BIG DATA ESTRATÉGICO:
UM *FRAMEWORK* PARA GESTÃO SISTÊMICA DO ECOSSISTEMA BIG DATA

Luciana Sodré Costa

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

Orientador: Marcos do Couto Bezerra Cavalcanti

Rio de Janeiro

Abril de 2016

BIG DATA ESTRATÉGICO:

UM *FRAMEWORK* PARA GESTÃO SISTÊMICA DO ECOSSISTEMA BIG DATA

Luciana Sodré Costa

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

Examinada por:

_____

Prof. Marcos do Couto Bezerra Cavalcanti, D.Sc.

_____

Prof. Marcus Vinícius Fonseca, D.Sc.

_____

Profa. Liz Rejane Issberner, D.Sc.

RIO DE JANEIRO, RJ - BRASIL

ABRIL DE 2016

Costa, Luciana Sodré

Big Data Estratégico: Um *Framework* para Gestão Sistêmica do Ecossistema Big Data / Luciana Sodré Costa. – Rio de Janeiro: UFRJ/COPPE, 2016.

XIII, 112 p.: il.; 29,7 cm.

Orientador: Marcos do Couto Bezerra Cavalcanti

Dissertação (mestrado) – UFRJ/ COPPE/ Programa de Engenharia de Produção, 2016.

Referências Bibliográficas: p. 86-92.

1. Big Data. 2. Big Data Estratégico 3. Ecossistema Big Data 4. Ciência das Redes. 5. Complexidade. 6. Tomada de decisão. I. Cavalcanti, Marcos do Couto Bezerra. II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia de Produção. III. Título.

Dedico esse trabalho às minhas três filhas: Isabel, Joana e Laura.

Esse trio é, de longe, o meu melhor projeto.

Agradecimentos

Às minhas filhas Isabel, Joana e Laura, por todos os *“boa sorte”, “vai dar certo”, “não se preocupe”, “não precisa pegar a gente na escola”, “deixe que eu faço”, “vou desligar a TV para não atrapalhar”, “ainda tá aí, mamãe?”*... Pequenas frases doces, soltas, repetidas, que me trouxeram tanto conforto...

À minha família, especialmente meus pais, pelo legado, amor, confiança e apoio de sempre. Jamais retribuirei o suficiente por tudo que recebi.

Ao meu orientador, Marcos Cavalcanti, pelas longas horas de conversas e discussões empolgadas, ambiciosas e utópicas, por todas as oportunidades oferecidas, pela generosidade com que se dispôs a me orientar e pelas sempre muito bem vidas desorientações providenciais.

Aos professores da Coppe, Carla Cipolla, Marcos Cavalcanti, Marcus Vinícius Fonseca e Roberto Bartholo, pelas inspirações que recebi de suas aulas e de suas visões de mundo.

Aos tantos amigos que assumiram parte da minha carga horária de mãe, para que eu pudesse me dedicar mais a esse trabalho.

Aos amigos que contribuíram com suas expertises, comentários, críticas, responderam meus pedidos de sugestão de bibliografia, de nomenclatura, de abordagem e de tantas outras maneiras. Não cairei na armadilha de citar nomes, mas vocês sabem quem são.

Aos meus alunos, que comentaram, criticaram e testaram meu trabalho.

PRÓLOGO

Uma pesquisa é como uma obra de arte:
nunca fica pronta, apenas é chegado o dia da entrega.

Alguns podem se dar ao luxo de ignorá-lo:
"I cannot live under pressures from patrons, let alone paint[1]*"*,
dizia Michelangelo para quem quisesse seus serviços.
Havia duas opções para quem desejasse ter uma obra do gênio dos séculos XV e XVI:
esperasse pelo tempo que ele achasse necessário,
ou não tivesse uma obra de Michelangelo.

Muito longe de possuir a genialidade de um dos maiores artistas do mundo ocidental de todos os tempos, me dou conta de que preciso para de criar e entregar o trabalho.

Não é o fim da jornada. Essa dissertação é apenas o diário de uma parte da viagem.
Tenho convicção de que trabalharei por toda a vida para que cada vez mais dados fiquem disponíveis e que cada vez mais pessoas sejam convencidas e preparadas
para abraçar a vital tarefa de dar sentido a esses dados.

Vamos começar o passeio.

[1] Se eu não aguento viver sob pressão dos meus clientes, imagine a pintura (Tradução nossa)

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

BIG DATA ESTRATÉGICO:
UM *FRAMEWORK* PARA GESTÃO SISTÊMICA DO ECOSSISTEMA BIG DATA

Luciana Sodré Costa

Abril/2016

Orientador: Marcos do Couto Bezerra Cavalcanti

Programa: Engenharia de Produção

A Internet já é parte incontestável da civilização humana e também já se tornou um marco divisório dela. Ainda é cedo para conhecer seus impactos, mas podemos afirmar com segurança que existe o antes e o depois do seu advento. A maior revolução que a Internet pode causar ainda estar por vir porque seu uso intensivo gera bilhões de gigabytes de dados diariamente, que permanecem subutilizados. Conhecido como *big data*, esse fenômeno de digitalização nos legou dados em volume, variedade e velocidade suficientes para torná-los confiáveis. Uma vez estruturados, esses dados permitem visualização de padrões, relações e correlações inusitadas fundamentais para a criação de modelos descritivos, diagnósticos e preditivos de alta performance usados como empoderador do tomador de decisão. Ainda existem muitos obstáculos para que isso seja algo comum nas instituições e a curva de aprendizagem está apenas começando. Entre os motivos mais relevantes estão a falta de informação e de metodologia. Essa pesquisa apresenta um *framework* simples, embasado por princípios da Ciência das Redes, e funciona a partir da análise sistêmica do ecossistema *big* data como guia para a adoção modelos de negócio e processos de decisão baseada em dados.

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

BIG DATA:
A FRAMEWORK FOR SYSTEMIC MANAGEMENT OF THE BIG DATA ECOSYSTEM

Luciana Sodré Costa

April/2016

Advisor: Marcos do Couto Bezerra Cavalcanti

Department: Production Engineering

The Internet is already an undeniable part of human civilization and has become one of its landmarks. It is still too early to know its impact, but we can safely say that there is a ‘before’ and ‘after’ Internet. The greatest revolution that the Internet can cause is yet to come because its intensive use generates daily, billions of gigabytes of data, which remain underutilized. Known as big data, this phenomenon of digitalization has given us a substantial amount of data with enough volume, variety and speed to make them reliable. Once structured, these data allow us to observe unusual patterns, relationships and correlations that are fundamental for the creation of descriptive, diagnostics and predictive models, which empower the decision makers. There are still many obstacles before this becomes a common use in institutions and the learning curve is still in the beginning. Amongst the most important reasons for the resistance are the lack of information and methodology. This research presents a simple framework, based on principles of the Complex Network Science, and starts with the systemic analysis of the big data ecosystem as a guide for the adoption of business models and data driven decision process.

SUMÁRIO

1. Introdução 1

1.1. Serendipidade ou a história de um caminho 2

1.2. A melhor época para estar vivo 6

1.3. O problema 10

1.3.1. Em busca de evidências 10
1.3.2. Relevância 16
1.3.3. Definição 16
1.3.4. Objetivos 17
1.3.5. Resultados esperados 17

1.4. Metodologia 17

1.5. Estrutura do texto 21

2. Referencial Teórico 22

2.1. Ciência das Redes 26

2.2. Problemas complexos 42

2.3. *Big Data* 46

Big data como conjunto de dados 47
Big data como disciplina 49
Big data como tecnologia 49
Afinal, o que é big data? 53

2.4. Ecossistema *big data* 55

3. *Framework*: *Big Data* Estratégico.....58

3.1. Visão 59

3.2. Ecossistema *Big Data* Crie: delimitação e descrição.....60

3.2.1. Delimitando o ecossistema: escopo do *framework* 61
3.2.2. Descrição dos *stakeholders* 63
a. Geração de informações 63
b. Geração de problemas.....64
c. Proposição de problemas.....64
d. Dataficação do problema.....64
e. Digitalização das informações.....65
f. Governança de dados 66
g. Estratégia de dados.....66
h. Analítica de dados.....66
i. Visualização de dados.....67
j. Tomada de decisão 67
k. Execução da decisão.....68
l. *Feedback*.....68

3.3. Big Data Estratégico: um *framework* para o Ecossistema *Big Data* Crie.....69

3.4. Como usar o *framework*? - *Check list*.....74

4. Conclusões 82

5. Recomendações para estudos futuros.....84

7. Bibliografia.....86

ANEXO I - FIGURA INTEGRANTE DE MATERIAL EXPLICATIVO DO INSTITUTO GARTNER SOBRE A HYPE CYCLE CURVE.....93

ANEXO II - FIGURA INTEGRANTE DE RELATÓRIO 2014 DO INSTITUTO GARTNER SOBRE TECHNOLOGY HYPE CYCLE.....94

ANEXO III - FIGURA INTEGRANTE DE RELATÓRIO 2014 DO INSTITUTO GARTNER SOBRE TECHNOLOGY HYPE CYCLE.....95

ANEXO IV - ORIENTAÇÕES PARA DESENVOLVIMENTO DE UM *FRAMEWORK* – *COMMUNITY TOOL BOX*.....96

ANEXO V - GOOGLE TRENDS PARA AS BUSCAS POR "BIG DATA" NO BRASIL EM 2010.....101

ANEXO VI - BIG DATA ESTRATÉGICO - QUADRO DE TRABALHO.....102

APÊNDICE I – CRIE: CENTRO DE REFERÊNCIA EM INTELIGÊNCIA EMPRESARIAL (LABORATÓRIO DE INOVAÇÃO DA COPPE/UFRJ).....103

APÊNDICE II – UNIDADES DE MEDIDA DE INFORMAÇÃO.....104

APÊNDICE III – GOOGLE TRENDS PARA AS BUSCAS POR *"BIG DATA"* NO BRASIL EM 2010.....105

APÊNDICE IV – CAPA DA REVISTA VEJA COM MATÉRIA SOBRE BIG DATA PUBLICADA NA SEMANA DE 12 A 18 DE MAIO DE 2012.....106

APÊNDICE V – *OCTALYSIS:* FRAMEWORK PARA DRIVES DE MOTIVAÇÃO PARA JOGAR.....107

APÊNDICE VI – *FRAMEWORK* PARA DEFINIÇÃO DE CAPACIDADE ANALÍTIVA.....110

APÊNDICE VII – *FRAMEWORK CLOUD SERVICE*.....111

APÊNDICE VIII – FRAMEWORK PARA DEFINIÇÃO DE TIPO DE VISUALIZAÇÃO DE DADOS.....112

LISTA DE FIGURAS

Figura 1 - Buscas para big data 2004-out 2012 (Brasil, Reino Unido, Estados Unidos e Índia) 4

Figura 2 - *Technology Hype Cycle* ...…….....…….....11

Figura 3 - Curva de Adoção de Inovação + *Chasm*.....13

Figura 4 – *Big Data: Hype Cycle Graph* + Curva de Adoção de Inovação 14

Figura 5 - Projeção para trajetória de *big data* no *Hype Cycle* 15

Figura 6 - *Google Trends* para *big data* (Brasil e EUA - 2012 a 2013) 23

Figura 7 – Curva Normal ou de Sino (eventos aleatórios) 30

Figura 8 – Curva de grau em lei de potência (redes sem-escala) 30

Figura 9 - *Big Data* Estratégico - *framework* para gestão estratégica do Ecossistema *Big Data*.....71

LISTA DE TABELAS

Tabela 1 - Problemas complicados X problemas complexos 42

Tabela 2 – *Big Data* estratégico – Quadro resumo..... 72

Tabela 3 - *Check List* para uso do *Big Data* Estratégico..... 75

1. Introdução

Internet, redes sociais, aplicativos, dados digitais... tudo isso já faz parte do nosso cotidiano, não apenas nos aspectos práticos, mas também em conversas informais entre pessoas de todas as idades, formações e ideologias. Esse é o lado banalizado e descomprometido com que a sociedade lida com a mais disruptiva inovação de toda história da humanidade.

O conceito de inovação disruptiva foi introduzido por Clayton Christensen (1997) para se referir a inovações que criam novos mercados consumidores. O autor adota a expressão inovação disruptiva em vez de tecnologia disruptiva[2], porque entende que o que provoca a disrupção de mercados não é uma tecnologia em si, mas o modelo de negócio associado a ela. A Internet é um exemplo perfeito disso já que a tecnologia que permitiu a sua criação data da década de 60, mas apenas no século XXI foi utilizada para criar novos modelos de negócio em qualquer setor produtivo. Em 2015, Eric Schimidt, então diretor do Google, causou alvoroço no Fórum Mundial de Davos, na Suíça ao declarar que Internet iria desaparecer. O que ele queria dizer era que a Internet estaria tão presente nas nossas vidas que não seria mais percebida. Cair na normalidade até se tornar invisível é a maior prova do caráter disruptivo de alguma coisa: significa que a inovação incluiu de tal maneira diferentes grupos de consumidores, que não há a sensação de desejo de fazer parte de algo que é exclusivo para poucos. O que poderia ser mais disruptivo?

Falar da Internet de modo banal e totalmente descomprometido, como se fosse o ar que respiramos, reflete essa sensação de inclusão e não há nada de errado nisso. Mas existe o outro lado dessa moeda: para muita gente a Internet passou a ser o eixo central de suas vidas profissionais, produzindo produtos, serviços e conhecimento, justamente porque veem a Internet para além do que a banalizou.

Aconteceu comigo.

[2] Tecnologia disruptiva era a expressão comumente usada, inclusive pelo próprio Crhistensen, no artigo *Disruptive Technologies: Catching the Wave*, publicado em 1995 na Harvard Business Review.

1.1. Serendipidade ou história de um caminho

> Você não consegue conectar os fatos olhando para frente. Você só os conecta quando olha para trás. Então tem que acreditar que, de alguma forma, eles vão se conectar no futuro. Você tem que acreditar em alguma coisa: sua garra, destino, vida, carma ou o que quer que seja. Essa maneira de encarar a vida nunca me decepcionou e tem feito toda a diferença para mim.
>
> Steve Jobs

Serendipidade é o que explica todos os caminhos de crescimento pessoal. A minha história não seria diferente. A palavra, criada no século XVIII[3], foi inspirada em um antigo conto persa sobre três príncipes de Serendip[4] que, por serem muito observadores, viajados e experientes, resolviam dilemas considerados insolúveis. Parecia sorte ou acaso, mas a moral da história é que o conhecimento que eles produziam era simplesmente o *output* das suas trajetórias pessoais. Serendipidade é o que explica um *insight* que temos quando juntamos os pontinhos das informações adquiridas na nossa trajetória, que é sempre única e intransferível.

Minha primeira formação foi em Administração de Empresas. Gostar de todas as disciplinas escolares e não ter preferência específica por nenhuma delas me afastou de escolher formações muito centradas em um único domínio do conhecimento. Depois de formada trabalhei com planejamento, RH, importação e exportação, marketing, logística e ainda passei pelo mercado financeiro, educação e produção de cinema e TV. Gostei de tudo que fiz, mas nunca consegui evitar uma inquietação para cruzar fronteiras e contaminar minha área com informações e inovações de outras áreas. Tudo indicava que eu não sabia mesmo o que queria fazer. Minhas publicações preferidas eram, e ainda são, aquelas que falam sobre tudo: de moda a projeto genoma, de mecânica a comportamento humano, de astronomia a arte. Quanto mais adentrava por esse caminho, mais se abria o *gap* entre duas sensações: eu estava finalmente me encontrando ou definitivamente me perdendo.

[3] 1754, pelo escritor britânico Horace Walpole
[4] Cidade imaginária localizada no antigo Ceilão, atual Sri Lanka.

Foi numa dessas publicações[5] que tive acesso a um texto intitulado *Thinking in Network Terms*[6], do cientista Albert-László Barabási. Barabási pesquisa sistemas complexos, aqueles cujas estruturas são dinâmicas e emergentes, típicas de sistemas vivos. Segundo ele, ver o mundo em termos de rede e entender as suas propriedades é a única maneira de gerenciar aqueles problemas insolúveis por definição: os problemas complexos[7]. Mas *Thinking in Network Terms* não é uma conversa sobre redes complexas e sim um alerta para a necessidade de compartilharmos dados digitais. Segundo Barabási (2012), dados digitais permitem a modelagem de problemas complexos que, formulados matematicamente, ganham poder preditivo podendo ser gerenciados e controlados. No entanto, ele continua, os dados digitais necessários para fazer isso estão em mãos privadas e nós precisamos de algum acordo social para compartilhar esses dados, permitindo que diferentes *stakeholders* criem valor com eles.

Era 26 de setembro de 2012. Nesse texto ouvi falar pela primeira vez em *big data.* Fosse por outra fonte, o assunto teria passado despercebido. Tendo nenhum *background* em tecnologias da informação, saber que uma grande quantidade de dados digitais desafiava as tecnologias de armazenagem e processamento da época[8], me teria sido uma informação pouco interessante. No entanto, *big data* nesse texto era a possibilidade de usar dados digitais para saber mais sobre tudo: de moda a projeto genoma, de mecânica a comportamento, de astronomia a arte. Mais que isso, a possibilidade de enxergar como essas coisas estão relacionadas. Não seria correto dizer que naquele momento encontrei um caminho (afinal se li aquele texto, já devia estar caminhando), mas minha trajetória passou a ter nome e destino conhecidos.

Quando li *Thinking in Network Terms* eu cursava o MBA em Gestão do Conhecimento (MBKM[9]) da Coppe/UFRJ, coordenado pelo professor Marcos Cavalcanti com quem compartilhei meus *insights* sobre *big e open data* como plataforma de inovação. As ideias caíram em campo fértil e fui incentivada a fazer o trabalho de conclusão de curso nessa

[5] edge.org: um site que promove entrevistas com pessoas, de qualquer domínio e especialidade, que estejam pensando e pesquisando na "fronteira do conhecimento". A única regra na entrevista é que o especialista fale de modo que o público não iniciado no assunto possa entender.
[6] https://edge.org/conversation/albert_l_szl_barab_si-thinking-in-network-terms
[7] No item 2.2. esse conceito voltará a ser abordado.
[8] Uma das primeiras definições para *big data* mais difundida e, para muitos, válida até hoje
[9] Master on Business and Knowledge Management

direção. Pesquisas exploratórias no *Google Trends*[10] em outubro de 2012, mostravam que, embora o fenômeno big data já tivesse sido detectado desde o início do século, como veremos mais adiante, dez anos depois quase não se falava do assunto no Brasil. A ferramenta do Google só começa a captar sinais de interesse no Brasil em 2011 mesmo assim, baixo se comparado a outros países.

A figura 1 mostra a evolução histórica[11] dessa tendência no Brasil até a data da pesquisa (outubro de 2012) comparada com Reino Unido, Estados Unidos e Índia[12].

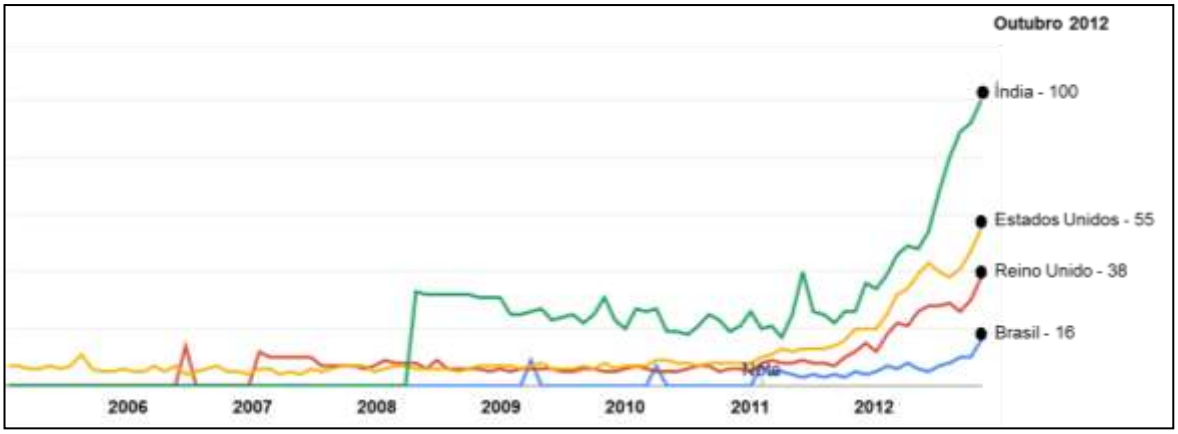


Figura 10 - Buscas para *big data* 2004-out 2012 (Brasil, Reino Unido, Estados Unidos e Índia) – Fonte: GoogleTrends

O Google Trends atribui 'índice 100' ao país com a maior quantidade de busca no período pesquisado e aos demais países, índices relativos ao máximo 100. Podemos observar que a Índia, país que comumente usamos para comparar o Brasil em diferentes indicadores, liderava o indicador do Google Trends para *big data* em 2012. Os Estados Unidos e Reino Unido apresentavam indicadores 55 e 38, respectivamente, enquanto que o Brasil havia começado há poucos meses uma escalada no gráfico que o levaria ao indicador 16, o máximo já atingido em sua série. Embora esses resultados não sejam conclusivos, era intrigante saber que um assunto tão paradigmático tinha pouca representação nos temas considerados tendências no Brasil. Não haviam livros em

[10] Ferramenta do Google que mostra os principais termos buscados, podendo-se restringir a pesquisa a períodos e locais específicos.
[11] Desde 2004 quando o Google passou a medir tendências de busca pelo Google Trends
[12] Justificativa da escolha desses três países: Estados Unidos: por ser sempre uma referência e por ter sido o berço das pesquisas de Barabási, autor que fundamenta boa parte desse trabalho; Reino Unido: por ter um bom histórico de adesão à causa digital; Índia: por ser um país com características econômicas mais próximas ao Brasil.

português e quase não haviam publicações sobre *big data* em português nas páginas brasileiras da *web*. Uma busca no Google por *"big data"* (entre aspas), solicitando apenas resultados em português de páginas no Brasil, retornavam algumas centenas de documentos, a maioria traduções de sites americanos.

Desse estudo exploratório, que associei a existência de pouca massa crítica sobre o assunto no Brasil, surgiu o foco do trabalho final do MBKM: uma análise das forças, fraquezas, ameaças e oportunidades do Brasil para se beneficiar da analítica digital de alta performance, conhecida como *big data.* O resultado do trabalho[13] foi uma espécie de matriz SWOT[14] do ambiente digital do país e uma matriz SWOT Cruzada onde procura-se neutralizar fraquezas com oportunidades e ameaças com forças, gerando uma proposta de planejamento estratégico factível. O objetivo era levantar argumentos para mobilizar agentes econômicos de diversos setores e criar massa crítica que demandasse pesquisa, eventos, publicações, regulamentação, produtos e serviços de projetos em *big data.*

Demos um passo criando o Instituto Big Data Brasil, associado ao Centro de Referência em Inteligência Empresarial (Crie)[15], também coordenado pelo professor Marcos Cavalcanti. Dessa parceria surgiram outras iniciativas relacionadas à geração de conhecimento a partir de dados digitais: nos últimos três anos desenvolvemos duas metodologias[16] em um trabalho para o centro de pesquisa da EMC, localizado no Parque Tecnológico da URFJ, concebemos e lançamos um curso de pós-graduação em *big data* para gestores[17], promovemos dois seminários[18] sobre o assunto e, em 2014, firmamos uma parceria com o *Open Data Institute*[19] de Londres, criando o ODI Rio, com a missão de alavancar a cultura dos dados abertos. O ODI Rio foi o primeiro nó do Brasil na rede internacional do *Open Data Institute*.

[13] Big Data: Estudo do Ambiente, Desafios e Analise Estratégica para o Brasil (2013)
[14] SWOT: do inglês *strengths, weaknesses, opportunities* e *threats.* Ferramenta utilizada para fazer análise de cenário, sendo usada como base para planejamento estratégico.
[15] Laboratório de pesquisa em inovação da Coppe/UFRJ. Mais informações no Apêndice I
[16] Prospecção de Oportunidades em *Big Data*; Grau de Prontidão para *Big Data*
[17] Web Intelligence & Digital Ambience (WIDA)
[18] Seminário Big Data Brasil I e II (2014 e 2015)
[19] theodi.org

Em 2013 o professor Marcos me incentivou a continuar os estudos sobre *big data* estratégico[20], em um Mestrado na Coppe, esse que termino agora.

Ainda em 2013 tive a oportunidade de conhecer Albert-László Barabási e de lhe dizer que *Thinking in Network Terms* tinha sido, para mim, um divisor de águas, ao que ele respondeu: *"It is odd... It is my worst talking ever...! I wasn't very convincing...*[21]*". Essa* passagem pitoresca, quase lendária por se tratar do maior cientista em ciência da complexidade do mundo, só corrobora o fenômeno da serendipidade: não foi o melhor trabalho de László Barabási que chegou em minhas mãos, aliás em suas próprias palavras, foi o pior, mas foi suficiente. Eu não sabia o que estava procurando, mas reconheci imediatamente quando encontrei. Todos os valores que cultivei desde então a respeito do ambiente digital vieram daquele 26 de setembro quando li Barabási pela primeira vez. Todos estão presentes nesse trabalho.

Você sabe que está se preparando para algo quando alguém lhe diz: *"você precisa me dizer o que quer!"* E você se sente confortável para responder: "*mostre-me algo, eu vou saber quando olhar"*[22]*.*

Serendipidade é sempre uma questão de tempo.

1.2. A melhor época para estar vivo

> Pela primeira vez na história nossa capacidade
> está começando a alcançar nossas ambições
>
> Peter Diamandis

Uma das questões mais importantes da atualidade diz respeito à utilização dos dados digitais gerados pelo uso intensivo da internet nos últimos 20 anos. É impossível estimar o volume de bytes acumulados, mas especialistas arriscam-se a fazer estimativas em bases

[20] Big data estratégico: termo que adotei para fazer referência à abordagem de *big data* pelo viés do empoderamento do tomador de decisão de uma organização.
[21] Estranho... foi a minha pior entrevista! Não acho que fui muito convincente...
[22] Diálogo atribuído a Steve Jobs, fundador da Apple (ele na segunda fala)

diárias: aproximadamente 2,5 quintilhões de bytes[23] de dados são criados diariamente a partir de postagens em redes sociais, upload de fotos e vídeos, registros de transações comerciais, sinais de GPS, rastros de navegação e sensores dos mais diversos tipos[24].

Esse volume tende a crescer exponencialmente com a ampliação do contingente de coisas conectadas, com a migração de mais atividades para a rede digital e com o aumento da capacidade tecnológica para aquisição e transmissão de dados. Provavelmente o fenômeno que melhor representa essa afirmação é a chamada Internet das Coisas (IoT[25]) da qual falamos ainda muito pouco, mas que promete ter um grande impacto não só na quantidade de dados, mas também em sua diversidade. Em termos genéricos, trata-se de uma rede sem fio[26] conectando sensores colocados nos mais diferentes tipos de objetos, animais e plantas que geram, trocam e armazenam dados através da internet a partir de um sistema eletrônico de informação. Os registros gerados pela IoT abrangem desde dados estáticos inerentes ao objeto, como descrição, modelo, características físicas, local e data de fabricação, até dados dinâmicos sobre sua utilização, relação com seres humanos, com o meio ambiente ou com outros objetos.

Em 2011, um relatório da CISCO[27] apontava que o número de 'coisas' conectadas à internet havia ultrapassado em 2008 o número de humanos na terra (em torno de 7 bilhões). No mesmo relatório a empresa também estimou que em 2020 um trilhão de objetos estariam conectados à internet. Confirmadas as previsões, em breve teremos um trilhão de objetos se comunicando de modo contínuo, gerando uma quantidade de dados incalculável.

Convencionou-se chamar tudo isso de *big data*, e vários indícios levam a crer tratar-se de um fenômeno progressivo, irreversível e vital[28]. 'Progressivo' porque, desde o surgimento

[23] Ou 2,5 bilhões de *gigabytes*. Para ver as relações entre as unidades de medidas de volume de informação, ver Apêndice II.
[24] Estimativa da IBM publicada em http://www-01.ibm.com/software/data/bigdata/ (página visitada em 05/07/2013)
[25] IoT, do inglês *Internet of Things*, a rede que interliga objetos no mundo inteiro. Algumas empresas como a CISCO já utilizam a nomenclatura *Internet of Everything* (IoE) - Internet de Todas as Coisas - para evitar a setorização da internet.
[26] *Wireless Sensor Network*, um protocolo de comunicação criado pelo Auto-ID Laboratory do MIT.
[27] http://www.cisco.com/web/about/ac79/docs/innov/IoT_IBSG_0411FINAL.pdf (página visitada em 25/02/2013)
[28] "Progressivo, irreversível e vital" é o modo como nós, pesquisadores do Crie, nos referimos ao fenômeno *big data*.

da internet, a quantidade de dados que geramos e acumulamos cresce constantemente e, por tudo o que já foi dito aqui, não há sinais de arrefecimento. 'Irreversível' porque, uma vez digitalizado, um processo não volta a ser desenvolvido de modo não digital, já que a plataforma digital garante aos processos mais eficiência, segurança e melhor gestão. Finalmente, 'vital'. A definição mais difundida é a de que *big data* é um banco de dados tão gigantesco que desafia as tecnologias atuais de armazenagem e processamento. Fosse isso, *big data* seria um efeito colateral negativo do uso da internet que gera custos indesejados e estaria, inevitavelmente, associado à palavra 'estorvo'. No entanto, mesmo quem nada entende sobre o assunto sabe que *big data* gera novos negócios, aumenta as receitas e os lucros e é uma expressão definitivamente associada à palavra 'inovação'.

