

Informação, Conhecimento, Ação Autônoma e

# Big Data

continuidade ou revolução?

Daniel Martínez-Ávila  
Edna Alves de Souza  
Maria Eunice Quilici Gonzalez  
Organizadores

*FiloCzar*

CULTURA  
ACADÊMICA  
Editora



Estudos interdisciplinares sobre implicações éticas, epistemológicas, estéticas e políticas do desenvolvimento e uso de recursos de tecnologias de comunicação informacional (TIC), na ciência e na ação cotidiana, estão na agenda quase que obrigatória de cientistas, filósofos e artistas contemporâneos. No universo acadêmico brasileiro, estudos sobre essa problemática vêm sendo desenvolvidos, não sem controvérsia, há mais de duas décadas. Por um lado, os recursos da tecnologia, dentre os quais se destacam atualmente os de Big Data, propiciam a comunicação sem fronteiras e o acesso a fontes informacionais relevantes, outrora inacessíveis e quicá inimagináveis. Por outro lado, esses mesmos recursos tecnológicos vêm conduzindo à aceleração vertiginosa das ações humanas, com consequências (muitas vezes indesejáveis) políticas, ambientais, éticas, psicológicas, entre outras.

É no cenário das tecnologias de informação digital que a presente obra se insere; seus vários capítulos buscam atualizar e aprofundar a análise da temática do uso de recursos de TIC, na ciência e na ação cotidiana, a partir de

**INFORMAÇÃO,  
CONHECIMENTO, AÇÃO  
AUTÔNOMA E *BIG DATA***

DANIEL MARTÍNEZ-ÁVILA  
EDNA ALVES DE SOUZA  
MARIA EUNICE QUILICI GONZALEZ  
*ORGANIZADORES*

**INFORMAÇÃO,  
CONHECIMENTO, AÇÃO  
AUTÔNOMA E *BIG DATA*:  
CONTINUIDADE OU  
REVOLUÇÃO?**

Marília/Oficina Universitária  
São Paulo/Cultura Acadêmica

2019



*FiloCzar*

**CULTURA  
ACADÊMICA**  
*Editora*

**UNIVERSIDADE ESTADUAL PAULISTA**  
FACULDADE DE FILOSOFIA E CIÊNCIAS - FFC  
UNESP - campus de Marília

*Diretor*

Prof. Dr. Marcelo Tavella Navega

*Vice-Diretor*

Dr. Pedro Geraldo Aparecido Novelli

*Conselho Editorial*

Mariângela Spotti Lopes Fujita (Presidente)

Adrián Oscar Dongo Montoya

Andrey Ivanov

Célia Maria Giacheti

Cláudia Regina Mosca Giroto

Marcelo Fernandes de Oliveira

Neusa Maria Dal Ri

Renato Geraldi (Assessor Técnico)

Rosane Michelli de Castro

Dados da Editora FiloCzar

**FiloCzar**

Rua Durval Guerra de Azevedo, 511 – Parque

Santo Antônio

São Paulo – SP

CEP: 05852-440

Tels.: (11) 5512-1110 - 985246099

E-mail: cesar@editorafiloczar.com.br

www.editorafiloczar.com.br

*Ficha catalográfica*  
*Serviço de Biblioteca e Documentação - FFC*

---

I43 Informação, conhecimento, ação autônoma e big data : continuidade ou revolução? / Daniel Martínez-Ávila, Edna Alves de Souza, Maria Eunice Quilici Gonzalez, organizadores. – Marília : Oficina Universitária ; São Paulo : Cultura Acadêmica, 2019.

278 p. : il.

Coeditora: FiloCzar.

Inclui bibliografia

ISBN 978-85-7249-054-2 (Impresso)

ISBN 978-85-7249-055-9 (Digital)

DOI <https://doi.org/10.36311/2019.978-85-7249-055-9>

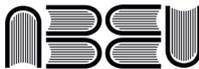
1. Big data. 2. Filosofia. 3. Ciência da informação. 4. Tecnologia da informação. 5. Abordagem interdisciplinar do conhecimento. I. Martínez-Ávila, Daniel. II. Souza, Edna Alves de. III. Gonzalez, Maria Eunice Quilici.

CDD 004.6

---

Copyright © 2019, Faculdade de Filosofia e Ciências / FiloCzar

Editora afiliada:



Associação Brasileira de  
Editoras Universitárias

Cultura Acadêmica é selo editorial da Editora UNESP

Oficina Universitária é selo editorial da UNESP - campus de Marília

# SUMÁRIO

## PREFÁCIO

*Maria Eunice Quilici Gonzalez* ----- 9

## APRESENTAÇÃO

*Daniel Martínez-Ávila; Edna Alves de Souza;*

*Maria Eunice Quilici Gonzalez* ----- 13

## *BIG DATA* NA FILOSOFIA E NA CIÊNCIA

### ***Big Data* e Autonomia: Continuidade ou Revolução?**

*Edna Alves de Souza; Maria Eunice Quilici Gonzalez* ----- 25

### **O contexto da produção científica de *Big Data*: análise cientométrica**

*Ely F. Tannuri de Oliveira; Rafael Castanha;*

*Rene F. Gabriel Junior; Leilah Santiago Bufrem* ----- 47

### **Intersecção temática de programas de pós-graduação brasileiros: considerações sobre *Big Data***

*Jacquelin Teresa Camperos-Reyes ; Luiza de Menezes Romanetto;*

*Ricardo Cesar Gonçalves Sant'Ana;*

*Plácida Leopoldina Ventura Amorim da Costa Santos* ----- 67

## INFORMAÇÃO, TRANSDUÇÃO E METADADOS: FENÔMENOS DE EMERGÊNCIA?

**Informação como ação significativa em processos semióticos emergentes baseados em multiagente**  
*Davidson Bruno da Silva; Tiago Campos Ferreira;*  
*Leonardo Lana de Carvalho ----- 85*

**Transdução Informacional: impactos do controle sobre os dados**  
*Ricardo César Gonçalves Sant'ana ----- 117*

**Metadados de preservação digital na era do *Big Data***  
*Jean Fernandes Brito; Daniel Martínez-Ávila ----- 129*

**Curadoria digital e representação de acervo museológico no AtoM: um estudo da customização do *software***  
*Lais Alpi Landin; Nandia Leticia Freitas Rodrigues;*  
*Lucinéia da Silva Batista; Maria José Vicentini Jorente ----- 143*

### IMPACTOS DE *BIG DATA* NA CIÊNCIA E NA INDÚSTRIA 4.0

**Redes sociais digitais e a esfera pública: “fake news” e a manipulação da opinião coletiva**  
*Magaly Prado ----- 165*

**O *Big Data* no desenvolvimento da indústria 4.0: novas perspectivas para o empreendedorismo acadêmico**  
*Selma Leticia Capinzaiki Ottonicar; Gisele Rodrigues Atayde;*  
*Luis Antonio de Santa-Eulalia ----- 185*

**Algoritmos genéticos e aprendizagem: quem, de fato, aprende?**  
*Monica Aiub ----- 203*

**Regime de informação e análise do discurso: aproximações  
teóricas e conceituais na era *Big Data***

*Mariana da Silva Caprioli; Larissa de Mello Lima;*

*João Batista Ernesto de Moraes ----- 219*

**HOMENAGEM A MARIANA BROENS**

Conversa com o professor Lauro Frederico Barbosa da Silveira

*Edna Alves de Souza ----- 233*

Pequenas palavras para tanto afeto e admiração ----- 251

**SOBRE OS ORGANIZADORES ----- 261**

**SOBRE OS AUTORES ----- 267**

# PREFÁCIO

*É no junto do que sabe bem, que a gente aprende o melhor.*

João Guimarães Rosa

Estudos interdisciplinares sobre implicações éticas, epistemológicas, estéticas e políticas do desenvolvimento e uso de recursos tecnológicos de comunicação, na ciência e na ação cotidiana, estão na agenda quase que obrigatória de cientistas, filósofos e artistas contemporâneos. No universo filosófico brasileiro, estudos sobre essa problemática vêm sendo desenvolvidos, não sem controvérsia, há mais de duas décadas. A contribuição de Mariana Claudia Broens, a quem dedicamos esta obra, para celebrar o seu sexagésimo aniversário, não poderia passar despercebida.

Batalhadora incansável em defesa dos direitos humanos, Mariana Claudia Broens vem desenvolvendo profícuas reflexões sobre implicações éticas, políticas e pragmáticas do emprego desenfreado de tecnologias de computação ubíquas na ação cotidiana. Em particular, ela vem investigando, em conjunto com membros do Grupo Acadêmico de Estudos

Cognitivos (GAEC) da UNESP, possíveis impactos do uso de técnicas de *Big Data* na dinâmica da opinião pública, da linguagem e das emoções humanas.

Na obra *Por uma outra Globalização*, Milton Santos (2001, p. 17) nos presenteia com uma análise lúcida e instigante do cenário contemporâneo, moldado pelo progresso das ciências e das técnicas, cuja utilização permite que o nosso mundo se torne, entre outros, “[...] esse mundo confuso e confusamente percebido”. Essa confusão parece decorrer de um paradoxo: por um lado, os recursos da tecnologia, dentre os quais se destacam atualmente os de *Big Data*, propiciam a comunicação sem fronteiras e o acesso a fontes informacionais relevantes, outrora inacessíveis e quicá inimagináveis. Por outro lado, esses mesmos recursos tecnológicos vêm conduzindo à aceleração vertiginosa das ações humanas, com consequências (muitas vezes indesejáveis) políticas, ambientais, éticas, psicológicas, entre outras.

