificamente pelo Centro de Integração de Dados e Conhecimentos para Saúde (CIDACS)

Este trabalho terá como foco de análise o CIDACS, tendo em vista os princípios que norteiam as diretrizes dos trabalhos deste órgão e também por conta da já utilização de Big Data para suas políticas que envolvem a saúde pública e a gestão dos dados a ela pertinentes. O CIDACS é uma instituição vinculada a uma rede na qual participam outras instituições importantes, como a Fiocruz da Bahia (Instituto Gonçalo Muniz), Instituto de Pesquisa Econômica Aplicada (Ipea), Fundação Universidade de Brasília (FUB), Fundação Getúlio Vargas (FGV) e Fiocruz-Brasília. O CIDACS é quem realiza estudos e pesquisas com base em projetos interdisciplinares que possuem origem na integração de grande volume de dados. O Centro busca contribuir com a produção de conhecimento científico para fornecer auxílio na tomada de decisão em relação às políticas públicas na área de saúde.

Cabe à fundação FIOCRUZ “promover a saúde e o desenvolvimento social, segundo a própria, gerar e difundir conhecimento tecnológico e científico para ser assim um agente da cidadania”⁵. Partindo-se de tal princípio, a fundação tem realizado cursos sobre Big Data para o aprimoramento da tecnologia de dados na área da saúde, justamente para antecipar a ocorrência de doenças e vírus, promovendo assim a prevenção. Esses cursos começaram com o projeto “Big Data em saúde”, projeto este que procurou ampliar a quantidade de dados que

⁵ Portal Institucional da fundação FIOCRUZ. Disponível em: <https://portal.fiocruz.br/pt-br/content/funda%C3%A7%C3%A3o>

eram utilizados pela instituição e aumentar o leque do tipo de dados utilizados, adotando também o uso de mídias sociais, dados não estruturados e dispositivos conectados à internet.⁶

Tendo como base a Teoria do Ator-Rede para observar as conexões entre atores, vale a pena frisar o organograma da FIOCRUZ, que será trabalhada aqui

Segundo o que se verifica no Relatório de Atividades Icict 2013-2014, a Fiocruz é, em conjunto com a Funasa, uma fundação pública e que, por sua vez, em conjunto com as autarquias Anvisa e ANS, com a empresa pública HEMOBRAS e com as sociedades de economia mistas hospitais Cristo Redentor, Fêmina e N.S da Conceição, mais uma das instituições que ajudam a formar o que se denominou por Entidades Vinculadas (EV) que compõem o Ministério da Saúde.

⁶ BEZERRA, André. Portal de notícias da FIOCRUZ. “*Grupo de pesquisa da Fiocruz se dedica à estratégia Big Data em Saúde*”. Disponível em: <https://portal.fiocruz.br/pt-br/content/grupo-de-pesquisa-adota-estrategia-big-data-em-saude-sistema-que-facilita-o-acesso>

Objetivo Geral

A partir do problema de pesquisa exposto, espera-se ter, ao final deste trabalho, reflexões a respeito do que seja Big Data para o setor público, analisando uma iniciativa que é produto de uma rede interorganizacional da qual a Fundação Oswaldo Cruz (Fiocruz) participa. O objetivo geral, portanto, consiste em analisar os processos sociotécnicos e políticos que envolvem o uso de Big Data para o desenvolvimento de políticas públicas na área de saúde e suas implicações legais e éticas.

Pode-se elencar como objetivos específicos:

- Identificar os usos e aplicações de big data para pesquisas;
- Analisar em que medida essas pesquisas vem sendo usadas, posteriormente, para formular políticas públicas ou outras ações governamentais;
- Discutir as limitações e potencialidades do uso de big data para a formulação, implementação e avaliação de políticas públicas;
- Analisar os desafios éticos e legais relacionados ao consentimento do uso de dados pessoais;
- Identificar a percepção dos cientistas em relação ao não-consentimento dos cidadãos que, muitas vezes, tem seus dados pessoais coletados e tratados.

2. ABORDAGENS TEÓRICO-METODOLÓGICAS

Ferrarezi e Tomacheski sugeriram, em 2011, a discussão necessária a respeito da sustentabilidade de inovações premiadas no país e como elas se enraizavam nas instituições governamentais mudando, muitas vezes, o comportamento não só do sistema, como também da própria gestão pública e dos servidores que com ela trabalham. Como esses autores assinalam, a inovação não era algo, de início, próprio da gestão pública, mas sim algo que tinha mais relação com a lógica do sistema privado:

Os estudos sobre inovação, em sua origem, foram desenvolvidos no âmbito das políticas de ciência e tecnologia. O conceito tinha caráter restrito à inovação tecnológica e era equiparável à invenção, implicando originalidade. O estímulo à inovação era considerado como inerente às organizações privadas, movido pela competição e busca da maximização de lucros vigentes no ambiente de mercado (FERRAREZI e TOMACHESKI, 2011, p.3).

A lógica até então adotada abordava mais o âmbito da originalidade e da tecnologia quando se tratava de inovação. As necessidades estatais alteraram tal concepção. Ficou clara a importância de considerar a inovação no setor público como necessária e produtora de valor público, percepção distinta da lógica focada em originalidade para geração de valor de mercado. Conforme explicam os autores,

com a emergência nos anos 1990 dos princípios da Nova Gestão Pública, propondo tornar as organizações públicas mais *accountables*, por meio de mecanismos de controle de resultados e uso de indicadores para mensuração de desempenho e eficiência, a inovação passou a ser percebida como algo compatível com a gestão pública, sob motivações distintas dos estímulos do mercado (FERRAREZI e TOMACHESKI, 2011, p.3).

Nesse sentido, big data representa uma inovação atual para a gestão pública vigente no sentido de potencializar as possibilidades de geração de valor público. O conjunto de tecnologias que moldam Big Data baseiam-se em alguns dos pressupostos comuns à Nova Gestão Pública, como a busca por eficiência e eficácia. Schonberger e Cukier (2013) chamam a atenção para a importância de se ter acesso a uma variedade de dados e de sua consequente organização para uso posterior. Em seu livro *“Big Data - como extrair volume, variedade, velocidade e valor da avalanche de informação cotidiana”*, os autores demonstram como a riqueza de dados e informações, principalmente na área da saúde, podem salvar vidas. Os autores explicam que, à época do surto do vírus H1N1 nos Estados Unidos, as autoridades de saúde pública não conseguiam acompanhar o quadro de pandemia que se instalava na população norte-americana, os casos aumentavam cada vez mais e mais e, sendo assim, as informações eram computadas com severa defasagem. Entretanto, funcionários da Google já

havia coletado informações sobre a gripe e seus supostos locais de incidência com base nas buscas realizadas pelos seus usuários na ferramenta. Como explicam Schonberger e Cukier:

Os autores explicaram como a Google pode “prever” a disseminação da gripe de inverno nos Estados Unidos, não apenas nacionalmente, mas em regiões específicas e até mesmo em estados. A empresa obteve essa previsão ao analisar os termos mais pesquisados na internet. Como o Google recebe mais de três bilhões de pesquisas por dia e as salva, a empresa tinha muitos dados com os quais trabalhar.

A Google pegou os 50 milhões de termos de busca mais comuns dos americanos e os comparou à lista com dados do CDC sobre a disseminação da gripe de 2003 e 2008. A ideia era identificar pessoas infectadas pelos vírus da gripe de acordo com o que pesquisavam na internet. Outros haviam tentado fazer o mesmo com os termos de busca, mas ninguém dispunha de tantos dados, poder de processamento e conhecimento estatístico como a Google (SCHONBERGER e CUKIER, 2013, p.2).

O sistema foi capaz de combinar/correlacionar termos de busca com casos reais, o que ajudou a verificar os números oficiais demonstrados pelo governo americano. Assim foi possível identificar características da gripe, os principais locais de sua disseminação e por onde a gripe havia se espalhado de maneira mais específica:

Apesar de os funcionários da Google pensarem que as buscas poderiam ser usadas para obtenção de informações sobre a gripe – ao escrever frases como “remédio para tosse e febre”-, essa não era a questão: eles não sabiam e criaram um sistema sem muita utilidade. Tudo o que ele fazia era procurar correlações entre a frequência de certas buscas e a disseminação da gripe em relação ao tempo e espaço. No total, eles processaram impressionantes 450 milhões de modelos matemáticos diferentes a fim de testar os termos de busca, comparando suas previsões com casos reais de gripe registrados pelo CDC em 2007 e 2008. E tiveram sorte: o programa descobriu uma combinação de 45 termos de busca que, quando usados juntos num modelo matemático, tinham forte correlação entre a previsão e os números oficiais. Com o CDC, eles podiam ver por onde a gripe havia se espalhado, mas, ao contrário do CDC, podiam apontar a disseminação quase que em tempo real e não com uma ou duas semanas de atraso” (SCHONBERGER e CUKIER, 2013, p.2).

Olhando para o futuro, esse simples fato talvez possa vir a servir como referência para muitos outros casos onde haverá trabalho conjunto entre administração pública e ente privado. Esse trabalho conjunto pode servir para prevenir situações trágicas.

Big Data tem ajudado a definir demandas para grandes corporações e ao mesmo tempo demandas sociais para o governo, para que assim este possa focar suas diretrizes de maneira mais adequada e específica possível, tendo em vista situações trágicas e outras que exigem também diligências.

Mas o que seria Big Data? Temos já a noção da importância da variedade na coleta de dados e da compactação desse grande volume para um futuro uso, mas o que de fato vem a ser esse termo que, a grosso modo, pode muito bem ser traduzido como “grande volume de dados” (grande volume, grande variedade, muitos dados, etc.).

A verdade é que não há um consenso na literatura que conceitua o termo ainda, o que temos são conceitos que às vezes podem parecer difusos mas, com um olhar mais atento, podemos sim pelo menos tentar chegar a algum denominador comum para ajudar à nossa compreensão quanto ao tema.