Por que um banco de dados teria esse poder? Aqui entra a dimensão 'vital' desse fenômeno: dados são a matéria prima do conhecimento que, por sua vez, é a única forma de garantir a sobrevivência. Isso sempre foi assim: captamos dados do mundo através dos nossos órgãos dos sentidos, relacionamos com outros dados, formulamos hipóteses, criamos modelos para compreender e antecipar cenários, tomamos decisões, verificamos o impacto dessas decisões e recomeçamos tudo em novas bases[29]. Quanto maior a quantidade e variedade dos dados que analisamos e quanto maior a sua confiabilidade, maiores as nossas chances de tomarmos boas decisões e de sobreviver. Nesse contexto, o verbo sobreviver pode ser entendido no seu sentido denotativo, relativo à vida, ou conotativo, relativo à falência ou insucesso. Assim, quando escolhemos um tratamento para uma doença ou o melhor momento para atravessar uma rua, sobreviver significa ficar vivo. Quando escolhemos nossos fornecedores, público alvo ou modelo de negócio, sobreviver é evitar a falência.

Nossos sensores e processadores naturais (órgãos dos sentidos e cérebro, respectivamente) são essenciais nas nossas decisões e têm sido usados com grande sucesso ao longo de toda a história da humanidade, no entanto, eles possuem suas limitações. Se existem performances cognitivas humanas que um computador não consegue imitar, o contrário também é verdadeiro: sensores e processadores digitais superam em muitas vezes certas capacidades humanas. Um exemplo simples disso é a percepção de temperatura. Nosso tato consegue dizer se uma superfície está quente ou

[29] Mais adiante veremos que esse é exatamente o ciclo de geração de conhecimento a partir de dados digitais

fria, e até percebe variações quando as temperaturas sobem ou caem, mas só um sensor digital diz exatamente qual a temperatura e percebe e quantifica imediatamente qualquer variação. Nosso cérebro também atua aquém da capacidade de qualquer computador quando a função a ser desempenhada é, por exemplo, correlacionar uma grande quantidade de dados. A memória é outra função cognitiva em que os computadores conseguem superar o cérebro humano.

As tecnologias digitais empoderam de diversas maneiras o ser humano no seu desafio constante de resolver problemas para sobreviver. Empresas, governos e outras organizações já perceberam isso e têm aumentado seus investimentos para desenvolver e adquirir essas tecnologias. No entanto, diversos especialistas têm alertado para o fato de que a maior barreira para gerar valor a partir de dados digitais não é a tecnologia e sim o fato da maioria das empresas ainda não terem compreendido o verdadeiro impacto dos ambientes digitais na sociedade, nem saberem de fato como transformar grande volume de dados em melhores decisões[30].

Dados e tecnologias de armazenagem e processamento, são condições necessárias para qualquer projeto em big data, mas não suficientes. É preciso compreender que o ambiente digital é composto por outros agentes que se inter-relacionam e impactam os processos analíticos e que, por isso mesmo, precisam ser gerenciados[31]. Além disso, a falta de conhecimento das partes e de compreensão da dinâmica do ambiente digital leva a crer que apenas grandes empresas podem se beneficiar de analítica em *big data.* Isso é um equívoco, justificado pela falsa premissa de que *big data* é 'dados e ferramentas', portanto só beneficia os grandes *players* do mercado como Google, Facebook, Amazon, entre outros, que controlam grandes volumes de dados e têm acesso a igual volume de recursos financeiros para investir em ferramentas inacessíveis ao capital mais restrito.

Uma nova revolução cognitiva está em curso. O ambiente digital possibilita um avanço exponencial nas nossas capacidades de apreender, entender, aprender, correlacionar, diagnosticar, predizer, prognosticar e decidir. Fazer parte disso não será opcional. De

[30] Isso nos leva de volta ao conceito de Christersen de inovação disruptiva, em vez de tecnologia disruptiva.
[31] Nesse trabalho usamos o verbo gerenciar como sinônimo de governar. Embora governar seja conceitualmente mais adequado, já que envolverá agentes sobre o qual não se tem ingerência, gerenciar é

modo ativo ou passivo todos serão impactados pelas tecnologias digitais. Às gerações que ainda estão por nascer, provavelmente estará garantido um *mindset*[32] ajustado para esse novo paradigma, no entanto, existem ainda algumas gerações de migrantes digitais, que atuam de modo importante e impactante no mundo que precisarão se converter da melhor maneira possível ao novo cenário de produção de conhecimento e de riqueza.

É nesse contexto de uso intensivo da Internet, geração de grande volume de dados e de tomada de decisão a partir de dados digitais, fenômeno conhecido hoje como *big data*, que está situado essa pesquisa. Ela utiliza e articula os conceitos de ciberespaço e cibercultura[33], complexidade e ciência das redes, buscando oferecer uma sistemática simples e replicável para planejamento do uso de dados como empoderador do tomador de decisão.

1.3. O problema

1.3.1. Em busca de evidências

Segundo o Instituto Gartner[34], toda tecnologia inovadora percorre uma trajetória de atenção publicitária mais ou menos comum, desde que sai dos laboratórios de pesquisa e desenvolvimento (P&D) até atingir a maturidade, identificada como a fase em que a tecnologia atende plenamente à sociedade. A figura 2 representa essa curva teórica que o Gartner chama de *Technology Hype Cycle*[35]. Não existe uma tradução oficial para o português, mas algumas pessoas chamam de curva de adoção de inovação, cometendo um equívoco conceitual já que as duas curvas, embora se correlacionem de alguma forma, representam informações diferentes, como veremos.

---

mais utilizado na língua portuguesa no contexto corporativo. Exceção se faz quando usamos a palavra na sua forma substantiva, onde governança faz mais sentido que gestão, quando se trata de ecossistemas.

[32] Palavra adotada de modo generalizado no Brasil significando mentalidade

[33] Conceitos do filósofo francês Pierre Lévy, 2011, citado na seção 2, item 2.4

[34] Empresa americana sediada em Stamford, Connecticut, presente em mais de 85 países do mundo, líder em pesquisa e consultoria de tecnologias da informação (do site da empresa)

[35] Conceito desenvolvido pelo Instituto Gartner em 1995.

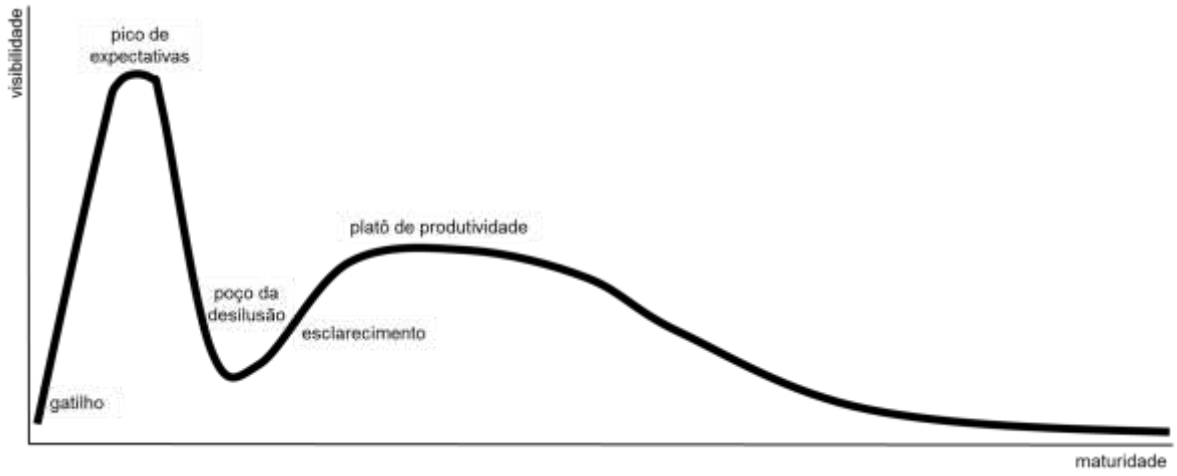


Figura 11 - *Technology Hype Cycle*[36] - Fonte: Instituto Gartner

O ***Hype Cycle*** mostra que logo que consegue resultados positivos em ambientes controlados (P&D), dispara-se o gatilho[37] da publicidade[38]. A tecnologia recebe muita publicidade feita pelos próprios fornecedores, pela mídia especializada e pelos veículos destinados ao público geral. Como existe pouca maturidade em relação ao assunto e quase nenhum resultado em ambientes não controlados, o foco da publicidade é nos aspectos positivos, ainda potenciais, da tecnologia. O clima geral é de *'isso muda tudo!'.* Esse tipo de visibilidade inflaciona as expectativas em relação à inovação e leva rapidamente a curva até o pico de expectativas[39]. Com o aparecimento dos resultados dos entusiastas que adotam primeiro a tecnologia, dá-se início a fase de ponderação das promessas onde percebe-se rapidamente que *'não é bem assim'.* A mídia coloca em pauta o lado negativo da inovação e as expectativas caem até atingir o poço da desilusão[40]. Essa fase é mais que um ajuste entre expectativa e realidade do mercado, porque arrasta consigo analistas superficiais que acreditam estarem diante de uma desqualificação da tecnologia. Nesse momento a ideia *'sai de moda*, a mídia para de falar ou quando fala apresenta um tom negativo ou pelo menos cauteloso. A partir daí especialistas buscam corrigir os rumos tomados, surgem novas gerações de aplicações e a exposição a diferentes olhares cria novas oportunidades de negócio. A tecnologia entra

[36] www.gartner.com/technology/research/methodologies/hype-cycle.jsp
[37] Tradução nossa para *technology trigger*, termo em inglês utilizado pelo Gartner.
[38] Entendida aqui como a exposição pública de algo, através de diferentes meios e formatos e sem nenhuma restrição quanto aos objetivos.
[39] Tradução nossa para *peak of inflated expectations*, termo em inglês utilizado pelo Gartner.
[40] Tradução nossa para *trough of disillusionment*, termo em inglês utilizado pelo Gartner.

na escalada de esclarecimento[41] até atingir o platô de produtividade[42], onde o objetivo e benefícios já estão mais claros, já existem metodologias e processos de implantação, uma rede confiável de fornecedores de serviços e pós-venda e indicadores de sucesso. A mídia volta a dar atenção ao assunto com um enfoque mais ponderado e a tecnologia permanece mais uma vez em foco durante algum tempo até que sua adoção entra no objetivo comum de todos fazendo a mídia perder o interesse.

Todo esse movimento define a curva mostrada na figura 2. Como já foi dito, a curva de Gartner é teórica e cada tecnologia a percorre no seu tempo e com picos e depressões também peculiares, mas a trajetória da curva costuma ser idêntica a essa.

A outra curva citada, a de Adoção de Inovação, também conhecida como curva de adoção de tecnologia, foi introduzida pelo sociólogo norte-americano Everett Rogers (1982). Refere-se ao comportamento da adoção de uma inovação pelos seus usuários potenciais que são classificados em cinco grupos: entusiastas, visionários, pragmáticos, conservadores e céticos. É representada pelo formato de sino da curva normal onde os entusiastas estão no começo da cauda anterior. Esse grupo é o primeiro a adotar a novidade assumindo os riscos da inovação ainda sem utilidade e qualidade comprovadas, e correspondem a aproximadamente 2,5% do mercado potencial. Logo depois, ainda na cauda anterior da curva, aparecem os visionários. Eles estão na fronteira da investigação sobre inovação em seus respectivos setores de atuação e rapidamente percebem uma conexão entre a nova tecnologia e as suas próprias necessidades. É importante ressaltar que os visionários, nesta classificação de Rogers, são adotantes com capacidade de investimento já que, geralmente, uma tecnologia nova tem custos de adoção elevados. Os visionários correspondem a 13,5% do mercado. Os pragmáticos ocupam a primeira metade do centro da curva, representando 34% do mercado. São aqueles que visam a obtenção de receita e por isso só adotam a inovação depois que identificam um histórico seguro de sucesso que garanta um retorno sobre o investimento em médio prazo. A segunda metade do centro da curva é ocupada pelos conservadores. Eles são 34% do mercado que só adotam inovação tardiamente, depois que a ideia passa a ser consenso geral e os preços de adoção já caíram bastante. Os conservadores são mais imediatistas que os pragmáticos na expectativa da relação custo-benefício. No final da curva, na cauda posterior, estão os céticos, aqueles resistem à mudança e só a fazem quando a

[41] Tradução livre de *slope of enlightenment*, termo em inglês utilizado pelo Gartner.

tecnologia anterior não está mais disponível ou quando ficar de fora se configura em ameaça a sobrevivência da organização.

Geoffrey Moore (2001), um teórico das organizações, percebeu que a adoção de uma tecnologia não era tão contínua como representada pela curva, mas apresentava quebras[43] entre um tipo adotante e outro, que eram vácuos na adoção da tecnologia em relação à linha do tempo. Especificamente entre os visionários e os pragmáticos, o autor destacou uma "quebra" que merecia maior destaque: **"**A quebra que existe entre esses dois mercados (visionários e pragmáticos) é tão significativa a ponto de se justificar chama-la de abismo"[44] (MOORE, 2001, p. 5). A figura 3 apresenta a curva de adoção de inovação de Rogers com o *"abismo"* proposto por Moore.

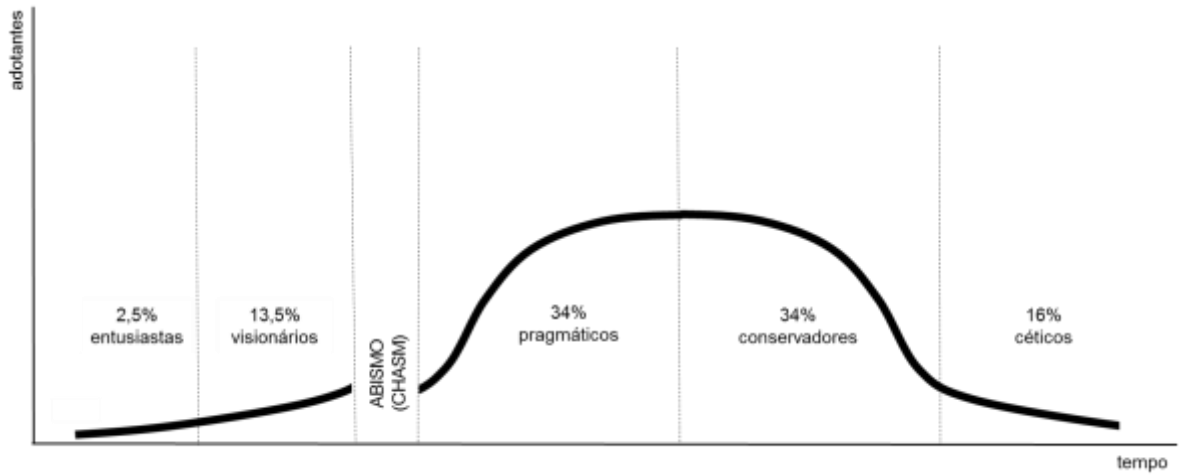


Figura 12 - Curva de Adoção de Inovação + *Chasm* - Fonte: Moore, 2001

Os gráficos do Gartner e de Rogers/Geoffrey possuem eixos diferentes, portanto não podem ser sobrepostos. No entanto, podemos inferir a passagem de tempo no *Hype Cycle*[45] e mesclar as informações dos dois. A ideia do movimento *hype* dialogar com o de adoção de tecnologia não é original. Várias consultorias de TI fizeram isso e uma busca simples na Internet mostra que a maior parte dos modelos propostos alinham o 'poço das desilusões' do *Hype* Cycle com o 'abismo' da curva de adoção da inovação. A Figura 4

[42] Tradução livre de *plateau of productivity,* termo em inglês utilizado pelo Gartner.
[43] Tradução nossa para o termo em inglês *crack in the bell curve,* utilizado pelo autor
[44] Tradução oficial do termo em inglês *chasm*, utilizado pelo autor
[45] A passagem do tempo não é explícita no *Hype Cycle*, mas pode ser naturalmente inferida

apresenta uma proposta diferente gerada a partir de uma figura (ver no Anexo I) integrante de um material explicativo do Gartner sobre a *Hype Cycle* de 2014.

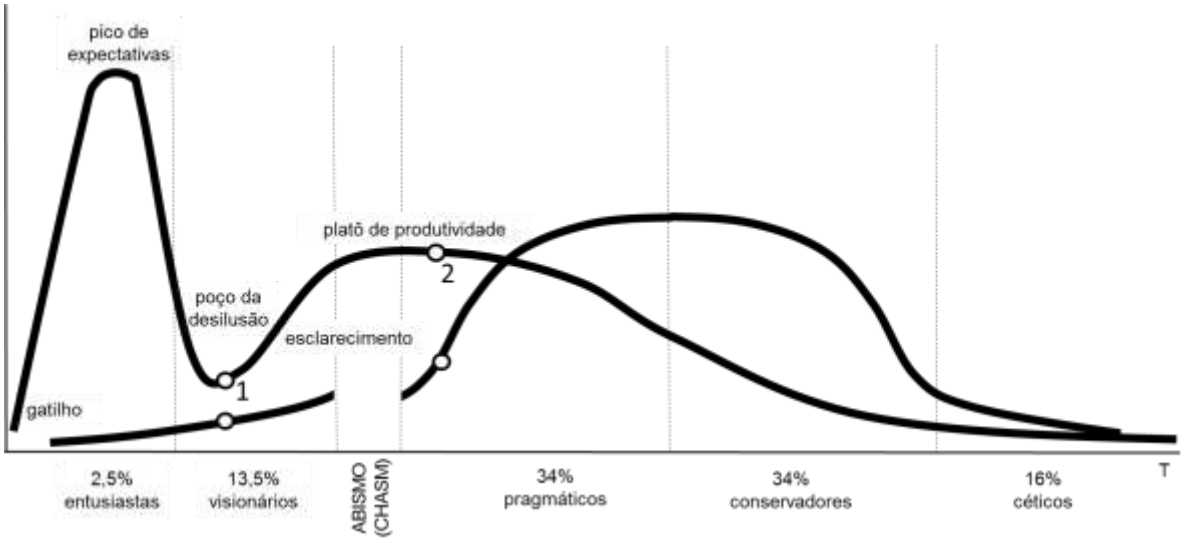


Figura 13 – *Big Data: Hype Cycle Graph* + Curva de Adoção de Inovação – Fonte: própria

Segundo o relatório, quando a tecnologia inicia a escalada de esclarecimento (ponto 1), atingiu apenas 5% do mercado potencial e no topo do platô de produtividade (ponto 2), algo entre 20 e 30%. Marcamos os pontos na curva *Hype* e traçamos a curva de adoção de tecnologia para atender a essas informações do relatório: 5% do mercado corresponde a todo o mercado de entusiastas (2,5%) mais uma parte do mercado dos visionários; e 20 a 30% do mercado, já incorporou todos os visionários e já alguns pragmáticos.

Essa sobreposição das curvas não entra em confronto com as definições de Rogers para cada tipo de adotante: os entusiastas adotam a tecnologia antes de virar moda, os visionários percebem a importância antes da inovação voltar a ter visibilidade na mídia, os pragmáticos aguardam o assunto se estabilizar e garantir o mínimo de segurança, os conservadores entram quando a mídia já começa a não ter mais novidades para divulgar e os céticos adotam a inovação quando ela já é corriqueira que não pauta mais as matérias da imprensa.

Uma última informação foi útil para a concepção do problema dessa dissertação: o Gartner, além de posicionar as tecnologias em algum ponto da do *Hype Cycle*, faz

projeções para o atingimento do platô de produtividade. O relatório de 2014, exatamente quando comecei esse trabalho, colocava *big data* começando a descer o "pico das expectativas" e estimava que o "platô" seria atingido em cinco ou dez anos. A figura integrante desse relatório que contém essa informação está no Anexo II.

Voltemos mais uma vez aos gráficos combinados para ver esses pontos identificados na curva *Hype Cycle* e associá-los à projeção de adoção da inovação (figura 5).

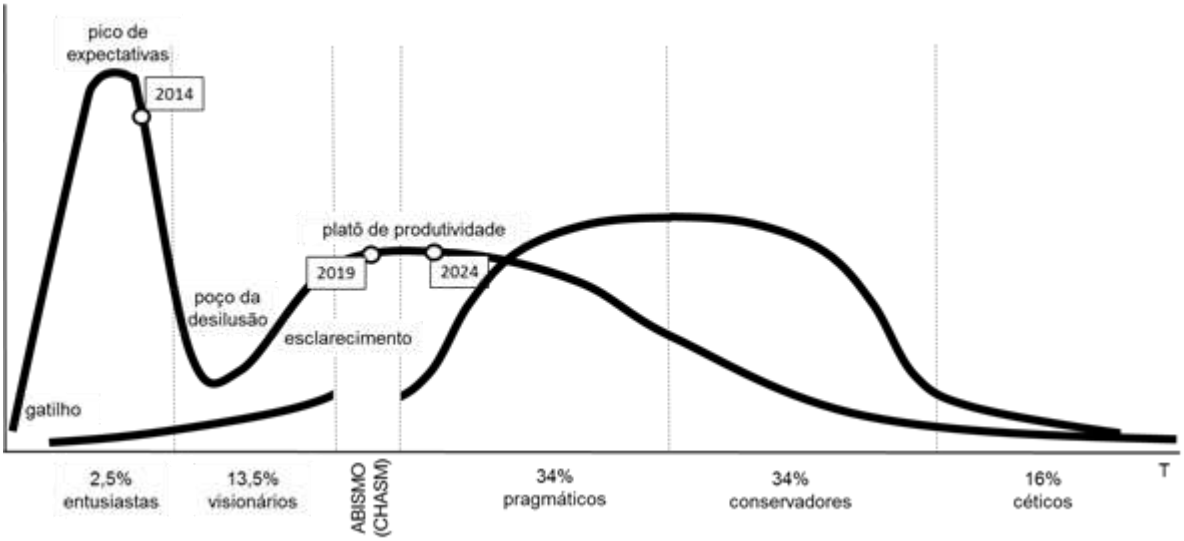


Figura 14 - Projeção para trajetória de *big data* no *Hype Cycle* – Fonte: própria

A partir de 2015 o instituto parou de representar '*big data'* na curva *Hype Cycle*[46] e em seu lugar passou a colocar as tecnologias associadas[47] separadamente, cada uma delas em pontos diferentes da curva[48]. Por isso não podemos comparar os relatórios para saber o deslocamento de *big data* pela curva em 2015, mas o bom senso e a percepção que temos do interesse da mídia pelo assunto, nos diz que '*big data'* já desceu o 'pico das expectativas' e aparentemente também já superou o 'poço das desilusões'. Confirmadas as previsões do Gartner, em breve enfrentaremos o 'abismo' que precede o afluxo dos pragmáticos ao mercado (34% do mercado potencial).

[46] Esse movimento do Gartner foi a concretização da noção de que big data não é uma tecnologia, e sim um fenômeno e quando passa a ser considerado como vital, passa a ser uma filosofia, como será defendido na sessão 2.3. Existem várias tecnologias que permitiram o aparecimento do fenômeno e outras tantas que permitem a geração de valor a partir de dados digitais.
[47] Analítica preditiva, Internet das coisas, *Cloud computing, Linked data*, Linguagem natural, etc
[48] Esta figura é apresentada no Anexo III

1.3.2. Relevância do problema

Em seu livro *Crossing the Chasm* Moore (2001), afirma que a chave para ir além do mercado dos entusiastas e visionários é mostrar que a nova tecnologia permite um salto estratégico e para isso deve-se evitar o apelo exclusivamente técnico. É preciso mostrar aos pragmáticos e conservadores mais que máquinas, equipamentos e softwares de última geração.

Sabemos que as forças motoras do fenômeno *Big data* foram a Internet, a personalização dos computadores, a tecnologia de comunicação móvel, evolução dos sensores e processadores e a computação em nuvem. Talvez isso justifique o fato de que o esforço de evangelização do mercado para geração de valor a partir de dados digitais tenha se concentrado no desenvolvimento e venda desses produtos e serviços. Tudo isso é bastante compreensível, no entanto precisamos de uma nova abordagem se desejamos que dados digitais estejam mais presentes na cadeia produtiva de conhecimento e no empoderamento das tomadas de decisão em nossas organizações. Entendemos que em torno dos dados e da infraestrutura tecnológica que os transforma em informação existem agentes não digitais (pessoas, objetos, meio ambiente e instituições) que dão contexto, significado, direção e utilidade a esse processo. Tudo isso junto, dados, infraestrutura tecnológica e agentes não digitais formam uma rede que chamo nesse trabalho de ecossistema *big data*[49]. Conhecer essa rede, em suas partes e relações, e saber como gerenciá-la será fundamental para que organizações adotem as tecnologias associadas a ela com garantia de retorno sobre o investimento.

1.3.3. Definição do problema

Diante de tudo isso, o problema que essa dissertação procura resolver é: Como estruturar e apresentar um *framework* para servir como ferramenta de diagnóstico, projeto, comunicação e gestão efetiva de processos de tomada de decisão baseada em dados digitais (DDD)[50]?

[49] Na literatura encontraremos a expressão ecossistema *big data* com outros significados, como veremos na sessão 2.4.
[50] **DDD** é a sigla para ***data driven decision***, a expressão em inglês para **tomada de decisão baseada em de dados digitais**. Nesse trabalho usaremos alternadamente as três designações.

1.3.4. Objetivos

A pergunta-problema dessa pesquisa vai gerar três *outputs*:

Output* 1.* Uma visão sobre ecossistema *big data* compatível com o escopo estratégico de uma organização. A visão construída servirá para gerar o *output* 2.

***Output* 2**. A definição de um ecossistema *big data* ideal para a governança de processos de decisão baseada em dados digitais*.* O ecossistema será usado para gerar o *output 3.*

***Output* 3**. Um *framework* útil para diagnóstico, projeto e comunicação de processos DDD.

O *output* 3 é o objetivo específico dessa pesquisa enquanto que os dois primeiros são rastros que ela necessariamente deixará pelo caminho.

1.3.5. Resultados esperados

Ao criar um *framework* para gestão eficaz de um processo DDD, esperamos contribuir para o desenvolvimento de boas práticas de governança do ambiente digital de modo que se potencialize e realize toda expectativa em torno de *big data*.

1.4. Metodologia

> Aquele que só aceita a viagem quando conhece uma boa estrada não me interessa. Eu quero aquele que vai, mesmo que não exista o caminho.
>
> David Livingstone

O caminho se faz ao caminhar, portanto nem toda metodologia desse trabalho foi definida a priori.

O que já estava no mapa desde a partida:

Esse trabalho tem três *outputs* bem definidos sobre o ecossistema *big data*: uma visão, um ecossistema e um *framework*. Para todos eles, os aspectos centrais estão na dinâmica das suas relações constituintes e não em suas variáveis quantificáveis (SILVEIRA e CORDOVA, apud GERHARDT e SILVEIRA, 2009 p. 32). Por esse motivo, quanto à sua abordagem essa pesquisa é qualitativa. Isso quer dizer que as informações necessárias são coletadas sem instrumentos formais ou estrutura pré-definida e a compreensão do objeto de estudo é feito a partir de uma análise sistêmica dessas informações, produzindo uma narrativa intuitiva, não exclusiva e não definitiva.

A pesquisa é aplicada quanto à sua natureza, sem pretensão de gerar conhecimentos novos, nem alcançar verdades universais. Os conceitos e proposições construídos aqui são focados em problemas objetivos e permitem matização por parte de quem os utiliza de modo a manter sua vocação pragmática e utilitária.

Quanto ao seu objetivo, o trabalho se enquadra na definição de pesquisa exploratória através da qual buscamos proporcionar maior familiaridade com o conceito *big* data, seu valor potencial e desafios para adoção, com vistas a torná-lo mais explícito para seus *stakeholders* ou a construir hipóteses sobre a melhor maneira de obter resultados com as tecnologias a ele associadas (GIL, 2007, apud GERHARDT e SILVEIRA, 2009 p. 35).

Finalmente, quanto aos procedimentos adotados, é basicamente uma pesquisa bibliográfica ventilada com outros rituais metodológicos a saber:

Os *outputs* visão e ecossistema, em conjunto, aproximam-se em seus objetivos aos da construção de uma ontologia, uma vez que busca uma “especificação formal e explícita de uma conceituação compartilhada” (BORST, 1997, apud SILVA, SOUZA e ALMEIDA, 2008). Em outras palavras, buscamos inicialmente uma conceituação sobre *big data* que, embora original, possa ser compartilhada por outros pares a partir da sua divulgação. A isso daremos o nome de visão e a partir dela definimos nosso ecossistema de modo a deixar a menor margem de ambiguidade possível. Não adotamos o rigor de uma ontologia no que diz respeito à designação de nomenclaturas, o que ficou como sugestão para futuros trabalhos.

A conceituação compartilhada (visão) é construída a partir do resgate dos trabalhos de cientistas da complexidade, que atuam na recém-criada Ciência das Redes. Essa comunidade foi a primeira a identificar que dados gerados em rede refletem o comportamento de sistemas complexos e, portanto, dar sentido a esses dados é mais que ser moderno e adotar uma nova tecnologia. Trata-se de uma missão que todo ser humano deverá abraçar com cuidado porque eles são o código que nos mostrará como dar equilíbrio ao nosso planeta.

Tudo indica que o referencial teórico da Ciência das Redes se perdeu conforme o assunto escalou para a curva da visibilidade em direção ao “pico das expectativas”, quando a falta de embasamento teórico domina o tom das publicações sobre um determinado assunto.

Esse resgate teórico que embasará a nossa visão de *big data* é feito das fontes das mais importantes instituições de pesquisa sobre o tema: *Northeastern University* em Boston-EUA, *Santa Fé Institute* em Santa Fé-EUA, *Intitute for Advanced Studies* (IMT) em Lucca na Itália, *Massachusetts Institute of Technology* (MIT) em Massachussets-EUA e *University of Notre Dame* em Indiana-EUA de onde saiu o artigo mais citado em toda história da ciência moderna[51].