É nesse cenário que a presente obra se insere; seus vários capítulos buscam atualizar e aprofundar a análise desse mundo “confuso e confusamente percebido”, a partir de uma perspectiva interdisciplinar, falibilista, tendo como leme a interrogação “*Big Data*: reforma ou revolução?”. Na esteira de Milton Santos, entendemos que no passado os colonizadores enviavam exércitos e religiosos para conquistar territórios e apoderar-se de recursos ambientais considerados preciosos; hoje, com recursos avançados de tecnologias digitais, além dos exércitos, os novos colonizadores financiam técnicos qualificados para minerar e comercializar dados de fontes variadas, incluindo aquela de usuários de redes sociais. As consequências, positivas e negativas, dessa complexa atividade clamam por pensadores que, como Mariana Claudia Broens,

ajudam a ampliar o campo de reflexões sobre o presente e o futuro das novas gerações influenciadas pelo acelerado avanço da tecnologia.

Inspiradas na sensibilidade poética de Guimarães Rosa (Grandes Sertões: Veredas, p. 487), expressa através do pensamento que: “É no junto do que sabe bem, que a gente aprende o melhor”, queremos manifestar a nossa profunda admiração por essa pessoa que com sentimento ético e social vem reforçando o “exército de Branca Leoni” na difícil tarefa de agregar trabalho, alunos, amigos, para abrir caminhos com tino e generosidade, disseminando, com coragem e ousadia, sempre que possível, reflexão ampla sobre os saberes e seres essenciais que nos ajudam a viver melhor.

A questão chave sobre o livro, “*Big Data*: continuidade ou revolução”, provoca o leitor a investigar (não pacificar nossos conflitos; posto que o seu eterno retorno é inevitável), aproveitando o extremo limite do possível, a escavar, no mundo em que estamos, o mundo que queremos e devemos criar.

Obrigada Mariana! No meio desse mundo confuso, e confusamente percebido, resta ainda o celebrar da marchinha de carnaval: “Quem sabe, sabe/Conhece bem/Como é gostoso/Gostar de alguém” (Carvalhinho, 1956).

Marília, 09 de agosto de 2019.

MARIA EUNICE QUILICI GONZALEZ

## REFERÊNCIAS

CARVALHINHO. *Quem sabe, sabe*. São Paulo: Odeon, 1956.

GUIMARÃES ROSA, João. *Grande Sertão: Veredas*. São Paulo: Nova Aguilar, 1994.

SANTOS, Milton. *Por uma outra Globalização: do pensamento único à consciência universal*. Rio de Janeiro, São Paulo: Record, 2001.

## APRESENTAÇÃO

*Big Data*: Continuidade ou Revolução? Essa é a questão que direciona a presente obra, dividida em três partes temáticas, que agrupam trabalhos da Filosofia e da Ciência da Informação, entre outras ciências. Sem o objetivo de fechar a questão que preocupa não apenas cientistas e filósofos contemporâneos, os participantes do X Encontro Internacional de Informação, Conhecimento e Ação (EIICA 2018) discutiram, principalmente, os seguintes temas: I - *Big Data* na filosofia e na ciência; II - Informação, transdução e metadados: fenômenos de emergência?; III - Impactos de *Big Data* na Ciência e na indústria 4.0.

Discussões sobre o tema I estão concentradas nos primeiros três capítulos, que focalizam implicações centrais do uso de *Big Data* na Ciência e na ação cotidiana. No primeiro capítulo, intitulado “*Big Data* e Autonomia: Continuidade ou Revolução?”, Edna A. de Souza e Maria Eunice Q. Gonzalez centram suas investigações em dois problemas:  $P_1$  “Estamos vivenciando uma revolução ou uma mera reforma social com o emprego de *Big Data* na ciência e na ação cotidiana?”;  $P_2$  “Quais as possíveis implicações éticas da manipulação de *Big Data* na ação autônoma?”. Elas apresentam uma contextuali-

zação do que vem sendo chamado de *Big Data*, em termos de seu desenvolvimento histórico. Essa contextualização ressalta a dificuldade de uma caracterização exaustiva sobre o tema de *Big Data*. Tal dificuldade é compreensível uma vez que se considere *Big Data* enquanto elemento resultante de uma imbricada relação cultural, acadêmica e científico-tecnológica. Em suas várias dimensões política, econômica, acadêmica, social e cultural, estudos sobre *Big Data* estão frequentemente vinculados às posições extremistas de otimismo exagerado ou pessimismo exacerbado. Relacionada a essa dualidade polarizada na consideração dos *Big Data* (de um lado as apostas miraculosas dos entusiastas e de outro os alertas para o extremo perigo dos críticos radicais) estão as perspectivas de sua inovação ou continuidade, que as autoras, em uma perspectiva faliblista, consideram criticamente. Elas argumentam que o apelo aos recursos de análise de *Big Data*, mesmo que somatório e não revolucionário, no paradigma metodológico vigente, nos remete para além de considerações epistemológicas, àquelas também de cunho ético.

No segundo capítulo, intitulado “O contexto da produção científica de *Big Data*: análise cientométrica”, Ely F. Tannuri de Oliveira, Rafael Castanha, Rene F. Gabriel e Leilah Santiago Bufrem, analisam as características da produção científica sobre *Big Data*. Esse trabalho identifica as áreas do conhecimento, as temáticas candentes e o perfil das publicações periódicas que mais produzem sobre o tema de *Big Data*. Além disso, os autores apresentam os países mais produtivos e as coautorias internacionais relacionadas a esse tema, por meio de procedimentos cientométricos, a partir da base de dados *Scopus*. Eles concluem que a temática está impregnando a ciência mundial e os veículos de comunicação científica mais utilizados na ciência.

No capítulo intitulado “Intersecção temática de programas de pós-graduação brasileiros: considerações sobre *Big Data*”, Jacquelin Teresa Camperos-Reyes, Luiza de Menezes Romanetto, Ricardo Cesar Gonçalves Sant’Ana e Plácida Leopoldina Ventura Amorim da Costa Santos, investigam o tipo de relação teórica existente entre os programas de Ciência da Informação, Ciência da Computação e Matemática, no Brasil, ao abordar temáticas relacionadas com *Big Data*. Os autores reconhecem que na conjuntura dos *Big Data* a informação é resultado da geração, processamento e análise de dados, ciclo esse realizado por diferentes áreas do conhecimento, como a ciência da computação, a matemática/probabilidade e a estatística, entre outras. O estudo identifica, no âmbito de programas de pós-graduação dessas três áreas, quais características, temáticas e disciplinas oferecidas podem conter similaridades pertinentes ao tema dos *Big Data* no Brasil. Trinta e nove programas de pós-graduação de reconhecimento no Brasil, com avaliação CAPES 5, 6 e 7, foram considerados, identificando os conteúdos abordados nesses programas, vinculados aos *Big Data*, assim como a interrelação entre as áreas e disciplinas abordadas.

A segunda parte do livro, intitulada “Informação, transdução e metadados: fenômenos de emergência?”, inicia-se com o capítulo “Informação como ação significativa em processos semióticos emergentes baseados em multiagente”, de Davidson Bruno da Silva, Tiago Campos Ferreira e Leonardo Lana de Carvalho. Os autores analisam o conceito de informação, entendido como ação significativa em sistemas complexos adaptativos. Considerando informação em termos de ações envolvidas em processos semióticos, eles defendem a hipótese que os significados se encontram nos usos coletivos das palavras e expressões. Para dar suporte às suas análises,

eles utilizam modelagem e simulação computacional de processos cognitivos, baseadas em multiagente e computação com inspiração biológica. A hipótese é a de que informação é ação significativa que ocorre em processos semióticos emergentes, implementáveis em sistemas complexos baseados em multiagentes. Um Simulador de Processos Semióticos (SIMPS) foi concebido, cujos resultados indicam que ações significativas podem ser falsas e que essas podem ser mais prejudiciais ao coletivo de agentes do que as comunicações sem convergência de relações semióticas.

No capítulo: “Transdução Informacional: impactos do controle sobre os dados”, Ricardo César Gonçalves Sant’ana argumenta que a crescente participação das Tecnologias da Informação e Comunicação no acesso a dados tem originado mudanças e desequilíbrios nas relações entre os atores envolvidos. Ele ilustra, de forma didática, um cenário em que suportes analógicos propiciam relativa autonomia aos usuários na coleta e registro de dados, sem, no entanto, atender à eficiência desejada nos processos de recuperação de dados. Nesse cenário, passa-se a um novo quadro em que as tecnologias digitais medeiam todo o ciclo de vida dos dados, proporcionando novos patamares de volume, velocidade e variedade dos dados. Contudo, para que esse processo seja possível, os conteúdos precisam sofrer transformações que vão além do simples tratamento, ou operações relacionadas a sua organização, requerendo processos de geração e regeneração de novos conteúdos, que envolvem formatos e energias distintas da original, propiciando a emergência do que o autor denomina *transdução informacional*. Tais transformações ocorrem em camadas de abstração, envolvendo opacidade para a percepção dos atores, que ignoram como se inserem nessas relações entre os diversos ciclos de vida dos dados, em especial para aqueles que não detêm o controle sobre os pro-

cessos envolvidos. Tal opacidade gera o fenômeno de insciência do usuário sobre o campo informacional a que está submetido, gênese da assimetria nas relações de acesso aos dados e que, com o senso comum sobre uma pressuposta eficiência intrínseca de agentes maquínicos, agrava ainda mais a situação.

O capítulo intitulado “Metadados de preservação digital na era de *Big Data*”, de Jean Fernandes Brito e Daniel Martínez-Ávila, apresenta os metadados de preservação digital e sua relação com os *Big Data*. O objetivo dos autores é apresentar uma reflexão sobre os metadados de preservação digital à luz dos *Big Data*. A pesquisa é descritiva, teórica e qualitativa. O capítulo apresenta subsídios para, a partir dos padrões de metadados, desenvolver uma política de preservação digital, além de indicar como o seu processo de armazenamento deve ser realizado.