De acordo com o *GartnerGroup* (2012), “Big Data, em geral, é definido como ativos de alto volume, velocidade e variedade de informação que exigem custo-benefício, de formas inovadoras de processamento de informações para maior visibilidade e tomada de decisão”. Já, segundo a IDC (International Data Corporation), “... as tecnologias de Big Data descrevem uma nova geração de tecnologias e arquiteturas projetadas para extrair economicamente o valor de volumes muito grandes e de uma variedade de dados, permitindo alta velocidade de captura, descoberta, e/ou análise”. Cezar Taurion (2013) afirma que Big Data equivaleria à soma de volume, variedade e velocidade. E, por fim, temos o conceito da IBM (2014), como nos demonstram Rufino, Volpato e Dias (2016). Segundo os autores, “Big Data é o termo utilizado para descrever grandes volumes de dados e que ganha cada vez mais relevância à medida que a sociedade se depara com um aumento sem precedentes no número de informações geradas a cada dia”⁷.

Quando tratamos de grandes volumes de dados gerenciados por empresas ou órgãos governamentais, estamos tratando de dados estruturados e não estruturados. Como Huacarpuma (2017) explica, as formas que esses dados podem ser encontrados são das mais diversas. Independente da estrutura, podemos ter dados em áudio, vídeo, imagem, entre outras formas, o que denota a complexidade desses dados e a variação em relação ao seu conteúdo.

Esses dados, como Huacarpuma (2017) preceitua, por conta do quanto é variável o conteúdo, tornam-se assim heterogêneos. Após a sua coleta, precisam ser analisados, e é aí que se verifica a novidade trazida pelo Big Data, nas possibilidades de análise e tratamento dos dados. As formas mais tradicionais eram, geralmente, com base no uso de sistemas de computação centralizados para a análise desses dados, sistemas de computação tradicional com servidores centralizados, o que ainda é possível, porém trazendo óbices como a capacidade de processamento e o volume de armazenamento possível. O processamento paralelo, ou seja, a análise em paralelo de dados com conteúdos diferentes entre si acaba encontrando uma vantagem em relação aos sistemas mais tradicionais.

⁷ VOLPATO, et al. Big Data- Transformando Dados em Decisões.2016. Disponível em :<http://web.unipar.br/~seinpar/2014/artigos/graduacao/Tiago_Volpato.pdf>.

Segundo Silva (2017), big data insere-se no contexto em que já não mais preceitua a privacidade apenas do ponto de vista moral, mas também como um direito. O uso de big data serve para buscar as correlações de dados entre si e também, como aponta Alves (2016), a relação desses com os usuários da internet que os disponibilizaram. Convive-se com dados diuturnamente, desde um post feito para uma rede social ou até mesmo e-mails enviados de uma pessoa para outra, tudo gera rastros, caminhos, pegadas digitais que dão pistas acerca de quais sejam esses mais variados sentidos que estão postos ali, em um emaranhado denso e consistente de informação que evolui a todo o tempo.

Alves (2016) então conceitua *big data* de acordo com suas características. Assim determina-se o mesmo como um conjunto de procedimentos de análise, de formas de processamento e de armazenamento das informações disponíveis, pelas novas ferramentas para se manipularem os dados, pelos novos algoritmos, tendo como um todo o objetivo de, após feita a coleta e manutenção desses, de se buscar transformar as informações em ganhos de produtividade no setor privado, no setor público; o objetivo é gerar valor público sob a forma de serviços e bens públicos.

2.1 Big Data na área de Saúde

Podemos observar pelo exemplo de Schonberger & Cukier (2013), que Big Data serviu de certa forma para um diagnóstico mais atualizado em relação ao surto da pandemia que acometeu a população norte-americana à época, ou seja, pode vir a se tornar uma possibilidade para medidas mais eficazes para a prevenção de problemas na área e para o aprimoramento da saúde pública. Mas o que seria a saúde pública?

Um dos conceitos mais difundidos de saúde pública que temos seria provavelmente o de Winslow (1920) que define saúde pública como

a ciência e a arte de prevenir doenças, prolongar a vida e promover saúde física e eficiência através de esforços comunitários comuns pelo saneamento do meio ambiente, o controle de infecções comunitárias, a educação do indivíduo em princípios de higiene pessoal, de organização de serviços médicos e de enfermagem para recente diagnóstico e tratamento preventivo para doenças, e o desenvolvimento do maquinário social que irá assegurar para cada indivíduo dentro da comunidade um padrão de vida adequável para a manutenção da saúde. (WINSLOW, 1920, p.30).

Depreende-se de tal conceito que a saúde pública não seria somente um empreendimento estatal, mas também um empreendimento da sociedade a partir do momento em que se concebe que são necessários esforços comunitários comuns, ações tanto individuais como também coletivas, partindo-se da atividade da sociedade civil organizada para que

padrões de higiene e saneamento sejam verificados no dia-a-dia da população. O curioso de tal conceito é justamente o misto entre arte e ciência, colocando ambas as esferas do intelecto humano em um patamar de equidade quando se refere ao uso tanto de uma quanto de outra ou até mesmo do uso das duas juntas para se atingir o que se entende como saúde pública. A importância denotada ao social seria justamente em relação ao maquinário formado. O conceito abrange a ideia de que nós, como indivíduos coletivos, acabamos gerando um maquinário, um maquinário social, analogamente a um sistema que, composto por suas peças, possui um certo tipo de funcionamento que não depende apenas daquele que realiza o controle direto dele, mas justamente também dos indivíduos que nele se inserem e que indiretamente ou não ajudam em tal controle.

2.2 A Teoria do Ator-Rede (ANT- Actor Network Theory)

A Teoria do Ator-Rede foi escolhida então justamente por esse caráter descritivo ou interpretativo em relação às relações acerca das inserções associativas entre os indivíduos e como ocorre a ajuda no controle anteriormente citado, ou seja, como ocorre justamente a formação desse maquinário social a partir de uma perspectiva sociológica um pouco diferente da convencional.

Para essa teoria, haveriam duas sociologias. Em um primeiro momento, aquela dos sociólogos, que trata do social como algo homogêneo; os vínculos sociais seriam a solução para uma série de questionamentos sobre a conexão existente entre uns e outros. A segunda alternativa trataria do social não se referindo às conexões homogêneas como aquelas variáveis que explicam os fenômenos observados; tais conexões, na verdade, seriam os problemas, não as soluções. Interessa compreender os questionamentos que trazem à tona a heterogeneidade daquilo que chamamos por social. Não há um todo homogêneo, mas várias associações (e aqui a teoria procura se aproximar mais do significado etimológico da palavra). A primeira seria a “sociologia do social” e a segunda a “sociologia das associações” (LATOUR, 2012, p.27).

As associações formam o que atribuímos como o social, e essas associações têm sim seus denominadores comuns, mas não podemos atribuir a um todo unitário chamado de sociedade tudo o que há de comum entre todos os grupos que a compõe, a teoria trata de evitar as generalizações e a buscar as reais conexões entre os atores, sejam eles humanos ou não-humanos (LATOUR, 2012, p.29-30).

Nesse sentido, não há só um simbolismo atrelado ao big data ou à tecnologia em si, mas os próprios caminhos que o objeto percorre são de interesse da ANT; como Latour (2012) explica, a ANT busca algo que vai à contra mão do que a maioria das teorias pós-modernas procuram trazer consigo: aqui não se busca a desconstrução de conceitos, a destruição de termos construídos e constituídos no passado para vigorarem no presente; na teoria do ator-rede, busca-se a reconstrução social daquilo que se tem como novo e impactante em relação à associação a qual pertence. Trata-se justamente de buscar novas instituições, novos procedimentos e novos conceitos que reagrupem o que se entende por big data, que reagrupem de maneira mais realista o quanto possível o aspecto sociológico atrelado ao termo.

O empirismo realista de que trata essa teoria busca interpretar quem são os atores e analisar o que eles fazem, analisar os processos que estão atribuídos a cada um ou que, ainda que não seja exclusividade deles uma ou outra coisa, qual a realidade da ação desses atores propriamente dita, como o movimento se dá e principalmente como cada outro ator corresponde a tal movimento na medida em que gera impacto para si e para a associação.

As lentes da ANT nos oferecem tais pressupostos. Como Latour explica que seriam eles desde a natureza dos grupos, ou seja, haver várias formas contraditórias de se atribuir identidade aos atores, como quando se trata deles serem humanos ou não-humanos por exemplo; ou até a natureza das ações dos mesmos, ou seja, como em cada curso de ação, há uma variedade de agentes que pode se deslocar dos objetivos originais; os movimentos são assim tão ou mais importantes do que os próprios atores em si.

Os objetos, as instituições que os atores representam, que eles participam, tanto a natureza dos atores quanto a natureza das instituições às quais pertencem importa para a rede, importa para a indicação do tom que os mesmos imputarão aos debates, às ideias, em todos os movimentos que serão conectados entre si. E, por fim, mas não menos importante, o pressuposto de que o vínculo das ciências naturais com a sociedade é sempre ávido de constante fonte de controvérsias, havendo assim o pressuposto da natureza dos fatos.

3. METODOLOGIA

Em pesquisas acadêmicas, geralmente o pesquisador tem duas opções que ditam o modo como os seus objetivos serão analisados e as observações serão feitas. As duas grandes vertentes em questão são a da pesquisa quantitativa e a da pesquisa qualitativa. A pesquisa quantitativa tem um foco maior em modelos numéricos, em levantamentos e modelos estatísticos. A pesquisa qualitativa, por sua vez, não se atém a números necessariamente, mas foca, muitas vezes, no caráter subjetivo do objeto pesquisado. Muitas vezes, busca abordar questões que vão se tornando claras no decorrer da própria investigação feita, partindo-se de uma dúvida inicial. Como preceituam Lara e Molina (2001), toda pesquisa precisa ter clareza, rigor e domínio de teorias e métodos a serem utilizados:

Para fazer pesquisa é necessário ter uma dúvida, um questionamento, uma pergunta. Fala-se do problema, o que se quer investigar? É a partir desta dúvida ou desta pergunta inicial, que parte do senso comum, que se procura a teoria e o método que fundamentarão a pesquisa. Parece simples colocar as coisas nestes termos, tenho um problema, procuro uma teoria, uma metodologia e está resolvido o meu projeto de pesquisa. Bom, é claro que o conhecimento científico não é elaborado de forma simplista, este entendimento seria temerário e porque não dizer ingênuo. Ao propor uma discussão de base científica são necessários: clareza, rigor, domínio de conceitos, teorias e métodos (LARA.; MOLINA. 2001, p.2).