O que foi incorporado no caminho:

Não foi encontrado na literatura um ecossistema que atendesse à visão para *big data* que definimos. Tornou-se necessária a definição de um ecossistema próprio. Para isso partimos da definição original do termo ‘ecossistema’ no campo da biologia, passamos para a definição clássica do termo adaptada para o mundo dos negócios na década de 80, quando Edward Freeman popularizou o termo *stakeholder* e, a partir daí, entendendo a importância de um ecossistema e de um *stakeholder* para o planejamento de governança, definimos o que estava em nosso ecossistema.

As definições e descrições dos elementos do ecossistema são feitas pelo senso comum. Segundo Silva *et al* (2008), apesar do grande número de ontologias já desenvolvidas, não há um padrão na sua construção, embora eles identifiquem alguns processos que têm

sido comumente seguidos como usar o conhecimento consensual, extraído da literatura e linguagem do cotidiano. Segundo o autor, esse processo é especialmente útil comum para ontologias criadas para sistemas relativamente novos ou de escopo bastante específicos. Acreditamos que o ecossistema big data se enquadre nas duas condições.

Para a proposição do *framework* partimos da sua definição. Existem diversas definições para framework, mas adotamos a mais abrangente: “Um framework é uma abstração da realidade que se propõe a identificar classes de elementos relevantes e suas interrelações” (GALDO, p. 16, no prelo). Isso quer dizer que ele está diretamente relacionado com os objetivos e com a visão de realidade que o proponente possui.

A metodologia para construção foi o senso comum inspirado por outras experiências. Apesar de estar no horizonte da pesquisa desde o início, não definimos metodologia a priori porque esperávamos encontrar facilmente essas orientações. No entanto, para nossa surpresa, não encontramos nada muito estruturado.

Buscas no Google para “como fazer um *framework*” (entre aspas) retornaram dois resultados. A busca em inglês “*how to do a framework*” retornaram quatro resultados. Os seis documentos referiam-se a *frameworks* de *software,* algo específico entre profissionais das TI. Sem aspas, a pesquisa em português retornou aproximadamente 570 mil resultados e em inglês, 300 milhões. As primeiras 10 páginas de resultados nas duas línguas também eram sobre *softwares.* As buscas por variações tipo ‘como fazer um *framework* para negócio, ou para *business*, também não foram bem-sucedidas. Encontramos dois documentos, ambos em inglês, que deram alguma direção. O primeiro era a página pessoal de James Booth[52], um desenvolvedor de *softwares* que criou um guia de *design* de *frameworks* para *softwares*. O segundo era a *Community Tool Box*[53], uma plataforma colaborativa de ferramentas onde encontramos uma orientação, um conjunto de passos, para criar um *framework,* o que pode ser visto no Anexo IV. Esse documento apresenta uma semelhança considerável com a percepção que tínhamos sobre o que precisava ser feito. Juntamos informações generalizáveis das duas fontes e seguimos a seguinte orientação:

[51] *Emergence of Scaling in Random Networks* (1999, BARABÁSI, Albert-László e ALBERT, Réka)
[52] http://www.jamesbooth.com/designing_your_own_framework.htm
[53] http://ctb.ku.edu/en/4-developing-framework-or-model-change

Princípios qualitativos de um bom *framework* (BOOTH):

- Simplicidade
- Clareza
- Limites
- Expansibilidade

Passos para construir um *framework* (*Community Tool Box)*

- Definir objetivo
- Definir o *output* esperado com seu uso
- Definir uma visão
- Definir escopo
- Identificar todos os componentes dentro do escopo
- Faça um desenho com uma organização lógica
- Teste a lógica e a abrengência do desenho com exemplos conhecidos
- Use/Feedback/Refaça

O *Business Model Generation*[54] de Alex Ostervalder, não foi uma inspiração a princípio, pelo menos não consciente. Certa vez, apresentando o *Big Data Estratégico* para alunos de graduação em Engenharia de Produção da Coppe, alguém reparou na semelhança entre os *frameworks.* Como já conhecia e admirava o BMG Canvas, uma inspiração inconsciente pode ter ocorrido. Buscamos a tese de doutorado de Ostervalder no entanto, nela o autor faz apenas a ontologia de um modelo de negócio, deixando para apresentar o *framework* depois em um contexto comercial, não deixando os rastros da sua metodologia. De qualquer modo, a partir de então consultamos várias vezes o trabalho de Ostervalder sobre o BMG (a tese e o livro).

1.5. Estrutura do texto

Esse documento está estruturado em cinco seções:

[54] Também conhecido como BMG Canvas.

A seção 2 apresenta o referencial teórico que foi subdividido por assunto: 2.1. Ciência das Redes; 2.2. Complexidade e problemas complexos; 2.3. *Big Data* e 2.4. Ecossistema *Big Data*.

A seção 3 apresenta O *framework* objetivo dessa pesquisa. A seção está subdividida em: 3.1. Visão, 3.2. Ecossistema *Big Data* Crie, 3.3. Big Data Estratégico: um *framework* para o Ecossistema *Big Data* Crie e 3.4. Como usar o framework? *Check List*.

A seção 4 traz as conclusões da pesquisa e a seção 5, as recomendações para futuros trabalhos.

A seção 6 lista a bibliografia consultada e em seguida são apresentados os Anexos e Apêndices.

2. Referencial teórico

> Uma das razões pelas quais o futuro não pode ser previsto é porque ele pode ser influenciado.
>
> Adam Kahane

Até 2010 *big data* não era um assunto de grande interesse no Brasil. *"Big data não foi pesquisado com frequência suficiente para figurar nesse gráfico"*[55] foi a resposta que o *Google Trends*[56] quando perguntamos sobre o volume de pesquisa pelo termo feito no Brasil em 2010. O resultado dessa busca pode ser visto no Anexo V, na impressão da tela feita na época. Apenas em 2012 o *Google Trends* captura esse interesse e aponta um pico de interesse pelo assunto em maio de 2013. A curva tendia a voltar à normalidade quando registrou um novo pico em junho do mesmo ano. A Figura 6 mostra esse gráfico com as duas datas em destaque, juntamente com a curva para o mesmo período nos Estados Unidos, apenas para efeito de comparação[57].

[55] *Not enough search volume to show graphs*
[56] https://www.google.com/trends/ - ferramenta do Google que mostra os principais assuntos pesquisados na Internet através do Google
[57] Experiência feita em 2015 e repetida em fevereiro de 2016 com o mesmo resultado. Em 17/03/2016 a experiência foi repetida encontrando resultados levemente diferentes, como pode ser visto no Apêndice III, provavelmente por alteração de algoritmo do Google. No entanto, a diferença não inviabiliza as conclusões tiradas aqui.

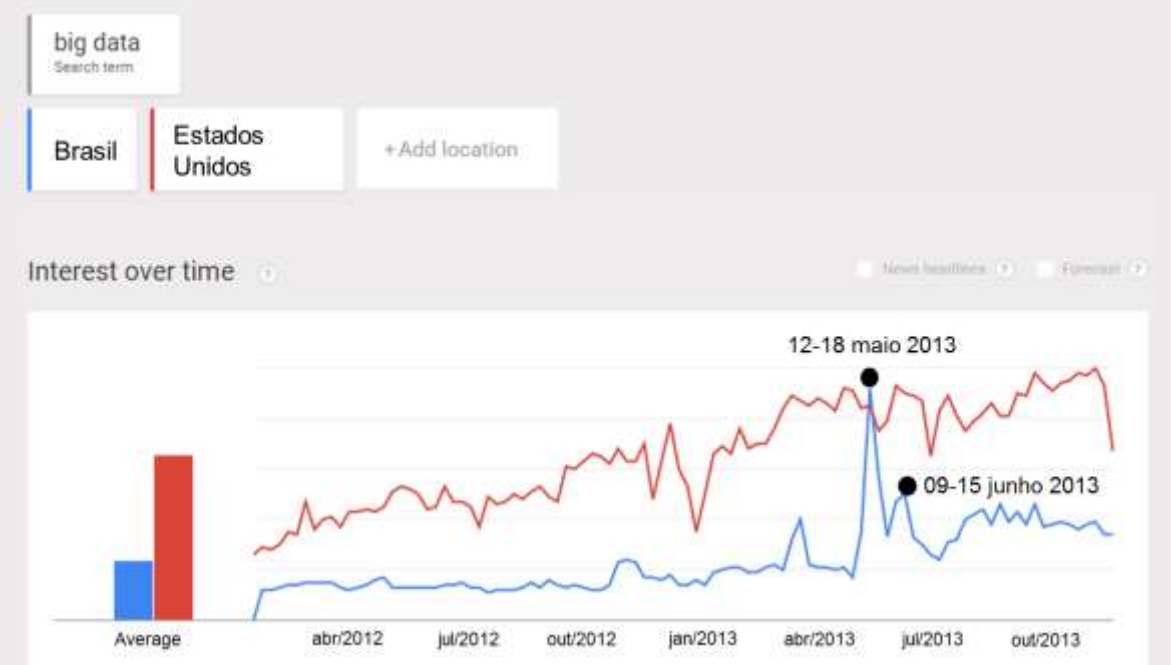


Figura 15 - Busca para *big data* (Brasil e Estados Unidos - 2012 a 2013) - Fonte: *GoogleTrends*

Pesquisas na Internet sobre notícias nessa época revelaram que a Revista Veja, uma publicação de grande circulação do Brasil, fez uma reportagem de capa[58] sobre o assunto, exatamente na semana de 12 a 18 de maio. A matéria da revista não era negativa mas chamava atenção para algo que mudaria tudo radicalmente. Apontava o uso de *big data* em diversas áreas de atuação humana e apresentava casos reais de outros países.

Alguns dias depois, especificamente em 5 de junho do mesmo ano, Edward Snowden[59] revelou ao jornal britânico *The Guardian*[60] as espionagens que a NSA, Agência de Segurança Nacional dos Estados Unidos, fazia utilizando dados de ligações telefônicas e outros dados de navegação na internet de pessoas em todo mundo. As matérias sobre esse vazamento foram publicadas no Brasil em 9 de junho pelas revistas Veja[61] e

[58] Ver capa da revista no Apêndice IV
[59] Ex-funcionário da CIA e da NSA, Agência de Segurança Nacional dos Estados Unidos, que tornou públicos detalhes do sistema de vigilância global da agência americana (Wikipedia)
[60] http://www.theguardian.com/world/2013/jun/06/nsa-phone-records-verizon-court-order
http://www.theguardian.com/world/interactive/2013/jun/06/verizon-telephone-data-court-order
[61] http://veja.abril.com.br/noticia/mundo/ex-tecnico-da-cia-se-responsabiliza-por-vazamentos-sobre-espionagem-nos-eua

Exame[62]. Logo depois, o jornal O Globo publicou[63] que a NSA teria espionado também e-mails e ligações telefônicas de brasileiros[64].

Existem fortes indicações de que essas matérias tenham provocado o súbito interesse pelo assunto e o afluxo de curiosos para a Internet em busca de mais informações. Infelizmente, as matérias não explicaram o motivo pelo qual os dados digitais gerados pelo uso da Internet geram modelos descritivos e preditivos de valor, muito menos destacaram que utilizar esses dados é algo vital para que possamos dar o salto de conhecimento que precisamos para resolver os problemas sócio-econômicos-ambientais que enfrentamos. É natural, portanto, que grande parte das pessoas não esteja atenta à importância da analítica digital para o alcance dos objetivos da sociedade, no âmbito provado e público.

Também não foi dito que usar dados digitais não significa necessariamente quebra de anonimato ou privacidade. Por isso, compreendemos que exista uma certa resistência de gestores na adoção de *big data*, seja por convicção pessoal de que é algo ameaçador, seja por resistência do seu cliente, no caso de empresas, ou do cidadão, no caso dos governos.

A informação é a única maneira de reverter esse quadro de desinteresse, descrença ou medo, que justificam os atrasos na adoção de uma inovação tão promissora.

A principal fundamentação teórica para a proposta desse trabalho são os princípios da Ciência das Redes. Complementam essa seção referencial teórico sobre complexidade e problemas complexos, *big data* e ecossistema *big data*.

Acreditamos que com esse embasamento, construiremos uma visão consensual e uma ferramenta útil para auxiliar gestores públicos e privados a incorporarem o uso de *big data* em seus processos de tomada de decisão.

[62] http://exame.abril.com.br/mundo/noticias/edward-snowden-o-novo-delator-do-governo-obama
[63] http://oglobo.globo.com/mundo/eua-espionaram-milhoes-de-mails-ligacoes-de-brasileiros-8940934
[64] Veja, Exame e O Globo são publicações de grande visibilidade no país

O tema ‘tomada de decisão’ é um campo especifico e já bastante aprofundado pela ciência, no qual não nos cabe uma incursão por não fazer parte do escopo desse trabalho. No entanto, como a tomada de decisão é o objetivo central do *framework* que será produzido, achamos importante esclarecer alguns pontos essenciais:

A Internet foi o gatilho para uma nova revolução cognitiva na história da humanidade, o que acontece quando surge uma tecnologia capaz de ampliar consideravelmente nossas funções cognitivas[65]. Consideraremos tomada de decisão como a função cognitiva (KELLOG, 2007) responsável pelo processo de elaboração de um plano de ação a partir da consideração de informações disponíveis, visando um determinado resultado (definição nossa). Fica claro que a internet, apenas com o grande volume de informação que armazena e circula, empodera a nossa capacidade de decidir, mas o que está em jogo quando se fala em *big data* é algo de outra dimensão já que com a digitalização da informação temos acesso aos dados brutos que podem ser combinados e recombinados criando novas informações e meta informações. Fica claro ainda, pela definição acima, que consideramos a tomada de decisão como uma tarefa exclusivamente humana já que que exige visão de passado e de futuro, capacidades específicas da espécie humana.

Na descrição do ecossistema e do *framework* propostos, voltaremos a falar da importância desses pontos destacados no processo de tomada de decisão.

Todos os assuntos do referencial a seguir possuem literatura acadêmica, no entanto em alguns momentos fazemos referência a publicações do setor produtivo que, no contexto digital, muitas vezes tem sido mais prolífico e inovador que a academia.

[65] NEPOMUCENO notas de aula, 2012 e blog do autor, diversos. Carlos Nepomuceno é pesquisador especialista em Internet e suas causas e consequências na sociedade. Segundo ele, citando Pierre Lévy (notas de aula), a última vez que tivemos uma revolução cognitiva foi no século XV com a invenção da prensa tipográfica por Gutemberg que permitiu que mais pessoas tivessem acesso direto a diversos tipos de informações e ideias, mas a revolução só se deu porque a tecnologia também permitiu que as pessoas se comunicassem.

2.1. Ciência das Redes

> Você pode chamar de Ciência das Redes, dinâmica humana, computação social ou big data. Não importa o nome que você dê, existe uma quantidade imensa de informações coletadas sobre nós e temos a obrigação de dar um sentido a tudo isso.
>
> Albert-László Barabási – cientista

A informação de que existe um gigantesco banco de dados inéditos já é suficiente para revolucionar qualquer área do conhecimento porque aumenta exponencialmente as possibilidades de descoberta de informações e correlações inusitadas.

No entanto, a promessa inovadora do que chamamos de big data vem da possibilidade de tornar os sistemas complexos não apenas mais compreensíveis, mas também mais previsíveis e controláveis.

Os sistemas complexos, como o código genético, o sistema nervoso e a sociedade por exemplo, são aqueles cuja estrutura é formada por um grande número de partes que se relacionam de forma complexa e que, muitas delas são desconhecidas até hoje. São sistemas complexos todos os sistemas e subsistemas naturais.

A ciência usa rede como uma ferramenta para visualizar e modelar um sistema qualquer. Uma rede é graficamente representada por pontos ligados por linhas, onde os pontos são os objetos e as linhas os relacionamentos entre eles.

As redes podem ser artificiais ou reais. Até o final do século XX, só conseguíamos modelar matematicamente redes artificiais, aquelas cuja estrutura (nós e links) eram fixas e conhecidas, como por exemplo as estradas que ligam cidades, linhas de metrô ligando estações ou movimentos das peças do xadrez no tabuleiro. Essas redes são objetos de estudo de um campo da Matemática que surgiu no século XVIII, a Teoria dos Grafos.

Sistemas complexos também são representados em forma de rede e por representarem sistemas naturais, essas redes, chamadas de redes reais, ajudam a visualizar e guiar o estudo dos sistemas complexos, no entanto, até o início do século XXI, os cientistas não conseguiam formular matematicamente o comportamento desses sistemas.

Durante muitos anos, o desconhecimento das propriedades dos sistemas complexos fez a ciência acreditar eles se comportavam de modo randômico ou aleatório, o que significava dizer que não existia nada que explicasse a formação, crescimento, reorganização e comportamento dos sistemas naturais. Seriam como uma caixa preta, definitivamente inacessível ao conhecimento humano. Aos olhos do observador, esses sistemas tinham comportamento completamente aleatórios, e por isso eram estudados pela Teoria das Redes Randômicas.

Alguns cientistas discordavam desse posicionamento científico da época e estavam convencidos da existência de leis e princípios que organizavam e governavam os sistemas naturais, mas não tinham como provar. Numa frase que ficou imortalizada Einstein, por exemplo, afirmou intuitivamente que “Deus não jogava dados com o universo”. A Teoria do Caos e a Teoria dos Fractais, surgiram nesse contexto de negação do comportamento randômico dos sistemas naturais e como uma tentativa de atribuir alguma lógica a eles.

Uma experiência de Stanley Milgran sobre a interconectividade social é uma das primeiras tentativas de encontrar provas para a intuição de Einstein. Milgran (1967) solicitou a várias pessoas que fizessem uma correspondência chegar às mãos de uma determinada pessoa escolhida por ele. A única orientação para fazer isso era: se você conhece essa pessoa, envie o envelope diretamente a ela; se não, envie a alguém que você conhece que pode ter mais chance que alcançar o destinatário. Assim o experimento seguia até que o envelope chegasse a seu destino. Analisando posteriormente o trajeto de cada envelope que chegou ao destino, Milgran percebeu que a média de intermediários era seis e formulou a hipótese de que a rede social era totalmente interconectada e que, por isso mesmo, para conectar um nó da rede a qualquer outro, as pessoas sempre procurariam o melhor caminho, em vez de agir aleatoriamente.

Especulações a partir dessa experiência levaram à teoria dos seis graus de separação que afirmou que não só todas as pessoas de uma determinada população estão de alguma forma interligadas, mas também que o número máximo de caminhos necessários para relacionar quaisquer duas pessoas era seis. Isso poderia ser o primeiro passo para modelar e compreender a topologia da complexidade, mas a pesquisa, embora tenha causado sensação e grandes especulações, não foi levada muito a sério pela falta de rigor científico com que foi realizada. A experiência foi feita com 296 cartas e apenas 64 chegaram ao destino. Além disso, todas partiram de Omaha, uma cidade no interior do estado de Nebraska nos Estados Unidos, e tinham como único destino a cidade de Boston, no estado de Massachussets. Certamente poderia haver um viés na amostra e no processo que impediria a generalização das conclusões.

Alguns cálculos matemáticos endossavam a teoria da total interconectividade social. De fato, se cada pessoa conhecesse apenas duas outras no mundo estava garantido que, de algum modo, todo nó da rede estaria conectado a todos os outros. No entanto, a teoria dos seis graus de separação ainda era algo a se provar. A experiência precisava ser repetida em uma escala global, com uma amostra bem maior e testando conexões de qualquer ponto a qualquer ponto, não apenas um único trajeto, uma empreitada muito difícil de conseguir.

No início da década de 80, Tim Berners-Lee criou um programa que permitia que computadores compartilhassem informações. Ele não tinha consciência disso, mas estava criando as condições para comprovar as ideias de Milgran: em 2008 a Microsoft repetiu a experiência de utilizando 30 bilhões de mensagens eletrônicas de 180 milhões de usuários do MSN Messenger chegando surpreendentemente ao mesmo resultado. Se a experiência de Milgran tinha sido desacreditada pela comunidade cientifica os resultados da Microsoft parecem endossá-la.

A rede de comunicação, postal ou eletrônica, parece ser uma boa representação do sistema social. De fato, durante a década de 90 e início dos anos 2000, a obsessão pela teoria dos seis graus de separação inspirou diversos experimentos semelhantes, com populações distintas e amostras diferentes, que obtiveram como resultado sempre a mesma ordem de grandeza de seis graus. Por isso, além de definir o grau de separação

máximo dos nós do sistema social, as experiências parecem também comprovar a propriedade fractal nesse sistema: as propriedades do todo estão presentes também nas suas partes.

Apesar das críticas a Milgran, a teoria dos seis graus de separação nunca foi refutada o que sugere, segundo os pesquisadores da Ciência das Redes, algum tipo de princípio envolvido.

Em 1999, quando a web já contabilizava algo em torno de 1 bilhão de documentos, cientistas da universidade de Notre Dame, nos Estados Unidos decidiram testar se a propriedade da total interconectividade se aplicava também à essa rede. Barabási e seus colaboradores utilizaram dados de navegação cedidos por amigos que administravam sites de grande tráfego, e concluíram não só que a web era totalmente interconectada, mas também que entre quaisquer dois documentos dessa rede, o grau de separação era no máximo 11. Ou seja, para chegar a qualquer documento da internet, a partir de um ponto qualquer, não seriam necessários mais que 11 links. A semelhança com o número sugerido pela experiência de Milgran realizada trinta anos antes levou ao surgimento da teoria dos “poucos graus de separação”.

Quando as experiências de Milgran e Barabási apontam um número máximo de caminhos para ligar dois nós de uma rede complexa, está implícito o caráter não randômico dessas redes. Ao contrário, o menor caminho jamais seria encontrado se a escolha e a ordem dos nós intermediários fossem aleatórias.

Barabási contabilizou também a quantidade de links que cada nó da www apresentava em uma determinada amostra. A hipótese era a de que todos os nós dessa rede (no caso os sites) tinham as mesmas chances de serem vistos e acessados, o que caracteriza uma rede randômica. Numa representação gráfica, uma rede randômica teria sua distribuição de nós formando uma curva normal, com formato de sino (Figura 7). A curva normal sempre carrega a informação de que “a maioria dos nós (da rede) possui o mesmo número de links e não existem nós com grandes quantidades de links” (BARABÁSI, 2001, p.64).

Figura 16 – Curva Normal ou de Sino (eventos aleatórios) – Fonte: Linked, 2002

No entanto, a curva encontrada para a distribuição da quantidade de links para cada nó da www foi a curva de grau em lei de potência (Figura 8).

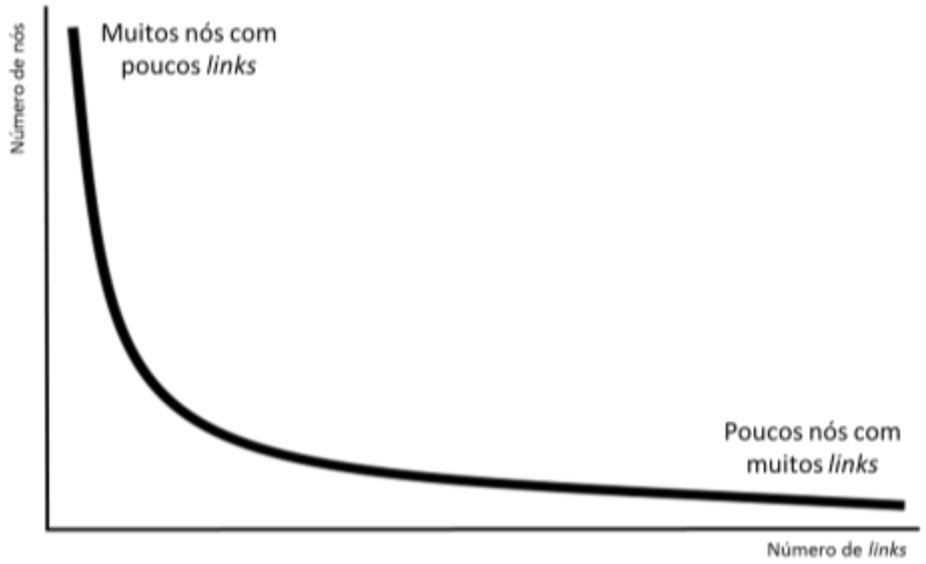


Figura 17 – Curva de grau em lei de potência (redes sem-escala) - Fonte: Linked, 2002

Essa curva mostra que alguns poucos nós são altamente conectados enquanto que a maioria apresenta poucas conexões. Eventos com nós altamente conectados são exclusividade de sistemas complexos, emergentes e auto organizados. Esses eventos

simplesmente inexistem em fenômenos aleatórios. As redes que apresentam esse tipo de configuração são chamadas redes sem-escala.

O conhecimento que temos hoje da www nos levaria a esse conhecimento intuitivo. Claro que sites como o Google ou a Amazon, tem muito mais links com outros nós da rede do que o blog de um professor de uma universidade, ou de uma pequena rede de comércio local, por exemplo. Mas é importante lembrar que a experiência foi feita em 1999 e nessa época ainda se acreditava que a rede era completamente igualitária. Se qualquer pessoa pode criar um site na Internet e estabelecer os links que desejar com outros sites, o que explica o fato de a www não apresentar uma distribuição normal onde a maioria dos sites teriam mais ou menos a mesma quantidade de links?

> Os *hubs* são os mais fortes argumentos contra a visão utópica de um ciberespaço. Sim, nós todos temos o mesmo direito de colocar qualquer coisa que desejemos na Web. Mas, todos os documentos serão notados com a mesma ênfase? Se a Web fosse uma rede randômica nós teríamos a mesma chance de sermos vistos e ouvidos (BARABÁSI, 2002, posição 983, tradução nossa)

Segundo Barabási, apesar de ser uma rede artificialmente construída, a www se comporta como uma rede real, justamente porque a sua configuração é livre e espontânea, refletindo o comportamento humano no mundo físico.

O *Hypertext Transfer Protocol* (http), acordo social que foi criado para a utilização da Internet e que felizmente vigora até hoje, é o protocolo base da comunicação na Internet e permitiu que a *world wide web* surgisse, se desenvolvesse e se comportasse como uma rede real ou natural se assemelhando muito com o sistema social que já conhecíamos fora do ambiente digital, ou seja, exercemos nossas escolhas e preferências.

> Quando escolhemos os links que faremos na Web, estamos seguindo nossas preferencias de conexão. De um modo geral, uma página que é mais visível e visitada tem muito mais chances de receber novas conexões do que uma página menos conectada. Embora estejamos livres para fazer as conexões que desejamos, e isso torna um comportamento individual imprevisível, como grupo nós seguimos padrões bem rigorosos, portanto previsíveis (BARABÁSI, 2002, posição 1409, tradução nossa).

A descoberta de que as redes naturais não eram randômicas era o início de uma nova era para a ciência. Segundo Barabási (2002), “nas últimas décadas os cientistas reconheceram que ocasionalmente a natureza gera grandezas que seguem a distribuição da lei de potência em vez da curva de sino”.

O resultado dessa pesquisa, intitulado *Emergence of Scaling in Random Networks* foi publicado na Science Magazine, dando início à era das pesquisas digitais sobre a complexidade.

Isso abre um mundo de possibilidades para o estudo de redes complexas naturais, já que sugere que resultados obtidos facilmente em estudos de dados digitais podem ser analogamente aplicados à outras redes que apresentam naturalmente maior dificuldade na coleta e análise de dados, facilitando assim a compreensão do seu surgimento, organização e comportamento.

Os avanços da Ciência das Redes podem revolucionar outras ciências que estudam sistemas complexos específicos. Como desconheciam as propriedades que governavam as redes complexas, essas ciências no século XIX e grande parte do século XX reduziram os sistemas estudados às suas mínimas partes na tentativa de compreendê-los. Quase todo sistema complexo conhecido já foi descrito através da enumeração e codificação das suas partes. Até hoje, esse método é utilizado e tem grande importância na ciência, mas não é suficiente para os problemas complexos.

Cada descrição que a ciência faz das partes de um sistema complexo cria uma expectativa na sociedade de que todos os problemas que envolvem esse sistema possam ser resolvidos. No entanto, frustrações sucessivas nesse sentido têm nos mostrado que esse reducionismo, embora necessário, não é suficiente, porque alguns problemas e consequentemente suas soluções, estão nas relações estabelecidas na rede, e não em suas partes isoladas.

Um bom exemplo disso é Projeto Genoma Humano. Ao identificar e descrever todos os genes humanos, a ciência e a sociedade esperavam um súbito aumento na eficácia dos tratamentos de algumas doenças. No entanto, o que se viu no ano seguinte foi uma

redução de mais de 80% na quantidade de registros de medicamentos comparados com anos anteriores. Essa redução é um alerta de que soluções não foram encontradas nas partes do sistema, ou seja, nos genes (Barabási, 2012).

> Uma lista de genes supostamente envolvidos (no surgimento de determinada doença) não é suficiente. Para encontrar a cura para a maioria das doenças precisamos compreender o sistema vivo em sua integridade. Precisamos decifrar como e quando diferentes genes trabalham juntos, como suas mensagens são transmitidas nas células, que reações estão acontecendo, ou não, em um determinado momento e como os efeitos de cada reação se espalham por todo organismo. (BARABÁSI, 2002, posição 2848, tradução nossa)

Ainda segundo o autor, a boa notícia é que finalmente os cientistas estão aprendendo a mapear interconectividade e esses mapas estão colocando uma nova luz nos estudos de sistemas complexos, revelando surpresas e desafios nunca imaginados há até poucos anos (2002, posição 179, tradução nossa).

O resultado do projeto Genoma não é para ser desprezado, mas não é suficiente. Um modo de analisar positivamente a redução dos lançamentos de remédios é que, por causa das informações que ficaram conhecidas pelo sequenciamento genético, a ineficácia de muitas drogas pôde ser apontada antes delas alcançarem o mercado, mas trata-se de um logro econômico que não atenua os desafios no domínio da medicina.

A Ciência das Redes tem se dedicado então a encontrar propriedades na www que possam ser generalizadas para outras redes complexas.