A segunda parte da coletânea encerra-se com a contribuição de Nândia Leticia Freitas Rodrigues, Lucinéia da Silva Batista, Laís Alpi Landim, Maria José Vicentin Jorente, intitulada “Curadoria digital e representação de acervo museológico no AtoM: um estudo da customização do *software*”. Esse trabalho propõe uma análise da museologia, na segunda década do século XXI. Ao assumir novas perspectivas no contexto da digitalidade e a adoção de espaços híbridos para o compartilhamento de informação dos acervos musealizados, estudos na área de museologia enfrentam uma nova problemática a ser solucionada pelas instituições museológicas no seu movimento em direção à promoção de acesso às suas coleções. Em países em desenvolvimento, como o Brasil, a falta de um *software* aberto e customizável específico para realização do tratamento descritivo de objetos musealizados pode dificultar a representação, recuperação, interoperabilidade de informação e importação de dados entre sistemas com diferentes padrões de

normas, estrutura e metadados usados por museus. As autoras argumentam que as descrições carecem dessa padronização e de *software* que promovam um maior detalhamento descritivo, e que os acervos musealizados, por serem compostos por variados tipos de suportes documentais, demandam atenção às suas peculiaridades. A hipótese levantada nesse capítulo é a de que o *software* de descrição arquivística AtoM pode ser aplicado no processo de tratamento descritivo.

A terceira parte temática do livro, “Impactos de *Big Data* na Ciência e na indústria 4.0” começa com o capítulo “Redes sociais digitais e a esfera pública: *fake news* e a manipulação da opinião coletiva”, de Magaly Prado. A autora discute a interferência de algoritmos de *fake news* nas redes sociais e o seu dano provocado na esfera pública, investigando de que forma notícias, documentos e dados são utilizados nas notícias fraudulentas. Entre as questões tratadas no capítulo está a investigação de como se identificam mentiras publicadas por meio de filtros e o discernimento sobre as estratégias implementadas para engajar a opinião coletiva em determinados públicos. A autora argumenta que a cacofonia provocada pelas manipulações assola as redes sem que se tenha uma clara definição, tampouco uma solução para os problemas por ela provocados; esse é o principal problema, pois, até agora não se desenvolveu um mecanismo eficiente para validar e contrastar conteúdo das notícias. Ela ressalta ainda a importância da reflexão sobre o processo de mediação de notícias na democracia, desejando que se resguarde o zelo para com a informação, que envolve significado, confiança e credibilidade.

O capítulo “O *Big Data* no desenvolvimento da indústria 4.0: novas perspectivas para o empreendedorismo acadêmico”, de Selma Leticia Capinzaiki Ottonicar, Gisele Rodrigues

Atayde, Luis Antonio de Santa-Eulalia, consiste em uma análise de *Big Data* e da indústria 4.0. Os autores ressaltam que a quarta Revolução Industrial, também conhecida como Indústria 4.0, tem trazido mudanças para as organizações contemporâneas; sendo que a principal transformação é a produção massiva de dados, a partir de fontes diversificadas. No contexto da Indústria 4.0, as pessoas e os objetos produzem grande quantidade de dados, processo esse que exige cuidado e atenção dos profissionais no seu uso em tomadas de decisões. Eles entendem que as transformações da Indústria 4.0 podem impulsionar o empreendedorismo entre alunos universitários, na medida em que surgem novas oportunidades de negócios; ênfase é dada à maneira pela qual os *Big Data* podem contribuir para a aplicação da Indústria 4.0. O trabalho sugere que a produção massiva de dados pode estimular o desenvolvimento de negócios e o aumento de procura por novas profissões, e habilidades competentes para interpretar a informação disponível em bancos de dados. Os autores defendem a hipótese de que, apesar da quarta Revolução Industrial ter como consequência o desaparecimento de algumas profissões, há novas demandas profissionais que estão cada vez mais relacionadas com a busca, interpretação, análise, organização e disseminação de dados e informação.

No capítulo “Algoritmos Genéticos e Aprendizagem: Quem, de fato, aprende?”, Monica Aiub discute alguns estudos sobre algoritmos genéticos, com ênfase em hipóteses do Darwinismo Neural, proposto por Edelman (1987; 1988; 1992), e o conceito de hábito presente na obra de Peirce. A autora questiona quem é o sujeito/agente da aprendizagem e discute se há continuidade ou revolução no uso dos algoritmos genéticos. Algoritmos genéticos simulam/replicam processos de evolução biológica, a partir de cálculos de probabilidade. Algumas pesquisas nesse campo têm gerado algoritmos capazes

de desenvolver atividades consideradas de exclusividade humana, como identificar personalidade, diagnosticar depressão, produzir arte criativa, entre outras. Algoritmos genéticos têm sido utilizados em educação, gerando padrões específicos de associação de dados que, repetidos, podem se tornar hábitos incorporados. Além disso, cada vez que um aplicativo é utilizado, novos dados dos usuários são coletados, incluindo as formas de associação de dados; tais dados constituem “capital” da empresa proprietária do aplicativo. A autora indaga: considerando as recentes mudanças na Lei de Diretrizes e Bases da Educação Nacional (LDBN), a qual permite que parte da carga horária seja oferecida à distância, inclusive na Educação Básica; as discussões no Senado acerca de educação domiciliar (*homeschooling*); a 4ª Revolução Industrial, com a aposta na paulatina substituição de profissionais (professores, médicos, advogados, engenheiros etc.) por *softwares*, que rumos estaríamos traçando para a educação? Seriam tais rumos uma revolução educacional ou a continuidade de um processo de precarização da educação?

Por último, o trabalho de Mariana da Silva Caprioli, Larissa de Mello Lima e João Batista Ernesto de Moraes, intitulado “Regime de informação e análise do discurso: aproximações teóricas e conceituais na era *Big Data*”, focaliza ideias de Bernd Frohmann, em especial, o conceito de ‘Regime de Informação’. Esse conceito foi proposto originalmente em 1984, como uma genealogia de políticas informacionais, ou seja, como uma alternativa aos estudos de política de informação e, também, como crítica ao reducionismo de abordagens da política, praticadas na Ciência da Informação e na Biblioteconomia. Os autores ressaltam que a noção de ‘Regime de Informação’ se aproxima daquela de poder proposta por Michel Foucault, que discute o conceito de ‘Regime de Poder’, referente às formas de estruturação política das ciências que têm

como objetivo a regulamentação das linguagens e dos enunciados científicos, considerando obras no espaço heterológico dos saberes. Os conceitos de formação discursiva e ideológica estão muito presentes na Análise do Discurso, sendo a Formação Discursiva construída pela relação de sentido entre as posições ideológicas, inseridas em um processo sócio-histórico em que os sentidos e as palavras são produzidas. Eles argumentam que, com a Formação Discursiva, pode-se entender que as palavras não têm sentido nelas mesmas; elas derivam o seu sentido do meio em que se inserem, estando diretamente relacionadas com a Formação Ideológica, pois se representam por meio desta última. Os autores consideram ainda que é importante entender que o acervo da unidade de informação, os discursos e ideologias formam uma massiva quantidade de dados, constitutivos dos *Big Data*, e que as formações discursivas e ideológicas, em que esses dados circulam, dão origem a um novo Regime de Informação.

Os organizadores deste volume esperam que a leitura dos capítulos acima sumarizados possa contribuir para uma visão esclarecida, falibilista, e quem sabe crítica, do emprego de técnicas de *Big Data* na ciência, na filosofia e no cotidiano. Afinal, consequências éticas, políticas, epistemológicas, estéticas e semióticas desse emprego direcionarão, em grande parte, o futuro de nossa cultura!

Marília, 10 de agosto de 2019.

*DANIEL MARTÍNEZ-ÁVILA*

*EDNA ALVES DE SOUZA*

*MARIA EUNICE QUILICI GONZALEZ*

***BIG DATA* NA FILOSOFIA E NA CIÊNCIA**

# ***BIG DATA* E AUTONOMIA: CONTINUIDADE OU REVOLUÇÃO?**

*Edna Alves de Souza  
Maria Eunice Quilici Gonzalez*

O cenário inusitado da contemporaneidade, caracterizado pelo crescimento exponencial do volume de dados coletados, armazenados e organizados em uma variedade de formatos digitais e a velocidade de coleta, nunca antes vistos, ilustra aspectos do que atualmente se compreende, vagamente, por *Big Data*. A utilização de dados massivos, por vezes sem a permissão dos usuários, apresenta desafios à ciência e à vida cotidiana. Um desses desafios, de especial interesse para a presente investigação, consiste na capacidade de agentes racionais se adequarem a esse cenário, mantendo-se em sintonia com ele sem prejuízo de sua autonomia.

Neste contexto, dois problemas centrais direcionam a presente investigação:

P<sub>1</sub>: Estamos vivenciando uma revolução ou uma mera reforma social com o emprego de *Big Data* na ciência e na ação cotidiana?

Discutimos também possíveis implicações dos impactos dos *Big Data*, já sentidos, na ação autônoma. Interessa-nos, em particular, elaborar considerações éticas da forma com que instituições públicas e privadas, fazendo uso de técnicas de análise de *Big Data*, atingem pessoas, das mais diversas formas. Nesse contexto, o segundo problema pode ser assim formulado:

P<sub>2</sub>: Quais as possíveis implicações éticas da manipulação de *Big Data* na ação autônoma?

Não temos respostas definitivas, nem mesmo consensuais, para os problemas P<sub>1</sub> e P<sub>2</sub>, mas julgamos relevante a reflexão filosófica sobre eles, pois as implicações para a ciência e a vida cotidiana do que podemos chamar de Era de *Big Data* são amplas e, aparentemente, graves.