Materiais textuais também podem ser fontes para análises. Segundo Bauer e Gaskell:

Os pesquisadores sociais têm a tendência de subestimar materiais textuais como dados. Os métodos de pesquisa passam por ciclos de moda e de esquecimento, mas a World Wide Web (www) e os arquivos on-line para jornais, programas de radio e televisão, criaram uma grande oportunidade para os dados em forma de textos. À medida que o esforço de coletar informações esta tendendo a zero, estamos assistindo a um renovado interesse na análise de conteúdo (AC) e em suas técnicas, em particular em técnicas como auxílio de computador (BAUER;GASKELL, 2003, p.190).

Com a redução do esforço para coletar informações é cabível falar aqui sobre as diferentes análises que podem ser feitas sobre um texto, um vídeo ou sobre uma música, um filme. Quando pesquisamos um pouco mais sobre a Análise de Conteúdo entendemos ser um termo atribuído a um determinado tipo de pesquisa que geralmente é feita cobrindo textos segundo um determinado lapso de tempo ou alguma outra característica que os textos tenham em comum e que, assim sendo, possam ser arquivados/agrupados em um mesmo compartimento por tal congruência/ semelhança de característica. Já o objetivo de uma Análise de Discurso geralmente é entender a contextualização, ou seja, entender aquilo que não está escrito, mas que permeia o texto todo, como por exemplo, como os jornais veiculam notícias em tempos de guerra e em tempos de paz? Quais as permissões de fala e linguagem

no século XX em contraste com as do século XXI quando analisamos a mídia das duas épocas?

A análise aqui proposta se preocupa com a opinião em relação a esses diferentes conteúdos e muitos outros, ou seja, em relação à opinião pública, às diferentes visões que permeavam à época e com os preconceitos, estereótipos, valores e sentimentos que cada texto traz em suas palavras. De acordo com Bauer e Gaskell:

Podemos distinguir dois objetivos básicos da análise de conteúdo ao refletir sobre a natureza tríplice da mediação simbólica: um símbolo representa o mundo; esta representação remete a uma fonte e faz apelo a um público (Buehler, 1934). Através da reconstrução de representações, os analistas de conteúdo inferem a expressão dos contextos, e o apelo através desses contextos. Se enfocarmos a fonte, o texto e um meio de expressão. Fonte e público são o contexto e o foco de inferência. Um corpus de texto e a representação e a expressão de uma comunidade que escreve. Sob esta luz, o resultado de uma AC e a variável dependente, a coisa a ser explicada. Textos atribuídos contêm registros de eventos, valores, regras e normas, entretenimento e traços do conflito e do argumento. A AC nos permite reconstruir indicadores e cosmovisões, valores, atitudes, opiniões, preconceitos e estereótipos e compará-los entre comunidades. Em outras palavras, a AC é pesquisa de opinião pública com outros meios. (BAUER,M.W.;GASKELL,G, 2003, p. 192)

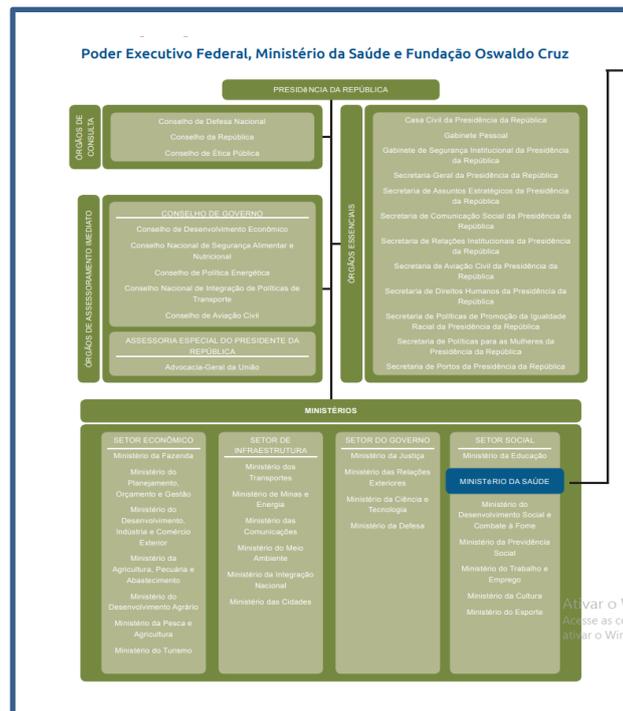
Tendo como objetivo, neste trabalho, a análise de processos sociotécnicos que envolvem o uso de Big Data pela FIOCRUZ, mais especificamente pelo Centro de Integração de Dados e Conhecimentos para Saúde (CIDACS), essa análise utilizará de uma das técnicas de pesquisa qualitativa: a análise de discurso.

Foram encontrados muitos documentos que dizem respeito às ações já implementadas pela Fiocruz e pelo CIDACS, inclusive, ações estas utilizando-se da tecnologia de Big Data, como também documentos que dizem respeito à formulação do que a Fiocruz e o CIDACS realizam. Com essas duas perspectivas se deu início aqui à análise do material disponível. Por meio dos relatórios e outros documentos, a pesquisa também foi acompanhada de entrevistas, utilizando-se como critério de escolha para as mesmas profissionais da Fiocruz e do CIDACS, funcionários que trabalham nesses institutos e que tenham conhecimento da tecnologia de grandes bases de dados que está sendo implementada. Com a entrevista realizada conseguiu-se apoio ao resultado pretendido no objetivo geral deste trabalho como também para se ter uma resposta mais satisfatória à pergunta norteadora utilizando-se de toda a metodologia aqui prevista e proposta.

4. ANÁLISE DE DADOS

Pelo o que podemos ver pela imagem extraída do Relatório Iciat 2013/2014, a primeira subdivisão importante é a de “ Setor Social”. Quando o governo estabelece setores como os de Infraestrutura, o Econômico e o “Do Governo” propriamente dito, isso corrobora com a ideia que temos do vínculo associativo entre os entes, ou seja, tomando por base a ANT temos que cada um desses setores irá determinar o “tipo” de associação que fará parte e consequentemente o tipo de objetivo que cada uma terá. Tendo em vista o Setor Social, que é justamente o setor no qual se insere nosso estudo, temos que Educação, Saúde, Previdência, Trabalho e Emprego, Cultura e Esporte são assim “associados”, ou seja, pertencem à uma rede (*network*) na qual são atores que, dependendo de suas ações, podem gerar impacto bastante expressivo.

Figura 1: Organograma Institucional do Ministério da Saúde



Fonte : Relatório Iciat 2013-2014

Como Latour explica, quando determinamos o conjunto que compõe uma associação, não podemos apenas analisar isoladamente cada um dos atores, como se a mera presença deles ali definisse aquela associação como tal, mas são os movimentos, as interações

propriamente ditas que geram essa definição, que permitem que se chegue próximo. Isso nos retoma à passagem em que o autor faz alusão a um supermercado:

[..] Retomando a metáfora do supermercado, chamaremos de social não uma gondola ou ala específica, mas as várias modificações feitas no lugar para exibir os produtos – embala-los, etiqueta-los, colocar-lhes preço – porque essas pequenas alterações revelam ao observador quais combinações novas foram exploradas e que caminhos serão seguidos (aquilo que, mais tarde, definiremos como “rede”)
(LATOUR. B, 2012, p. 99).

Logo, essas alterações, a definição e redefinição de competências, as razões das causas, as micropráticas e relações de poder que irão compor o social são as questões que importam para entender os caminhos que o governo e, conseqüentemente, a população irão seguir; haverá, assim, a compreensão das associações feitas que explicam, por sua vez, o social.

A especificidade que se dá em relação ao CIDACS é justamente o fato de ter sido uma instituição criada para utilizar, tratar e analisar grandes bases de dados. São missões do CIDACS conduzir estudos e pesquisas desenvolvendo novos modos de se investigar, além de promover treinamento científico e profissional baseado em projetos interdisciplinares e envolvendo a integração de grandes bases de dados relacionadas a big data. Também a promoção da integração do conhecimento em níveis distintos, chamados por níveis populacional, individual e sub-individual para poder aprimorar as Ciências da Saúde, como a Medicina e a Biologia por exemplo.

Mais ainda, quanto ao conhecimento, é missão do centro garantir um ambiente seguro e de alta performance para que se produza um conhecimento inovador, um conhecimento que se traduza em instrumento útil para a tomada de decisão em políticas públicas. Para essa missão são eleitas ainda parcerias nacionais e internacionais a fim de que se favoreça e estimule a produção na área de saúde e principalmente a produção científica para o SUS.

4.1 A produção de big data no CIDACS : A “Plataforma de Dados” e a Produção dos dados

O instrumento utilizado pelo CIDACS para tratamento de big data é a “Plataforma de Dados”⁸. Por meio dessa plataforma são feitos todos os procedimentos para o armazenamento, o processamento, a preservação e a disponibilização dos dados que são recebidos do governo brasileiro, além da Produção e da Curadoria. Os dados que são encaminhados para o centro

⁸ Portal Institucional Cidacs. Disponível em: <<https://cidacs.bahia.fiocruz.br/plataforma-de-dados/>>.

são dados administrativos relacionados à Saúde e a Programas Sociais; uma vez recebidos esses dados são então encaminhados para uma fase de produção.

É na fase da Produção que os dados vão receber valor, ou seja, vão ser “tratados” para a eles serem atribuídos diversos processos e funções que vão torná-los úteis para os pesquisadores responderem suas questões científicas. Os procedimentos realizados nessa fase são os de limpeza, padronização e harmonização para que os dados sejam devidamente tratados e sirvam de fato para as pesquisas. Depois de terem recebido tratamento, os dados são então agrupados em *datasets*.

*Dataset*⁹ é, por definição, uma representação de dados residentes na memória que fornecem um modelo de programação relacional consistente e de maneira independente em relação à sua origem, ou seja, em relação à origem dos dados contidos nele. Ele representa um conjunto completo de dados aonde estão contidas tabelas que contêm, que pedem e que restringem os dados, tornando-os mais específicos (Microsoft, 2017).

Na produção feita pelo CIDACS, os dados recebidos tanto de Programas Sociais como de áreas relacionadas à Saúde são agrupados nesses *datasets*; cada grupo recebe um nome segundo a fonte dos dados ali agrupados; se os dados vieram do programa Bolsa Família, por exemplo, logo serão agrupados em um *dataset* denominado “Bolsa Família”. Se vierem do Minha Casa Minha Vida, serão então agrupados no *dataset* “MCMV- Minha Vida” e assim por diante. Esses *datasets* são chamados também de “bases de dados”.