Embora usemos os termos redes e sistemas como sinônimos, aqui é importante que uma distinção seja feita: a rigor, rede é uma "abstração que nos permite codificar relacionamentos entre pares de objetos" (FIGUEIREDO, 2013, material de aula[66]) de um sistema e não pode ser confundida com o próprio sistema.

Barabási (2002), define redes complexas (sistemas) como aquelas que se reconfiguram continuamente não só através do surgimento e desaparecimento de nós, mas também pela reorganização de links. O cientista observou que essa reconfiguração é feita através

[66]http://www.land.ufrj.br/~daniel/grafos/slides/aula_1.pdf

de duas capacidades que se complementam: a capacidade preferencial e a capacidade competitiva.

A capacidade preferencial pode ser ilustrada na experiência de Milgran, pelo fato de cada pessoa escolher com autonomia para quem enviaria o envelope. As remessas não eram feitas de modo aleatório, ou os envelopes, se algum dia chegassem ao destino, jamais seria em apenas seis remessas apenas.

A capacidade preferencial de certa forma explica porque sair na frente atribui uma vantagem ao nó. Muitas das concentrações de links em um determinado nó de um sistema são explicadas por essa capacidade de ser eleito como preferencial por ter se estabelecido primeiro. Na experiência postal, caso fosse feita com uma amostra bem maior, seríamos capazes de perceber que algumas pessoas da sociedade receberiam mais cartas do que outras porque as pessoas a identificavam como a pessoa com mais probabilidade de encontrar o destinatário. Essas pessoas seriam os hubs da rede. Aquele nó que tem mais conexões e, portanto, têm mais chances de estar conectados com o nó desejado.

A capacidade competitiva está naquele nó que, não raramente, rouba a cena e passam a atrair mais links que outros nós, mesmo os mais antigos. Essa característica de um nó explica formação dos hubs, ou seja, a dominação de um novo entrante em um ambiente aparentemente estável.

As duas capacidades, preferencial e competitiva, atendem à necessidade de adaptação dos sistemas complexos visando a própria sobrevivência e elevam o nível das incertezas da ciência para um novo patamar: como se dá esse alinhamento entre a necessidade de sobrevivência de um sistema e o comportamento da sua estrutura? Ou o que faz um nó competir e ter sucesso no estabelecimento de relações? O que faz os outros nós elegerem esse nó como relevante e estabelecerem preferencialmente links com ele?

Enquanto não se compreende o que leva ao sucesso ou à performance diferenciada de cada link, a Ciência das Redes identifica propriedades comuns às redes complexas. É

como se a explicação não fizesse muita diferença se podemos provar que a topologia de rede de um sistema complexo tem um padrão que determina seu comportamento.

O comportamento global de um sistema complexo é o resultado coletivo das ações dos seus componentes, mas isso está longe de ser uma simples soma das partes. Não existe um controle central das ações, mas existem propriedades que configuram o comportamento desses sistemas (MITCHELL, 2006, p. 2, tradução nossa)

Hidalgo (2015, p. 93, tradução nossa) aponta que estabelecer e manter cada relação de um sistema complexo tem um esforço associado. Os nós envolvidos avaliam que esforço vale a pena ser desprendido e essa avaliação está diretamente associada com a importância que cada relação tem com perpetuação do sistema. Essa capacidade que um sistema complexo tem de discernir e empreender esforços para sobreviver é uma propriedade muito importante quando pretendemos controlar sistemas complexos. Para ficar mais claro, apresentamos um exemplo bem conhecido, de como a propriedade de perpetuação ajudou a Organização Mundial de Saúde a decidir sobre procedimentos para tentar controlar a de Ebola que havia surgido na África em 2013. Em setembro de 2014 o Ebola se tornou oficialmente global com a confirmação do primeiro caso de transmissão da doença fora dos limites geográficos de sua origem. Alguns países sugeriram a proibição de voos entre os países da África contaminados e o resto do mundo. Tom Frieden, diretor do Centro de Prevenção e Controle de Doenças nos Estados Unidos manifestou-se contra:

> Paradoxalmente, isolar esses países aumentará o risco de que a doença se espalhe para o resto do mundo. As pessoas continuarão a se deslocar entrando e saindo dos países afetados, a despeito de qualquer proibição. E esse tipo de deslocamento será difícil de rastrear.
> (http://www.forbes.com/sites/jvchamary/2014/10/13/ebola-travel/)

Em outras palavras, proibir os voos faria com que indivíduos desesperados usassem rotas alternativas, por terra ou mar, para escapar e esses links ficariam invisíveis aos olhos de quem está tentando fazer o controle.

Em outro contexto, a governadora do estado do Arizona, responsável por fazer cumprir as leis federais de imigração, disse a respeito da dificuldade de conter os imigrantes ilegais

que cruzavam a fronteira do México com os Estados Unidos: “Mostre-me um muro de 15 metros e eu lhe mostro uma escada com 16 metros[67]” (NAPOLITANO[68], 2005, apud Arthur, 2014, p. 2, tradução nossa).

Esse comportamento dos agentes da rede real contornando um obstáculo para benefício próprio é o que pode ser reconhecido como *hackear* o sistema. A palavra é originalmente do domínio da Informática, mas tem sido usada em qualquer contexto para descrever modificações e manipulações não previstas ou não autorizadas pelo sistema[69]. Por essa capacidade de reconfiguração constante esses sistemas são conhecidos como sistemas complexos adaptativos, ou CAS, do inglês *complex adaptive systems* (MITCHELL, 2011).

A ciência das redes encontrou ainda outras coisas interessantes sobre os sistemas complexos. Eles estão interligados com outros sistemas complexos, formando em sua representação uma rede de redes. Muitas dessas interconexões ainda são completamente desconhecidas, mas existem indícios de que temos um universo totalmente conectado, mesmo que ainda não existam provas concretas disso.

As conexões inter redes são feitas por links específicos que chamamos de links pontes ou links fracos. O termo “fraco” pode nos iludir e fazer pensar que sejam pouco importantes, no entanto esses links são fundamentais no estudo das redes. Eles garantem a oxigenação dos sistemas trazendo conteúdo de outros. *Links* fracos têm um papel essencial na nossa capacidade de comunicação com mundos externos (...). Para conseguir novas informações precisamos ativar nossos *links* fracos. (BARABÁSI, 2002, posição 751, tradução nossa)

Num sistema social, o link fraco é aquele que permite, por exemplo, que encontremos oportunidades (pessoais e profissionais) fora do nosso círculo menor (GRANOVETTER, 1973, apud BARABÁSI, 2002). Num sistema de conhecimento, o link fraco nos garante maiores possibilidades de correlação entre informações de diferentes domínios. Em última instância, os links fracos aumentam a chance de sobrevivência de um sistema já que são as ligações que garantem a conexão vital com todo ambiente do qual faz parte.

[67] *Show me a 50-foot wall and I’ll show you a 51-foot ladder*.

Os links de uma rede complexa possuem conteúdo, direção e peso. “O peso captura o valor da relação e a direção aponta do provedor para o recebedor do *link”.* (BARABÁSI, 2002, posição 3273, tradução nossa).

Numa relação de amizade, por exemplo, o peso se dará pela quantidade e qualidade de ocorrências que fortalecem a relação como troca de mensagens, quantidade e duração de encontros pessoais, diversificação de assuntos tratados, etc. Numa relação comercial o peso se mostrará nas quantidades de compras e vendas feitas e no volume e valor das transações. (BARABÁSI, 2002)

A direção refere-se ao caminho que pode ser percorrido na rede. Em uma determinada rede pode-se partir de um nó para outro, mas pode ser impossível fazer o caminho inverso. Exemplos disso são páginas na internet que possuem links que levam a sites maiores cujo caminho inverso inexiste, cadeia de predadores, onde o teor alimentar (energia) segue numa só direção e rede de fãs de um determinado artista, em que um fã sabe tudo do seu ídolo e no sentido contrário a relação quase inexiste.

Conhecer a direção dos *links* de uma rede é fundamental para alguns processos de tomada de decisão. Se você deseja ser apresentado ao presidente do país, por exemplo, você precisa encontrar uma pessoa cuja relação com esse presidente permita que ela o aborde, ou seja, o *link* entre eles tem que necessariamente ter essa direção. Se essa pessoa conhece pessoalmente o presidente, mas a relação existente só permite que ela seja solicitada e não que ela solicite um encontro, essa pessoa não logrará sucesso na pretensa apresentação.

Outro exemplo simples é a rede de conexões de voos. Uma determinada companhia pode fazer um voo direto da cidade A para a cidade B, mas o contrário pode não ser verdadeiro. Para ir de B a A o passageiro precisa fazer conexão ou escala em outra cidade.

[68] Janet Napolitano, ex-governadora do estado do Arizona, EUA
[69] Wikipedia

A rigor, uma rede complexa bem representada deve identificar não apenas a quantidade de links, mas também a direção (geralmente representada através de setas) e o peso (representado pela espessura da linha ou números indicativos) de cada um deles. Para estudos específicos essa diferenciação pode ser fundamental para se compreender a dinâmica de uma rede.

No entanto, para grande parte dos estudos de redes complexas, a visualização da existência dos links, apenas as quantidades sem nenhum conteúdo, direção ou peso, é suficiente para compreender muita coisa do comportamento dessa rede. Nesse caso a representação é simplificada com a utilização de nós de igual tamanho e de linhas de igual espessura e sem setas direcionais. Usando o exemplo da rede de voos, saber que existe uma linha direta entre duas cidades, mesmo que não saibamos a direção e a frequência dos voos (peso), já nos informa que determinada companhia aérea atua nesse trajeto.

Outro exemplo é o estudo que as telefônicas fazem para alguns casos de investigação para ver se e quando alguém que está sendo investigado ligou para um possível cúmplice ou para sua vítima. Quando a autorização judicial para quebra de sigilo é feita tardiamente, não sendo possível gravar o conteúdo da conversa (peso da relação), a recuperação dos dados de conexão já serve de indício para confirmação ou afastamento de algumas suspeitas.

Outro conceito importante em redes complexas é o que o sociólogo Peter Marsden chamou de argumento central[70] core para referir-se ao conteúdo ou tipo de relação que estamos mapeando. Isso é importante porque cada nó de um sistema complexo faz parte de mais de uma rede ao mesmo tempo. Uma rede de pessoas, por exemplo pode ser definida pelas relações profissionais, de amizade, de parentesco etc. Ao mapear uma rede é importante que o argumento central esteja bem definido e que somente as relações associadas a ele sejam representadas[71].

[70] Tradução nossa para a expressão em inglês *core discussion* utilizada pelo autor
[71] Para representar diferentes relações de um mesmo nó, é preciso usar a sobreposição de mapas

Os sistemas complexos possuem duas propriedades cujo conhecimento é vital para gerenciar seu desenvolvimento e/ou evitar ou provocar seu colapso total: robustez e vulnerabilidade.

A robustez é a capacidade que um sistema tem de resistir a ataques randômicos, ou seja, é a sua resiliência. “Aparentemente os sistemas naturais conseguem sua robustez por causa da alta interconectividade das suas estruturas. Essa escolha universal que a natureza faz na arquitetura de suas redes certamente não é uma mera coincidência” (BARABÀSI, 2002, posição 1800, tradução nossa). Como a dinâmica dos sistemas naturais é sempre em busca da perpetuação, a arquitetura *hiper* conectada foi a que mais se adequou a esse propósito.

A vulnerabilidade também é um princípio resultante da estrutura altamente interconectada dos sistemas naturais. Robustez e vulnerabilidade como princípios decorrentes da mesma causa é apenas o paradoxo aparente. Se a robustez é a resistência a ataques randômicos, a vulnerabilidade é a suscetibilidade a ataques específicos. Isso quer dizer que, uma vez conhecida a estrutura e topologia de uma rede, ataques cuidadosamente planejados passam a ser potencialmente danosos à sua perpetuação. “Apesar das redes com configuração sem-escala serem vulneráveis a ataques, é preciso que um grande número de relações ou nós específicos sejam removidos simultaneamente para causar o colapso. Isso significa, por exemplo, destruir ao mesmo tempo de 5 a 10% dos *hubs* da rede (BARABÁSI, 2002, posição 1910, tradução nossa)

Robustez e vulnerabilidade não são intrinsecamente características boas ou ruins. Tudo depende da rede da qual estamos falando e dos objetivos de quem deseja mantê-la ou destruí-la. O ecossistema natural é resiliente aos constantes ataques que os humanos fazem contra sua estrutura. Como os ataques são randômicos, ou seja, ninguém tem intenção de destruir a natureza, ela tem resistido bem. Cientistas cuidam para identificar ações humanas que signifiquem golpes fortes na estrutura do ecossistema, que exponham sua vulnerabilidade. Uma rede terrorista é um exemplo de rede que desejamos conhecer bem seu funcionamento para destruí-la. Ataques randômicos a grupos terroristas têm se mostrado, como sabemos, ineficazes no controle do desenvolvimento desses grupos.

Os registros digitais representam uma grande oportunidade para a ciência, porque oferecem acesso inédito aos dados de estrutura e dinâmica de uma rede complexa por excelência, a internet, que codifica os relacionamentos entre pares de objetos de um sistema real, a sociedade.

> No estudo dos sistemas complexos, emergentes e auto organizados (a moderna Ciência da Complexidade), as redes se tornando cada vez mais importante como uma ferramenta matemática universal, especialmente porque uma massiva quantidade de dados está disponível. (CALDARELLI e CATANZARO, 2012, posição. 334, tradução nossa)

Além disso, as características e propriedades inferidas desses estudos podem ser de grande utilidade para compreender outras redes complexas conhecidas.

Admitindo a existência de padrões na dinâmica humana, com dados suficientes é possível descrever e quantificar um problema e formular hipóteses. Um problema matematicamente formulado ganha poder preditivo e, a partir daí é possível desenvolver mecanismos de controle que é o objetivo maior de toda ciência (BARABÁSI, 2012, tradução nossa). “A última prova do nosso entendimento de um determinado sistema, seja ele tecnológico ou natural, está refletida na nossa habilidade em controlá-lo” (LIU, SLOTINE e BARABÁSI, 2011, tradução nossa).

A Ciência das Redes usa a palavra controlar com naturalidade e sem pudor porque não atribui a essa ação um juízo de valor. Isso porque as propriedades das redes complexas se manifestam de modo diferente nos diferentes tipos de sistemas. O esforço de estabelecer um link, por exemplo, pode ser um custo financeiro, o dispêndio de tempo, o esforço físico, o desgaste emocional ou o esforço que um vírus faz para vencer um sistema imunológico.

A capacidade de *hackear* um sistema pode se manifestar na fraude de um sistema fiscal sem cometer ilícito ou na espantosa capacidade humana de reaprender a fazer coisas que foram “desaprendidas” quando uma parte do cérebro sofre uma lesão irreversível.

Um link fraco pode ser aquela pessoa que sempre traz novidades de outros grupos ou lugares que apenas ela frequenta, ou pode ser aquele médico que foi contaminado como vírus do Ebola na África e voltou ao seu país na Europa.

A robustez de uma rede pode ser a resistência de um vírus que se transforma em epidemia ou a resiliência do ecossistema natural de não sucumbir mesmo com os constantes ataques às suas espécies e intervenções no seu equilíbrio.

Se as propriedades de um sistema não são necessariamente boas ou ruins, também não têm esse juízo de valor a gestão que podemos fazer nesses sistemas. Controlar uma rede pode ser negativo se falamos de espionagem do governo à vida privada dos cidadãos, mas é extremamente positivo se estamos falando de impedir ataques terroristas ou catástrofes naturais.

“Sistemas não são imunes a manipulações feitas por um grupo pequeno que joga para tirar vantagens”, (ARTHUR, 2014), mas temos que saber que em ciência das redes ‘manipulações’, ‘jogar’, ‘tirar vantagem’ e outras expressões não expressam necessariamente algo bom ou ruim.

O entendimento que teremos das redes complexas a partir do estudo dos dados digitais certamente poderá ser usado para controle de todo tipo. Por mais que isso pareça assustador, é bom lembrar que esse fenômeno não é exclusividade do ambiente digital. Todo sistema complexo tem nós e relações lutando para a sua perpetuação, assim como sempre está ameaçada por outros nós e relações que insistem em destruí-lo. Abrir mão da ferramenta por receio de empoderar os destruidores de redes é fechar os olhos para o fato de que, em alguns casos o destruidor pode ser o bem-intencionado.

2.2. Problemas complexos

> Para cada problema complexo existe uma solução clara, simples e errada.
>
> H. L. Mencken

A Ciência das Redes Complexas mostrou no início do século XXI que a analítica digital em *big data* é de grande utilidade para o entendimento da complexidade. Dados digitais permitem a modelagem de problemas complexos que, formulados matematicamente, ganham poder preditivo podendo ser gerenciados e controlados (BARABÁSI, 2012)

Como podemos definir e abordar os problemas complexos?

Usualmente consideramos complexo como sinônimo de complicado, mas a distinção entre os dois conceitos é fundamental quando estamos falando de problemas e precisamos eleger abordagens de atuação.

Problemas complicados e complexos têm em comum o fato de envolverem muitas variáveis a serem consideradas, mas as similaridades acabam aqui. Uma boa revisão da literatura sobre as diferenças entre problemas complicados e complexos foi feita por Glouberman e Zimmerman, que resumimos na Tabela 1:

Tabela 3 - Problemas complicados X problemas complexos[72]
Fonte: Glouberman e Zimmerman (2002)

	PROBLEMAS COMPLICADOS	PROBLEMAS COMPLEXOS
Importância das fórmulas, receitas ou manuais	Crítico e necessário	Aplicação limitada
Relação da experiência com o sucesso	Aumenta a chance de sucesso	Não tem uma relação direta com o sucesso / Não é necessário nem suficiente
Importância da Expertise e superespecialização	Imprescindível	Pode contribuir, mas o melhor é a diversidade de olhares
Resultado buscado	Alto grau de certeza. Busca de uma solução	Alto grau de incerteza. Busca-se apenas o controle através da gestão do problema.

[72] Traduzido e adaptado

Em resumo, problemas complicados são finitos, previsíveis e pertencem a domínios de conhecimento definidos. Sabemos que habilidades são necessárias para resolvê-los, e a dedicação constante ao assunto garante vantagem competitiva na solução do problema. Já os problemas complexos são o resultado de tantas outras variáveis, muitas delas ainda desconhecidas, que fica difícil ter segurança da abordagem a ser feita e da resultante das ações.

Adam Kahane, um dos pesquisadores mais respeitados no campo da complexidade, faz uma descrição diferente de problemas complexos, mas chega a conclusões semelhantes. Para ele um problema complexo apresenta sempre três dimensões: social, generativa e dinâmica (2008, p. 24). A social é a dimensão do olhar: cada um vê o problema de um jeito diferente porque tem conhecimento e experiências diferentes e por isso é impactado de modo diferente por ele. A dimensão generativa significa que ninguém sabe como começou, porque problema complexo não tem causa única e não se conhece exatamente as relações entre as causas. A dimensão dinâmica tem a mesma explicação que a generativa: é impossível dizer exatamente como um sistema complexo vai se comportar e a única certeza que temos é que qualquer interferência terá ressonância contínua, passando de nó em nó através dos links *hiper* conectados.

Essas dimensões não apenas caracterizam os problemas complexos, mas trazem também algumas implicações. Alguns dados que poderiam ser relevantes podem estar tão distantes em relação ao problema que se tornam invisíveis. “Geralmente apenas os sintomas estão disponíveis, de onde o tomador de decisão deve inferir as causas do problema” (FUNKE, 1991, tradução nossa). Outra implicação é o fato de que, lidar com um problema complexo significa tentar alcançar vários objetivos ao mesmo tempo e, como se a multiplicidade de gols não bastasse, é comum haver conflito de interesses entre eles. Isso requer uma grande habilidade de equalização e capacidade de *trade-off.*

As duas maiores implicações das dimensões de um problema complexo estão ligadas aos resultados esperados. A primeira é que um problema complexo não pode ser resolvido. À primeira vista isso pode parecer fatalista e desencorajador, mas o que significa de fato é que um problema complexo é um fenômeno que deve ser abordado continuamente. A segunda é que não podemos planejar com segurança um *outcome* de uma ação. Tudo o

que podemos fazer é intervir no sistema, ir avaliando os resultados e reorientando as ações.

Educar um filho é um excelente exemplo de problema complexo. Nunca sabemos de onde vêm determinados comportamentos e não sabemos exatamente qual será o resultado das nossas ações. Não raro enfrentaremos situações de conflito de interesse como por exemplo na hora de deixar um filho sair sozinho; porque queremos sua segurança, mas também queremos que ele seja independente. Não existe um manual que se seguido garantirá o sucesso da empreitada e ter um primeiro filho tão pouco garante sucesso na educação do segundo. Por falar em sucesso, como avaliamos o sucesso de uma educação? Cada agente social julgará seu trabalho de acordo com valores e expectativas diferentes.

É preciso ressalvar que a distinção entre problema complicado e complexo é mais didática do que real. Uma doença por exemplo, fica no limiar entre as duas definições. É um problema complexo na medida em que vemos os sintomas, mas causas estão em outras partes sendo que algumas são desconhecidas; um mesmo procedimento pode não garantir cura em dois casos diferentes e a interferência no problema pode gerar consequências inesperadas. No entanto tratar uma doença se aproxima muito de um problema complicado quando a expertise e experiência têm forte relação com o resultado.

A Ciência das Redes tem se dedicado a compreender as propriedades das redes complexas para orientar esse movimento. No entanto, profissionais que atuam em algum setor produtivo, ansiosos por melhorar a efetividade[73] de suas ações se perguntam se é possível agir com racionalidade em ambientes complexos para conduzi-lo na direção de resultados desejados.

Irene Ng, especialista em marketing e serviços e pesquisadora do *Advanced Institute of Management* (AIM)[74], tem se dedicado à questão de projetar para resolver problemas complexos. Segundo ela, a dificuldade é em parte consequência do nosso legado de mais

[73] Efetividade = eficácia e eficiência. Em outras palavras significa melhores resultados com menos recursos.
[74] Instituto de pesquisa em gestão sediado no Reino Unido

de 100 anos projetando para problemas complicados, com grande taxa de sucesso (2011, tradução nossa).

Ng desenvolve sua pesquisa na zona cinzenta entre complicado e complexo; aquele espaço onde não tem como separar um do outro nem conceitualmente nem na prática, o que ela chama de sistemas sócio técnicos.

A pesquisadora reconhece que existe uma interação entre os dois (complicado e complexo) e que nem sempre conseguiremos distinguir um do outro e que, mas afirma que abordagem que busca performance de sistemas previsíveis (complicados) só garante resultados na performance desses sistemas. Se esse sistema colabora para obter soluções para problemas complexos, nem sempre a melhor performance levará a isso. Ng usa como exemplo exatamente questões ligadas a saúde e bem-estar que citamos anteriormente. Para o NHS[75], o indicador de sucesso é basicamente associado à eficiência: quanto mais pessoas forem atendidas, melhor. Se toda a gestão do sistema for voltada para melhorar os processos, podemos garantir que esse resultado será alcançado e isso seria uma abordagem típica de um problema complicado. Pode ser isso leve também a bons resultados para o problema complexo (saúde e bem-estar), mas também pode ser que haja uma correlação inversa nos resultados.

Sabe-se que as metodologias de abordagem de problemas complicados não são desprezíveis, mas já existe um consenso entre gestores públicos e privados de essas metodologias são inadequadas para problemas complexos.

Will Allen, especialista em gestão colaborativa afirma que,

> Em situações complexas, é importante resistir à ideia de encontrar ‘a solução’ que vai resolver o problema e, em vez disso, buscar pontos de alavancagem que podem acionados para que o sistema conduza melhor a situação. Mudar o comportamento de pessoas, por exemplo, pode exigir uma mudança nas regras ou nos acordos tácitos, nas relações formais, nos indicadores de sucesso ou até nas ferramentas e instrumentos usados. (2013, tradução nossa).

[75] *National Healthy Security.* Sistema público de saúde do Reino Unido

Ainda segundo Allen, trabalhar em problemas complexos exige uma aprendizagem contínua sobre o contexto onde eles se manifestam e um ambiente que promova a comunicação entre *todos os* envolvidos e interessados[76] (2001).

Se contexto e dinâmica dos agentes é fundamental para mapear e gerenciar problemas dessa natureza nos ambientes físicos, não seria diferente no ambiente digital. Construir algoritmos capazes de descrever, prever e gerenciar digitalmente problemas complexos exige o conhecimento anterior de toda a topologia da rede onde o problema se desenvolve.

Entendemos que conhecimentos de propriedades de redes e da dinâmica dos problemas complexos são competências essenciais para projetos de tomada de decisão baseada em dados.

2.3. *Big Data*

In God we trust; all others must bring data
William Edwards Deming

É difícil encontrar a origem da expressão *big data* que já existia muito antes da era dos dados digitais. Mesmo se fizermos um recorte no tempo buscando o termo apenas dentro do contexto digital, o trabalho se mostra inconclusivo, como nos conta o jornalista Steve Lohr em seu artigo intitulado *The Origins of 'Big data': An Etymological Detective Story.* Lohr começou sua pesquisa em 2012 e contou com a ajuda de pesquisadores das universidades de Yale e da Pensilvânia e de um analista do Instituto Gartner, no entanto, mesmo com esse reforço, a parte mais assertiva do artigo diz que "o termo *big data* (...) provavelmente se originou em conversas na mesa de almoço na empresa *Silicon Graphic*, em meados da década de 90[77]" (2013, tradução nossa). Existem outras versões como a conta que a primeira utilização do termo foi feita no início da década de 90 pelos

[76] Também conhecidos como *stakeholders* e *shareholders*, respectivamente.
[77] Publicado em 2013 no New York Times: http://bits.blogs.nytimes.com/2013/02/01/the-origins-of-big-data-an-etymological-detective-story/?_r=0

astrônomos da NASA[78], quando se depararam com os grandes conjuntos de dados complexos que desafiavam os limites tradicionais de captura, análise e armazenagem.

Apesar da incerteza sobre a data e autoria, existe o consenso de que o uso do termo no contexto digital surgiu na década de 90. O que cada um chamava de *big data* era o grande volume de dados que conseguia ver na sua área de atuação e que representava uma preocupação com a capacidade tecnológica em lidar com isso. Naquela época, se houve algum vislumbre de inovação disruptiva a partir desses dados, certamente era circunscrito ao domínio de conhecimento onde os dados eram gerados, ou seja, grande volume de dados gerado na NASA possibilitaria novos conhecimentos para a astronomia.

O entendimento sobre *big data*, tal como temos hoje, só seria possível no início do século XXI, após a disseminação das ideias do artigo *Emergence of Scaling in Random Networks* do cientista Albert László-Barabási, já citado no item 2.1.

Não existe uma definição formal para *big data.* As principais utilizações do termo referem-se a um grande conjunto de dados, a uma disciplina ou a uma ferramenta da tecnologia da informação.

***Big data* como conjunto de dados:**

Geralmente, aqueles que definem *big data* dessa forma complementam a definição com as quatro características que diferenciam esse *dataset* dos tradicionais '*data warehouses'*[79]: volume, variedade, veracidade e velocidade conhecidas também como os 4 Vs[80] do *big data.*

A primeira característica, **volume** dispensa maiores explicações. Trata-se da chamada escala *web*[81]: 2,5 bilhões[82] de gigabytes diários de dados que geramos a partir de

[78] Sigla em inglês de *National Aeronautics and Space Administration* (Administração Nacional da Aeronáutica e do Espaço). Agência do Governo dos Estados Unidos da América, responsável pela pesquisa e desenvolvimento de tecnologias e programas de exploração espacial.
[79] Grande banco de dados que armazenam dados estruturados de uma empresa
[80] A primeira referência aos 4 Vs para *big data* foi feita por Doug Laney em um relatório para o Instituto Gartner
[81] Tradução livre da expressão em inglês *web scale.*

postagens em redes sociais, *upload* de fotos e vídeos, registros de transações comerciais, sinais de GPS, rastros de navegação e sensores dos mais diversos tipos.

Por serem capturados nas mais diversas áreas de atuação humana, representado todos os domínios de conhecimento existentes, e também por possuírem formatos diferenciados, costuma-se destacar também essa dimensão crítica da **variedade**. Afinal, onde mais encontraremos ao mesmo tempo dados de mobilidade, saúde, finanças, economia, consumo, etc., senão em um *dataset* em escala *web*?

A dimensão conhecida como **veracidade** é mais fácil de ser verificada quando pensamos nos dados de dinâmica humana que são registrados quando interagimos em redes sociais ou deixamos nossos rastros de navegação. Os dados registrados são fidedignos porque representam interações reais. Outros métodos de captação de dados de dinâmica humana como entrevistas e observação, podem conter desvios próprios do método como ruídos na comunicação, parcialidade, interferência no fenômeno observado, etc.

O mesmo raciocínio vale para dados aquisitados[83] por sensores e armazenados automaticamente. Nesse caso, a veracidade estará não apenas no método de coleta automatizado que evita erros, mas também na possibilidade de se capturar dados com mais granularidade (seja reduzindo a periodicidade da captura dos dados, seja fazendo cada sensor representar um espaço cada vez menor) tornando mais representativa a base de dados do fenômeno estudado. Um exemplo clássico é o registro de condições climáticas. Um sensor digital é capaz de registrar as variações de temperatura, por exemplo, em tempo real (periodicidade de aquisição próxima a zero). Além disso, podemos espalhar cada vez mais sensores em uma mesma área, fazendo com que a coleta seja cada vez mais hiperlocalizada.

Entende-se que, aumentando a granularidade nas dimensões tempo e espaço, teremos sim uma maior quantidade de dados, mais a dimensão significativa deixa de ser o volume e passa a ser a veracidade já que o conjunto de dados terá maior fidedignidade ao fenômeno observado.