Consideramos a postura falibilista, caracterizada pela abertura à possibilidade de reformulação, ou abandono de opiniões e crenças previamente estabelecidas, diante do reconhecimento de erro ou de novidade relevante, um desiderato científico-epistemológico de grande valor prático. Nesse sentido, sugerimos que posicionamentos extremos, sejam eles pessimistas ou otimistas, em relação às implicações éticas e epistemológicas da Era de *Big Data*, sejam colocados em análise da forma mais crítica possível.

O capítulo está estruturado em duas partes. Na primeira, apresentamos uma contextualização dos *Big Data*, em termos de seu desenvolvimento histórico. Essa contextualização conduz a uma reflexão sobre a dificuldade de uma carac-

terização exaustiva sobre o tema de *Big Data*. Tal dificuldade pode ser compreendida ao considerarmos *Big Data* enquanto elemento resultante de uma imbricada relação cultural, acadêmica e científico-tecnológica. Em suas várias dimensões política, econômica, acadêmica, social e cultural, estudos sobre *Big Data* estão frequentemente vinculados às posições extremistas de otimismo exagerado ou pessimismo exacerbado. Relacionada a essa dualidade polarizada na consideração dos *Big Data* (de um lado as apostas miraculosas dos entusiastas e de outro os alertas para o extremo perigo dos críticos radicais) estão as perspectivas de sua inovação ou continuidade. Na segunda parte, refletimos sobre possíveis implicações do emprego de recursos de *Big Data* na ação autônoma. Por fim, argumentamos que o apelo aos recursos de análise de *Big Data*, mesmo que somatório e não revolucionário, no paradigma metodológico vigente, nos remetem para além de considerações epistemológicas, àquelas também de cunho ético.

## 1 DOS DADOS AOS *BIG DATA*

A criação de “bancos de dados” é uma prática antiga e bancos de dados digitais já existem há um bom tempo. Como Mayer-Schonberger & Cukier (2013) ressaltam, o registro de dados mais antigo parece remontar a pelo menos 8.000 a.C., quando os Sumérios usaram continhas de barro para controle de bens comercializados. A contagem em larga escala pode ser exemplificada com o censo praticado pelos antigos governos egípcios e chineses, a fim de manterem o controle governamental por meio da coleta de dados sobre a população. O surgimento do que hoje denominamos ciência ocidental (a antiga Filosofia Natural), segundo Zingano (2005), também é marcado pela coleta, organização, armazenamento e análise

de dados. A exploração, o registro e a classificação, de forma metódica, de dados sobre o mundo natural, empreendidos por Aristóteles e a fundação da Biblioteca de Alexandria, são exemplos emblemáticos da relação existente entre bancos de dados e a prática científica. A era digital, por sua vez, veio facilitar e acelerar o registro e o processamento de dados.

De acordo com Mayer-Schonberger & Cukier (2013), por volta de 1880 deflagrou-se nos Estados Unidos a necessidade de novas técnicas de processamento de dados, uma vez que a Constituição Americana prescrevia a realização do censo a cada década. Porém, dada a aceleração do crescimento populacional e a demora em se completar o censo, a informação obtida ficava obsoleta antes mesmo de se tornar disponível. Assim, eles relatam que foi preciso oito anos para que o censo de 1880 fosse concluído e a expectativa era a de que o término do próximo censo ultrapassaria dez anos, o que contrariaria, inclusive, a própria exigência legislativa. Eles ressaltam ainda que, dada essa situação problemática, o inventor americano Herman Hollerith foi contratado para usar sua ideia de cartões perfurados e máquinas de tabulação para a elaboração do censo de 1890. Ele, em um feito memorável, conseguiu diminuir o tempo de tabulação de oito anos para menos de um ano, marcando, assim, o início do processamento automatizado de dados, que serviu de base para o que mais tarde se tornou a IBM (*International Business Machines*). Mas, tal feito, como método de coleta e análise de grande quantidade de dados, ainda era lento, complexo e dispendioso.

Na chamada Era de *Big Data*, a novidade que merece maior reflexão e discussão diz respeito à crescente escala, sofisticação e onipresença da coleta e análise de dados, bem como a utilização por pesquisadores de técnicas avançadas de análise

de correlações para identificação de padrões de informação, potencialmente úteis, para subsequente inferência ou tomada de decisão. Tudo isso realizado em tempo real!

Tecnologias informacionais de comunicação (TICs), como as redes móveis e a computação em nuvem ou névoa, deram origem a mundos de informação incompreensivelmente grandes, tornando mais complexos os métodos de coleta e análise de *Big Data*. Os analistas de dados (humanos e máquinas), a partir do uso de técnicas avançadas de correlação, podem escrutinar enormes quantidades de dados, em pouco tempo, para prever comportamentos, situações e eventos de maneiras inimagináveis há bem pouco tempo. Estamos falando de dados armazenados em *terabytes*, *petabytes*, *exabytes* ou *yottabytes* (um trilhão de *terabytes*!).

Nesse cenário, definir ‘*Big Data*’ não é uma tarefa simples e parece exigir considerações interdisciplinares. São levantadas questões que oscilam sobre a origem do termo à imprecisão de seu significado, passando por aplicações distintas em áreas diversas. O rastreamento das primeiras ocorrências do termo diz pouco a seu respeito além de seu forte apelo comercial. Ibekwe-SanJuan & Bowker (2017, p. 192) observam que um analista de dados, Douglas Laney, no entanto, contribuiu decisivamente para a atual caracterização dos *Big Data* “[...] ao cunhar os ‘três Vs’ populares e cativantes dos *Big Data* (volume, variedade e velocidade)”.

O exponencial aumento de volume, velocidade e variedade de dados processados, exige um planejamento, em especial, por parte dos usuários neste novo cenário que está se instalando no século XXI. Nas palavras de Laney (2001, p.1): “O e-comércio, em particular, explodiu os desafios do geren-

ciamento de dados em três dimensões: volumes, velocidade e variedade”.

A digitabilidade, o volume, a velocidade e a variedade, na coleta e análise de dados, parecem ser os aspectos reconhecidos cientificamente como característicos dos *Big Data*, apesar da persistente imprecisão quanto à natureza dos dados. Aos três Vs iniciais, têm sido acrescidos muitos outros Vs, que não tão distintivos como os originais, mas que os orbitam, como o V de valor e o de veracidade. Em um espírito entusiasta e de divulgação, há quem diga que “estamos agora a caminho dos 100 Vs dos *Big Data* e da Ciência dos Dados!” (SHAFER, 2017, p. 3).

Do ponto de vista físico (quantitativo), esclarecem Ibekwe-SanJuan & Bowker (2017, p. 192):

[...] os *Big Data* representam volumes de dados que os algoritmos tradicionais de bancos de dados são incapazes de lidar e que requerem infraestruturas e algoritmos de computação mais robustos e distribuídos, como *hadoop clusters*, *grid infrastructure* e *cloud clusters*.

O crescimento constante de bancos de dados, de computação ubíqua e de desenvolvimento de *softwares* capazes de, dentre outras coisas, interconectar sistemas tecnológico-informacionais e proceder à análise de dados, com recursos algorítmicos variados, parece indicar que uma “revolução silenciosa” está a caminho. A confluência de diversos fatores está tornando os *Big Data* uma força poderosa por si só, geradora de um tipo radicalmente novo de “infraestrutura de conhecimento” comum e científico. Daí o questionamento, expresso em  $P_1$ , se a Ciência que hoje faz uso de *Big Data* em suas investigações representa uma continuidade, uma convergência, de práticas

científicas tradicionais ou se está nos conduzindo a uma revolução científico-tecnológica, no sentido kuhniano.

Não é nosso objetivo discutir aqui as várias interpretações da concepção kuhniana de revolução científica, mas apenas ressaltar um de seus sentidos, qual seja, que após uma revolução, um novo paradigma surge sob o qual se assentará a pesquisa científica futura. Nesse sentido, cabe questionar se a sociedade contemporânea está passando por uma revolução, que reconfigura hábitos e relações, desde as mais básicas, existentes entre seus elementos, dando lugar a um novo modelo de organização (que se poderia chamar de universo da cultura digital), ou se está passando apenas por uma reforma, com a adição de um novo espaço público, da realidade virtual, que, nem sempre (ou quase nunca) está em sintonia com a realidade em seu sentido tradicional, ou seja, o mundo natural.

A literatura sobre os *Big Data* é permeada de exemplos, como os apresentados por Mayer-Schonberger & Cukier (2013), que pretendem, dentre outras coisas, ilustrar o aspecto revolucionário ocasionado por essa nova forma de lidar com informação disponível através de computação ubíqua, que atua ininterruptamente na coleta de dados sobre o cotidiano das pessoas, através de celulares, cartões de crédito, câmeras, entre outros. Vejamos alguns deles: (a) Google, por meio dos seus mecanismos de pesquisa, considerando o momento e a localização das buscas de usuários, consegue prever fenômenos, como surtos de violência e tendências de desemprego, antes mesmo que os resultados estatísticos de pesquisas oficiais sejam oficialmente divulgados pelo Estado. (b) Técnicos de empresas de cartões de crédito registram e analisam vastas quantidades de dados com informação sobre hábitos e ações financeiras pessoais para tentar detectar fraudes e identificar

tendências de compra do consumidor. (c) Pesquisadores da área médica analisam os registros de saúde de milhares de pessoas para tentar identificar correlações úteis entre tratamentos médicos e resultados de saúde. (d) Empresas realizam estudos de “mineração de dados” sobre informações pessoais em sites de redes sociais, na tentativa de identificar preferências sutis do consumidor e criar estratégias eficientes de *marketing*. Os dados de “geolocalização” permitem que instituições públicas e privadas, como grandes empresas, analisem dados de dispositivos móveis para fazer inferências intrigantes sobre a vida das pessoas e a economia: dados coletados do rastreamento da localização de celulares – por exemplo, a periodicidade com que as pessoas vão aos *shoppings* – pode ser um indicador econômico para medir a demanda dos consumidores.