Dois dos tratamentos realizados aos dados durante a Produção são o *data linkage* e o *record linkage*. A partir desses dois procedimentos, as informações que estão em bases de dados diferentes, mas que tratam de um mesmo indivíduo, são então integradas, ou seja, são qualificadas não tão somente quanto ao tipo de informação, mas também sobre de quem se trata. Se ambas as informações tratarem de um mesmo indivíduo, elas serão integradas para esse indivíduo. Posteriormente são calculadas as informações similares, seja por uma identificação única (utilizando o Número de Identificação Social das pessoas nos cadastros sociais, por exemplo) ou por probabilidades utilizando-se o nome da mãe, a data de nascimento e o próprio nome da pessoa.

4.2 A Curadoria

Aqui ocorre um maior comprometimento com a segurança das informações disponíveis. Na curadoria, os dados são recebidos de suas bases de dados originais e depois

⁹ Portal de Informações da Microsoft. *DataSets, DataTables e DataViews*. Disponível em: <https://docs.microsoft.com/pt-br/dotnet/framework/data/adonet/dataset-datatable-dataview/>

são criptografados para que se busque manter um sigilo das informações ali contidas. Ainda para garantir o sigilo, esses dados são levados para o armazenamento em um disco externo especializado que implementa a criptografia feita, a criptografia nativa, com um código de acesso que é de conhecimento apenas do portador do disco.

O disco externo então é entregue ao responsável pela Curadoria Digital do Centro. É gerado um termo acusando o recebimento desse disco para, depois, leva-lo a um cofre com proteção contra arrombamento, incêndio e inundação. Esse ambiente aonde o disco fica guardado e sendo monitorado é denominado de Sala Segura.

Pode se observar aqui, já na Curadoria, que as bases de dados recebem tratamento em relação à segurança da informação. Se não bastasse já a criptografia feita para garantir a segurança, a mesma é ainda reforçada para que depois as informações coletadas sejam enviadas para um ambiente propício para a segurança.

Pela ótica da ANT, da teoria do ator-rede, podemos até mesmo classificar essa Sala Segura como um *actante* não-humano. Desde a produção dos dados, passando pela criptografia, curadoria e depois armazenamento, temos uma interação de processos anteriores ao armazenamento propriamente dito nessa sala que se submetem à tecnologia contida ali, na Sala Segura.

Nota-se, então, um dos pontos que Latour discute sobre os efeitos que os *actantes* geram na rede e como podemos aprender não tão somente o que eles fazem como também compreender os motivos que levam determinadas coisas a serem feitas. A Sala Segura acaba sendo uma resposta à segurança exigida em relação aos dados ali contidos no centro. Para que as pesquisas e os trabalhos do centro ocorram sem maiores transtornos ou problemas, ou seja, para que eles tenham certeza da segurança e proteção dos dados pessoais, as leis e os limites que impõem sobre o que pode ser feito com as informações pessoais acabam sendo *actantes* também não-humanos que exercem papel de suma importância na rede.

4.3 A privacidade e a ética quanto ao tratamento dos dados

Em relação à ética ao tratamento desses dados e à privacidade dos indivíduos, o CIDACS enuncia que:

[...] a plataforma de dados emprega alto nível de segurança na produção de dados para garantir a confiabilidade, integridade e disponibilidades dos dados, além de garantir a privacidade e o uso ético dos dados.¹⁰

Quando observamos o que o CIDACS enuncia para as bases legais que temos, podemos correlacionar o que acima está escrito com o que temos no Marco Civil da Internet (MCI). No próprio Marco Civil (lei 12.965/2014), temos o estabelecimento de limites de até onde os dados podem ser analisados; no Capítulo 2 da lei, que trata dos “Direitos e Garantias dos Usuários”, a definição dos tipos de inviolabilidade e a forma como as informações devem ser utilizadas por aqueles que as detém, acaba tendo relação direta com as práticas do Cidacs com big data. A Lei do Acesso à Informação (LAI), por exemplo, possui um artigo que permite o uso de dados pessoais para fins de pesquisa:

Art. 31. O tratamento das informações pessoais deve ser feito de forma transparente e com respeito à intimidade, vida privada, honra e imagem das pessoas, bem como às liberdades e garantias individuais.

§ 1º As informações pessoais, a que se refere este artigo, relativas à intimidade, vida privada, honra e imagem: II - poderão ter autorizada sua divulgação ou acesso por terceiros diante de previsão legal ou consentimento expresso da pessoa a que elas se referirem.

§ 3º O consentimento referido no inciso II do § 1º não será exigido quando as informações forem necessárias: II - à realização de estatísticas e pesquisas científicas de evidente interesse público ou geral, previstos em lei, sendo vedada a identificação da pessoa a que as informações se referirem;¹¹

As leis exercem papel fundamental em torno das diretrizes dos trabalhos relacionados ao acesso à informação e ao tratamento que se dá às tecnologias de informação. As leis estabelecem um tipo de hierarquia, uma relação sobre o outro daquilo que simplesmente deve ser cumprido, que está sendo regulamentado para o bem não tão somente do próprio indivíduo, mas também da sociedade.

Quando pensamos em leis, pensamos em relações de subordinação a regras que mostram como as coisas devem ser. Entretanto, quando vamos além da interpretação legal, ou seja, em sentido mais comum e que já está posto para todas as leis, podemos observar que as leis podem ser sim *actantes* que não simplesmente criam uma hierarquia entre o que nelas se

¹⁰ Portal Institucional do CIDACS. Acesso em 2018 (Disponível em: <<https://cidacs.bahia.fiocruz.br/plataforma-de-dados/>>).

¹¹ BRASIL. LEI Nº 12.527, DE 18 DE NOVEMBRO DE 2011. Lei de Acesso à Informação, Brasília, DF; novembro de 2011. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/112527.htm>. Acesso em: 12 jun. 2018.

está escrito e o que na realidade está posta como é o caso da relação entre big data, conjunto de tecnologias de informação do centro e o que está escrito por exemplo no MCI, mas também, além de estabelecer hierarquia, ajudam nos resultados finalísticos, ajudam nos produtos do social para permitir o bem estar social a partir da regulamentação e do oferecimento de segurança jurídica.

Ainda que no MCI exista um artigo que enuncie os tipos de inviolabilidade para garantir os direitos dos usuários da internet, esses tipos de inviolabilidade traduzem também o pensamento que já estava posto na relação de big data com a sociedade antes do surgimento da lei propriamente dita. Da mesma forma, já havia um pensamento posto previamente para o CIDACS, apontando os limites do que poderia ser acessado em relação a dados pessoais e os dilemas éticos em torno disto. Esse pensamento do social seria fruto de visões distintas, ou seja, visão de diferentes categorias relacionadas ao tema e que muitas vezes são visões conflitantes, porém visões que se colidem, se ajustam para futuramente serem transformadas no pensamento no qual se constituirá a lei que, no caso, felizmente constituiu o MCI, uma conquista da sociedade civil.

A lei afirma que são direitos dos usuários a inviolabilidade da intimidade e da vida privada, com a consequente indenização por dano material ou moral que ocorram por conta da sua violação; a inviolabilidade e sigilo tanto dos fluxos das comunicações pela internet dos usuários quanto das comunicações privadas armazenadas, salvo por ordem judicial em relação às duas inviolabilidades. Ao mesmo tempo em que isto está posto, sobre a inviolabilidade e a segurança que devem ser dadas às informações disponíveis sobre os usuários, temos na própria lei, ainda no próprio capítulo, o inciso V que trata do fornecimento de dados pessoais a terceiros. No inciso está que não é permitido o fornecimento a terceiros de dados pessoais, incluindo os registros de conexões feitas e do acesso a aplicações de internet, salvo mediante o livre consentimento, expresso e informado, ou ainda nas hipóteses já previstas em lei, como posto:

VII - não fornecimento a terceiros de seus dados pessoais, inclusive registros de conexão, e de acesso a aplicações de internet, salvo mediante **consentimento livre, expresso e informado** ou nas hipóteses previstas em lei;

VIII - informações claras e completas sobre coleta, uso, armazenamento, tratamento e proteção de seus dados pessoais, que somente poderão ser utilizados para finalidades que:

a) justifiquem sua coleta; b) não sejam vedadas pela legislação; e c) estejam especificadas nos contratos de prestação de serviços ou em termos de uso de aplicações de internet;

Pelo que pode ser observado, em ambos os incisos e nas alíneas, não há uma vedação absoluta em relação ao uso dos dados pessoais, das informações dos indivíduos, mas há uma permissão desde que se cumpra com certos requisitos. Ainda assim, a interpretação não é tão objetiva, na medida em que o poder em relação ao uso deve atender a “finalidades que justifiquem sua coleta” e que “não sejam vedadas pela legislação”, subentendendo-se assim que seriam aquelas que não conflitam diretamente com o que está posto nas normas.

O MCI e a LAI configuram-se assim como *actantes* não-humano. Se observarmos com atenção, vemos que há permissões sobre uso, tratamento dos dados pessoais e das informações. O que se enuncia sobre produção e curadoria dos dados disponibilizados para o CIDACS, além também de outros objetos e também *actantes* – como a Sala Segura – é contemplado pela base legal que surge após o que já anteriormente foi posto nos documentos do Centro. Tais leis apresentam o que na ANT se entende por Simetria, que é a ausência de hierarquia entre atores, os atores não podendo ser hierarquizáveis, mas possuindo as mesmas possibilidades de interferirem de igual modo sobre a questão¹²; ou seja, no nosso caso, sobre como deve ocorrer o tratamento dos dados, do que seria ético e em respeito às normas legais e do que não seria. Há uma mediação entre esses *actantes*, tanto o que está posto nas leis quanto ao que está posto no CIDACS, há uma intermediação daquilo que pode ser permitido ao Centro fazer até mesmo em relação a terceiros, utilizando-se dos dados pessoais desde que seja de acordo com as hipóteses permitidas em lei, hipóteses essas que contemplam o que o Centro já faz, não havendo, assim, uma vedação absoluta do que se pode fazer com os dados pessoais de milhões de pessoas cadastradas nos programas governamentais, mas sim uma permissão em observando-se os usos e tratamentos que já estavam sendo realizados ao tempo de tais leis.