---

[82] Estimativa da IBM publicada em http://www-01.ibm.com/software/data/bigdata/ (página visitada em 05/07/2013)

Por fim, a dimensão mais característica do chamado *big data* é a que a literatura convencionou chamar de **velocidade.** Trata-se da possibilidade de se obter dados sobre um determinado fenômeno em tempo real, ou seja, no mesmo momento em que ele ocorre. Isso é fundamental para o valor das informações que podem ser geradas a partir desses dados, já que podem reduzir drasticamente o tempo entre a obtenção dessa informação e a tomada de decisão. Essa dimensão também define o processo de governança de dados, onde novas tecnologias são desenvolvidas para que essa característica fundamental não seja perdida. Gerar dados em tempo real tem mais valor se a tecnologia permitir sua utilização imediata da cadeia produtiva de soluções.

Big data **como disciplina**

Quando é utilizado para se referir a uma disciplina, o termo *big data* corresponde ao campo empírico de estudo de uma ciência. A ciência que utiliza dados digitais em suas pesquisas é a Ciência das Redes, que surgiu há pouco mais de 10 anos, como já relatado no item 2.1.

Big data **como tecnologia**

Muitas vezes ouvimos falar de *big data* como tecnologia (metodologia ou ferramenta) capaz de extrair valor[84] de uma grande quantidade de dados digitais.

Big data como tecnologia equivale à expressão **analítica de dados digitais**[85] que pode ser definida como o uso da tecnologia para a compreensão, comunicação e utilização inteligente dos dados digitais[86].

---

[83] Aquisitar é um neologismo já incorporado no vocabulário da TI significando capturar dados
[84] Alguns autores consideram o *valor* como o quinto **V**, ou seja, mais uma dimensão crítica para *big data.* Nesse trabalho optamos por desconsiderar essa dimensão por não ser uma característica intrínseca aos dados digitais e sim uma dimensão que se alcança após o esforço analítico exercido sobre eles.
[85] Em inglês, *data analytics* ou simplesmente *analytics*
[86] Definição nossa.

Existem basicamente quatro tipos de Analítica (CAVALCANTI, 2013)[87]:

Analítica Descritiva: se encarrega de analisar o que aconteceu;

Analítica Diagnóstica: analisa porque determinado evento aconteceu;

Analítica Preditiva: aponta o que irá (ou poderá) acontecer (predição);

Analítica Prescritiva: indica o que pode ser feito para alcançar um objetivo (reforçar, modificar ou evitar uma predição).

A rigor, analítica de dados pode ser feita em ambiente digitais ou não. Mesmo antes da existência dos computadores, gestores já praticavam os quatro tipos de analíticas. A analítica preditiva sem o uso de dados digitais é feita a partir de projeções dos dados disponíveis utilizando processos como proposição de cenários, simulações, *brainstorms* e também com o uso da intuição, permitido àqueles que possuíam mais experiência.

Conhecer o ambiente é fundamental para a sobrevivência em sistemas complexos, sejam espécies vivas no planeta, sejam empresas no sistema econômico. Todas as espécies vivas aprendem sobre o seu ambiente para continuar sobrevivendo, mas a espécie humana é a única que sobrevive para além da reação: o ser humano aprendeu a controlar os sistemas complexos. Para isso, a descrição e o diagnóstico são extremamente importantes, mas a efetividade do controle vem da capacidade de antecipar fenômenos, movimentos, comportamentos, e de desencadear ações adequadas. Ou seja, controlar depende da nossa capacidade preditiva e prescritiva.

A ideia de controlar sem ter descrição e o diagnóstico completos vai de encontro à máxima de que é impossível controlar o que não se pode medir. Trabalhar com *big data* desafia a lógica com que tentamos apreender o mundo. Geralmente nossa relação com o conhecimento é caracterizada pela busca obsessiva de descrições e explicações, mas em analítica preditiva digital, encontrar correlações nem sempre significa compreender o

[87] José Carlos Cavalcanti (UFPE) http://creativante.com/new/index.php/2013-02-03-19-36-05/2013-02-04-18-

fenômeno. “Uma correlação nem sempre nos diz exatamente porque alguma coisa está acontecendo, mas ela nos alerta para o fato de que ela está acontecendo e em muitas situações isso é suficiente” (SCHÖNBERGER e CUKIER, 2013, posição 221, tradução nossa). Com a analítica em *big data* podemos nos contentar com a ordem de grandeza das coisas.

Em *big data* os quatro tipos de analíticas são possíveis embora os maiores valores agregados aos dados estejam nas duas últimas (preditiva e prescritiva) já que as duas primeiras (descritiva e diagnóstica) já eram amplamente utilizadas e com sucesso antes do fenômeno *big data.* Isso porque, para descrever e diagnosticar, dados sobre o passado são extremamente úteis enquanto que para prever e prescrever, são necessários dados de tendências e dinâmica, que só foram possíveis de obter com o advento da Internet. “Em última instância, big data refere-se à possibilidade de antecipar o futuro” (SCHÖNBERGER e CUKIER, 2013, posição 184, tradução nossa)

Nesse ponto, aparece uma das principais diferenças entre os dados em *big data* e bancos de dados estruturados de grande porte.

Quando falamos em analítica preditiva, estamos falando de uma metodologia completamente diferente das previsões que estamos acostumados a ver a partir de *data warehouses* ou de pesquisas de opinião. O que determina a mudança na metodologia é a natureza dos dados utilizados.

A previsão (ou *forecast*) é feita a partir da utilização de grandes quantidades de dados de mesma natureza. A metodologia consiste em uma projeção, em que os dados viajam para dimensões maiores de tempo ou espaço.

O exemplo mais conhecido disso é a previsão do tempo. Dados de um mesmo domínio, que têm uma relação direta de causa e efeito com o fenômeno, são transferidos para o futuro (viagem no tempo) gerando um resultado com grau de incerteza conhecido.

19-49/155-big-data-e-analitica-ferramentas-e-modelos-de-negocios-revolucionarios

Outro exemplo é a pesquisa de opinião. Dados de uma mesma natureza, coletados em uma amostra, são transferidos para o total da população (viagem no espaço) gerando também um resultado confiável.

"Trabalhar com amostras em vez de usar o todo implica em um *trade-off:* a organização pode encontrar o que está procurando mais rapidamente e mais barato, no entanto ela jamais insights para questões que não considerou previamente" (SCHÖNBERGER e CUKIER, 2013, posição 385, tradução nossa). Não existe nada de errado nisso e os métodos de previsão baseados em amostras vão continuar existindo e sendo importantes. No entanto é preciso ficar atento para o fato de que desse modo só se obtém respostas para as perguntas que foram feitas.

O resultado de uma previsão é uma estimativa e a ciência que fundamenta a metodologia é a Estatística.

A predição (analítica preditiva ou *predictive analytics*) é feita a partir de uma grande quantidade de dados de domínios diferentes (*big data*). A metodologia é um julgamento baseado na experiência e no aprendizado[88] que se dá quando dados de um domínio do conhecimento viaja para outros domínios.

Em 2008, por exemplo, o Google conseguiu antecipar em 7 a 10 dias, a dinâmica da contaminação pelo vírus da gripe H1N1, baseado nos dados de utilização da ferramenta de busca dos seus usuários. Rastreando o conteúdo das buscas, a empresa conseguiu identificar correlações inusitadas entre as pesquisas e a contração da doença em um determinado grupo social. É importante entender que o modelo antecipava não apenas a quantidade de casos, mas também o deslocamento do vírus. O modelo foi tão bem-sucedido que fundamentou as ações preventivas do governo americano para impedir uma pandemia da doença.

Boa parte do sucesso preditivo desse projeto vem de percepção que os cientistas do Google tiveram de que os dados de busca da ferramenta deles era um repositório das

[88] Predição pode ser feita com dados digitais ou não. No contexto de dados digitais (*big data*), experiência e aprendizado correspondem ao auto ajuste do algoritmo feito por inteligência artificial.

intenções e curiosidades humanas (BATTELLE, 2006, tradução nossa). Portanto, quando a curiosidade sobre H1N1 em determinado local crescia muito, era indicativo de que a doença estava por perto e uma epidemia podia estar à vista. Pesquisamos algo no Google quando esse algo chega muito perto de nós, uma notícia boa ou ruim, uma ameaça, uma oportunidade. Além de saber que o número de casos estava aumentando naquele local, o Google sabia também para onde as pessoas desses locais poderia carregar o vírus. Isso se dava a partir da utilização das intenções de compras de passagens para deslocamentos entre cidades.

Além disso, diferente da previsão, a analítica preditiva é capaz de antecipar as mudanças nas tendências porque considera fatores que não têm uma relação causa-efeito imediata com o fenômeno. O que buscamos são correlações e "correlações não implicam em relação causal. A descoberta de uma relação entre dois fatores A e B não significa que que um cause o outro, nem mesmo indiretamente" (SIEGEL, 2013, tradução nossa). Como não há a necessidade de entender lógica causal, a analítica preditiva desafia o senso comum e permite que se encontrem respostas para perguntas que nunca foram ou serão feitas.

Afinal, o que é *big data*?

Os quatro Vs em *big data* foram descritos por um cientista do Instituto Gartner em um relatório feito em 2001. Naquela época fazia sentindo entender *big data* como um imenso conjunto de dados que apresentava essas características. Hoje, no entanto, essa definição é imperfeita e diz muito pouco sobre o real conceito de *big data*.

> Apesar do nome que carrega, o tamanho do *dataset* não é o que de fato define *big data.* O que classifica os projetos de empresas como Google como projetos de *big data* é que, em vez de usar amostras randômicas em suas analíticas, elas usam dados verossímeis ou mais aderentes a complexidade do problema que estão tentando modelar (SCHÖNBERGER e CUKIER, 2013, posição 429, tradução nossa)

O nome *big data* vem do fato de que dados verossímeis e aderentes só podem ser conseguidos a partir de uma larga escala de dados. Além disso, que fazemos com esses dados, não podem ser feitos com dados em pequena escala.

Na impossibilidade de definir *big data* Peter Norvig, então pesquisador de inteligência artificial na empresa Google, faz uma analogia com imagens. Para ele, um desenho na caverna, uma pintura ou uma fotografia, embora sejam obtidos através de diferentes técnicas, são essencialmente a mesma coisa: uma imagem. No entanto, diversas imagens capturadas e exibidas em velocidade transformam-se em um filme. E um filme é essencialmente diferente de uma fotografia. A conclusão é que sempre que algo muda significativamente em suas dimensões, muda também em sua essência (apud SCHÖNBERGER e CUKIER, 2013, posição 164, tradução nossa). Se fosse apenas um conjunto de dados digitais, *big data* seria como uma coleção de fotografias, mas as tecnologias de comunicação, armazenagem e processamento mudam as dimensões dos dados de modo tão desproporcional a tudo que já havia antes, que inevitavelmente gera algo de outra essência.

Schönberger e Cukier definem *big data* como a "habilidade da sociedade de aproveitar as informações disponíveis em novas maneiras de produzir conhecimento, produtos ou serviços de valor significativo" (2013, posição 45, tradução nossa)

> A mesma tecnologia será aplicada para diagnosticar doenças, recomendar tratamentos, talvez até identificar 'criminosos' antes deles cometerem crimes. Assim como a Internet mudou radicalmente o mundo adicionando a capacidade de comunicação aos computadores, *big data* transformará visões fundamentais que temos da vida ao dar a ela uma dimensão quantitativa que ela jamais teve (SCHÖNBERGER e CUKIER, 2013, posição 194, tradução nossa)

O potencial revolucionário em *big data* é a sua própria definição. A dimensão real desse potencial nunca será conhecida porque seria o equivalente a colocar um limite nas capacidades humanas de conhecer e inovar. "De fato, com mentalidade e atitudes adequadas, dados podem ser sabiamente usados e reusados como uma fonte inesgotável de inovação. Dados podem revelar segredos àqueles que tenha humildade, desejo e ferramentas para escutá-los" (SCHÖNBERGER e CUKIER, 2013, posição 90, tradução nossa).

O que nos leva à resposta de Bill Franks, CAO (*Chief Analytics Officer*) de uma grande empresa de analítica de dados digitais, que respondeu da seguinte maneira quando perguntado sobre qual era a definição de *big data:* "Definição de big data? Que

importância isso tem? O que importa é o que você faz com esses dados[89]" (tradução nossa). Usando uma metáfora, Siegel concorda com Franks quando diz que "quando dados se acumulam, dá-se início à corrida do ouro. Dados não são ouro, eu sempre repito. Dados em sua forma crua não significam nada. O ouro é o que pode ser feito com ele" (2013, tradução nossa)

Embora não seja uma definição, a resposta de Franks é de extrema sabedoria porque captura exatamente a essência de *big data* como desencadeador de ações. Significa dizer que *big data* só tem importância quando empodera o tomador de decisão e isso coloca o ser humano onde ele deve estar nessa discussão: no comando. Franks ainda tangencia a questão da ética na utilização desses dados. Bom ou ruim, para o bem ou para o mal, *big data* de fato será o que nós fizermos com ele.

2.4. Ecossistema *Big data*

Ecossistema é uma palavra que tem origem no campo da biologia mas tem sido exaustivamente usada para se referir a sistemas produtivos. A ideia é relativamente recente e é uma influência que os estudos sobre sistemas complexos adaptativos (CAS) feito pelas ciências naturais têm feito nas ciências sociais (PELTONIEMI e VUORI, 2004).

O conceito refere-se à visão do ambiente produtivo dentro de um contexto maior, para além das suas divisas proprietárias. Os proprietários de um sistema são os *shareholders.* Para além das divisas proprietárias estão os *stakeholders*, aqueles que são partes interessadas do sistema produtivo. Um sistema produtivo pode ser uma pessoa exercendo uma atividade, uma área dentro de uma organização, a organização como um todo, um setor econômico e até uma cadeia produtiva. A depender da unidade considerada os tipos de *stakeholders* variam.

O conceito de ecossistema vai além da soma de partes (*stakeholders*) e leva em conta as relações estabelecidas entre elas. "Um *stakeholders* é, por definição qualquer grupo ou indivíduo que afeta ou é afetado pelas realizações dos objetivos de uma organização"

[89] *"Definition of big data? Who cares? It is what we are doing with it".* Disponível em www.opentracker.net/article/definitions-big-data

(FREEMAN, apud FREEMAN, 2010, p. 207, tradução nossa). Afetar e ser afetado pode englobar relações de vários níveis entre numerosos *stakeholders*, que se forem todos incluídos, se perderia o caráter objetivo da definição de um ecossistema produtivo, que é o planejamento de sua governança. Se definirmos governança como o modelo de gestão que promove a redução ou eliminação dos conflitos de interesses entre as partes, a definição dos *stakeholders* que comporão um ecossistema é uma mensagem clara do gestor do que será considerado nessa tarefa.

Um *shareholder* é um tipo especial de *stakeholder* por isso, a partir de agora usaremos apenas *stakeholders* para nos referir às partes que afetam ou são afetadas pela realização dos objetivos de um projeto em *big data*

Independente de um maior ou menor escopo, o objetivo de se definir um ecossistema *big data* é dar uma visão sistêmica sobre a analítica digital que possibilite decisões mais estratégicas, seja na definição de um modelo de negócio, seja na criação de um produto ou serviço, ou seja na gestão do processo de decisão *data driven* da própria organização.

Não existe consenso sobre o que deve ser considerado no ecossistema *big data,* mas a ideia é considerar um ambiente para além dos dados digitais que são o coração do sistema. Uma revisão da literatura mostra que existem várias visões, dependendo da complexidade de se dá ao tema e do que se pretende incluir no projeto de governança.

O conceito mais simples diz que o ecossistema *big data* possui os seguintes componentes: infraestrutura de armazenagem, estruturação e modelagem de dados, segurança dos dados e *softwares* para analítica. Em outras palavras, “todo complexo composto por equipamentos para armazenagem, segurança, processamento, visualização e entrega de resultados” (DEMCHENKO, NGO e MEMBREY, 2013, p. 6). Claramente a descrição feita representa a visão de engenheiros da computação preocupados com a capacidade e performance das ferramentas de TI.

Peter Hirshberg, CEO do *Re-imagine Group* e ex executivo da Apple tem viajado o mundo difundindo a ideia de que *big data* não é só tecnologia. Uma revisão das suas entrevistas e palestras revela que, para além da tecnologia de governança e analítica de dados já

citada, o especialista acrescenta os sensores capazes de digitalizar informações de diversas naturezas e o tomador de decisão que deverá ter capacidade de transformar a informação gerada em insight e este em decisão.

A visão mais difundida de um ecossistema *big data* é representada pelo acrônimo SMAC que significa *social, mobile, analytics* e *cloud.* A ideia do SMAC veio do estudo das forças motoras que promoveram a analítica digital à categoria de um fenômeno irreversível e vital. Trata-se da cadeia de suprimentos das tecnologias da informação para *big data.* Nesse ecossistema são consideradas: **S**: a ferramentas de interação social entre os *stakeholders;* **M**: a comunicação móvel que permite a aquisição de dados em qualquer hora e lugar; **A**: a capacidade analítica para interpretar dados complexos, também conhecida como inteligência artificial; **C**: do inglês *cloud*, a armazenagem e computação em nuvem, que permite escala e agilidade aos projetos de *big data*. O ecossistema SMAC tem sido usado com sucesso para melhorar os processos analíticos principalmente no que diz respeito à redução de custos, mas inclui também a integração digital dos agentes da cadeia de valor (GOHEL e GONDALIA, 2014).

A visão mais inovadora é a que defende que o ecossistema *big data* é formado por "aqueles que geram os dados (de modo consciente ou não), aqueles que tem a capacidade de coletá-los e aqueles que têm a expertise de analisa-los" (MANOVICH, 2011[90], apud BOYD e CRAWFORD, 2012, p. 675, tradução nossa). De toda literatura visitada, essa foi a únicas que, citando Manovich, incluiu no ecossistema *big data* os agentes do mundo físico. Apesar de inovadora, ficaram de fora agentes que consideramos de grande importância no ecossistema *big data* como por exemplo os que formulam o desafio a ser analisado e os que tomam decisão DD.

Não existe certo ou errado nas visões de ecossistema apresentadas. Como já foi dito, ecossistema do contexto produtivo serve para visualização dos agentes e das forças envolvidas de modo a guiar planos e ações. Cada ecossistema contém o universo onde se pretende agir.

[90] 'Trending: the promises and the challenges of big social data', in Debates in the Digital Humanities, ed. M. K. Gold, The University of Minnesota Press, Minneapolis, MN, [Online] Available at: http://www.manovich.net/DOCS/Manovich_trending_paper.pdf (15 July 2011).

3. *Framework*: Big Data Estratégico[91]

A riqueza não provém do domínio das fronteiras, mas do controle dos fluxos

Pierre Levy

A proposição de um *framework* para gestão do ecossistema *big data* já seria um desafio de sintaxe e semântica se todos os elementos e interrelações já estivessem bem definidos. No entanto, a própria definição de *framework* se contrapõem à ideia de predefinição.

"Um framework é uma abstração da realidade que se propõe a identificar classes de elementos relevantes e suas interrelações" (GALDO, p. 16, no prelo). Isso quer dizer que, ao cria-lo, toda subjetividade da visão do proponente e dos seus objetivos estarão definitivamente presentes. É interessante notar que o caráter subjetivo e particular de um *framework* não o impede de se tornar consensual e compartilhável. Em alguma instância, onde se compartilha a visão e objetivos que embasaram o modelo proposto, a ferramenta será aplicável e útil.

Também por isso foi feita a opção por *framework* e não por um modelo. Diferente do modelo, que pretende ser a forma ideal, o paradigma (GOUVEIA Jr., 1999), um *framework* pode até ser "agnóstico[92] sobre a forma particular de relações teóricas que possam existir" (TEECE, apud GALDO, p. 16, no prelo). Não é a proposição de uma solução, mas a estrutura a partir da qual a solução será desenhada. Ele deve se limitar à representação do comportamento geral do objeto estudado.

Como já foi dito, o objetivo desse *framework* é servir como ferramenta de diagnóstico, projeto, comunicação e gestão efetiva de processos (eficientes e eficazes) de tomada de decisão baseada em dados digitais (DDD). O usuário da ferramenta deverá ser capaz de planejar e definir práticas de governança do ambiente digital de modo que a potencializar e realizar sua expectativa em torno de um determinado projeto em *big data*. Começamos definindo a visão particular de *big data*, que vai gerar nosso conceito e descrição do

[9191] Nome dado ao *framework* resultante desse trabalho
[92] Aquele que reconhece que não conhece tudo (nota de GALDO, no prelo)

ecossistema *big data,* finalizando com a proposição do *framework* para gestão sistêmica desse ecossistema. Desse modo, teremos cumprido os objetivos desse trabalho.

3.1. Visão

Uma visão é a descrição objetiva da perspectiva pela qual se vê o mundo e a partir da qual nos posicionamos perante os fenômenos e agimos diante de desafios. Especificamente nesse trabalho, a visão foi construída a partir da perspectiva que temos sobre complexidade e sobre *big data*, já devidamente explanadas no referêncial teórico.

O principal pilar desta visão específica sobre *big data,* que vai fundamentar a parte conceitual e descritiva do ecossistema, é a convicção de que o ambiente digital começa e termina no mundo físico. Ignorar isso pode ser a diferença entre gerar ou não valor a partir de dados digitais. Essa ideia corrobora com a definição de ciberespaço e cibercultura de Pierre Lévy:

> Ciberespaço, que também chamarei de rede, é (...) não apenas a infraestrutura material da comunicação digital, mas também o universo oceânico de informação que ela abriga, assim como os seres humanos que navegam e alimentam esse universo. Quanto ao neologismo 'cibercultura', especifica o conjunto de técnicas, práticas, atitudes, modos de pensamento e valores que se desenvolvem juntamente com o crescimento do ciberespaço. " (2011, p.17)

O outro pilar, que vai fundamentar a parte prática do ecossistema, é um conjunto de cinco convicções sobre dados digitais:

- Iluminam a rede vital que as pessoas estabelecem entre si e com todas as outras coisas do mundo;
- Vêm de todos os lugares e devem ser enviados para todos os lugares;
- Só têm valor quando acionados para resolver problemas
- Aceitam diferentes olhares por isso, funcionam melhor em *crowdsourcing*[93]
- Empoderam indiscriminadamente todo tomador de decisão.

[93] Processo de obtenção de ideias mediante solicitação de um grande número de pessoas.

Somando os pilares conceituais e práticos, elaboramos a visão de que *big data* é o fenômeno de digitalização do mundo a partir do uso intensivo da Internet, gerando dados que não possuem valor intrínseco, que podem ser usados infinitamente gerando conhecimento para todos.

Alinhando o norte da bússola com essa visão, partimos para o escopo do *framework.*

3.2. Ecossistema *Big Data* Crie[94]: delimitação e descrição

Como já foi visto no referencial teórico[95], um ecossistema é definido por seus *stakeholders*, aqueles, estando dentro ou fora dos limites proprietários de um sistema produtivo, agregam valor a esse sistema. Dito isto, para definir um ecossistema temos que definir de que sistema produtivo estamos falando e, a partir daí, sair em busca dos seus *stakeholders.*

No entanto, apenas definir *stakeholders* não nos levará ao objetivo de fazer uma gestão efetiva do ecossistema. Isso porque, concretamente falando, a gestão se faz sobre agentes e não sobre *stakeholders.*

Consideramos importante, antes de prosseguir, deixar bem clara as definições de *stakeholders* e agentes, tal como usaremos no *framework* proposto.

Um *stakeholder* agrega valor a um sistema produtivo, mas sua figura é etérea, é algo que se define conceitualmente, mas não está concretizado em lugar nenhum. O valor que o define está difuso na cadeia produtiva, sem que seja de propriedade exclusiva de um só agente produtivo. Com o não existe concretamente, um *stakeholder* não pode ser gerenciado e sobre ele não podem recair práticas de governança.

[94] O ecosssitema proposto não tem nome definido. Crie é o Centro de Referência em Inteligência Empresarial da Coppe/UFRJ, onde o conceito foi criado e vem sendo testado, juntamente com o desenvolvimento de metodologias e ferramentas
[95] Seção 2.4.

Um agente é aquele que com suas intervenções no sistema, produz um determinado valor. Ele se materializa de alguma maneira no sistema produtivo de modo que podemos gerenciá-lo e estabelecer práticas de governança que incluam alterar seu comportamento em prol de um objetivo.

Feitas as distinções entre os dois conceitos, podemos concluir facilmente que um *stakeholder* pode ser representado por mais de um agente. Isso quer dizer que, para agregar um determinado valor a um sistema, vários agentes trabalham articulados.

De modo análogo, as intervenções de um determinado agente podem agregar valores diferentes ao processo.

Essas definições e conclusões ficam mais claras quando os *stakeholders* e agentes do Ecossistema *Big Data* Crie são apresentadas. Neste momento o que é importante saber é que não há correspondência biunívoca entre agentes e *stakeholders*.

3.2.1. Delimitando o ecossistema: escopo do *framework*

Quando definimos os componentes que consideramos relevantes envolver na governança de um sistema produtivo, estamos definindo um ecossistema.

O sistema produtivo que queremos representar é o de soluções de problemas a partir de dados digitais ou, em outras palavras, o sistema produtivo de *data driven decision* (DDD). Os *stakeholders* são aqueles que agregam valor a esse sistema produtivo, estando dentro ou fora de suas fronteiras proprietárias.

Quem são os *stakeholders* de um processo de resolver problemas a partir de dados digitais?

É importante esclarecer que a palavra ‘problema’ não está necessariamente associada a coisas negativas. A necessidade ou desejo de criar um produto ou serviço novo é considerado problema assim como uma epidemia que ameaça uma população. Do

mesmo modo, problema não é algo cuja solução sempre nos agrada. Um terrorista escolhendo o local do próximo ataque é um problema legítimo nesse contexto. Usar *big data* para visualizar problemas e encontrar soluções é apenas uma etapa. O que se faz com isso é outra questão que fica tecnicamente fora do ecossistema *big data.* Genericamente um problema é tudo o que pode ser colocado em forma de pergunta e solução é tudo que se propõem a responder à pergunta feita.

Até pouco tempo atrás, o *stakeholder* mais importante de qualquer ecossistema *big data* era aquele que coletava e/ou controlava grande quantidade de dados. Primeiro porque se pensava que dados digitais funcionavam com a lógica dos ativos tangíveis, onde existe valor na posse e, segundo, porque se acreditava que só proprietários de dados poderiam se beneficiar da analítica digital. Hoje existem visões mais abrangentes que consideram, genericamente, três grupos de *stakeholders* para *big data*: aqueles que atuam agregando valor aos dados, os que empoderam as habilidades técnicas e os que contribuem com ideias em todo sistema produtivo DDD (SCHÖNBERGER e CUKIER, 2013, posição 1876).

Destrinchando esses três grupos propostos por Schönberger e Cukier, encontramos *stakeholders* que no processo de produção de decisões *data driven:*

a. geram de informações;
b. geram de problemas;
c. propõem desafios;
d. formulam e testam hipóteses sobre o desafio proposto;
e. digitalizam informações;
f. garantem a estrutura e processos para uso dos dados;
g. definem a estratégia de compartilhamento de dados e uso de *open data*
h. fazem a analítica digital;
i. entregam o resultado da analítica;
j. tomam decisões;
k. executam a decisão tomada e
l. fornecem *feedbak*.

Esses *stakeholders* foram escolhidos para este *framework*, porque agregam valor a cadeia produtiva de DDD, sem conflitar com visão e objetivos propostos. Além de necessários os *stakeholders* citados têm se mostrado suficientes, já que em testes teóricos e práticos, nenhum outro *stakeholder* precisou ser adicionado para produzir os resultados esperados com o uso da ferramenta[96].

3.2.2. Descrição dos *stakeholders*

Descrever os *stakeholders* ajuda entender como eles se relacionam, que valores agregam ao ecossistema e que agente é responsável por cada valor.

Cada um dos 12 *stakeholders* enumerados no item anterior foi citado pela ação (verbo) específica que explica o valor agregado. Para nomeá-los, utilizaremos a versão nominal desses verbos, obedecendo à lógica que complementa a expressão:

***stakeholder* que faz a/o...**

a. Geração de informações

São agentes da dinâmica do mundo físico. Pessoas, objetos, animais, plantas, condições ambientais e instituições existindo, agindo, interagindo e se relacionando com o meio ambiente no plano físico, de diversas maneiras, gerando as informações relevantes para tomadas de decisões. Ainda não estamos falando de ambiente digital. Apreender bem esse conceito exige uma abstração para um mundo pré-tecnologias digitais. É daqui que os dados digitais serão capturados.

Ex. Pessoas se comunicando, se deslocando pela cidade, usando objetos do dia-a-dia, animais sendo alimentados ou medicados, empresas vendendo para consumidor final ou para outras empresas, plantas fazendo fotossíntese, chuva caindo etc.

[96] Definido no item 1.3.3.

b. Geração de problemas

São aqueles com problemas a serem resolvidos, mesmo que não se proponham a resolvê-los ou nem tenham consciência deles. São os mesmos agentes de geração de informação (pessoas, objetos, animais, plantas, condições ambientais e instituições), agora como *stakeholders* de outro valor agregado à cadeia: eles desafiam a capacidade analítica do ecossistema.

Ex. Pessoas precisando produtos ou serviços (mesmo que não tenham consciência disso), prefeitura controlando trânsito, produtora decidindo o tema do próximo filme, detetive investigando um crime, médicos decidindo fazendo um diagnóstico, cientistas em busca da redução de CO_2 na atmosfera, vírus da gripe querendo se espalhar em uma população etc.

c. Proposição de problemas:

São pessoas, físicas ou jurídicas, que por cidadania, necessidade de sobrevivência ou busca de oportunidades de negócio, formulam os desafios a serem resolvidos. Podem ou não ser os causadores do problema ou impactados por eles; podem ou não ser os responsáveis pela solução ou impactadas por elas. Esses *stakeholders* têm o relevante papel de direcionar a analítica[97]: todo processo DDD, desde o estudo do problema até a tomada de decisão, será influenciado pelo modo como o problema for formulado.