As situações indicadas acima são amplamente comentadas por Mayer-Schonberger & Cukier que, não por acaso, intitulam a sua obra *Big Data: A revolution that will transform how we live, work, and think*. Elas ilustram aspectos impactantes dos *Big Data* para questões da vida em sociedade. O alegado caráter revolucionário acarretado pelos *Big Data* na vida cotidiana, e também na ciência, estaria na subjacente metodologia adotada e desenvolvida pelos estudiosos de *Big Data*, ou seja, o foco implícito nos exemplos acima: o trato, já em curso, dos dados a partir da identificação de *correlações* consideradas relevantes para a análise e previsão do fluxo de eventos no mundo.

Independente de os *Big Data* representarem uma revolução ou uma continuidade (com algumas novidades impactantes) para atividades científicas e cotidianas, concordamos com Bollier (2010, p. 1) que suas “[...] implicações para os negócios, o governo, a democracia e a cultura são enormes”. Daí

a dificuldade de separar caracterizações de *Big Data* de seus elementos valorativos mais amplos que, comumente, conduzem às posturas dicotômicas, conforme veremos a seguir.

Para além da perspectiva meramente quantitativa, as técnicas de análise de *Big Data* podem ser vistas, em uma perspectiva metodológica, como um conjunto de métodos de análise que incorpora uma diversidade de conhecimentos, técnicas de programação e tecnologias. Desse modo, Boyd & Crawford (2012, p. 663) caracterizam as suas diferentes dimensões cultural, tecnológica e acadêmica que, em suas palavras, “[...] repousam na interação de tecnologia, análise e mitologia”. Elas sugerem que a tecnologia desenvolvida, visando a aplicação na análise de *Big Data*, permite maximizar o poder de computação e de precisão algorítmica para coletar, analisar, vincular e comparar grandes conjuntos de dados. A análise, no contexto dos *Big Data*, possibilita a extração de padrões em grandes conjuntos de dados, a fim de se fazer intervenções econômicas, sociais, técnicas e legais. A mitologia, por sua vez, consistiria na crença generalizada de que os *Big Data* oferecem uma forma superior de inteligência e conhecimento, com a aura de verdade, objetividade e precisão, que pode gerar percepções antes impossíveis.

Em síntese, indicamos até aqui concepções de *Big Data* que enfatizam o seu papel impactante na contemporaneidade. Nossa tendência, entretanto, é admitir que sem uma clara definição atual de seu escopo, os recursos de *Big Data* promovem apenas uma reforma, antes que uma revolução, nos modos de organização de hábitos individuais e coletivos existentes a partir de uma longa história evolutiva humana. Sem dúvida, técnicas de análise e de organização de dados podem auxiliar a pesquisa literária, bibliográfica, política, acadêmica, ao otimizar a reu-

nião e o cruzamento de dados massivos, em diferentes épocas e contextos; indivíduos e grupos podem aumentar suas tendências consumistas e exacerbar suas tendências agressivas com os recursos das TICs. Mas é principalmente na ciência que os *Big Data* estão trazendo novidades consideradas revolucionárias.

Anderson (2008), por exemplo, decreta o fim da teoria científica tradicional, ao afirmar, no subtítulo de seu artigo, que “o dilúvio de dados torna o método científico obsoleto”. Para ele, em uma era de computação em nuvem e conjuntos de dados massivos, o verdadeiro desafio não é criar novas taxonomias ou modelos científicos, mas peneirar os dados de novas maneiras para encontrar correlações significativas.

Em contraste com a visão otimista *do* uso de *Big Data* na ciência, críticos dos *Big Data* insistem que esse instrumento de manipulação de dados massivos poderia ser visto, no máximo, como um complemento possível e não uma alternativa ou um substituto da investigação científica tradicional. Embora técnicas de *Big Data* sejam boas para detectar correlações, não são elas, por si só, que especificam quais correlações são significativas e o porquê de o serem. Ilustrativamente, o site *Spurious Correlations*, de Tyler Vigen (2015), ficou conhecido por trazer uma série de correlações espúrias engraçadas, mas com o propósito de denunciar o uso irresponsável de correlações estatísticas. Um programa computacional que registra correlações em grandes bases de dados foi desenvolvido, com resultados, no mínimo provocantes, que relacionam, por exemplo, a redução do consumo per capita de margarina à diminuição de divórcios.

Radicalismos expressos por Anderson (2008) e outros entusiastas e promotores dos *Big Data*, bem como, em outro extremo, por Vigen (2015) e demais críticos aguerridos dos *Big Data*, indicam uma tendência na literatura corrente sobre

o tema: perspectivas de inovação ou continuidade na consideração de *Big Data* parecem estar relacionadas à dualidade polarizada – de um lado o otimismo exagerado de seus entusiastas e de outro o pessimismo exacerbado de seus críticos radicais.

Consideramos que o problema da produção e da organização acelerada de dados massivos e do uso de técnicas de análise de *Big Data* na investigação científica, enquanto envolve em questões e análises do tipo “sociotecnológicas”, pode conduzir a retóricas utópicas ou distópicas (de acordo com as propensões dos envolvidos), que se confundem com a própria caracterização do que é entendido por *Big Data*.

Além disso, talvez uma das chaves para o entendimento de possíveis efeitos dos *Big Data* na Ciência, bem como na ação autônoma, esteja na análise do conceito de dado. Essa análise da dinâmica do conceito de dado parece ter o potencial de revelar mudanças sutis na ação coletiva que estão ocorrendo no, ou com o, uso de *Big Data*, não sendo obscurecido de antemão pelos vieses utópicos ou distópicos de questões sociotecnológicas.<sup>1</sup>

Na próxima seção refletimos a respeito de algumas possíveis implicações do emprego de recursos de *Big Data* na ação autônoma.

## 2 INFORMAÇÃO, *BIG DATA* E GRAUS DE AUTONOMIA DA AÇÃO

Em seus instigantes estudos sobre a natureza da informação, Gregory Bateson (1972, p. 460) a caracteriza como: “[...] *uma diferença que faz uma diferença*”. Essa aparente enig-

---

<sup>1</sup> Essa hipótese é apresentada e discutida em nosso artigo *Big Data: Philosophy of Science Contemporary Issues* (no prelo).

mática caracterização expressa um dos pilares da chamada Escola de Palo Alto, que inaugurou a *Sistêmica Qualitativa*. Em contraste com os pressupostos da teoria matemática da comunicação (SHANNON & WEAVER, 1949) – que propiciou, entre outros, subsídios formais para o desenvolvimento de métodos analíticos de *Big Data* –, ênfase é dada à abordagem sistêmica ao estudo do *significado* presente na comunicação.

A *Sistêmica Qualitativa* começou a se estruturar na década de 1950, quando Bateson desenvolve um estudo interdisciplinar sobre a natureza da informação significativa, com a ajuda de filósofos, psicólogos, antropólogos, biólogos, neurocientistas, entre outros (destacando-se Birdwhistell, Goffman, Hall, Jackson, Scheffen, Sigman, Watzlawick). O que interessa à escola de Palo Alto é o entendimento da dinâmica constitutiva de *padrões informacionais* prenhes de significado (que não se reduzem a enormes conjuntos de dados) e de seus possíveis efeitos na ação de organismos contextualmente situados.

Inspirado em hipóteses da Cibernética, em especial na noção de *feedback circular* (WIENER, 1950), e na Teoria de Tipos Lógicos (WHITEHEAD & RUSSELL, 1910), Bateson (1972) sugere uma ponte explicativa entre o que ele denomina *fatos da vida e do comportamento* e a *dinâmica estrutural* constitutiva de padrões de ordem diversas que “ligam” seres vivos, em várias escalas, possibilitando a comunicação de *informação significativa*.

Em sua formulação mais geral sobre a natureza da informação, Bateson argumenta que *informação significativa*, presente em inúmeras formas de comunicação entre seres vivos, emerge de relações *qualitativas* (e não apenas numéricas) estabelecidas entre seres vivos e meio ambiente. A complexa dinâmica de relações contextualmente situada engendraria ti-

pos de *padrões informacionais*, estruturados em várias escalas. Tais padrões, por sua vez, emergem de ações habituais comuns aos organismos que compartilham, não apenas vínculos informacionais e emocionais, mas também aspectos de estruturas biológicas. Bateson (1979) menciona, por exemplo, o padrão que liga funcionalmente as pinças dos caranguejos e os dedos polegar e indicador nas mãos de humanos; o padrão constituído de dois olhos e nariz no centro do rosto, que liga seres humanos a outros animais, como elefantes, macacos e felinos em geral, permitindo um tipo de vínculo fundamental entre diferentes espécies. Uma vez que a comunicação significativa extrapola o domínio das estruturas linguísticas, estas serão analisadas como um caso particular de tipos de padrões informacionais no domínio da pragmática humana.

Com o advento da linguagem proposicional humana, estruturas dinâmicas, ordenadas em redes de padrões informacionais, possibilitaram a emergência de inúmeros planos de comunicação, gerando *disposições* para a ação de agentes morais, com diferentes graus de autonomia. Nesse contexto informacional, diferenças (entre elementos constitutivos de padrões informacionais) podem fazer diferença (não apenas para a comunicação verbal, mas também no plano da ação individual e coletiva).