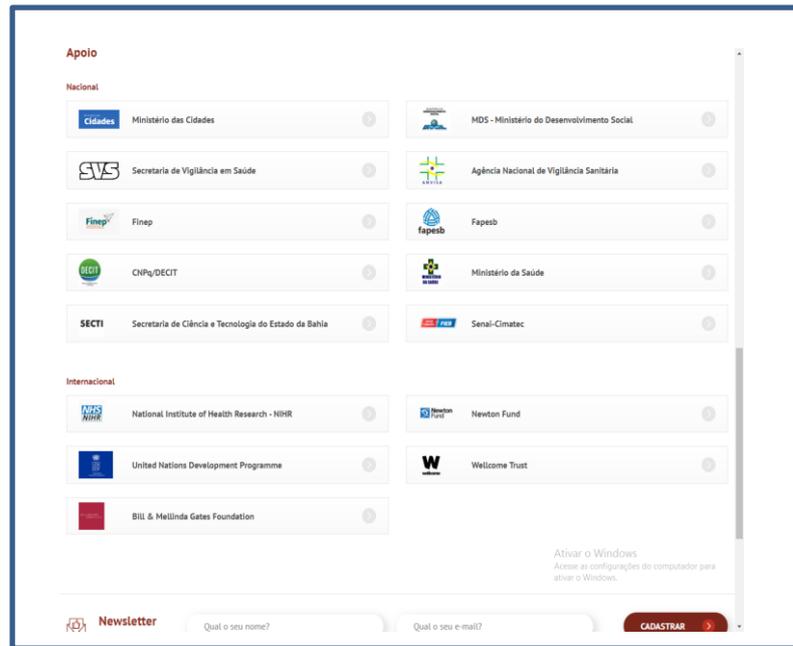
4.4 Links entre actantes

Além da simetria, temos também o conceito de tradução. Pela ANT, tradução seriam os links propriamente ditos entre *actantes*, ou seja, remete à ideia de comunicação, de contato ou até mesmo da criação de novos contatos que antes não existiam, mas que agora passam a existir dada a associação que ali está configurada, que ali está posta. A criação de um novo link, de um novo contato entre *actantes* faz com que haja modificações em todos os agentes e que gera associações que podem ser observadas.

¹² CAVALCANTE, R.. et al. *A Teoria Ator-Rede como Referencial Teórico-Metodológico em pesquisas em Saúde e Enfermagem*. Minas Gerais: Universidade Federal de São João Del Rei, 2017.

No caso do Cidacs, isso não é diferente. O centro atualmente possui links com instituições nacionais e internacionais como, por exemplo, a *Bill & Melinda Gates Foundation*, o *United Nations Development Programme*, dentre outros que coadunam com a ideia do centro quanto ao tratamento de dados e tratamento da saúde pública.

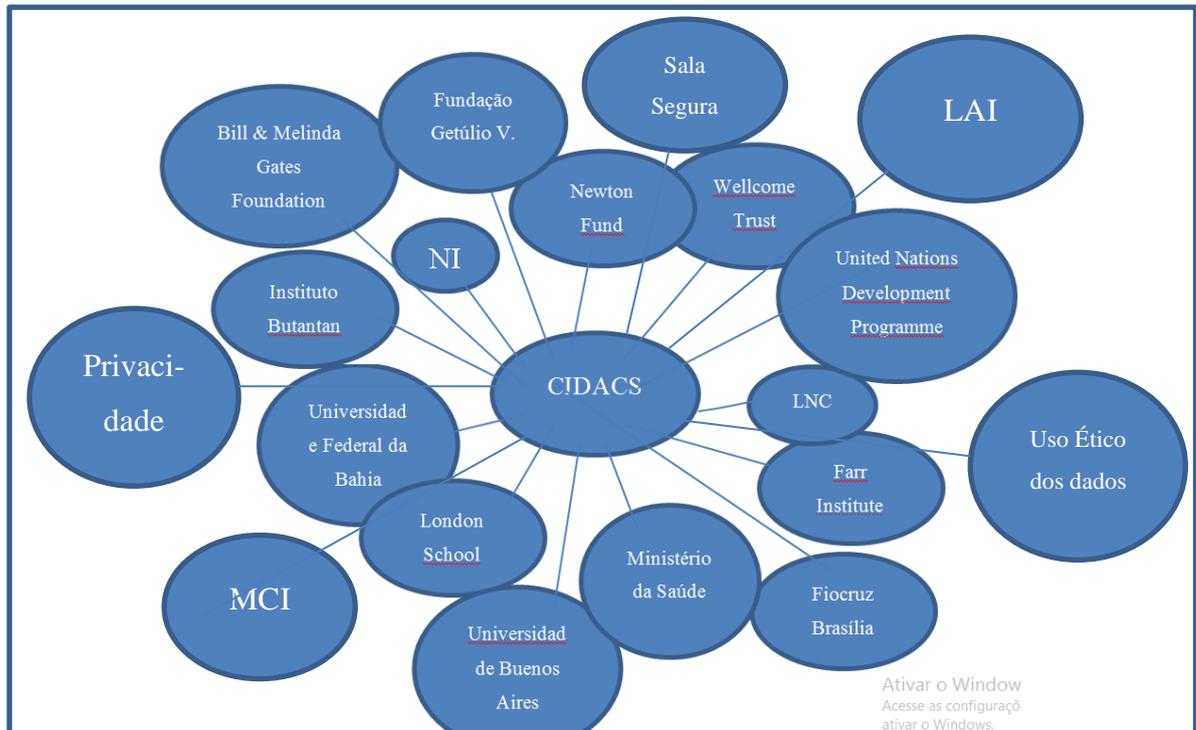
Figura 2: Lista de Colaboradores e Apoiadores do Cidacs



Fonte: Portal Institucional do Cidacs

Podemos compreender esses colaboradores como actantes também para compor uma rede associativa na qual o Cidacs estaria inserido:

Figura 3: Desenho de Rede de Atores Institucionais e *actantes* não-humanos correlacionados ao Cidacs



Fonte: Próprio Autor

Esta seria então, como podemos ver pela figura, uma maneira de entendermos o linkage entre as instituições, tanto nacionais como internacionais, que traduzem a colaboração e apoio ao Cidacs.

As linhas não são apenas linhas de conexão entre os atores pela teoria do ator-rede, mas são de fato movimentos que decorrem do vínculo entre esses atores. A associação entre o Cidacs e a Fiocruz Brasília, em um primeiro momento, poderia até ser mais previsível do que uma associação internacional como a da Fundação Bill & Melinda Gates, por exemplo. Por ser uma instituição internacional, poderíamos inferir que haveria uma maior dificuldade de conexão e de movimento pendente para o Cidacs ou do Cidacs para a fundação internacional. Só que, como a teoria indica, não podemos inferir isso logo tão cedo, o mais correto é buscarmos entender a complexidade desse movimento, ou seja, não tentarmos descrever quais são os atores e nos restringirmos apenas a eles, a multiplicidade pode ser quase que infinita ainda que com um tema restrito.

O que vem ao cuidado aqui é que, por meio da ilustração, busca-se entender os apoios e as colaborações feitas, tanto nacional quanto internacionalmente. Não se trata de tentar

entender tudo o que concerne a todas as redes das quais o Cidacs acaba fazendo parte. A teoria do ator-rede, ao nos indicar que o ávido pesquisador deve buscar não as certezas mas as incertezas das relações, aponta as controvérsias que podem existir nos movimentos entre atores.

Para a compreensão de todas as controvérsias entre os mais diferentes actantes humanos e não-humanos que possam existir, desenvolve-se a ideia da Cartografia de Controvérsias como forma de análise dos diferentes interesses envolvidos com o tema em questão.

4.5 Cartografia de Controvérsias: o comum e o diferente entre os atores para as políticas públicas

A Cartografia de Controvérsias, como assim é explicado por Cavalcante et al. (2017, p. 5), seria de fato a verdadeira operacionalização da teoria do ator-rede. É a consideração das técnicas para se explorar e visualizar os problemas, as controvérsias existentes e que levam em conta o debate social, os dilemas éticos em torno dos problemas técnicos e científicos que se tem à disposição. A observação deve não ser estática, mas sim uma observação que acompanhe o movimento dos atores e relações, ou seja, que acompanhe tudo aquilo que de certo modo indique mudança, transição de um pensamento, de uma concepção que se tem acerca de uma controvérsia latente, pujante e que está ali, está para poder ser tratada pelo pesquisador considerando os pontos de vista e mudanças que os cercam.

Uma controvérsia possível entre os atores seriam os dilemas éticos sobre o uso de dados pessoais, ou seja, um debate social que é importante e que tem relação com questões como a proteção que é feita dos dados pessoais dos usuários e as limitações que existem até para o centro em sua liberdade de tratar aquilo que armazena.

Como é assinalado ainda por Cavalcante et al. (2017, p. 5), o que se observa é um jogo pelo poder, disputas políticas que envolvem razões diversas, muitas vezes contraditórias em um primeiro momento (por representar grupos de interesse distintos). Nessa contradição, pode-se evidenciar o que mais tarde irá se configurar em desdobramentos e negociações para o alcance de um objetivo comum; o resultado não é algo neutro, mas algo que advém desses embates.