Ex. Alguém buscando uma oportunidade de negócio, alguém querendo acabar com a corrupção, uma empresa desejando atender melhor seu cliente ou um novo mercado consumidor etc.

d. Datificação do problema

Datificar[98] é 'decupar' um problema em dados da melhor maneira possível para obter o melhor com a analítica desses dados (COSTA, FERNANDES e ZUPPO, 2013). O *stakeholder* de datificação é a ciência dos dados representada por pessoas (agentes) que usam os princípios dessa ciência para formular e testar hipóteses sobre o problema

[97] Direcionam a analítica, não a tomada de decisão. O que direciona a tomada de decisão são os objetivos, valores e princípio do tomador de decisão. Ressalva importante, mas fora do escopo desse trabalho.
[98] Definição nossa. Na literatura encontramos datificar como sinônimo de digitalizar

proposto ou sobre o determinado *dataset.* Isso significa que agregar valor por dataficação possui dois sentidos: **1.** do problema para os dados: a ação necessária para dataficar nesse sentido é ter *insights* sobre como funciona a rede do problema (quais são os nós e como eles se relacionam). Esses agentes podem ser um especialista no problema, se for assunto de muita complexidade, ou simplesmente alguém que se dedica minimamente a compreender o fenômeno. Também não é necessário que seja um profissional de ciência da computação[99]. A única exigência para ser considerado um *stakeholder* de dataficação é ter visão de complexidade e ferramentas mentais para pensar 'fora da caixa'; **2.** Dos dados para o problema: dataficar nesse sentido é criar *clusters* de *datasets*[100] e buscar neles padrões, desvios, correlações, sequências, qualquer coisa que possa sugerir algum *insight* gerando uma informação desconhecida. Também se datafica nesse sentido especulando em que outro domínio do conhecimento as informações de um *dataset* seriam relevantes.

Ex. Para dataficação no sentido 1. Pessoas, especialistas ou não, percebendo que escolha de destino de viagem em determinado período pode significar propagação de alguma doença viral; ou que chuva aumenta a demanda por serviços tipo *delivery*; ou que quantidade de carros + tempo de sinais + existência de escolas nas imediações pode provocar engarrafamentos. Para dataficação no sentido 2. Analistas, cientistas de dados, estatísticos minerando dados em busca de algo que possa ser um valor. Alguém perceber que os dados de consumo de energia de uma cidade podem indicar o poder aquisitivo hiperlocal[101]. Um geolocalizador no pulso de um idoso que vive sozinho pode servir para localizá-lo ou para indicar que, após muito tempo sem variação de posição, ele está necessitando de ajuda.

e. Digitalização das informações

São agentes que transformam informações em dados digitais. Podem ser sensores que transformam grandezas físicas (som, calor, movimento) em grandezas digitais, ou pessoas e ferramentas que codificam linguagem natural, imagens, valores, relações e

[99] Essa definição de cientista de dados não é encontrada comumente na literatura nem a com partilhada pelo mercado de tecnologias digitais. No Crie acreditamos que o cientista de dados é aquele curioso em relação ao que os dados podem produzir e formulam e testam suas hipóteses mesmo que precisem da ajuda em processos técnicos para os quais ele não tem habilidade.
[100] Encontrar um *cluster* de *datasets* pode ser entendido como sair do *big data* (universo de dados), que no fundo é apenas um conceito, e encontrar um *smart data*, ou seja, um *dataset* ou grupo de *datasets* onde efetivamente serão feitas as analíticas digitais.

outras informações do mundo físico. Em resumo, são responsáveis por preparar a informação para ser analisada por ferramentas de analítica digital.

Ex. Celulares com suas dezenas de sensores embarcados, sistemas de GPS, redes sociais, caixas eletrônicos, livros digitais, pessoas tirando fotos e colocando *tags*[102] em uma imagem etc.

f. Governança de dados

Agentes que cuidam da armazenagem, segurança, cópia, transmissão, codificação, etiquetagem dos dados. Esses agentes garantem o valor importantíssimo de disponibilizar de modo seguro e ágil, os dados para o processo analítico.

Ex. Os responsáveis por serviços de computação e armazenagem em nuvem, sistemas de anonimização de dados, sistema de 'tagueamento'[103] de imagens, etc.

g. Estratégia de dados

Representados por agentes que planejam e viabilizam a circulação dos dados com dois objetivos principais: a obtenção dos dados demandados pelo problema e o maior aproveitamento dos dados sob seu controle. Para o primeiro objetivo o estrategista se pergunta como ele pode encontrar os dados que precisa. Para o segundo, para quem mais os dados que ele controla têm valor. O valor que esse *stakeholder* agrega ao ecossistema está muito bem representado na afirmação de Schönberger e Cukier: "Muitas vezes o valor potencial de um *dataset* só pode ser alcançado quando combinado com outro *dataset*, talvez até bem diferente dele" (2013, posição 1616, tradução nossa).

Ex. Plataformas open data

h. Analítica de dados

São os agentes (pessoas, máquinas, softwares, linguagens de programação, algoritmos) envolvidos na aquisição, tratamento (limpeza, indexação, agregação, anonimização, etc),

---

[101] Um geopocicionamento bem específico num espaço relativamente bem menor que o todo.
[102] Anglicanismo. A tradução da palavra é etiqueta, mas nesse contexto só se usa a palavra em inglês *tag*
[103] Anglicanismo. Neologismo para o ato de colocar *tags* (etiquetas).

processamento e análise quantitativa desses dados gerando informações confiáveis para o tomador de decisão. A credibilidade é o valor que esse *stakeholder* agrega ao ecossistema.

Ex. Analistas de dados, desenvolvedores de API[104], computadores, códigos e linguagens de programação etc., qualquer agente que contribua de algum modo para transformar dados em informação.

i. Visualização de dados

Agentes que entregam o resultado da analítica. São a interface entre a analítica dos dados e a tomada de decisão e têm como objetivo uma rápida e segura apreensão das informações geradas. Convencionou-se chamar esse processo de 'visualização de dados', mas é importante esclarecer que no Ecossistema *Big Data* Crie ele pode ser mais que uma interface literalmente visual. Em um processo automatizado de tomada de decisão, por exemplo, dispositivos têm embarcado as tecnologias com funções de capturar (sensores), processar e analisar dados e disparar uma ação independente da ação humana. No entanto é preciso saber que existe uma interface de entrega entre as duas últimas funções, que é a programação humana feita para acionamento condicional (se, então). Separar essas etapas do processo é importante para a gestão do ecossistema como veremos na apresentação do *framework* na seção 4. O *Stakeholder* de visualização de dados agiliza o processo de tomada de decisão.

Ex. Uma planilha com números; gráficos de barra, pizza ou qualquer outro tipo; o *display* de um aplicativo de planejamento de trajetória como o Waze ou Google Maps; a programação (se, então) de um sistema automatizado de irrigação que determina o acionamento se algumas condições analíticas forem cumpridas.

j. Tomada de decisão

São agentes empoderados pela analítica digital em seu processo de escolha. Na literatura específica da área tecnológica, existe um consenso de que as decisões podem ser tomadas por pessoas, mas também por instituições, objetos, animais e plantas, por

automação. Nesse trabalho usamos a definição de tomada de decisão da Ciência da Cognição[105] e, portanto, esse *stakeholder* é invariavelmente um ser humano. Mesmo em processos automatizados, defendemos que a tomada de decisão é feita por seres humanos. Quando um dispositivo automático age, é porque antes um ser humano tomou a decisão e o programou para agir. O valor agregado pelos agentes que toma decisões é orientar (dar sentido e direção) a ação do executor.

Ex. Um médico que faz um diagnóstico após a análise dos exames; um cidadão que escolhe um trajeto de deslocamento depois de consultar o aplicativo Waze; o aplicativo de táxi que seleciona um carro para determinado chamado depois de analisar a localização do pedido.

k. Execução da decisão

São agentes que de fato mudam a dinâmica do problema. A ação pode ser desempenhada por pessoas ou máquinas. A execução bem feita muda essa dinâmica na direção e sentido planejados pelo tomador de decisão e esse é o valor que esses agentes agregam ao ecossistema.

Ex. Um médico que prescreve um tratamento após a leitura do diagnóstico; o motorista do carro que dirige o carro pela rota indicada pelo Waze; o motorista de taxi que se dirige ao endereço do chamado, depois de ser designado pelo aplicativo para fazer o serviço.

l. *Feedback*

Os agentes que agregam esse valor são os que sofrem as consequências da decisão DD. Não existe conotação negativa na palavra “sofrer”. Os impactos podem ser desejados ou não, mas é invariável que traga desdobramentos em outros contextos do agente[106]. Esses agentes são os únicos podem dar *feedback* ao processo de DDD. O *feedback* pode ser

[104] Acrônimo de *Application Programming Interface* ou, em português, Interface de Programação de Aplicativos. Trata-se de um conjunto de padrões de programação que permite a construção de aplicativos e a sua utilização de maneira não tão evidente para os usuários.
[105] Como já foi dito na introdução à seção 2 (referencial teórico)
[106] Por causa das dimensões de um problema complexo (generativa, dinâmica e social) mencionadas na seção 2, item 2.2.

feito ação consciente ou por algoritmos que analisam rastros digitais[107]. Nos dois casos os princípios são os mesmos de qualquer *feedback.*

Ex. *Feedback* consciente: o motorista de taxi que avalia o sistema e o usuário; o usuário que avalia o motorista e o sistema; o sistema que avalia motoristas e passageiros; o resultado dos exames de um paciente em tratamento; *Feedback* por algoritmo: A rua que recebe um fluxo de carros maior porque foi o trajeto mais indicado pelo Waze e um algoritmo a coloca na cor vermelha no mapa.

3.3. *Big Data* Estratégico: um f*ramework* para o Ecossistema *Big Data* Crie

O escossistema *big data*, descrito como foi na seção anterior, é um suporte teórico para adoção de projetos em *big data.* O que traz a necessidade de um *framework* é o desejo de facilitar essa apreensão teórica, agilizar a comunicação da ideia e a adoção de processos de DDD.

Seguimos os princípios qualitativos existentes nos guias de *design* de *frameworks* para *softwares*: simplicidade, clareza, limites e expansibilidade (BOOTH).

O princípio da simplicidade tem o objetivo de proporcionar a qualquer pessoa interessada uma rápida apreensão de sua ideia principal, mesmo que os detalhes sobre seu uso e propriedades demandem mais tempo e a leitura de material explicativo. O uso de recursos visuais é imprescindível.

O princípio da clareza refere-se aos aspectos comportamentais do framework. O usuário não precisa ser especialista, nem entender os detalhes e o funcionamento de cada parte específica da estrutura, para utilizá-lo. O que importa é o entendimento da dinâmica geral do ecossistema que o framework representa.

[107] O conceito de rastro digital (deixar informações na plataforma digital de modo inconsciente), foi abordado no item 2.3.1.

O princípio dos limites, clama pela não violação das fronteiras de responsabilidades da ferramenta. Abarcar mais do que o estritamente necessário para cumprir seus objetivos faz do framework uma ferramenta complexa e sem utilidade.

O princípio da expansibilidade visa garantir ao *framework* duas frentes de expansão: uma interna, que permiite que o usuário acrescente novos elementos à medida que o escopo definido evolui tecnologicamente; e outra expansão externa, que permite que o usuário acople, outros *frameworks* que estão fora do seu escopo.

O *framework* para *Big Data* Estratégico foi construído com o DNA de "código aberto" ou seja, prevê todas as edições que o usuário queira fazer, para adaptar às suas necessidades ou para atualizar às novas tendências. O que será apresentado nesse trabalho é o estado da arte do que consideramos relevante em nosso recorte.

Conceitualmente, o que chamamos de *Big Data* Estratégico é a ferramenta para diagnóstico, projeto e comunicação de um ecossistema *big data.*

Concretamente, é uma ferramenta composta de quatro partes[108]: **1.** o *framework* propriamente (Figura 9); **2.** um quadro resumo que mostra como os *stakeholdres* e agentes estão representados (Tabela 2); **3.** um Check List com perguntas e considerações para orientação sobre o uso da ferramenta (Tabela 3) e **4.** um quadro de trabalho, em branco, para ser preenchido pelo usuário da ferramenta (Anexo VI)

É importante destacar que os ícones representados na Figura 9 são agente do ecossistema. Os *stakeholders*, por representarem valores, estão dispersos pelo quadro e são reconhecidos quando percebemos a relação entre os agentes. Como já foi dito[109], um mesmo agente pode ser *stakeholder* de vários valores, assim como um mesmo valor (*stakeholder*) pode estar disperso por vários agentes.

[108] A necessidade de material complementar explicativo é intrínseca a uma ferramenta gráfica simplificada de gestão como outros exemplos conhecidos: modelo das Cinco Forças de Porter, *Business Model Generation Canvas* (BMG Canvas) de Ostervalder, Matriz SWOT de Humphrey entre outros.
[109] Item 3.2.

Os ícones da figura são auto explicativos, extraidos do senso comum e do uso cotidiano. As caixas de texto complementam as informações indicando a lógica da disposição do ícones. A opção por não usar setas no *framework* derivou de três considerações: 1. a dinâmica desse ecossistema é bem conhecida, mesmo entre não especialistas em tecnologias digitais; 2. para efeito de gestão, todos os agentes estão interrelacionados, portanto a informação seria redundante e 3. a colocação de setas de duplo sentido entre todos os agentes acarretaria em um problema estético que atrapalharia a visualização das informações importantes e consequentemente comprometeria a apreensão e entendimento da figura.

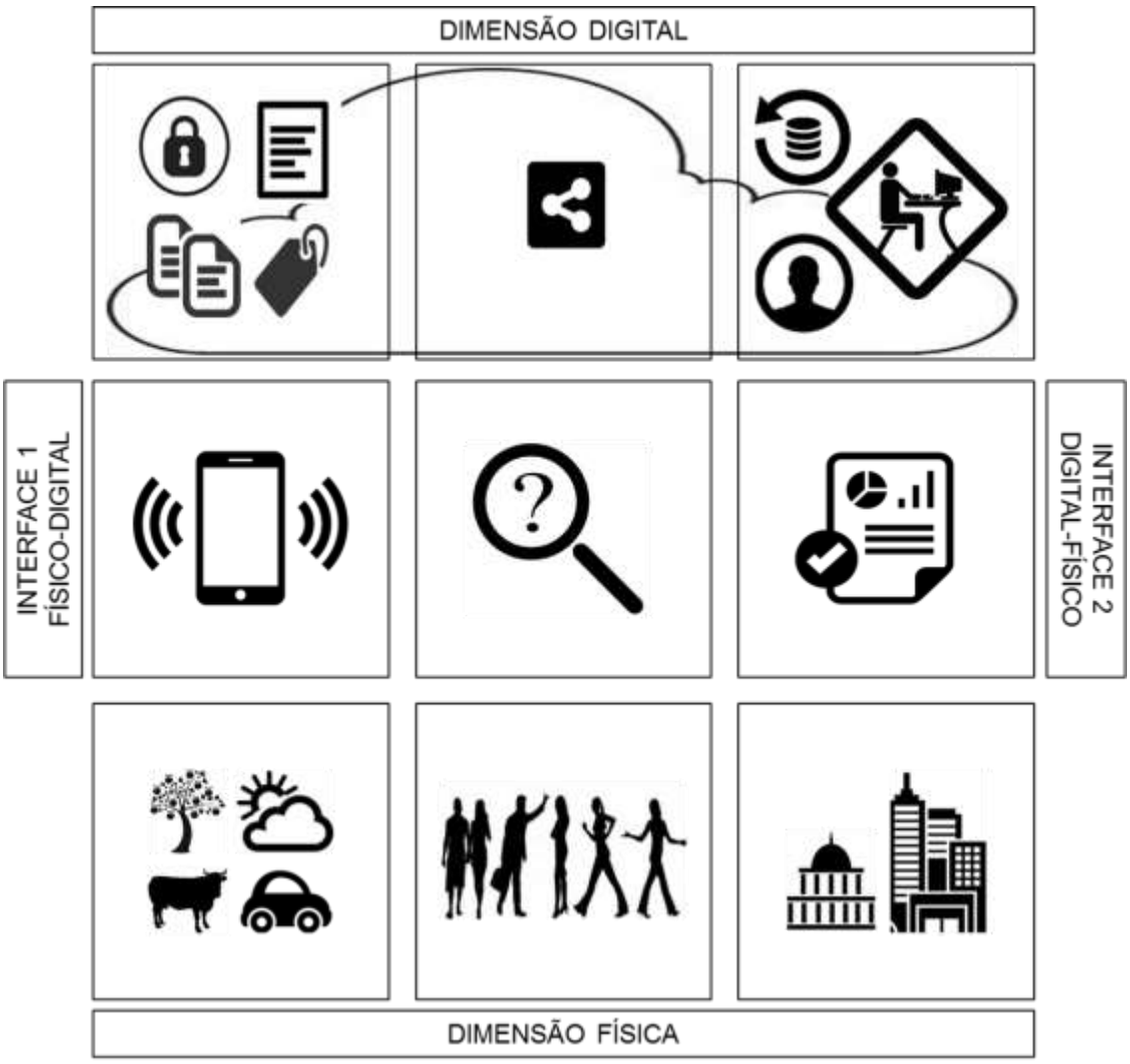


Figura 18 – *Big Data* Estratégico – *Framework* para gestão sistêmica do ecossistema *big data*
Fonte: própria

A Tabela 2, *Big Data* Estratégico – Quadro resumo é a reunião de todas as informações do *framework*, as explícitas e as implícitas. A tabela contém cinco colunas: a primeira coluna apresenta os ícones; a segunda, correlaciona o ícone com um agente do ecossistema; a terceira, associa ícone-elemento a algum tipo de *stakeholder*; a quarta coluna mostra o que cada *stakeholder* agrega à cadeia de valor de processos DDD; e a última coluna traz algumas notas explicativas.

Tabela 4 – *Big Data* Estratégico - Quadro resumo - (Fonte: própria)

STAKEHOLDER	**ÍCONE**	**AGENTE**	**VALOR**	**NOTAS EXPLICATIVAS**
a. Geração de informações		Objetos, animais, plantas e condições ambientais	Informações sobre objetos, animais e plantas	As informações sobre esses elementos e as relações entre eles são digitalizadas por sensores cuja rede é conhecida em seu conjunto como Internet das Coisas.
		Pessoas	Informações sobre as pessoas	As informações sobre esses elementos e as relações entre eles (grau de parentesco, amizades, profissionais) são digitalizadas pelas plataformas genericamente conhecidas como Redes Sociais (Ex. Facebook, Linkedin, Twitter, etc).
		Instituições	Informações sobre instituições	As instituições são empresas, governos, entidades sem fins lucrativos, associações etc. As relações entre esses agentes são institucionais e quando digitalizadas geram os dados conhecidos como transacionais.
		Objetos, animais, plantas e condições ambientais, pessoas e instituições	Informações sobre as relações entre objetos, animais, plantas, condições ambientais, pessoas e instituições	As informações sobre as relações entre eles entre esses elementos são digitalizadas por diversos sensores e plataformas de prestação de serviços (sites das instituições).
b. Geração de problemas		Objetos, animais, plantas e condições ambientais, pessoas e instituições	Desafio para a analítica	Os desafios do mundo real são a razão de existir da analítica digital.
c. Proposição de problemas		Pergunta	Direção para a analítica	Qualquer coisa que possa ser colocada em forma de pergunta para ser respondida.
d. Dataficação do problema		Ciência dos dados	Modelagem do um problema para ser transformado em algoritmo	Decupagem do problema em dados. Identificação dos nós e relações da rede do problema.

STAKEHOLDER	ÍCONE	AGENTE	VALOR	NOTAS EXPLICATIVAS
e. Digitalização das informações		Dispositivos digitais	Quantificam as informações	Primeira interface do mundo físico com o digital. Transformação de informação em dado. Qualquer dispositivo que transforme uma grandeza física em digital ou mapeie digitalmente relações entre nós de uma rede.
f. Governança de dados		*Cloud* ou nuvem	Redução do tempo entre coleta dos dados e disponibilização da analítica para tomada de decisão).	Dados ficam disponíveis em vários servidores permitido processos em tempo real.
		Segurança	Garantia de integridade aos dados	Restrição de acesso e uso da rede e dos dados.
		Cópia	Segurança e agilidade	Disponibilização do mesmo datasets por diferentes servidores
		Anonimização	Publicação	Desvinculação de dados de interesse público das informações sensíveis associadas a eles (pessoais, de segurança ou de competitividade). Esse stakeholder possibilita a publicação e compartilhamento de dados.
		Tags ou etiquetas	Identificação dos dados	As tags permitem o gerenciamento dos dados (busca, associação, organização), principalmente dos chamados dados não estruturados.
g. Estratégia de dados		Publicação e compartilha-mento	Mais correlações e insights	Disponibilização de dados para uso de terceiros e uso de dados de terceiros a partir de algum acordo social de compartilhamento (open data, troca ou venda de dados).
h. Analítica de dados		Indexação	Ordenação e gerenciamento aos dados	Inventário de dados.
		Preparação de datasets	Confiabilidade aos dados	Limpeza, agregação, anonimização
		Pessoas, máquinas, softwares, linguagens de programação, algoritmos, etc.	Visualização da rede do problema. Empoderamento do tomador de decisão	Aquisição, tratamento, processamento e análise quantitativa dos dados. Dados transformados em informação quantificada, que pode ser comparada, agregada, projetada e correlacionada.

STAKEHOLDER	ÍCONE	AGENTE	VALOR	NOTAS EXPLICATIVAS
i. Visualização de dados		Planilhas, dashboards e programação de automação	Rápida e segura apreensão de informações	Apresentação dos dados de modo a destacar o que é importante para o tomador de decisão.
j. Tomada de decisão		Pessoas	Sentido e direção a ação	Apenas pessoas tomam decisão. Uma automação apenas executa a decisão prévia de alguém. Essa decisão é programada e embarcada no dispositivo.
k. Execução da decisão		Objetos, animais, plantas e condições ambientais, pessoas e instituições	Mudam a dinâmica do problema	Objetos, animais e plantas executam decisão por automação. Uma boa execução segue a direção e o sentido dados pelo tomador de decisão.
l. *Feedback*		Objetos, animais, plantas e condições ambientais, pessoas e instituições	Retroalimentação do sistema	Agentes que são impactados por uma tomada de decisão DD. Esse agente tem valor para o ecossistema porque só ele pode dar feedback ao processo de DDD.

3.4. Como usar o *framework? – Check List*

Apenas para facilitar o uso e a rápida adoção do *framework Big Data Estratégico*, sugerimos uma sequência de perguntas que devem ser feitas para a composição do quadro. A Tabela 3 apresenta essa lista que chamamos de *Check Lis*t para uso do *Big Data* Estratégico.

As perguntas são uma orientação para quem está usando o *framework*. Nem todas são aplicáveis a todos os casos do mesmo modo que inúmeras novas perguntas podem e devem ser acrescentadas para tornar a ferramenta mais completa.

Optamos pela apresentação das perguntas na ordem lógica de um projeto que começa pelo *stakeholder* de geração de problemas, representado pelo ícone de todos os agentes do mundo físico na base do quadro. No entanto, qualquer *stakeholder* do ecossistema pode usar o *framework* e isso pode significar a entrada por diferentes partes do quadro.

Logo abaixo de cada pergunta existem algumas considerações relevantes sobre a sua importância, sobre tendências e ainda advertências e cuidados que devem ser tomados

para evitar erros graves e custos desnecessários. Indicaremos também sugestões de interface externa com outros *frameworks.*

Tabela 5 – Check List para uso do *Big Data* Estratégico – (Fonte: própria)

Como encontrar um problema?	
	O mundo está repleto de desafios para a analítica digital. A percepção disso se dá por necessidade, observação ou pesquisa. Não podemos esquecer que os problemas têm uma dimensão social: o que é problema para uns pode não ser para outros.
Qual é o problema?	
	Essa pergunta é a que vai ser resolvida pela analítica digital. Se a pergunta for: Como empoderar uma pessoa para escolher a melhor rota para um deslocamento urbano? A resposta desse *framework* será um valor para um negócio. Se a pergunta for: Qual a melhor rota para ir do ponto A ao ponto B, A resposta desse *framework* será *app* que aponta o melhor trajeto. Definir corretamente o problema é fundamental, pois todo o processo estará atrelado a isso. Mudar a pergunta no meio do processo acarreta custos desnecessários e levar a pergunta errada até o final não trará o *output* desejado. Aqui o *framework* pode fazer uma interface externa com os *frameworks* para *design thinking..*
O problema é conhecer, entender ou prever?	
	Essa definição leva a modelos descritivos (padrões), diagnósticos (relações) e preditivos (tendências) e ajuda a desenhar a rede do problema. Alguns problemas podem mais de um objetivo.
Que agentes do ambiente físico estão envolvidos no problema?	
	Início da etapa de datificação. Caso necessário, peça ajuda de um especialista no problema (consultor de conteúdo). Antes de tudo, dê um nome à sua rede (argumento de rede). Isso ajuda a manter o foco no problema. EX. Rede de pessoas que se relacionam profissionalmente no edifício X da rua A (isso quer dizer que nada fora desse argumento deve ser considerado). Liste todas as pessoas, objeto, animais e plantas e instituições envolvidos no problema. Considere além do primeiro grau de envolvimento. Distancie-se sem perder a relevância. O grau máximo de distanciamento do agente com o problema depende do quanto ele é relevante para entender ou prever algo, da capacidade de investimento e de captura dos dados que o proponente tiver. Contudo, nesse momento deve-se imaginar que nada é impossível. Em um modelo descritivo buscamos agentes que descrevem o problema. Para um diagnóstico, agentes que estejam relacionados com o problema e nos ajudem a entende-lo. Para a analítica preditiva, agentes correlacionados, ou com relação de causa-efeito com o problema. Desse modo encontraremos os nós da rede do problema. Caso o proponente não se sinta capaz é aconselhável procura de um especialista no assunto relacionado ao problema. Esses agentes serão o nó da rede do problema.

Que relações existem entre os agentes desse ambiente físico?	
	Essas relações são os *links* existentes entre os nós. Inicialmente, indique apenas a existência da relação (pessoa-objeto, planta-objeto, pessoa-instituição etc) nesse problema específico. A melhor maneira de fazer isso é desenhando 'nós' e *links*. Depois procure dar nomes às relações (o que relaciona um agente com outro?).
Quais os links mais relevantes para o seu problema?	
	Essa é a etapa de qualificação da rede para futuros cortes.
Qual a topologia dessa rede?	
	Identifique *hubs, links* fracos, densidade e direção dos links (ver as propriedades de redes no referencial teórico)
Qual o suporte dessas relações?	
	Supostamente, a rede desenhada é a ideal para fornecer informações que ajudem na tomada de decisão relacionada a seu problema. A partir de agora começa a etapa da estratégia de dados que define onde e como conseguir os dados necessários. Identifique os aparatos que suportam cada relação e identifique se a interface é física ou digital A relação entre duas pessoas (ou dois objetos) da rede é feita pessoalmente? Por telefone? Por aplicativo? Por sensor?
As relações digitais geram dados na qualidade necessária?	
	Comece analisando aquelas que têm suporte digital. Verifique estrutura, granularidade, periodicidade etc. dos dados. Se não estiverem na qualidade ideal, como adequar? Mesmo sendo dados controlados por você, analise questões de privacidade ou impedimento de uso secundário (quando um dado só pode ser usado para o que ele foi coletado). Caso necessário, utilize a ajuda de um especialista em direito digital. Para problemas mais complexos, o estrategista de dados trabalha junto com o analista e o arquiteto de dados. Muitas vezes os dados necessários são simples e estão tão facilmente disponíveis que não é necessário nenhum profissional de TI para traçar a estratégia de dados. Qualquer um sabe por exemplo que dados de GPS de ônibus na cidade do Rio de Janeiro estão livres em plataforma aberta e com várias APIs disponíveis.
Que dados existentes possuem certificação Open Data?	
	Identifique os dados que estão disponíveis, acessíveis e com licença para uso. Dados com certificação *open data* são mais fáceis de serem usados. A certificação classifica o *dataset* segundo alguns critérios que poupa o esforço de estudar as condições estruturais e de uso dele. Existem Instituições (como o *Open Data Institute* - www,theodi.org) que prestam serviços de certificação *open data.*

Que dados são livres, mas não possuem certificação Open Data?	
	Para os casos não enquadrados na questão anterior, identifique aqueles que carecem de formato adequado e/ou licença para uso
Que dados são controlados por terceiros?	
	Identifique os dados e os respectivos proprietários
Quem coleta/controla as informações relevantes ainda não digitalizadas?	
	Verifique o grau de dificuldade para acessar essas informações
Como digitalizar um link físico?	
	Enumere várias ideias para transformar uma relação física numa relação digital. Lembre-se que uma grandeza física pode ser indicativa de outra (ex. calor pode ser indicativo de presença, movimento pode indicar quantidade, etc.
O que o mercado oferece para renovação dessas interfaces digitais ou para a digitalização dos links que você precisa?	
	Faça uma investigação de aparatos que já existem, que existem para outros fins e podem ser adaptados, que estão em fase de teste, que estão em fase de patentes ou de pesquisa.
Os aparatos que digitalizam outras informações podem ser usados para suportar novos sensores/coletas?	
	Considere inicialmente utilizar interfaces existentes em outras relações.
É extremamente necessário introduzir novos aparatos para suportar novos sensores/coleta?	
	Considere introduzir novas interfaces apenas depois de responder à questão anterior. Um telefone celular, por exemplo, é um aparato digital que possui entre 12 e 20 sensores embarcados. Se ele já é o suporte para coletar dados de localização, pode também ser usado para coletar ruídos, temperatura, comportamentos, etc.
Se o agente que cede a informação é uma pessoa ou instituição, como você o convencerá a usar a interface digital que você oferece?	
	Considere servicização (oferecer algum serviço, como um aplicativo) e gamificação (o que motiva a pessoa ou a instituição a participar do seu jogo e lhe ceder dados?). Considere coleta de dados com mínimo de esforço (*idlesourcing*) com o uso de sensores, por exemplo. Aqui o *framework* pode fazer uma interface externa com os *frameworks* para gamificação.