Um pressuposto basilar da *Sistêmica Qualitativa*, de especial interesse para nós, é a conhecida hipótese batesoniana do *duplo-vínculo* (*double-bind*), segundo a qual relações de dependência entre seres vivos se expressam através de padrões informacionais, muitas vezes contraditórios e conflitantes, que condicionam, por meio da repetição, suas ações. Na Psicologia, o *duplo-vínculo* é caracterizado como: “[...] uma situação na qual, não importa o que uma pessoa faça, ela ‘não

pode vencer'. [...] uma pessoa aprisionada no duplo vínculo pode desenvolver sintomas esquizofrênicos” BATESON et al., 1956, p. 251). É a partir desse pressuposto que investigamos o problema  $P_2$ , sobre possíveis implicações da manipulação de *Big Data* na ação autônoma, focalizando os duplos vínculos que parecem estar sendo estabelecidos na contemporaneidade entre usuários das TICs.

De acordo com Anderson (2008, p. 1), a “[...] era do *Petabyte* é diferente porque *mais* é diferente”. Concordamos que “mais” pode fazer diferença quando se trata de acúmulos de padrões de dados correlacionados. Contudo, resta explicitar o papel da diferença que envolve graus de autonomia.

Entendemos que a ação de uma pessoa (P), ou grupo de pessoas (GP), expressa certo *grau* de autonomia se pelo menos duas condições forem satisfeitas: (1) P (ou GP) possui informação antecipadora de oportunidades de escolha, e (2) P (ou GP) possui capacidade de escolha, independente de pressão externa. Satisfeitas (1) e (2), consideramos que a ação de P (ou de GP) será tanto mais autônoma quanto mais independente de coação, ou imposição externa, for a sua escolha, e quanto mais informação, relacionada a cada escolha, estiver ao seu alcance.<sup>2</sup>

No que concerne aos graus de autonomia individual e coletiva, nossa hipótese é que vivenciamos atualmente uma relação de “duplo vínculo”. Por um lado, as TICs propiciam, no plano individual, inúmeras facilidades; o acesso atraente e eficiente à informação enciclopédica disponível na *Internet*, por exemplo, libera os indivíduos da penosa necessidade de memorizar dados necessários à resolução de problemas. Elas

---

<sup>2</sup> Uma análise mais detalhada do conceito de ação autônoma pode ser encontrada em Gonzalez (2017).

também possibilitam a desejada comunicação, quase imediata, entre pessoas de várias culturas, idades e gêneros. Por outro lado, vivenciamos a invasão de privacidade, de *bullying*, e uma uniformidade cansativa, pouco criativa, de comportamentos no ambiente digital, promovida pela ubiquidade das TICs que tendem a acelerar e padronizar as atividades cotidianas, sobrecarregando os indivíduos com massiva quantidade de informação alienante. É nesse contexto que recursos de análise de *Big Data* permitem que empresas como a Amazon e a Netflix, por exemplo, “criem necessidades” e antecipem os desejos de seus clientes, que nem sempre possuem condições econômicas de compra, comprometendo os graus de autonomia de suas ações.

Agências governamentais e empresariais, como a Google, Facebook, Instagram, estão descobrindo rapidamente usos estratégicos de algoritmos, estruturadores de grandes bancos de dados. Longe de serem neutros, esses algoritmos são empregados também para detectar padrões de ação, laços familiares, sonhos e medos, preferências de compra, entre outros, registrados em traços deixados por usuários das TICs, direcionando, muitas vezes de forma invisível, a dinâmica de formação de opinião dos indivíduos. Aprisionados em um duplo vínculo de “amor e ódio”, muitos indivíduos sentem-se obrigados a aceitar as condições impostas por essas agências para que possam usufruir de seus benefícios.

No plano coletivo, algumas das consequências práticas, positivas, da aplicação de resultados de análises de *Big Data*, dizem respeito, de um lado, à possibilidade de uso, desejável, de correlações de dados para prever e prevenir, idealmente, desastres ecológicos, bem como ações sociais indesejáveis relacionadas à segurança e aos direitos humanos. Por outro lado,

já vivenciamos, em várias sociedades, a crescente e conflituosa polaridade na dinâmica de opiniões em questões de gênero, de política, dentre outras. Estimulados pela influência das mídias digitais, grupos de amigos e familiares estão se desintegrando no confronto de opiniões muitas vezes apoiadas em notícias falsas. Novas formas de vigilância também já estão sendo aplicadas envolvendo algoritmos de rastreamento para extrair informações a partir de traços deixados por indivíduos e grupos sociais. Vivenciamos, também, aplicações de algoritmos manipuladores de grande quantidade de dados na robótica, em especial nas áreas bélicas, que vêm colocando em risco a autonomia da ação humana coletiva. O desconforto resultante desse “duplo vínculo” no emprego de técnicas de análise de *Big Data* – que por um lado otimista trazem benefícios atraentes aos usuários das TICs, mas por outro também os asfixiam – parece estar gerando uma situação cotidiana de “esquizofrenia” individual e coletiva.

No contexto científico, o emprego de algoritmos manipuladores de *Big Data* facilita a coleta de dados, que atualmente seriam impossíveis de serem coletados sem o subsídio desse tipo de recurso. Contudo, o emprego por vezes pouco crítico de técnicas de análise mecânica e de algoritmos que privilegiam a ocorrência de correlações, com prejuízo para a analiticidade da metodologia científica tradicional, indica uma possível revolução silenciosa nos modos de se fazer ciência. De forma ilustrativa, como explicam Santana e Fernandes (2016), na astronomia, o pesquisador recentemente tem passado de recursos do telescópio para os de *Big Data*; ou seja, mais do que perscrutar os céus, de forma mais ou menos direta e objetiva, com equipamentos como o telescópio de mão, em busca de algo novo que explique algum fenômeno ou ofereça resposta a algum problema, hoje, o astrônomo, com conhecimentos em estatística e programação,

emprega recursos de programas computacionais para minerar dados coletados por equipamentos de alta tecnologia, como as sondas espaciais. Os telescópios agora têm suas funções estendidas, pois são equipados com sensores que coletam, constantemente, enorme quantidade de dados o que torna praticamente impossível exaurir a capacidade humana de extração de conhecimento a partir dos dados coletados. Os dados permanecem valiosos não só para a Astronomia como para a Ciência em geral, devido a seu potencial não explorado, ensejando sua reutilização para outras finalidades. Atualmente, a teorização, por exemplo, sobre bilhões de galáxias, a partir da análise de *petabytes* de dados, exige infraestruturas e algoritmos de computação sofisticados. Contudo, é preciso saber que critérios de relevância são empregados para procurar o que se quer descobrir. Uma possibilidade seria observar tudo (o máximo possível), levantando uma quantidade massiva de dados, e tornar explícitos os critérios de relevância empregados para minerar esses dados em busca de correlações significativas para um determinado propósito. Ao se perscrutar “tudo”, corre-se o risco de se perder, em um primeiro momento, a delimitação do problema, do objeto de pesquisa, no sentido tradicional dos termos, e; em um segundo momento, ao ter no banco de dados o contexto da investigação, de se distanciar do contexto original da realidade.

O cenário acima esboçado requer reflexão epistemológica, metodológica e ética, e em certos casos intervenção, sobre possíveis perigos do emprego de recursos analíticos de *Big Data* no trato não só de imensa quantidade de dados, mas também de pessoas. Acreditamos que tal reflexão, se realizada a partir de uma perspectiva da sistêmica qualitativa, pode abrir horizontes para o desenvolvimento de uma ética que respeite a ação autônoma no exercício de maximização de interesses comuns.

## CONCLUSÕES PROVISÓRIAS

Iniciamos este capítulo com a formulação do problema  $P_1$ : Estamos vivenciando uma revolução ou uma mera reforma com o emprego de *Big Data* na ciência e na ação cotidiana? Ressaltamos que a Era de *Big Data* não consiste, simplesmente, no abandono da analiticidade, da metodologia científica tradicional, mas na articulação dessas, indicando um novo modo de se fazer ciência e de se agir no cotidiano, a partir de padrões que podem ser identificados por meio de algoritmos rastreadores de grandes volumes de dados.

O problema  $P_1$  nos remete à consideração de que, uma vez que agora dispomos de conjuntos massivos de dados e técnicas avançadas de manipulação dos mesmos, restaria ainda lugar para a metodologia científica tradicional, que visa a busca de explicações para a ocorrência dos fatos. Tradicionalmente, as explicações científicas estão fortemente apoiadas na noção de causalidade e não apenas em massivas correlações de dados detectadas por algoritmos, cuja estrutura pode depender de interesses, em geral comerciais. Embora Pietsch (2013; 2016) tenha sugerido, entusiasticamente, que a consideração de quantidade massiva de dados aumenta a chance de se identificar por meio de algoritmos de *Big Data* relações causais legítimas (e não meramente correlações), não estamos seguras do valor epistemológico dessa hipótese. Dentre outros problemas, o autor parece incorrer em uma falácia: a da petição de princípio, ou seja, assume de antemão que, no conjunto de dados massivos coletados com recursos de *Big Data*, há relações causais que os algoritmos de análise de *Big Data* seriam capazes de explorar e identificar (que é justamente o que se desejaria descobrir ou demonstrar).<sup>3</sup>

---

<sup>3</sup> Contribuímos para com o debate sobre correlação *versus* relação causal no já mencionado artigo *Big Data: Philosophy of Science Contemporary Issues* (no prelo).

Nas ciências exatas e biológicas, as contribuições de recursos de análise de *Big Data* parecem mais se somar à prática tradicional do que a revolucionar, em um sentido kuhniano de uma alteração radical e profunda em sua estrutura metodológica. Nas ciências humanas, o emprego otimista de técnicas de *Big Data* ainda encontra resistências (justificadas) quanto ao valor de cálculos estatísticos e de correlações descontextualizadas na análise de problemas, cuja compreensão requer informação de subsídios socioculturais, contextualmente situados.