Para a realização da cartografia foi acompanhado um evento de suma importância ocorrido na data de 22 de maio de 2018. Foi o seminário conjunto realizado na Câmara dos Deputados sobre “Proteção dos Dados Pessoais”. Para a cartografia, faz-se necessária a

determinação geográfica da rede, ou seja, os principais actantes, os verdadeiros porta-vozes, no caso em questão, do Cidacs e os porta vozes do direito relacionado à privacidade e aos dados pessoais. O evento realizado na Câmara dos Deputados teve como premissa a busca por uma convergência de ideias e interesses sobre o tema. Os participantes convidados foram Thiago Camargo Lopes, secretário de Políticas Digitais do Ministério da Ciência, Tecnologia, Inovações e Comunicações (Mctic), Torquato Jardim, Ministro de Estado da Justiça; Luiz Fux, Presidente do Tribunal Superior Eleitoral (TSE); Deputado Goulart, Presidente da Comissão de Ciência e Tecnologia, Comunicação e Informática da Câmara dos Deputados; Deputada Bruna Furlan, Presidente da Comissão Especial do PL 4.060/12- Tratamento e Proteção de Dados Pessoais. Para o Painel I, que tratava da “Abordagem regulatória para o tratamento de dados pessoais”, houve a participação dos seguintes palestrantes:

Figura 4: Lista de palestrantes do evento “ Abordagem Regulatória para o tratamento de dados pessoais”

Participantes	Nome	Função
	Thiago Camargo Lopes	Secretário de Políticas Digitais do Ministério da Ciência, Tecnologia, Inovações e Comunicações (Mctic)
	Luis Felipe Salin Monteiro	Secretário de Tecnologia da Informação e Comunicação do Ministério do Planejamento, Desenvolvimento e Gestão
	Frederico Meiberg Ceroy	Promotor de Justiça e coordenador da Comissão de Proteção dos Dados Pessoais do Ministério Público do Distrito Federal e Territórios (MPDFT)
	Ana Carolina Pinto Caram Guimarães	Diretora do Departamento de Proteção e Defesa do Consumidor (DPDC) Secretaria Nacional do Consumidor Ministério da Justiça (Senacon/MJ)
	Demi Getschko	Diretor-presidente do Núcleo de Informação e Coordenação do Ponto BR (NIC.br)
	Rafael Augusto Ferreira Zanatta	Advogado do Instituto Brasileiro de Defesa do Consumidor (Idec)
	Bruno Gencarelli	Diretor da Unidade Internacional de Proteção de Fluxos de Dados da Comissão Europeia (DG Justiça e Consumidores)
	Mauricio Barreto	Membro da Academia Brasileira de Ciências (ABC) e coordenador do Centro de Integração de Dados e Conhecimentos para a Saúde (Cidacs) da Fundação Oswaldo Cruz (Fiocruz)

Fonte: Próprio Autor

De antemão, frisa-se aqui a participação importantíssima de Maurício Barreto que é coordenador do CIDACS e um dos principais motivos para a realização da análise acerca desta palestra justamente por conta do objeto deste trabalho como também pela própria participação em si que, além de muito profícua, nos ajuda a entender não só seu posicionamento como também do próprio CIDACS em relação ao tratamento dos dados pessoais.

Primeiro trataremos da palestra em si e da ordem das falas. No início, o primeiro a falar foi o Sr. Thiago Camargo, Secretário de Políticas Digitais. O mesmo asseverou que agora é o momento para tratar de um assunto que ainda é muito novo: a exposição do consumidor/usuário ao ambiente digital. Os desafios referem-se à regulação e à legislação para a internet. Como até pouco tempo não se tinha ideia do *linkage* que poderia ser feito sobre informação de inúmeros indivíduos, o secretário defende que a inovação anda justamente na contramão do que as leis pregam, ou seja, legislar vai de encontro justamente com a ideia de se inovar e de se quebrar regras fixas sobre como as coisas devem ser, configurando-se assim, para o palestrante, em um dos maiores desafios – senão o maior – dessa nova geração. Implícita na fala do Secretário fica a ideia liberal de que o mercado deve ser livre para agir, sem que haja qualquer tipo de coação (como as leis) às suas práticas visando à inovação.

Já o Sr. Luis Felipe Salin, corrobora com essa visão liberal que aponta a necessidade de a inovação existente ser sim preservada. Salin explica que dado é o novo petróleo, ou seja, seria o novo bem essencial para o movimento da economia. Salin assevera que, por meio do uso de dados pessoais oferecidos pelo próprio cidadão, o custo de transações pode ser reduzido em até 97%, tornando-se assim de uso quase que obrigatório.

É dito que o Brasil é o 4º país do mundo em termos de uso de internet com base na quantidade de usuários conectados à rede. O assunto de proteção de dados pessoais já é exigido pelo governo. Salin explica como normas legais, como a Constituição Federal, o Código de Defesa do Consumidor e o Marco Civil da Internet são exemplos nítidos da preocupação acerca do tema.

Salin diz ainda que, a partir de julho, será lançado o Documento Nacional de Identificação (DNI), sistema que abarcará todos os documentos dos cidadãos. Haverá um agrupamento de todas as bases de dados que contemplam as documentações dos cidadãos e, na palma da mão (demonstra fazendo gesto com o próprio celular), os brasileiros poderão acessar todos os seus documentos via smartphones.

A única mulher na mesa, Sra. Ana Carolina Guimarães – Diretora do Departamento de Proteção e Defesa do Consumidor (DPDC) – explicou, primeiro, a sensibilidade do tema. O olhar, como a mesma diz, deve ser focado na defesa do consumidor. Contudo, esse olhar deve ir além disso e passar por direito à liberdade, à privacidade, direitos que influenciam a escolha do cidadão, direitos fundamentais. O consumidor, como é lembrado pela Sra. Guimarães, é a parte hipossuficiente na relação; ele, em diversos momentos, pode não ter ideia do que está acontecendo; ele é vulnerável e, por isso, precisa de uma proteção maior. O tratamento dos dados não pode ser utilizado distante da harmonia com as regulações que temos sobre defesa do consumidor. É citado o caso do Facebook e do vazamento dos dados. A Sra. Ana Carolina assevera que esse vazamento deve também ser analisado e lembrado aqui para ser tratado de modo que o mesmo não ocorra no Brasil, que campanhas eleitorais não sejam influenciadas por uso indiscriminado de dados, sem estar de acordo com a lei. Observa-se a preocupação da Diretora com a preservação dos direitos dos cidadãos, sem foco no livre mercado e seus interesses, percepção central nas observações dos dois primeiros atores analisados.

Mais próximo do fim da palestra foi passada a palavra para o pesquisador Dr. Maurício Barreto, pesquisador de suma importância para este trabalho, pois o mesmo é o próprio coordenador do Cidacs. É arguida a ideia de que há uma importância e um crescente consenso de que, para que se conheça cada vez mais os problemas relacionados à saúde não só no Brasil mas no mundo todo, faz-se necessário o uso de dados.

Os dados pessoais para pesquisa e estudos são importantes para o desenvolvimento econômico, ambiental, social e de saúde; as inovações emergem de uma dependência do uso de dados, agora massivos. O Sr. Maurício continua explicando que o processamento adequado é uma fonte de evidências sólidas para formuladores de políticas refinarem suas ações e lidarem com temas complexos. A adequada análise faz emergir soluções originais para muitos dos problemas existentes e para os quais a ciência convencional não tem apresentado soluções. Mais uma vez, nova perspectiva desponta. A percepção do pesquisador volta-se à preocupação com o uso adequado dos dados pessoais para incrementar programas governamentais e políticas públicas. Não menciona direitos e nem o papel da inovação para o mercado.

Observa-se a necessidade de avançar a base de conhecimento para que de fato hajam melhorias na sociedade. Para a área de saúde, os dados, não só os qualificados como de saúde, mas também os dados pessoais, tem sido úteis em pesquisas e práticas sobre saúde pública e epidemiologia; o esforço tem sido expresso nas denominadas medicina e prevenção

expressas com precisão; através da pesquisa e desenvolvimento com os dados pessoais, a precisão na aplicação de vacinas e remédios para a população tem sido efetiva.

Independente do avanço tecnológico exposto, os direitos previstos em lei devem ser respeitados, inclusive os da privacidade, evitando e impedindo o tratamento abusivo de dados pessoais. No que tange às pesquisas, os dados administrativos são preferíveis aos dados populacionais, principalmente porque as amostras passivas que acabam influenciando a capacidade de se conhecer detalhes ou eventos pouco frequentes – que não seriam possíveis em dados de inquéritos ou amostras pouco reduzíveis, por exemplo – podem ser conhecidas a partir dos dados administrativos.

Foram identificadas, nesse evento, visões bastante distintas sobre a posição que deve se ter em relação ao uso das novas tecnologias de dados pessoais. Pelo posicionamento do Sr. Thiago Camargo, observa-se a existência de grupos de interesse que podem tanto ver o Marco Civil e outras futuras proposições legais como barreiras ao ato de inovar em si e ao livre mercado, visão que aborda justamente a questão do empreendimento que se faz e que futuramente se pode ainda fazer, sem qualquer preocupação com os direitos dos cidadãos. É uma posição que se assemelha ao posicionamento do Sr. Salin ao reafirmar o valor econômico dos dados como “novo petróleo”. Há um *link* entre esses dois atores pelo menos com base em suas falas.

Pelos discursos da palestra, proferidos pelos atores que foram levados à Câmara para discutirem o tema, percebe-se desde o princípio uma busca por uma certa reafirmação da tecnologia de uso de dados como uma realidade intrínseca ao próprio desenvolvimento econômico do país, como se fosse um estado de coisas natural e não construído socialmente. Com isso, fecham-se as portas para as possibilidades de questionamentos a respeito desse estado de coisas. É inegável que dados tem sido usados para a obtenção de valor econômico por grandes empresas e corporações, desconsiderando valores morais e éticos em torno do tema e que, curiosamente, não foram abordados desde o início do palestra, mas da metade para o fim, pelos próximos outros palestrantes.

Ao terminar sua fala dizendo sobre o DNI, e como todos os brasileiros com acesso a smartphones poderiam acessar todos os seus dados por meio apenas de um dispositivo, novamente é ilustrado tal nova tecnologia como um avanço, uma quebra de antigos paradigmas existentes em relação até mesmo à burocracia que se fazia valer para tal função; pouco é dito em relação às implicações que, por exemplo, um ataque hacker poderia gerar nessas mesmas bases de dados em conjunto, agrupadas e com todos os dados de praticamente todos aqueles ali cadastrados; as inovações e as maneiras de se gerir os dados pessoais ou até

mesmo outros dados do governo já se mostraram falhas inúmeras vezes, sendo que tais falhas pouco foram apontadas de início por aqueles que estavam na palestra.

Mais tarde, pela voz da Sr^a Ana Carolina Guimarães, pode-se entender melhor as implicações morais e éticas relacionadas ao tema. Curiosamente a única mulher a compor a mesa (isso até foi apontado por um dos palestrantes), a diretora buscou explicar algo que ainda não tinha sido tocado, um tema latente que é justamente a questão do consumo, do direito ao consumo e conseqüentemente do direito do consumidor, de sua liberdade e privacidade. Ela cita o caso do Facebook, buscando referenciar um cuidado que deveria se ter futuramente com os dados pessoais, com a cautela e regulação pertinentes para tanto. Vale lembrar que, nesse caso, uma rede transnacional foi atingida; muitos brasileiros também tiveram seus dados violados, o que faz valer a ideia de que a ameaça a direitos que estão em pauta não é só futura, já é presente.