Qual o projeto para digitalização de um link físico?	
	Esse é o momento de ponderar desejo, possibilidade e viabilidade. Utilize ferramentas de *design thinking* para projetar, prototipar e testar o seu projeto. Considere: custo X necessidade (tipo de sensor, fiabilidade, rapidez, dimensão, estética)
Quem possui/controla as informações/dados que não são *open*	
	Verifique as restrições de uso dessas informações ou dados.
Que política de compartilhamento pode garantir o acesso a dados que você não controla?	
	Identifique a rede de valor dos dados que você controla e dos que você precisa ter acesso e considere trocas de valor das mais diversas naturezas. Considere sempre a possibilidade de tornar públicos os dados que você gera.
Que política pode garantir a segurança no uso de dados sensitivos? (associados a informações pessoais, de competitividade ou de segurança)	
	Considere como anonimizar, evitar acessos indevidos, estragos e perdas.
Qual a real necessidade infraestrutura de rede e dados do projeto?	
	Dimensione o projeto para definir se será necessária a propriedade da infraestrutura ou apenas contratar os serviços. Considere sempre infraestrutura e serviços em nuvem porque são mais ágeis, seguros e relativamente mais baratos, permitindo escalabilidade com elasticidade (a estrutura aumenta e diminui conforme o volume dos dados movimentados. Aqui o *framework* pode fazer uma interface externa com os *frameworks* para definição de serviços de nuvem.
Qual a real necessidade de capacidade analítica do seu projeto?	
	Analise se o projeto precisa de uma equipe fixa de cientistas de dados, desenvolvedores, analistas de sistemas. Considere a possibilidade de promover concursos e *hacktons,* ou de contratar esses serviços por demanda. Aqui o *framework* pode fazer uma interface externa com os *frameworks* para definição de capacidade analítica.
Os stakeholders do projeto têm competência para participar do processo analítico?	
	Capacite gestores, profissionais de TI e áreas de apoio para compreender a importância e os desdobramentos de um processo analítico.

Em que nível você precisa receber os resultados da analítica digital?	
	1. informação 2. conhecimento 3. sabedoria 4. decisão O nível de interferência humana que você precisa na tomada de decisão define o tipo de visualização de dados que o processo analítico vai entregar. Ela varia desde decisão totalmente humana até totalmente automatizada, passando por matizes intermediárias. Aqui o *framework* pode fazer uma interface externa com os *frameworks* para definição da visualização de dados a partir do perfil da tomada de decisão
O tomador de decisão está preparado para o processo DDD?	
	A tomada de decisão terá impactos em diversas interfaces com o problema. O tomador de decisão deve estar atento a todas elas e agir de acordo com a missão, princípios e valores da instituição. Aqui o *framework* pode fazer uma interface externa com os *frameworks* para tomada de decisão
Quem executará a decisão tomada?	
	A execução de uma decisão pode ser automatizada (feita por computadores a partir de parâmetros previamente definidos), semi-automatizada (parte executada por automação, parte por pessoas) ou humana. Considere prós e contras de cada uma levando em conta: escala, custo, conveniência, precisão, impactos em caso de erros, probabilidade de ser necessária uma redefinição de parâmetros com urgência.
Como o *feedback* será usado para melhorar o processo DDD?	
	Todos os agentes do mundo físico que sofrem os impactos de uma tomada de decisão podem fornecer *feedback* ao processo. Considere uma retroalimentação constante de modo que o processo analítico possa 'aprender' e tornar-se cada vez mais aderente com suas nas necessidades reais.

Ao ponderar e responder as perguntas do *check list*, estaremos definindo o ecossistema do processo DDD proposto não apenas com os agentes que precisam estar envolvidos, mas também como o modo como eles devem se relacionar e o como conseguiremos fazer com que tudo isso aconteça.

No Anexo VI está disponível um *framework* de trabalho que nada mais é do que o *framework* do Ecossistema *Big Data* apresentado aqui, mas completamente em branco para ser preenchido pelo usuário da ferramenta à medida que responde às perguntas do *check list.* Como já foi dito, essas perguntas não têm a pretensão de serem impositivas ou exclusivas. Cada usuário deve eleger aquelas que atendem ao escopo do seu projeto e,

ao respondê-las, terá tido uma visão geral do seu ecossistema: de como ele é composto e de como suas partes se relacionam. O resultado do uso do *framework* deverá ser a resposta para a pergunta feita pelo *stakeholder* de formulação do problema, resposta essa que poderá ser modelo de processo ou de negócio (que também é um processo), a depender do escopo definido.

O usuário perceberá também que nenhuma ferramenta, máquina, infraestrutura, metodologia ou competência humana deve ser adquirida ou definida de modo independente. Cada parte do ecossistema *big data* está intimamente relacionada com a outra e as decisões para montá-lo e fazê-lo funcionar deve levar isso em conta. Esperamos que o *Big Data* estratégico seja uma ferramenta de auxílio nesse processo.

Atendendo ao principio da expansibilidade (BOOTH), é importante notar que o *Big Data* Estratégico possui interfaces com outros *frameworks* importantes, alguns dos quais citamos a seguir:

Frameworks* para *design thinking: Ferramentas do campo do *Design* que podem ajudar especificamente na definição do problema, utilizando entre outras coisas a empatia (colocar-se no lugar de quem tem o problema) e a atitude de questionamento do *briefing* feito pelo proponente do problema.

***Frameworks* para gamificacão:** Ferramentas da Teoria dos Jogos fazendo interface com o *stakeholder* de digitalização de informações. O site O *Octalysis* proposto por Yukai Chou e apresentado no Apêndice V, por exemplo, é um *framework* que ajuda a definir qual a motivação que as pessoas têm para fazer parte de algo, ou na linguagem da Teoria dos Jogos, na motivação para jogar. O site *The Engagement Plataform,* disponível em apresenta diversos outros *frameworks* de gamificação para outros fins.

***Frameworks* para definição de capacidade analítica:** O Apêndice VI apresenta um *framework* desenvolvido pelo Crie para ajudar na definição da capacidade analítica necessária para o projeto.

***Frameworks* para definição de serviços de nuvem:** O uso de serviços de nuvem traz vários benefícios se adotado da maneira mais adequada ao projeto. O Apêndice VII apresenta um *framework* bastante popular no mercado para o entendimento dos tipos de serviços de nuvem e as vantagens e desvantagens de cada um.

***Frameworks* para definição da visualização de dados:** O Apêndice VIII apresenta um *framework* desenvolvido pelo Crie para ajudar na definição do tipo de visualização de dados associados ao perfil da tomada de decisão.

***Frameworks* para tomada de decisão:** O processo de tomada de decisão é um assunto ampla e profundamente estudado por diversos campos da ciência. Por isso mesm existem diversos *frameworks* que auxiliam na definição e dinâmica desse processo. Eles alertam para os viéses que um processo analítico (digital ou não) e o modo de apresentação dos dados podem ter, para as situação onde a máquina decide melhor e as que o ser humano é superior, entre outros alertas importantes.

Diversos outros *frameworks* estão disponíveis ou podem ser criados a partir das interfaces externas do *Big Data* Estratégico. O Instituto Gartner desenvolve e disponibiliza *frameworks* específicos para a área de tecnologia e *business*, que têm sido muito usados entre os gestores de todos os tipos de instituições.

O importante é sempre manter o foco no objetivo e só utilizar a extensão do *framework* quando necessário.

4. Conclusões

Não há explicações finais,
há explicações melhores

Marcelo Gleiser

A alardeada revolução dos dados digitais não acontecerá. Revolução é uma mudança radical por conspiração ou complô, que pressupõe que o novo tomará de assalto o cenário descontentando o que era velho.

Usar dados digitais para gerar conhecimento é uma evolução cultural e como tal se dará de modo paulatino, obedecendo à inexorável curva de aprendizagem. Só nos resta aceitar pacientemente o passo do seu curso ou incrementar o número de tentativas e erros que fazem a curva andar.

Essa pesquisa se inclina para a segunda opção. Inconformados com a dinâmica do movimento de adoção das tecnologias associadas a *big* data, buscamos algo que pudesse ser ao mesmo tempo evangelizador e prático. A proposta era de estimular o uso de dados digitais em processos de tomada de decisão sem, contudo, elevar demasiadamente o risco de insucesso. Os pragmáticos, aqueles que aceitam o novo quando percebem os benefícios já não demandam sacrifícios, eram o nosso alvo.

O primeiro desafio era colocar em palavras a visão para *big data* que desenvolvemos ao longo dos quatro anos em contato com o assunto. Precisava ser algo perene e que pudesse ser abraçada pelas pessoas que desejassem adotar a ferramenta prática que seria desenvolvida baseada nessa visão. Em retrospecto, podemos ver que tudo previsto na visão está presente no *framework* e reciprocamente nada relevante da ferramenta está ausente da visão proposta. Consideramos, portanto, o objetivo plenamente alcançado.

O segundo objetivo era definir um ecossistema, um conjunto de *stakeholders*, que esgotasse sem violar o escopo aceitável para a governança de processos de decisão *data driven.* O ecossistema com a configuração apresentada aqui já existe há três anos e foi usado em diversos casos teóricos e alguns reais, sem jamais ter sobrado ou faltado *stakeholders.* Ao longo dos três anos incluímos alguns agentes, como o de anonimização

de dados, por exemplo, mas o *stakeholder* de governança de dados associado a ele já estava previsto.

Enfim o terceiro e último objetivo. Criar um *framework* para adoção de *big data* com a fundamentação teórica da ciência das redes foi o maior desafio dessa pesquisa. Acreditamos que qualquer inflexão radical na curva de aprendizagem de *big data* não se dará sem o profundo entendimento de que os dados de que tanto falamos são a versão digital dos sistemas complexos que desejamos controlar.

Acreditamos ter alcançado esse objetivo ao propor uma etapa longa e detalhada de dataficação do problema a ser resolvido antes de qualquer outro investimento em maquinas, equipamentos, softwares, etc. O esforço de dataficação de problemas leva em duas direções igualmente desejáveis em um ecossistema *big data*: a digitalização de novos dados para cobrir digitalmente o problema a ser resolvido e a reutilização de dados coletados para determinado fim, em outros contextos. A dataficação é sem dúvida a parte mais importante do ecossistema *big data*, sem nenhum demérito para os demais *stakeholders* do ecossistema. O paradoxo de ter no mundo físico a parte mais importante da analítica digital é apenas aparente. Dados não têm valor intrínseco e o seu valor só é conhecido quando ele é utilizado na cadeia produtiva do conhecimento. Não é por acaso que não se vê comercio de dados por volume. Quando se atribui um valor a um determinado *dataset* é porque já se saber o que se pode fazer com ele. Para cada alocação, um valor diferente. A alocação de dados, que em última instância é responsável pelo seu valor, pertence ao mundo físico, não digital.

A pretensão de que o *framework* ajudasse na visualização sistêmica dos *stakeholders* de um sistema produtivo de conhecimento digital, também foi satisfatoriamente atingida. O planejamento de aquisição de cada agente do ecossistema é dimensionado a partir do valor que é agregado ao sistema por seus *stakeholders.* A ideia é de que nada deve faltar, mas também nada pode sobrar. Mais do que isso, que pode dar uma conotação apenas de quantidade, o correto é afirmar que nada deve ficar inadequado.

A pesquisa ficou aquém na sua expectativa de gerar nomenclatura consensual e compartilhável. Um detalhe que pode ser resolvido tardiamente, não sem uma certa frustração de nossa parte.

5. Recomendações para estudos futuros

> Essa é a melhor época possível para estar vivo, quando quase tudo o que você pensava está errado.
>
> Tom Stoppard

Nos últimos quatro anos, desde que comecei a me interessar por *big data* como empoderador de tomador de decisão, tenho me deparado com diversas questões para as quais ainda não encontrei respostas. Algumas têm relação direta com a proposta dessa pesquisa e por isso apresentarei aqui como sugestão para novos estudos.

Umas das grandes dificuldades na adoção de *big data*, como foi dito, é a compreensão de que os dados digitais representam sistemas complexos, portanto iluminam a rede dos problemas que desejamos resolver. O *design* de rede ainda é uma competência limitada. O que se faz em *design* de rede é a modelagem delas a partir de dados digitais, no entanto, como a proposta desse trabalho é partir do mundo físico para criar projetos de analítica digital, precisamos de uma metodologia para design de redes a partir de informações do mundo físico.

Outra questão importante e ainda não resolvida diz respeito à avaliação de eficácia de uma analítica preditiva. Como seu objetivo é desencadear ações, ao agirmos no mundo físico interferimos na dinâmica dos problemas e ficamos impossibilitados de saber se, caso a ação não tivesse sido tomada a predição realmente se realizaria. No Crie chamamos de ‘não evento’ o evento que não acontece (embora tenha sido previsto em um modelo preditivo) porque as ações cabíveis e necessárias foram tomadas. Se a previsão feita nunca chegou a acontecer (não evento), como sabemos então que o modelo estava correto? Como saber se foram de fato as ações empreendidas que levaram ao não acontecimento do evento previsto ou se simplesmente o modelo não funcionou? Um estudo aprofundado sobre isso seria bastante útil e importante.

Em relação mais direta com essa pesquisa, alguns trabalhos podem torná-lo mais completo:

- Uma apresentação visual mais adequada
- Definição de nomes mais simples e simbólicos, que possam ser uma referência em relação a ecossistema *big data.*
- Teste da ferramenta com diferentes entradas pelo *framework*
- Desenvolvimento de uma ferramenta em forma de aplicativo para facilitar a adoção do *framework* e coletar dados de *feedback*.

Diversas interfaces para pesquisas futuras podem ainda ser encontradas por pessoas que, tendo acesso a esse trabalho com diferentes olhares, experiências e objetivos, poderão contribuir para que o empoderamento das pessoas com o uso de dados digitais seja mais que uma bela utopia ou algo apenas potencial.

6. Bibliografia

ALLEN, Will J. Complicated or complex: knowing the difference is important for the management of adaptive systems. http://learningforsustainability.net/sparksforchange/complicated-or-complex-knowing-the-difference-is-important-for-the-management-of-adaptive-systems/. Acesso em 25 de ago. 2015

ALLEN, Will J. Working together for environmental management: the role of information sharing and collaborative learning. PhD (Development Studies), Massey University, 2001

ARTHUR, W. Brian. All Systems will be Gamed: Exploitive Behavior in Economic and Social Systems. Santa Fé: Santa Fé Institute, 2014

BARABÁSI, Albert-László; ALBERT, Réka. Emergence of Scaling in Random Networks. http://arxiv.org/pdf/cond-mat/9910332.pdf, 1999.

BARABÁSI, Albert-László. Linked: the new science of networks. Kindle Edition. Cambridge: Perseu Books Group, 2002.

BARABÁSI, Albert-László. Thinking in network terms. Edge Conversation. Set. 2012. Disponível em http://edge.org/conversation/thinking-in-network-terms. Acesso em 29 de set. 2012.

BATTELLE, John. The Search: How Google and Its Rivals Rewrote the Rules of Business and Transformed Our Culture. 2006

BERTOLUCCI, Jeff. Big Data´s New Buzzword: Datafication. Infomation Week. Fev.2013. Disponível em http://www.informationweek.com/big-data/news/big-data-analytics/big-datas-new-buzzword-datafication/240149288. . Acesso em 08 de jan. 2013.

BOOTH, James. Designing Your Own Framework. Disponível em: http://www.jamesbooth.com/designing_your_own_framework.htm. Acesso em: 18 mar 2016, 22:40:00

BOYD, Danah e CRAWFORD, Kate. Critical Questions for Big Data. In: Information, Communication & Society. Vol. 15:5, 662-679, 2012.

CALDARELLI, Guido; CATANZARO, Michele. Netowork: A very short introduction. Kindle Edition. Oxford: Oxford University Press, 2012.

CAVALCANTI, José Carlos. http://creativante.com/new/index.php/2013-02-03-19-36-05/2013-02-04-18-19-49/155-big-data-e-analitica-ferramentas-e-modelos-de-negocios-revolucionarios. Acesso em 2013

CAVALCANTI, Marcos; COSTA, Luciana. Big Data não é tecnologia

CHESBROUGHT, Henry W. Top 10 Lessons on the New Business of Innovation. MIT Sloan Management Review. North Hollywood, Winter 2011, 35-41 (The Era of Open Innovation)

COLLE, Simone de; FREEMAN, R. Edward; HARRISON, Jeffrey S.; PARMAR, Bidhan WICKS and Andrew C. Stakeholder Theory: The State of the Art. New York: Cambridge University Press, 2010.

COSTA, Luciana S., FERNANDES, Soraya e ZUPPO, Daniella. Big Data: Estudo do Ambiente, Desafios e Análise Estratégica para o Brasil. Rio de Janeiro: UFRJ/COPPE, 2013

CHRISTENSEN, Clayton M. The innovator's dilemma: when new technologies cause great firms to fail. Boston Harvard Business School Press, 1997

DEMCHENKO, Yuri; NGO, Canh; MEMBREY, Peter. Architecture Framework and Components for the Big Data Ecosystem. SNE technical report, 2013.

EVANS, Dave. The Internet of Things: How the Next Evolution of the Internet Is Changing Everything. Cisco Internet Business Group. Abr. 2011.Disponível em http://www.cisco.com/web/about/ac79/docs/innov/IoT_IBSG_0411FINAL.pdf. Acesso em 25 de fev. 2013.

FIGUEIREDO, Daniel. Disponível em http://www.land.ufrj.br/~daniel/rc/slides/aula_2.pdf

FREEMAN, R. Edward; HARRISON, Jeffrey S.; WICKS, Andrew C.; PARMAR, Bidhan and COLLE, Simone de. Stakeholder Theory: The State of the Art. New York: Cambridge University Press, 2010.

FRICK, Walter. What to Do When People Draw Different Conclusions From the Same Data. Harvard Business Review (digital edition). Março 2015. Disponivel em

FUNKE, Joachin. Solving Complex problems: exploration and Control of complex systems. 1991. Disponível em https://www.psychologie.uni-heidelberg.de/ae/allg/mitarb/jf/Funke_1991_CPS.pdf. Acesso em 24 de ago.2015

GALDO, Alessandra M. Ruiz. Capacidades Dinâmicas para Inovação Aberta: Análise com base no Capital intelectual, no prelo.

GERHARDT, Tatiana E. e SILVEIRA, Denise T. (org). Métodos de pesquisa. Porto Alegre: Editora da UFRGS, 2009.

GLOUBERMAN, Sholom; ZIMMERMAN, Brenda. Complicated and Complex Systems: What Would Successful Reform of Medicare Look Like? Commission on the Future of Health Care in Canada. 2002. Disponivel em http://c.ymcdn.com/sites/www.plexusinstitute.org/resource/collection/6528ED29-9907-4BC7-8D00-8DC907679FED/ComplicatedAndComplexSystems-ZimmermanReport_Medicare_reform.pdf. Acesso em 25 de ago. 2015.

GOHEL, Hardik; GONDALIA, Vivek. Role of SMAC Technologies in E-Governance Agility. Mumbai. CSI Communications. Vol. 38:7, 7-9, 2014.

GONZÁLEZ, Marta C.; BARABÁSI, Albert-László. From data to models. Nature. V.3, abr. 2007. Disponível em http://www.barabasilab.com/pubs/CCNR-ALB_Publications/200704-05_NaturePhys-DataModels/200704-05_NaturePhys-DataModels.pdf. Acesso em 15 de dez. 2012.

GOUVEIA Jr, Amauri. O conceito de modelo e sua utilização nas ciências do comportamento: breves notas introdutórias. Revista Estudos de Psicologia, PUC-Campinas, v. 16, n. 1, p. 13-16, janeiro/abril, 1999.

HIDALGO, Cesar. How information grows. London: Penguin Books, 2015

KAHANE, Adam. Como resolver problemas complexos. Ed. 1. São Paulo: Editora Senac, 2008.

KELLOG, Ronald T. Fundamentals of Cognitive Psychology, 2nd Edition. Sage Publications, 2007.

KLEINFELD, Judith. Could it be a Big World? http://www.judithkleinfeld.com/ar_bigworld.html. Acesso em

KORHONEM, Jouni. Four ecosystem principles for an industrial ecosystem. In: Journal of Cleaner Production Ed. 9. 2001. pg. 253–259

LÉVY, Pierre. Cibercultura. 3ª edição. São Paulo: Editora 34, 2011

MILGRAM, Stanley; JEFFREY, Travers. The Small-World Problem. 1967. Disponível em http://measure.igpp.ucla.edu/GK12-SEE-LA/Lesson_Files_09/Tina_Wey/TW_social_networks_Milgram_1967_small_world_problem.pdf. Acesso em 08 de jan. 2013

MITCHEL, Melaine. Complex Systems: Network Thinking. Santa Fé: Santa Fé Institute (2006).

MITCHEL, Melaine. Complexity: A guided tour. New York: Oxford University Press, 2011.

MOORE, Geoffrey A. Crossing the Chasm: Marketing and Selling High-Tech Product to Mainstream Customers. Adobe Acrobat E-Book Reader edition. Perfect Bound / HarperCollins Publishers, 2001

NEPOMUCENO, Carlos. Revolução Cognitiva e o macro aumento da taxa de circulação de ideias. http://nepo.com.br/2012/01/18/revolucao-cognitiva-e-o-macro-aumento-da-taxa-de-circulacao-de-ideias/#sthash.F2AseHJA.dpuf. Acesso em 04 de fev. 2016.

NG, Irene. Complicated vs Complex Outcomes. http://value-basedservicesystem.blogspot.co.nz/2011/08/complicated-vs-complex-outcomes.html. Acesso em 25 de ago. 2015

OSTERVALDER, Alexander. The Business Model Ontology: A Proposition in a Design Science Approach. PhD (*Development Studies*). l'Ecole des Hautes Etudes Commerciales de l'Université de Lausanne. 2004

PELTONIEMI, Mirva; VUORI, Elisa. Business ecosystem as the new approach to complex adaptive business environments. In: Proceedings of eBusiness research forum. 2004. p. 267-281.

ROGERS, Everett M. Diffusion of Innovations. Third edition. New York: The Free Press, 1982

RON, Adner. The Wide Lens: What Successful Innovators See That Others Miss. Nova York: Penguin, 2013

SIEGEL, Eric. Predictive Analytics. Hoboken: John Wiley & Sons, Inc, 2013

SILVA, Daniela L., SOUZA, Renato R. e ALMEIDA, Maurício M. Ontologias e vocabulários controlados: comparação de metodologias para construção. Ci. Inf., Brasília, v. 37, n. 3, p. 60-75, set./dez. 2008

SCHÖNBERGER, Viktor M.; CUKIER, Kenneth. Big Data: a revolution that Will transform how we live, work and think. Kindle Edition. New York: Houghton Mifflin Harcourt Publishing Company, 2013.

SANTOS, Augusto P.de M. Análise dos Impactos da Tecnologia da Informação nas Organizações: Um ensaio à luz da Teoria da Agência e da Teoria dos Custos de Transação. 2003. Dissertação (Mestrado em administração Pública e de Empresas) – Fundação Getúlio Vargas FGV, Rio de Janeiro. 2003. P. 23-32.

SLOTINE, Jean-Jacque; LIU, Yang-Yu; BARABÁSI, Albert-László. Controllability of complex networks. Nature. V.473, mai. 2011. Disponível em http://www.uvm.edu/~cmplxsys/newsevents/pdfs/2011/barabasi-network-controllability-nature-2011.pdf. Acesso em 15 de dez. 2012.

SITES VISITADOS

A Quinta Onda. www.aquintaonda.blogspot.com.br/2008/09/hype-cycles.html

Carlos Nepomuceno. Nepôsts - Rascunhos compartilhados. Disponível em: < http://nepo.com.br/>. Diversos acessos.

Community Tool Box. Developing a Framework or Model of Change. Disponível em: <http://ctb.ku.edu/en/4-developing-framework-or-model-change>. Acesso em: 18 mar 2016, 22:42:00

Computer World UK. http://www.computerworlduk.com/news/public-sector/3400618/maude-introduces-independent-reviews-for-open-data-countries/. Acesso em 25 de mar. 2013.

http://ctb.ku.edu/en/4-developing-framework-or-model-change

EMC. http://www.emc.com/leadership/digital-universe/2012iview/big-data-2020.htm e https://www.emc.com/collateral/analyst-reports/idc-digital-universe-united-states.pdf

Google Inc. www.google.com/intl/en/policies/privacy/ e http://support.google.com/a/bin/answer.py?hl=pt-BR&answer=60762. Acesso em 13 de mar. 2013.

IBM http://www-01.ibm.com/software/data/bigdata/. Acesso em 13/01/2013

http://www.jamesbooth.com/designing_your_own_framework.htm. Acesso em 09/03/2016.

http://www.jamesbooth.com/designing_your_own_framework.htm

Massachusetts Institute of Technology (MIT). Auto –ID LabsMIT. Grupo de pesquisa em indentificação por radiofrequência (RFID). http://web.mit.edu/newsoffice/2012/auto-id-cloud-of-things-big-data.html. Acesso em 17 de fev. 2013.

New York Times: http://bits.blogs.nytimes.com/2013/02/01/the-origins-of-big-data-an-etymological-detective-story/?_r=0. Acesso em 15 de ago. 2015

Open Data Institute. http://www.theodi.org. Acesso em 20 de mar. 2013.

Open Definition. http://support.google.com/a/bin/answer.py?hl=pt-BR&answer=60762. Acesso em 13 de mar. 2013.

Open Tracker. www.opentracker.net/article/definitions-big-data. Acesso em 22 de ago. 2015.

Personal Perspective on the Origin(s) and Development of "*Big data*": The Phenomenon, the Term, and the Discipline. http://www.ssc.upenn.edu/~fdiebold/papers/paper112/Diebold_Big_Data.pdf

Royal Pingdom. http://royal.pingdom.com/2013/01/16/internet-2012-in-numbers/. Acesso em 13 de mar. 2013.

Tech & Net. Divulgação de notícias de tecnologia e internet. http://www.techenet.com/2011/07/cisco-net-coisas-infografico/ Acesso em 25 de fev. 2013.

The Engagement Platafor. Gamification frameworks. Disponível em <http://www.hisocial.com/eng/blog/gamification-frameworks/>. Acesso em: 18 mar 2016, 23:00:00.

The Guardian. http://www.guardian.co.uk/technology/free-our-data Acesso em 25 de mar. 2013. http://www.theguardian.com/technology/2008/aug/03/internet.email. Acesso em

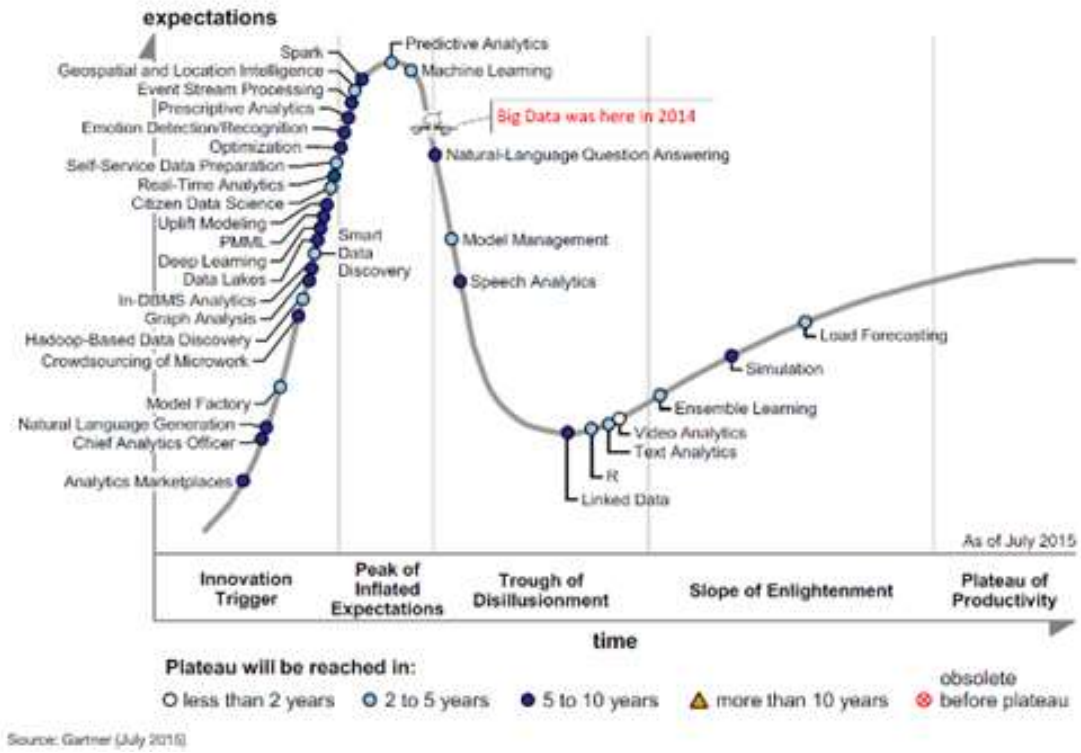
Trendwatching (setembro de 2012) http://www.trendwatching.com/trends/pdf/2012-08%20MINITRENDS.pdf. Acesso em 08 de jan. 2013

Wikipedia http://en.wikipedia.org/wiki/Main_Page. Diversos acessos.