Em ambos os casos, o ideal de inter/multi e transdisciplinaridade na pesquisa enfrenta dificuldades no estudo de problemas cuja solução demanda um diálogo ético, francamente cooperativo, entre pesquisadores que valorizem diferenças ambientais, biológicas, contextuais e culturais. Nesse sentido, considerações sobre a possibilidade de uma revolução científica silenciosa, que poderia estar a caminho na era de *Big Data*, requerem a colaboração filosófico-interdisciplinar voltada à atenção de suas consequências epistemológicas e éticas para as futuras gerações. Metaforicamente, poderíamos dizer que se, atualmente, na ciência, o que é procurado consiste em uma agulha digital perdida em um palheiro de *Big Data*, então, persiste pelo menos a necessidade do cuidado para evitar acidentes e erros, nessa tarefa, que possam trazer consequências indesejáveis.

No que diz respeito à vida cotidiana, embora o apelo aos recursos de *Big Data* possa ter efeitos considerados impactantes, reconfigurando hábitos, comportamentos e entendimentos de nós mesmos e da realidade que nos cerca, acreditamos que uma revolução ainda não ocorreu, não obstante o ufanismo de seus promotores. A herança da história evolutiva humana, geradora de hábitos e disposições de longa duração,

parece sustentar alguns dos passos da reforma de padrões informacionais, aparentemente em curso, na era de *Big Data*.

Julgamos que a escolha de critérios de relevância aplicados na elaboração de algoritmos, não neutros, para a seleção de correlações a serem privilegiadas em situações e estudos específicos, quando desvinculada de considerações éticas pode acarretar resultados indesejáveis a curto e longo prazo. Nesse contexto, sugerimos uma reflexão a partir da *Sistêmica Qualitativa*, destacando, dentre os pressupostos dessa abordagem, a *hipótese Batesoniana do duplo-vínculo*, sobre a dinâmica de relações de dependência que se expressam por meio de padrões informacionais ambíguos, propiciadores de situações conflituosas.

Este é o caso do problema  $P_2$  – sobre possíveis implicações éticas da manipulação de *Big Data* na ação autônoma –, para o qual sugerimos uma resposta não muito otimista, na perspectiva falibilista. Contudo, apenas o tempo histórico revelará, com mais propriedade, se de fato vivenciamos uma revolução ou uma mera reforma social com o emprego de *Big Data* na ciência e na ação cotidiana.

## REFERÊNCIAS

ANDERSON, C. The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. *Wired*, June 23, 2008. Disponível em: <https://www.wired.com/2008/06/pb-theory/>. Acesso em: 30/09/2018.

BATESON, G. *Steps to an Ecology of Mind*. Collected essays in anthropology, psychiatry, evolution, and epistemology. London: Jason Aronson Inc., 1972.

BATESON, G. *Mind and nature*. A Necessary Unity. New York: E. P. Dutton, 1979.

- BATESON, G.; JACKSON, D.; HALEY, J. & WEAKLAND, J. Toward a theory of schizophrenia. *Behavioral Science* [1956] 1(4): 251-264.
- BOLLIER, D. *The Promise and Peril of Big Data*. Washington: The Aspen Institute, 2010. (Report)
- BOYD, D. & CRAWFORD, K. Critical Questions For Big Data. Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society* Vol. 15, No. 5, June 2012, pp. 662–679.
- GONZALEZ, M. E. Q. Autonomous action in complex mechanical systems: A real dilemma? In: Adams, F.; Pessoa Jr., O. & Kogler Jr., J.E. (eds.) *Cognitive science: recent advances and recurring problems*. Wilmington (DE): Vernon Press, v. 1, p. 17-30, 2017.
- IBEKWE-SANJUAN, F. & BOWKER, G. C. Implications of Big Data for Knowledge Organization. *Knowl. Org.* 44 (2017).
- LANEY, D. 3D Data Management Controlling Data Volume Velocity and Variety. 2001. Disponível em: <https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>. Acesso em: 05/09/2018.
- MAYER-SCHONBERGER, V. & CUKIER, K. *Big Data. A Revolution That Will Transform How We Live, Work and Think*. Houghton Mifflin Harcourt, 2013.
- PIETSCH, W. *Big Data – The New Science of Complexity*. 2013. Disponível em: <http://philsci-archive.pitt.edu/9944/>. Acesso em: 24/06/2019.
- PIETSCH, W. The Causal Nature of Modeling with Big Data. *Philos. Technol.* (2016) 29: 137. pp 137–171. Disponível em: <https://doi.org/10.1007/s13347-015-0202-2>. Acesso em: 24/06/2019.
- SANTANA, E. S.; FERNANDES, I. F. Dados Astronômicos: uma proposta de implementação para banco de dados. *Caderno de Física da UEFS*. 16 (02): 2402.1-20 2018. Disponível em: <http://dfisweb.uefs.br/caderno/vol16n2/s4-a2-Dados%20Astronomicos.pdf>. Acesso em: 24/06/2019.

- SHAFER, T. The 42 V's of Big Data and Data Science. 2017. Disponível em: <https://www.kdnuggets.com/2017/04/42-vs-big-data-data-science.html>. Acesso em: 05/09/2018.
- SHANNON, C. E.; WEAVER, Warren. [1949] *A mathematical theory of communication*. Urbana: University of Illinois Press, 1998.
- SOUZA, E. A.; GONZALEZ, M. E. Q.; VICENTINI, M. R. *Big Data*, Pós-verdade e Novo Construtivismo: ainda há um lugar para o realismo? *Semeiosis – Semiótica e Transdisciplinaridade em Revista*, São Paulo, vol. 10, 2019 (no prelo).
- SOUZA, E. A.; GONZALEZ, M. E. Q.; VICENTINI, M. R. *Big Data: Philosophy of Science Contemporary Issues*. 2019 (no prelo).
- VIGEN, Tyler. Spurious correlations. New York: Hachette Books, 2015. *Spurious Correlations*. Disponível em: <http://tylervigen.com/spurious-correlations>. Acesso em 03/09/2018.
- WHITEHEAD, A. N. & RUSSELL, B. [1910]. *Principia Mathematica*. Cambridge: Cambridge University Press, 1963.
- WIENER, N. *Cibernética e sociedade*. O uso humano de seres humanos. Trad. de José Paulo Paes. São Paulo: Cultrix, 1950.
- ZINGANO, M. *Platão & Aristóteles*. O fascínio da Filosofia. 2. ed. São Paulo: Odysseus, 2005. (Imortais da Ciência)

# CONTEXTO DA PRODUÇÃO CIENTÍFICA DE *BIG DATA*: ANÁLISE CIENTOMÉTRICA

*Ely F. Tannuri de Oliveira*  
*Rafael Castanha*  
*Rene F. Gabriel Junior*  
*Leilah Santiago Bufrem*

A diversidade, a geração e o processamento de grande quantidade de dados disponíveis digitalmente são características associadas ao fenômeno *Big Data*, que despontou, mais significativamente, a partir do início do século XXI, com o crescimento exponencial da quantidade de dispositivos e de usuários conectados. *Big Data* consiste, primariamente, de conjuntos de dados, caracterizados pelo volume, variedade e velocidade (3Vs), cuja organização e gerenciamento demandam arquiteturas escaláveis para que os processos de armazenamento, manipulação e análise sejam eficientes.

Com o movimento pela ciência aberta e as formas de democratizar os saberes produzidos, surgem os modelos de produção de conhecimento interativos, abertos, colaborativos e livres para que pesquisadores possam deles dispor, reutilizar

e compartilhar. Renovam-se possibilidades graças à tecnologia e ao desenvolvimento de políticas que favorecem essa abertura para a construção científica, que tem sido chamada de “Revolução na sociedade”.

No contexto da produção sobre o tema, há uma surpreendente motivação para a busca por caminhos alternativos, não só na produção de modelos e estruturas propícias à interação entre pesquisadores e conhecimento, mas também nas formas de comunicação desses dados. Ilustra bem essa motivação o artigo de Roa-Martínez, Vidotti e Santana (2017), que propõe uma estrutura para o assim chamado artigo de dados, como uma modalidade de publicação científica. Se analisado o conceito de dado – atribuível a uma representação, uma informação, um evento ou uma coisa – decorre uma ampla possibilidade de definições. Daí a necessidade de sua especificação, ou seja, para o caso deste estudo, o foco é nos dados utilizados no processo de investigação científica, o que facilitará a discussão sobre o seu compartilhamento e publicação. A partir das definições correntes de dados, os autores procuram ir além, para especificar como caracterizar os dados de investigação ou científicos, aqueles de interesse para a comunidade de pesquisadores. Assim, indicam, para facilitar os estudos sobre o tema, a definição construída por Torres-Salinas e outros (2012):

[...] todo aquel material que ha sido registrado durante la investigación, reconocido por la comunidad científica y que sirve para certificar los resultados de la investigación que se realiza [...] que debe provenir de una fuente única», sin tener en cuenta otros elementos como las notas de laboratorio, los análisis previos, bocetos, informes o conversaciones informales con otros colegas (ROA-MARTÍNEZ; VIDOTTI; SANTANA, 2017, p. 2).

Com sólido material de pesquisa, os autores propõem uma estrutura, com os elementos necessários para permitir o compartilhamento e publicação de dados para que possam ser reutilizados. Essa estrutura de referência pode facilitar o processo de revisão por pares especialistas e as validações técnicas necessárias a partir de um conjunto de elementos ou parâmetros definidos, o que, por sua vez, segundo os autores, posiciona esse tipo de comunicação da produção científica como uma publicação de caráter científico. Defendem a possibilidade de que tenham o mesmo rigor científico e reconhecimento do que têm artigos tradicionais ou convencionais (ROA-MARTÍNEZ; VIDOTTI; SANTANA, 2017). Essa estrutura compõe-se de metadados do conjunto de dados de pesquisa publicados, tendo implícito que os dados não serão apenas compartilhados, mas que deverão ser publicados obedecendo a normas e padrões pertinentes ao processo de publicações que os dados requerem.