Não cabe aqui alegar que um ou outro palestrante poderia vir a desconhecer as possibilidades de invasão de privacidade de dados pessoais advinda de outros países, como por exemplo os Estados Unidos da América, mas é notável no decorrer da palestra como isso é algo ainda muito recente, muito novo para o conhecimento daqueles presentes na casa legislativa. A ideia de invasões virtuais advindas de outros países e que afetem o Brasil é algo que teve como fatos mais importantes casos ainda recentes e que talvez explique por que o direito internacional foi pouco explorado na palestra.

Já o discurso do cientista, Sr. Maurício Barreto, deixou clara a importância de big data para as políticas públicas e para a pesquisa e desenvolvimento como um todo. Ao chamar uma certa antiga e conhecida ciência como “ciência convencional”, nota-se o quão além o uso de dados chegou para o âmbito científico; fala-se agora em uma nova ciência, uma que já está sendo bastante explorada e que foge, escapa de convenções antigas, de paradigmas científicos mais usuais e tradicionais.

Além da referência ao valor científico que possuem tais dados, o sr. Maurício ainda traz em seu discurso um detalhe bastante curioso e interessante, justamente em relação aos tipos de dados e à tecnologia brasileira para o seu uso. Há uma certa dificuldade ainda no uso de dados populacionais com a mesma liberdade, capacidade que se tem para o uso de dados administrativos. Logo, não são todos os tipos de dados que são simplesmente usáveis da maneira que se queira utilizar, dependendo do tipo, certos dados são mais difíceis de serem traduzidos em usos efetivos do que outros o que reverbera com a ideia que se produz aqui, justamente do pouco conhecimento do alcance que se têm justamente quanto às possibilidades do conjunto de tecnologias que afetam os direitos de cada cidadão, assustadoramente, soa

como se até mesmo o uso pleno de dados populacionais, por exemplo, fosse apenas um etapa que eventualmente e inevitavelmente viesse a ser atingida, tendo logo o alcance do uso dos dados sendo praticamente infinito, de difícil determinação.

4.7 Percepção dos pesquisadores sobre o uso de big data

O CIDACS promove como objetivos políticos e missões que envolvem também em seu cerne a temática de big data. É no âmbito de questões relacionadas à tecnologia e à pesquisa e desenvolvimento que são direcionados os seus objetivos. São objetivos do CIDACS:

I-Conduzir estudos e pesquisa, desenvolver novas metodologias investigativas e promover treinamento científico e profissional baseado em projetos interdisciplinares e envolvendo a integração de grandes bases de dados ("big data"); II-Promover a integração de conhecimento em diferentes níveis (populacional, individual e sub-individual), com o propósito de engrandecer o escopo das Ciências da Saúde; III- Com o auxílio do conhecimento disponível e recursos computacionais de alta performance em um ambiente seguro, contribuir para a produção de conhecimento inovativo para alargar o entendimento dos problemas de saúde da população, tal qual suportar a tomada de decisões em políticas públicas, para o benefício da sociedade.¹³

Tendo como ponto de partida da análise os objetivos do CIDACS e os dilemas éticos que influenciam toda a integração das bases de dados que estão disponíveis para os diferentes programas governamentais em seus diferentes níveis, buscou-se, na entrevista realizada¹⁴, questionar o pesquisador acerca das questões pertinentes ao tema, não tão somente a maneira de trabalho do próprio centro, mas também, e mais ainda, quanto ao objetivo, segundo os aspectos que se configuram em dilemas ou não para a integração que é feita.

A primeira questão levantada foi quanto ao uso de big data. Foi questionado se big data era usado para fins de pesquisa e se esse mesmo uso era de fato importante. O entrevistado então afirma seu uso e importância. De acordo com o pesquisador, não é uma tecnologia meramente alegórica, mas importante e útil:

Eu uso big data para minha pesquisa; big data é importante para minha pesquisa e para qualquer pesquisa hoje que venha a ser realizada principalmente na área da saúde. Nós temos hoje grandes bases de dados que podem ser *linkadas* com diferentes bases de dados de diferentes políticas públicas e também com dados não estruturados que possam ser coletados, capturados, na internet; temos feito muitos trabalhos com essa lógica de organização de dados e, portanto, big data tem sido uma ferramenta importante para a gente.

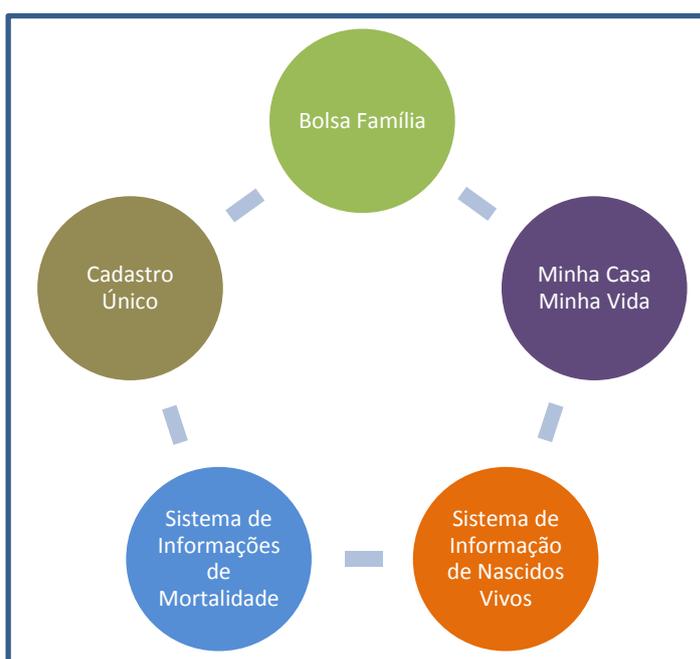
¹³ Portal Institucional do CIDACS: Quem Somos. Disponível em :<https://cidacs.bahia.fiocruz.br/sobre/quem-somos/>. Acesso em 2018.

¹⁴ Houve tentativa de realização de mais duas entrevistas, porém não houve respostas aos convites enviados por e-mail.

Corroborando o que já vem sendo analisado aqui, big data é de suma importância para os trabalhos que vem sendo feitos. O entrevistado fala de uma lógica de organização de dados, ou seja, tão logo se tem aqui não somente uma maneira de trabalho mas uma maneira de raciocínio em relação ao próprio trabalho, e como as políticas públicas também estão envolvidas, um modo de se pensar em políticas públicas utilizando-se big data.

Tomando por base a teoria do ator-rede, temos também algo que, pela resposta do entrevistado, tem relação com a teoria: a ideia de links. O linkage acaba se traduzindo também no link que é feito pelos próprios trabalhos com big data, ou seja, o linkage entre bases de dados de diferentes políticas públicas pode ser visto aqui como o linkage entre actantes não-humanos que a ANT preconiza como, por exemplo, o link entre cadastros do SUS com cadastros do programa Bolsa Família.

Figura 5: Desenho dos links entre bases de dados de diferentes programas sociais



Fonte: Próprio Autor

Quanto à formulação, implementação e avaliação de políticas públicas e as informações disponíveis por entidades governamentais, ou seja, como são pensadas as políticas públicas, o entrevistado demonstrou que havia uma certa dificuldade inserida nesse contexto:

No geral, as políticas públicas tem dificuldade de relacionar o seu desenvolvimento com pesquisas específicas. No nosso caso, na área de

gestão, a relação das pesquisas feitas aqui pelo nosso grupo com relação às políticas públicas, quanto a isso, eu tenho uma dificuldade de fazer essa conexão mesmo quando ainda há a citação de algumas coisas.

Ou seja, não se observa relação objetiva entre resultados de pesquisas e políticas efetivamente elaboradas. As conexões acabam sendo feitas mais em torno do linkage e cruzamento entre dados produzidos e trabalhados pelo centro do que elaboradas conexões que tratem das pesquisas feitas com relação às políticas públicas. Há uma dificuldade, como o entrevistado afirmou, do próprio desenvolvimento do programa ou da política pública serem relacionados com pesquisas e avaliações realizadas.

As possibilidades, ainda assim, são das mais variadas em relação ao uso de big data. Quando questionado em relação a isso, quanto às possibilidades e também quanto às limitações que poderiam existir no uso de big data para a formulação, implementação e avaliação de políticas públicas, o entrevistado afirmou que:

as possibilidades são cada vez maiores e crescentes... a gente acaba nem tendo dimensão do que pode ser alcançado com essa forma de utilização dos dados. A Fiocruz trabalhou o desenvolvimento do Centro de Integração de Dados e Conhecimentos (CIDACS) que é fundamentalmente trabalhar com grande massa de dados, e nós temos que, a cada dia, alcançar novos patamares de utilização disso... para políticas públicas temos certeza de que ele fornece evidências robustas e potentes para orientar e reorientar políticas públicas. O futuro irá dizer como podemos potencializar melhor o uso de grandes massas de dados disponíveis no governo e na sociedade em geral fazendo a conexão desses dados.

Acerca do consentimento dos cidadãos e do uso de seus respectivos dados, o entrevistado então revelou que há sim uma preocupação, um cuidado no processamento dos dados para que os mesmos não forneçam meios para que eles próprios possam ser controlados, ou seja, possam ter seus dados sensíveis, opiniões, opções sexuais ou religiosas, por exemplo, identificados e coletados. De acordo com o entrevistado, “nós estamos falando agora do patamar ético da utilização de dados relativos a questões sensíveis sobre o comportamento político, sexual, ou de qualquer outro do indivíduo”. Como o entrevistado assevera:

Nós temos que ter muito cuidado e capacidade de processar esses dados na perspectiva das políticas públicas, sem que forneça dados ou indícios sobre o comportamento individual de modo que as pessoas possam ser controladas; o sujeito tem que ter uma autorização de alguma forma do uso de seus dados. Hoje nós temos avançado nessas discussões a esse respeito, no Congresso Nacional tem o PL 4060/2016 que está em tramitação e que trata disso, a princípio não tratava do uso de dados de pesquisa, mas atualmente já trata disso, já foi inserido o artigo que fala

sobre a utilização de dados para pesquisa que é diferente do uso comercial ou do uso administrativo.