ANEXO I – FIGURA INTEGRANTE DE MATERIAL EXPLICATIVO DO INSTITUTO GARTNER SOBRE A HYPE CYCLE CURVE

ANEXO II – FIGURA INTEGRANTE DE RELATÓRIO 2014 DO INSTITUTO GARTNER SOBRE *TECHNOLOGY HYPE CYCLE*

ANEXO III – FIGURA INTEGRANTE DE RELATÓRIO 2014 DO INSTITUTO GARTNER SOBRE *TECHNOLOGY HYPE CYCLE*



Source: Gartner (July 2015)

ANEXO IV – ORIENTAÇÕES PARA DESENVOLVIMENTO DE UM *FRAMEWORK – COMMUNITY TOOL BOX*

http://ctb.ku.edu/en/4-developing-framework-or-model-change

Developing a Framework or Model of Change

1. **Describe the intended uses of your framework or model of change:**
    a. To convey the purpose and direction of your initiative or effort (i.e., the outcomes sought and how you will get there)
    b. To show how multiple factors interact to influence the problem or goal
    c. To identify actions and interventions more likely to lead to the desired result

    How will your organization or effort use its framework or model of change?

    Related resources:
    Developing a Logic Model or Theory of Change

2. **Outline your initiative or program's vision and mission:**
    a. Vision - summarize your statement of your initiative's dream for the future. It should be:
        i. Easy to communicate
        ii. Uplifting/inspiring to those involved in the effort
        iii. A reflection of the perspectives of the community it represents

        What's your group's vision for the effort?
    b. Mission - provide your group's mission statement. It should communicate:
        i. What the group is going to do (e.g., "...by connecting and supporting children and caring adults.")
        ii. What is going to do it (e.g., "Promote caring relationships...")

        What's your group's mission for the effort?

        Related resources:
        Proclaiming Your Dream: Developing Vision and Mission

Statements

3. **State the objectives of your initiative or effort:**
    a. Summarize all of the specific measurable results of your initiative or program that you anticipate. These should include behavioral changes and related community-level outcomes.
    b. State your assumptions and hypotheses regarding the personal and environmental factors contributing to the problem or goal. Discover these using multiple strategies:
        i. Forward logic (But why?) - ask yourself why this problem exists. What brought it about? What maintains it?
        ii. Reverse logic (But how?) - ask how this problem might be solved or goal accomplished?
        iii. Identify what personal factors (e.g., knowledge, belief, skills) contribute to the problem or goal
        iv. Identify the environmental factors (e.g., supports and services; access, barriers, and opportunities; consequences of efforts; policies and broader conditions) that contribute to the problem or goal.

Related resources:
Creating Objectives
Gathering and Using Community-Level Indicators
Community-Level Indicators: Some Examples
Analyzing Root Problems of Problems: The "But Why?" Technique
Defining and Analyzing the Problem

4. **Describe the appropriate scope or level of your framework or model of change:**
    a. The overall initiative - may include all strategies and relationships used to affect change and bring about improvement for the overall problem or goal (e.g., reduce violence; promote caring relationships)

b. A particular initiative or program - may include only the components and elements of a specific aspect of the overall effort (e.g., education programs; policy change)

c. A specific work plan for an action or model for cooperation among stakeholders or participating agencies

Which level will your model of change describe?

Related resources:

Identifying Action Steps in Bringing About Community and System Change

5. Identify ALL components to include in the logic model or model of change. Include:

a. Purpose or mission - what the group is going to do and why

b. Context and conditions under which the problem or goal exists and which may affect the outcome (e.g., history of the effort, broad cultural and environmental factors, political situation, economic conditions)

c. Inputs - resources and supports available, as well as constraints or barriers to meeting the initiative's objectives

d. Activities or interventions - what the initiative or program does to bring about change and improvement (e.g., enhancing support, modifying access)

e. Outputs - direct results or products of the group's activities (e.g., number of people trained or activities conducted)

f. Effects - more broadly measured outcomes or results (may include immediate, intermediate, and longer-term effects)

Related resources:

Generating and Choosing Solutions

Understanding Risk and Protective Factors: Their Use in Selecting Potential Targets and Promising Strategies for Interventions

Proclaiming Your Dream: Developing Vision and Mission Statements

Understanding and Describing the Community

Defining and Analyzing the Problem
Identifying Community Assets and Resources
Identifying Action Steps in Bringing About Community and System Change
Gathering and Using Community-Level Indicators
Community-Level Indicators: Some Examples

6. **Using the components, draft a picture of the framework or model of change. Include:**
    a. An expected time sequence (what occurs before what) to arrange the components and elements of the framework or model.
    b. Arrows or other methods to communicate directions of influence and sequences of events. Some arrows may point in both directions to show and interaction or mutual influence.

    Related resources:
    Developing a Logic Model or Theory of Change

7. **Check for the completeness of your logic model.**
    a. Select a case situation (real or hypothetical) in which you can obtain feedback about your logic model
    b. Check for the usefulness of the elements of the model (e.g., was it understandable?)
    c. Check for the completeness of the model (e.g., what was missing?)
    d. Revise and add to make it more complete.

    After testing the usefulness of the model with a case situation, what revisions did you make?

    Related resources:
    Developing a Logic Model or Theory of Change

8. **Once all current components and elements are identified and incorporated into the framework or logic model, put it to use. Uses may include:**

a. Orienting those doing and supporting the work - use to explain how the elements of the initiative or program work together, where contributors fit in, and what they need to be able to make it work.
b. Planning - use to clarify your initiative or program's strategies, identify targets and outcomes, prepare a grant proposal, identify necessary partnerships, and estimate timelines and needed resources for the effort.
c. Implementation - use to determine what elements you have and don't have in your initiative or program, develop a management plan, and make mid-course adjustments.
d. Communication and advocacy - use to justify to others why the initiative/program will work and to explain how investments will be used.
e. Evaluation - use to document accomplishments, identify differences between the ideal program and the currently operating one, determine which indicators will be used to measure success and frame questions about attribution (of cause and effect) and contribution of the program/initiative to the mission.

How might you put your model of change to work within your organization or community now? In the future?

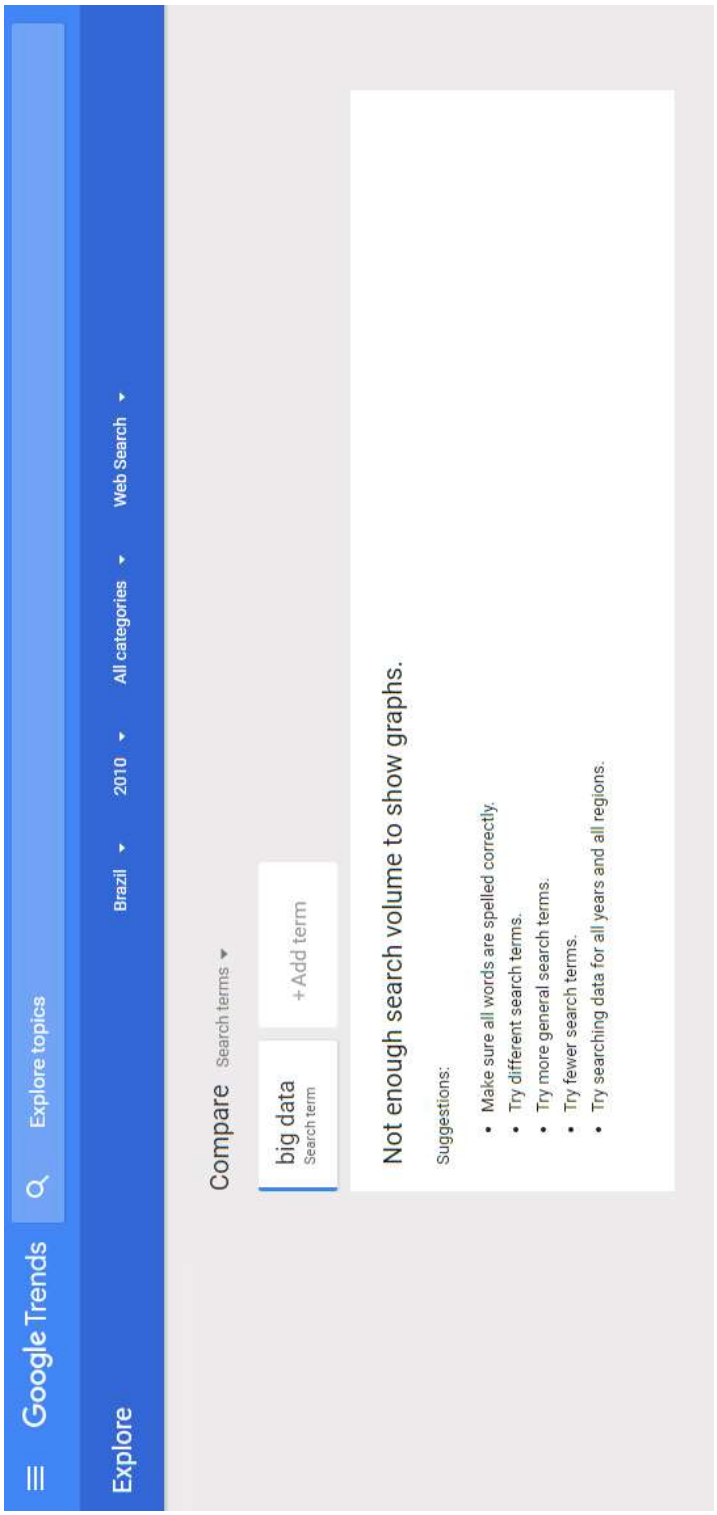
Related resources:
Providing Staff Orientation Programs
Providing Volunteer Orientation Programs
Providing Support for Staff and Volunteers

9. Revise the model (as needed) to adapt the elements and incorporate newly emerging ones. Using the model and seeing the interconnectedness of its components will allow you to:

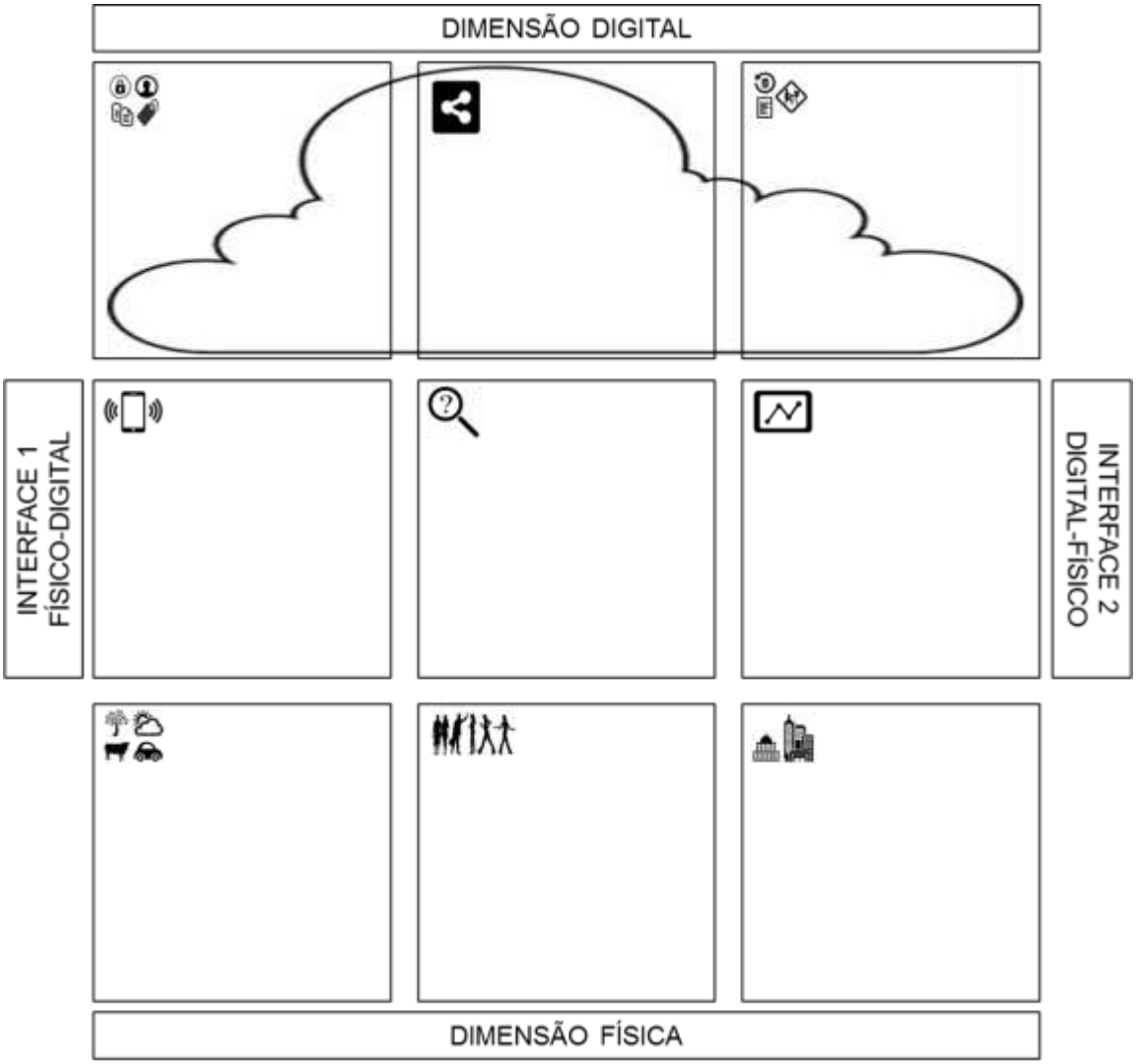
a. Link the path of activities to intended effects or outcomes
b. Plan expansion of activities to reach your goals
c. Understand the boundaries of your program or initiative
d. Adjust course to allow for unanticipated changes
e. Develop a new framework for an extended effort or new initiative

ANEXO V – GOOGLE TRENDS PARA AS BUSCAS POR "BIG DATA" NO BRASIL EM 2010

(CONSULTA FEITA EM 2015 E REPETIDA EM FEVEREIRO DE 2016)



ANEXO VI – *BIG DATA* ESTRATÉGICO - QUADRO DE TRABALHO



APÊNDICE I – CRIE: CENTRO DE REFERÊNCIA EM INTELIGÊNCIA EMPRESARIAL

Fundado em 1998, o Centro de Referência em Inteligência Empresarial (Crie) é o laboratório para empreendedorismo, gestão do conhecimento e inovação da Coppe/UFRJ.

O Crie desenvolve pesquisas, capacita e realiza projetos nas áreas de gestão do conhecimento, avaliação de intangíveis, big data, open data, gestão de redes e complexidade.

Em 2002 foi reconhecido pelo Ministério de Ciência e Tecnologia como um centro de referência em tecnologias de gestão.

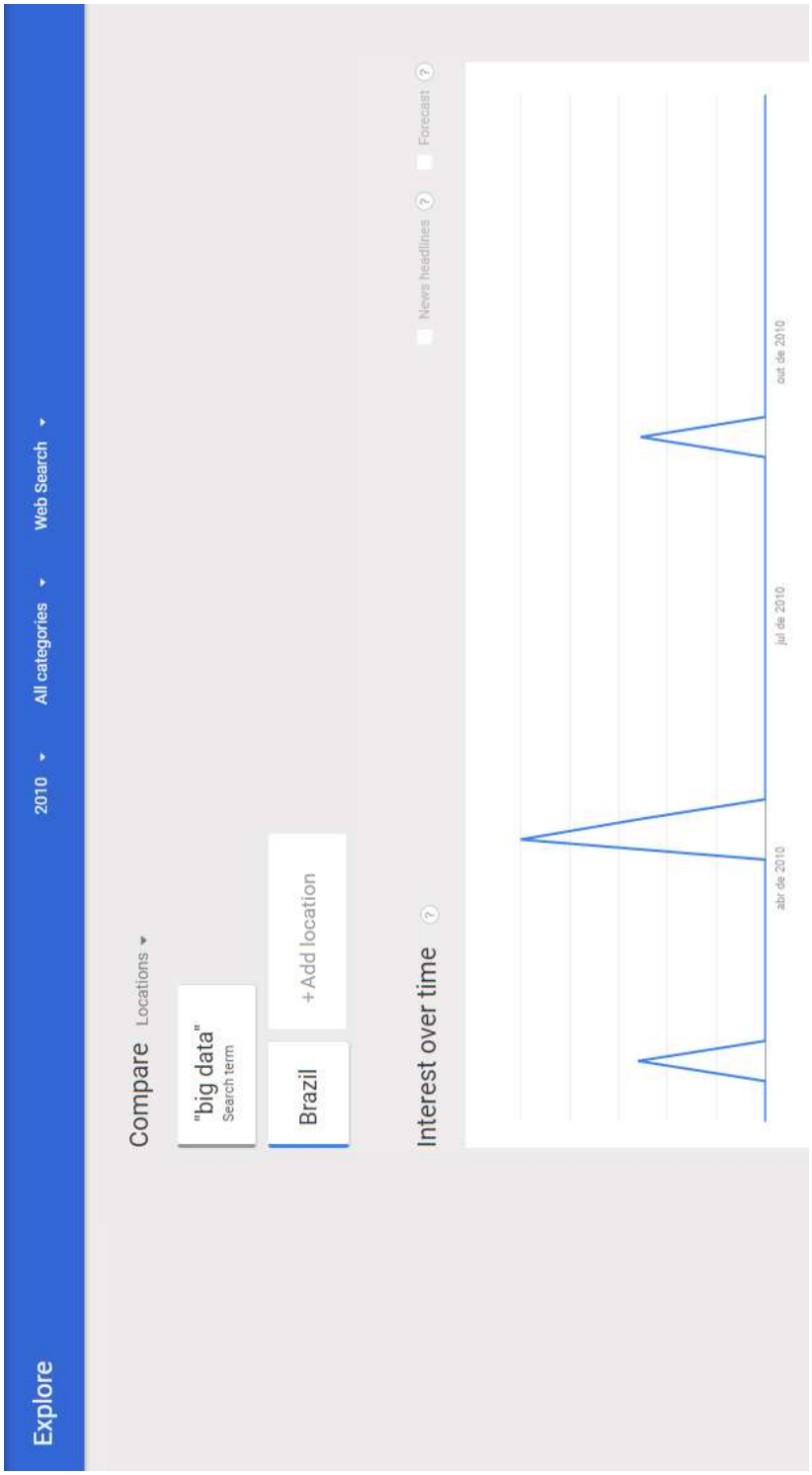
http://www.crie.ufrj.br/

APÊNDICE II – UNIDADES DE MEDIDA DE INFORMAÇÃO

Unidade	Corresponde a	Ou em bytes por extenso aproximado
Byte	1 byte	Um byte
Kilobyte	1.024 bytes	Mil bytes
Megabyte	1.024 kilobytes	Um milhão de bytes
Gigabyte	1.024 megabytes	Um bilhão de bytes
Terabyte	1.024 gigabytes	Um trilhão de bytes
Petabyte	1.024 terabytes	Um quatrilhão de bytes
Exabyte	1.024 petabytes	Um quintilhão de bytes
Zetabyte	1.024 exabytes	Um sextilhão de bytes
Yotabyte	1.024 zetabytes	Um septilhão de bytes
Brontobyte		
Geopbyte		

APÊNDICE III – GOOGLE TRENDS PARA AS BUSCAS POR *"BIG DATA"* NO BRASIL EM 2010

(CONSULTA FEITA EM MARÇO DE 2016)



APÊNDICE IV – CAPA DA REVISTA VEJA COM MATÉRIA SOBRE BIG DATA PUBLICADA NA SEMANA DE 12 A 18 DE MAIO DE 2012

APÊNDICE V – *OCTALYSIS:* FRAMEWORK PARA DRIVES DE MOTIVAÇÃO PARA JOGAR – Fonte: Yukai Chou

http://yukaichou.com/gamification-examples/octalysis-complete-gamification-framework/#.Vu7D3vkrLIU

The 8 Core Drives of Gamification

1) Epic Meaning & Calling

Epic Meaning & Calling is the Core Drive where a player believes that he is doing something greater than himself or he was “chosen” to do something. A symptom of this is a player that devotes a lot of his time to maintaining a forum or helping to create things for the entire community (think Wikipedia or Open Source projects). This also comes into play when someone has “Beginner’s Luck” – an effect where people believe they have some type of gift that others don’t or believe they were “lucky” to get that amazing sword at the very beginning of the game.

2) Development & Accomplishment

Development & Accomplishment is the internal drive of making progress, developing skills, and eventually overcoming challenges. The word “challenge” here is very important, as a badge or trophy without a challenge is not meaningful at all. This is also the core drive that is the easiest to design for and coincidently is where most of the PBLs: points, badges, leaderboards mostly focus on.

3) Empowerment of Creativity & Feedback

Empowerment of Creativity & Feedback is when users are engaged a creative process where they have to repeatedly figure things out and try different combinations. People not only need ways to express their creativity, but they need to be able to see the results of their creativity, receive feedback, and respond in turn. This is why playing with Legos and painting are fun in-and-of themselves and often become Evergreen Mechanics, where a game-designer no longer needs to continuously add more content to keep the activity fresh and engaging.

4) Ownership & Possession

This is the drive where users are motivated because they feel like they own something. When a player feels ownership, she innately wants to make what she owns better and own even more. Besides being the major core drive for wanting to accumulate wealth, this deals with many virtual goods or virtual currencies within systems. Also, if a person spends a lot of time to customize her profile or her avatar, she automatically feels more ownership towards it too. Finally, this is also the core drive that makes collecting stamps or puzzle pieces fun.

5) Social Influence & Relatedness

This drive incorporates all the social elements that drive people, including: mentorship, acceptance, social responses, companionship, as well as competition and envy. When you see a friend that is amazing at some skill or owns something extraordinary, you become driven to reach the same level. Also, it includes the drive we have to draw closer to people, places, or events that we can relate to. If you see a product that reminds you of your childhood, the sense of nostalgia would likely increase the odds of you buying the product.

This Core Drive is relatively well-studied too, as many companies these are days are putting a lot of priority on optimizing their online social strategies.

6) Scarcity & Impatience

This is the drive of wanting something because you can't have it. Many games have Appointment Dynamics within them (come back 2 hours later to get your reward) – the fact that people can't get something right now motivates them to think about it all day long. This is the Core Drive utilized by Facebook when it first started: at first it was just for Harvard. Then it opened up to a few other prestigious schools, and eventually all colleges. When it finally opened up to everyone, many people wanted to join because they previously couldn't get in it.

7) Unpredictability & Curiosity

Generally, this is a harmless drive of wanting to find out what will happen next. If you don't know what's going to happen, your brain is engaged and you think about it often. Many people watch movies or read novels because of this drive. However, this drive is also the primary factor behind gambling addiction. Also, this core drive is utilized whenever a company runs a sweepstake or lottery program to engage users. The very controversial Skinner Box experiments, where an animal irrationally presses a lever frequently because of unpredictable results, are exclusively referring to the core drive of Unpredictability & Curiosity, although many have misunderstood it as the driver behind points, badges, and leaderboard mechanics in general.

8) Loss & Avoidance

This core drive is based upon the avoidance of something negative happening. On a small scale, it could be to avoid losing previous work. On a larger scale, it could be to avoid admitting that everything you did up to this point was useless because you are now quitting. Also, opportunities that are fading away have a strong utilization of this Core Drive, because people feel like if they didn't act immediately, they would lose the opportunity to act forever.

APÊNDICE VI – *FRAMEWORK* PARA DEFINIÇÃO DE CAPACIDADE ANALÍTIVA

Fonte: Crie (Centro de Referência em Inteligência Empresarial)

Autoria: Luciana Sodré Costa e Marcos Cavalcanti

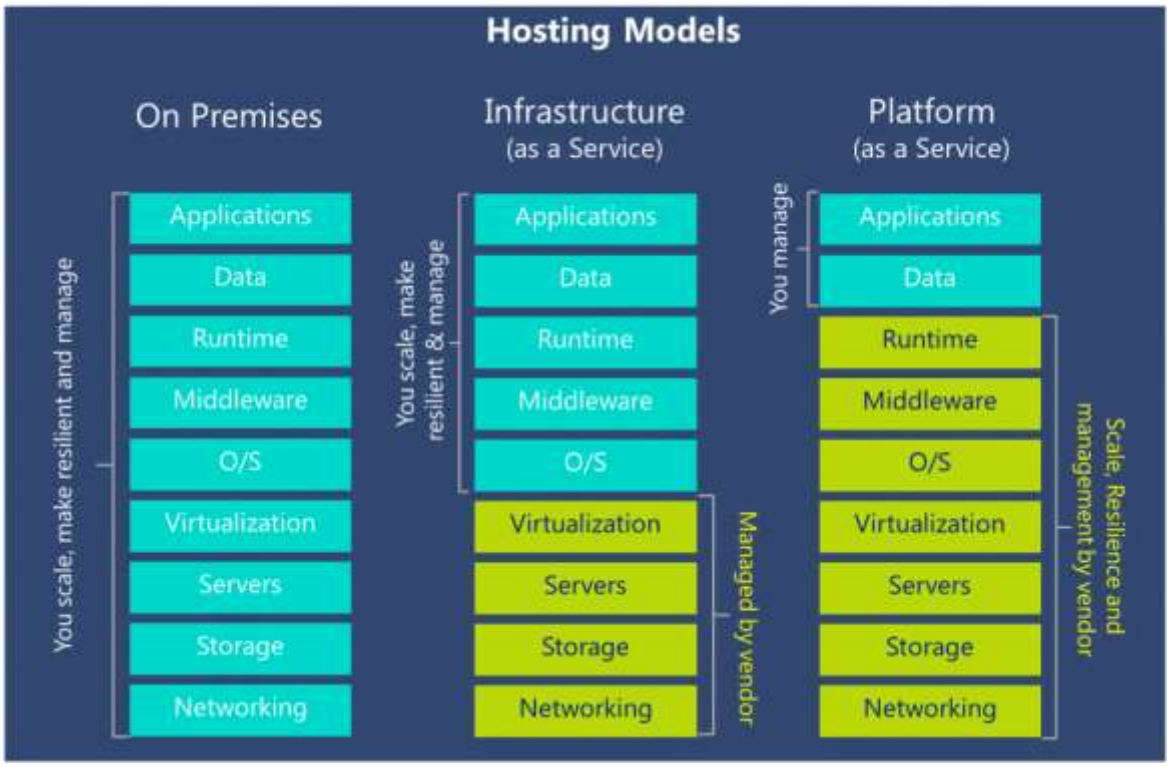
matriz de definição de capacidade analítica

PROBLEMA	CONTROLE DOS DADOS	CROWDSOURCING	FREQUÊNCIA	SOLUÇÃO PARA CAPACIDADE ANALÍTICA
SIM	SIM/NÃO	SIM	1	PRIZE
SIM/NÃO	SIM	SIM	1	HACKATHON
NÃO	SIM	SIM	∞	OPEN
SIM	SIM/NÃO	NÃO	1/ ∞	TERCEIRIZAÇÃO
SIM/NÃO	SIM/NÃO	NÃO	1/ ∞	IN HOUSE

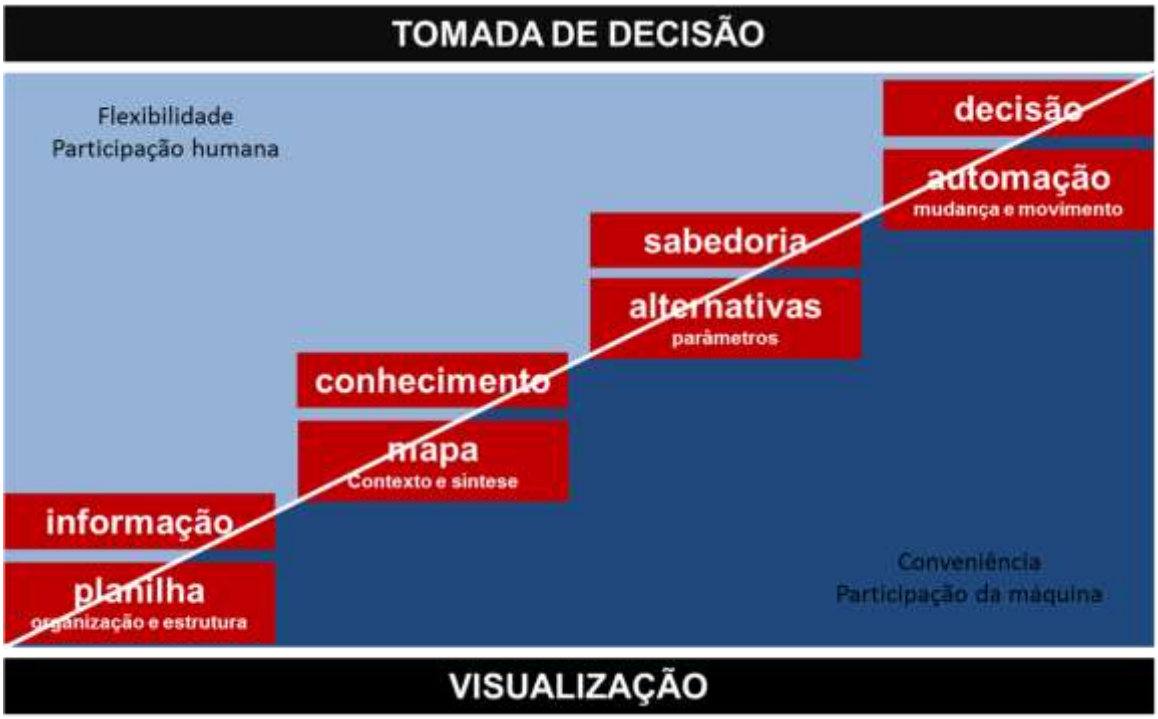
APÊNDICE VII – ***FRAMEWORK CLOUD SERVICE***

Fonte: https://www.linkedin.com/pulse/20140730172610-9679881-pizza-as-a-service

Autoria: Albert Barron



APÊNDICE VIII – *FRAMEWORK* PARA DEFINIÇÃO DE TIPO DE VISUALIZAÇÃO DE DADOS

Fonte: Crie (Centro de Referência em Inteligência Empresarial)

Autoria: Luciana Sodré Costa e Marcos Cavalcanti