Acerca da sociedade e ainda pensando nas inúmeras capacidades futuras que big data pode oferecer, o entrevistado revela que a não utilização da tecnologia que se possui, a tecnologia de dados, seria um erro. O uso dos dados já se insere na categoria de desenvolvimento nacional; eles tanto servem para o desenvolvimento social e econômico, como também para o político e científico. Segundo o entrevistado,

o não utilizar isso também tem um custo, então se deve utilizar, se deve prezar pela segurança dos indivíduos garantindo que eles não sejam usurpados no seu direito à privacidade mas que nós possamos, com o máximo de segurança, poder utilizar os dados que possam orientar o nosso desenvolvimento social e econômico, políticos, científicos.

Para o entrevistado, apesar de poder ser tentador o uso dos dados para se ter ciência de como o indivíduo pensa e se comporta, a ética, segundo o pesquisador – e pensando especificamente o campo científico, de produção de conhecimento – ainda prevalece. Isso porque a quantidade de informações é um bem material de suma importância para a ciência; o tratamento disto resta na base do sigilo, do bom uso. Interesses alheios aos da sociedade e do bem comum de todos não devem assim prevalecer, não devem ser contemplados:

A sociedade tem esse bem material que é a quantidade de informações que podem nos mostrar como a sociedade se organiza, se desenvolve e a utilização dos dados é importante. Agora nós temos que ter a preocupação ética de preservar o que é sigiloso, o que é sensível sobre os indivíduos, garantir a privacidade do sujeito é fundamental, não podemos deixar que, por interesses econômicos ou políticos, a sociedade passe a ser regulada a partir do uso de informações que o estado ou uma empresa possa deter sobre o comportamento individual”.

5. CONSIDERAÇÕES FINAIS

Tendo por base os objetivos dessa pesquisa e análise sob a perspectiva da formulação de políticas públicas, considera-se aqui que o trabalho conseguiu fazer uma análise bastante profícua apresentando como são feitos os trabalhos de produção e tratamento de dados do CIDACS, além da perspectiva do entrevistado que felizmente conseguiu fornecer a entrevista importante para a análise de fatores cruciais que cercam a visão e o pensamento de quem está envolvido com o uso da tecnologia e, também, por meio da análise da palestra que permitiu identificar atores importantes, como a rede é formada, os interesses existentes e os *links* que são produzidos.

Foi importante notar como existe um pensamento que também faz parte da rede em si, algo muito comum e forte nos discursos e que gera uma parcimônia quase que indubitável, que é a perspectiva jurídica sobre a questão do tratamento dos dados; é interessante notar como de fato há (pelo menos no discurso) uma preponderância daquele que analisa sob a perspectiva de que a lei deve ser respeitada e a validade que quaisquer instrumentos que norteiem a política pública, não importa o quão tecnológico sejam, deve estar sob o manto do viés constitucional para assim serem de fato válidos e poderem gerar produções que não gerem repercussões negativas, não sofram ameaças que culminem em seu fim.

Observou-se também como há, na produção e curadoria dos dados trabalhados no CIDACS, atenção a questões que poderiam por em cheque o trabalho dos cientistas, poderiam gerar dúvidas e questionamentos pertinentes quanto à discricionariedade no tratamento dos dados que ali se encontram. Por meio de fatores como a Sala Segura, actante não-humana, e a rede na qual está inserida, ou seja, os pensamentos, os grupos de interesse que cercam a ação e influenciam os movimentos feitos, observa-se que o próprio centro tem a noção do encaixe que deve fazer entre suas ações e o respeito a leis e valores subjetivos de dignidade e de responsabilidade para com os dados pessoais de milhões de indivíduos.

Este trabalho buscou contribuir permitindo que se observasse uma maior aplicação da TAR para questões pertinentes à produção e curadoria, ou seja, ao tratamento que é feito com os dados cadastrais de milhares de pessoas no Brasil. Sob a perspectiva da teoria, mostrou-se os grupos de interesse que abarcam a questão além das visões e dos posicionamentos-chave que importam para as tomadas de decisão que são feitas nas instituições aqui analisadas.

Conclui-se, também, que a neutralidade é pouca ou inexistente quando há defesa de um argumento. Dificilmente houve a alegação de neutralidade por parte daqueles que trataram

o tema; entretanto, se observou uma certa mediação nas falas, nos discursos, para atender ao mais diferentes interesses, para uma verdadeira comprovação de que aquela opinião se encaixava no pensamento dos demais, era válida como síntese. É inegável o valor econômico que o uso de big data possui. Entretanto, existem valores sociais, absolutamente mais importantes e que tratam de fatores intrínsecos ao ser-humano e cujo o valor monetário jamais irá alcançar que devem ser considerados quando da reflexão sobre o tema.

Assim como na teoria, as incertezas não são tão somente fatores que cercam a análise como também são pontos cruciais para a motivação da busca, para se valer a predominância da vontade em torno de se buscar análises futuras daquele tema. A incerteza se configura muito mais como a descoberta da teia complexa de movimentos que são as redes e da ilustração do quão complexas são elas; assim, tendo em vista a complexidade que cerca o tema o qual foi trabalho aqui e as inúmeras incertezas também geradas, espera-se que haja trabalhos futuros que amplifiquem mais o escopo do tema aqui analisado, que vão ainda mais fundo na complexidade aqui observada com os inúmeros pensamentos, grupos de interesse, parâmetros e desafios legais, morais, éticos que ampliam ainda mais a discussão e demonstram a necessidade de desenvolvimento de pesquisas futuras sobre o uso e o papel de big data nas sociedades contemporâneas.

REFERÊNCIAS

- ALVES. M. Big data e as pegadas do monstro: o digital na leitura da Análise do Discurso. 2016. Disponível em:
<<https://periodicos.sbu.unicamp.br/ojs/index.php/rua/article/view/8647448>>.
- ANGELONI. M. Elementos intervenientes na tomada de decisão. 2003.
- BAUER, M.W.; GASKELL, G. Pesquisa Qualitativa Com Texto Imagem e Som.. 2003.
- BERNARDES. G. Desenvolvimento de Software no Contexto Big Data. 2015.
- BRASIL. LEI Nº 12.527, DE 18 DE NOVEMBRO DE 2011. Lei de Acesso à Informação, Brasília, DF; novembro de 2011 . Disponível em:
<http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/112527.htm >.
- DINIZ. A. Tomar Decisão – uma arte a ser estudada. 2012. Disponível em:
<https://www.catho.com.br/carreira-sucesso/colunistas/arthur-diniz/tomada-de-decisao-%E2%80%93-uma-arte-a-ser-estudada>
- CAMARA DOS DEPUTADOS. Seminário Conjunto – Proteção de Dados Pessoais. Disponível em:<https://www.youtube.com/watch?time_continue=4627&v=wZHgfHZruHY>. Acesso em: 02 jun. 2018.
- DUARTE, Tomás. Mapeamento das Experiências do Consumidor. 2013. Disponível em:
<https://satisfacaodeclientes.com/mapeamento-experiencias-consumidor/>
- FERNANDES. F. O que, afinal, é o Indie ? Criatividade e comércio da produção musical contemporânea. 2004.
- FILHO. F. O que é Saúde Pública?. 1987.
- FRITZEN. F. Aspectos comportamentais de tomada de decisão: um estudo comparativo entre formandos de Administração e Contabilidade. 2007. Disponível em:
<http://tcc.bu.ufsc.br/Contabeis293851.pdf>
- GODOY. Arilda Schmidt. *Pesquisa Qualitativa: Tipos Fundamentais*. 1995. Disponível em:
<<http://www.scielo.br/pdf/rae/v35n3/a04v35n3>>
- GUIMARÃES. E. e ÉVORA. Y. Sistema de informação: instrumento para tomada de decisão no exercício da gerência. 2004.

JORGE. G. e MIYAKE. D. Estudo comparativo das ferramentas para mapeamento das atividades executadas pelos consumidores em processos de serviço. 2016.

LARA. A. e MOLINA. A. Pesquisa Qualitativa: Apontamentos, Conceitos e Tipologias. 2001. Disponível em : <<http://gepeto.ced.ufsc.br/files/2015/03/capitulo-angela.pdf>>

LATOUR. B. Reagregando o Social: uma introdução à teoria do Ator-Rede; ed. Salvador: Edusc, 2012.

LECHOPIER. N. Quatro tensões na saúde pública. 2015.

MOTA. P. Schumpeter: inovação, destruição criadora e desenvolvimento. 2016. Disponível em : <<http://terracoeconomico.com.br/schumpeter-inovacao-destruicao-criadora-e-desenvolvimento>>.

NUNES. E. Sobre a história da saúde pública: idéias e autores. 2000.

NUNES. T. .A Especialização em Saúde Pública e os serviços de saúde no Brasil de 1970 a 1989. 1998.

NUNES. S. .Diferenças e semelhanças entre a saúde coletiva e a saúde pública na percepção de discentes da Universidade de Brasília e da Universidade de São Paulo. 2016.

OLIVEIRA. F. Schumpeter: a destruição criativa e a economia em movimento. 2014. Disponível em: <http://www.ufjf.br/heera/files/2009/11/SchumpeterUFJF-2-para-pdf.pdf>

ROSEN. G. Uma história da Saúde Pública. 1994.

SCHUMPETER. J. Capitalism, Socialism and Democracy. 2003. Disponível em: https://books.google.com.br/books?id=ytrqJswRCoC&printsec=frontcover&dq=joseph+schumpeter&hl=pt-BR&sa=X&ved=0ahUKEwjO1sfaw__VAhUD-J8KHT-ID9wQ6AEIJzAA#v=onepage&q=joseph%20schumpeter&f=false

SILVA, C. Privacidade em Big Data: panorama e agenda de pesquisa. 2017. Disponível em: <<http://www.revistasg.uff.br/index.php/sg/article/view/1205/769>>

TANAKA. O e TAMAKI. E. O papel da avaliação para tomada de decisão na gestão de serviços de saúde. 2012. Disponível em: <http://www.scielo.br/pdf/csc/v17n4/v17n4a02>

TOMACHESKI. J. e FERRAREZI. E. Sustentabilidade de Iniciativas Premiadas no Concurso Inovação: Índícios de mudança da Gestão no Governo Federal ?2011.

WINSLOW, C. The Untilled Fields of Public Health. 1920. Disponível em: <http://www.jstor.org/stable/1645011?seq=1#page_scan_tab_contents>

VOLPATO, T. Big Data- Transformando Dados em Decisões.2001? Disponível em : <
http://web.unipar.br/~seinpar/2014/artigos/graduacao/Tiago_Volpato.pdf>.