Ao identificarem os principais nichos e vertentes de publicação sobre o *Big Data analytics*, Furlan e Laurindo (2017) admitem que a era do *Big Data* já é realidade para empresas e indivíduos, e a literatura acadêmica sobre o tema tem crescido rapidamente nos últimos anos. Realizando uma pesquisa bibliométrica na base de dados *ISI Web of Science*, utilizam-se do termo *Big Data* para focar as práticas de gestão e conseguem identificar cinco grupos distintos dentre os artigos encontrados: evolução; gestão, negócios e estratégia; comportamento humano e aspectos socioculturais; mineração dos dados (*data mining*) e geração de conhecimento; e Internet das Coisas. Concluíram que, embora o tema seja emergente e tenha crescido nos últimos anos, é pouco consolidado, apresentando grande variação nos termos empregados, o que influencia nas buscas bibliográficas. Como resultado complementar da pesquisa, foram identifica-

das as principais palavras-chave empregadas nas publicações sobre *Big Data analytics*, o que contribui para as pesquisas bibliográficas futuras.

A atual configuração da dinâmica relativa à produção e à comunicação científicas permite que se revele o protagonismo da Ciência Orientada a Dados, em concepção abrangente, representada principalmente por termos como *e-Science* e *Data Science*. Com o objetivo de apresentar a produção científica mundial relativa à Ciência Orientada a Dados a partir dos termos *e-Science* e *Data Science* nas bases de dados *Scopus* e *Web of Science*, entre 2006 e 2016, Bufrem *et al.* (2016) estruturaram uma pesquisa na qual buscam informações nas referidas bases para obtenção e representação dos registros bibliométricos, tendo como resultado o destaque, entre a produção científica analisada, para os termos *Distributed computer systems* (2006), *Grid computing* (2007 a 2013) e *Big Data* (2014 a 2016). Na área de Biblioteconomia e Ciência de Informação, a ênfase foi dada aos temas *Digital library* e *Open access*, evidenciando a centralidade do campo nas discussões sobre dispositivos para dar acesso à informação científica em meio digital. Concluem que, sob um olhar diacrônico, constata-se uma visível mudança de foco das temáticas voltadas às operações de compartilhamento de dados para a perspectiva analítica de busca de padrões em grandes volumes de dados.

Nesse contexto de transformações efetivas, Coneglian, Gonzalez e Santarem (2017) voltam-se aos profissionais da informação, desafiando-os a trabalhar com a rapidez imposta tanto pela geração de informação e de dados, quanto pela variedade de fontes em que se encontram. Assim, sua atuação apresenta-se como fundamental ao buscar a sinergia entre as necessidades de gestores e a expertise dos profissionais da computação.

Ao procurar identificar e discutir a intersecção entre as funções executadas pelos administradores (gestores), cientistas da computação e os profissionais da informação em ambientes de *Big Data Analytics*, os autores partem dos estudos da Ciência da Informação, do Profissional da Informação e do *Big Data*, para estabelecer um elo entre tais conceitos. Descrevem a atuação do profissional da informação em todas as fases da análise dos dados, verificando uma intersecção entre a proposta de relacionamento entre os atores “Administrador”, “Profissional da Informação” e “Profissional da Computação” com o chamado cientista de dados, concluindo que, devido a diferentes competências exigidas do cientista, tanto o profissional da informação quanto o profissional da computação podem se especializar nesta área, atuando em campos diferentes, mas complementares, fornecendo informações em cenários de *Big Data*, auxiliando a tomada de decisões e agregando valores para as organizações.

Pode-se questionar se com esse processo transformador ocorreria, de fato, uma cultura participativa, superando a questão da obsolescência das mídias, além da solução no armazenamento em nuvem (*in cloud*). A literatura parece indicar que o *Big Data*, embora enseje a recuperação e uso de um imenso volume de dados, hoje configura novos desafios, sendo o mais premente a longevidade digital, como sugeriu Besser (2010), ao defender a necessidade de “começar a pensar em um modo de assegurar que a informação digital continue acessível durante longo período”. Ele descreve o problema do desaparecimento da informação digital e analisa os cinco fatores causadores de problemas na sua longevidade: o da visualização, o do embaralhamento, o da inter-relação, o da custódia e o da tradução. Propõe ações para reduzir o problema de preservar os documentos digitais para o futuro e alerta sobre

a importância da cooperação internacional para a preservação da informação digital.

Katz e Martin (1997) apontam a coautoria como indicador da atividade de colaboração científica e apresentam algumas de suas vantagens: constitui-se de dados objetivos, podendo ser ratificada por estudos de outros pesquisadores; representa uma metodologia acessível e amigável para quantificar a colaboração. As pesquisas indicam que a coautoria incrementa a produção científica e a visibilidade do país, e tem promovido várias iniciativas governamentais dirigidas ao comportamento colaborativo, no âmbito de pesquisadores, instituições ou países.

Muitos estudos voltam-se para as relações existentes entre produtividade científica e coautoria científica. Entre eles, destaque-se os estudos de Eaton *et al.* (1999), que apontam uma forte relação entre o número de coautores e a produtividade científica, sinalizando que grupos que produzem com maior número de coautores tendem a publicar mais artigos.

Neste contexto, a análise de coautoria reflete um rol possível de intercâmbios e trocas entre os pesquisadores e constitui um procedimento significativo, sendo medida pelo número de publicações em colaboração entre autores, instituições ou países e empregada para identificar e mapear a cooperação regional, nacional ou internacional. A coautoria identifica a circulação do conhecimento e das novas ideias, dentro, especialmente, da comunidade mundial. A coautoria em âmbito internacional, no tocante à internacionalização, adquire relevância especial.

A questão desta pesquisa assim se explicita: quais as áreas e subáreas e quais as questões sobre o *Big Data*, enquanto objeto científico, estão sendo mencionadas em artigos científicos, no período de janeiro de 2016 a dezembro de 2017

registrados nos periódicos na base *Scopus*, levando em consideração as diferentes áreas do conhecimento? Pergunta-se, ainda, quais veículos de comunicação científica estão sendo mais utilizados?

A pesquisa tem como objetivo geral identificar a ênfase dada a artigos sobre *Big Data*. De forma mais especial, objetiva explicitar em quais áreas do conhecimento se destacam, quais as temáticas mais candentes e analisar o perfil das publicações periódicas que mais produzem sobre o tema. Por fim, apresentar os países mais produtivos e as coautorias internacionais, por meio de procedimentos cientométricos.

Esta pesquisa justifica-se pela necessidade de analisar este fenômeno nas diversas áreas do conhecimento e como tem se comportado a dinâmica dos estudos científicos sobre a temática em uma perspectiva sincrônica, em outras áreas, além da Ciência da Computação. Ainda, verificar em qual intensidade este tema está presente em grandes áreas do conhecimento, tais como, Ciências da Vida, Ciências Exatas, Tecnológicas e Multidisciplinar e Humanidades (CAPES, 2014). Pretende-se que este estudo se preste a subsidiar políticas de pesquisa sobre o tema.

## 1. METODOLOGIA

Essa pesquisa, de caráter exploratório e analítico, está situada no segundo grupo alvo de Glänzel (2003). O autor destaca três grupos-alvo em Bibliometria, a saber, *Bibliometria para profissionais da bibliometria* ( $G_1$ ), que busca desenvolver e debater a bibliometria como metodologia, isto é, está preocupada com o seu próprio desenvolvimento conceitual-teórico-metodológico; *Bibliometria aplicada às disciplinas científicas* ( $G_2$ ): esse é grupo de pesquisa bibliométrica “aplicada” e se constitui o maior e mais diversificado grupo de interesse na bibliometria e ainda, *Biblio-*

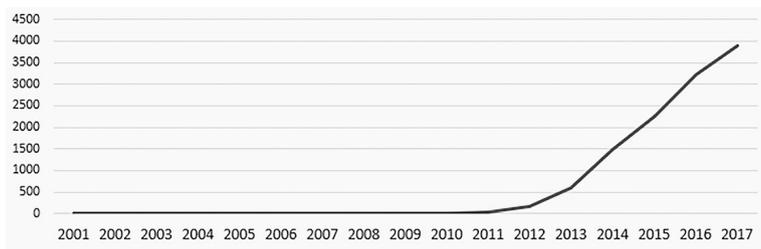
*metria para a política científica e gestão* ( $G_3$ ): esse é o domínio da avaliação da pesquisa com fins de orientar políticas científicas.

Assim, esta pesquisa situa-se em âmbito da cientometria e faz parte do grupo da *Bibliometria aplicada às disciplinas científicas*, considerando que serão utilizados procedimentos bibliométricos para análise dos objetivos propostos.

Como procedimento de pesquisa, em 25 de julho de 2018, consultou-se a base de dados *Scopus*, com o termo de busca “*Big Data*”, presente na tipologia artigos, em seus títulos, resumos e palavras-chave, publicados de 2016 a 2017. Justifica-se a análise desses dois anos por corresponderem a 60,9% de toda produção sobre o tema. Como resultado, foram encontrados 7.103 artigos publicados, sendo que 3.213 deles eram do ano de 2016 e 3.890 do ano de 2017, verificando-se um acréscimo de 21,07% de um ano para outro. Desse total de artigos, somente 436 (6,1%) estão em periódicos *open access*.

A produção sobre *Big Data* saltou de menos de 30 artigos publicados em 2011 para 3.890 em 2017, tendo um alto crescimento.

**Gráfico 1** - Representação Sincrônica da produção sobre o tema.



**Fonte:** Construção dos autores

Para verificação das temáticas mais usadas, utilizaram-se os dados da própria base, buscando-se as dez palavras-

chave mais presentes, adicionando-se às incidências de *Human* com as de *Humans*.

Apresentou-se a tabela dos periódicos que mais publicam no tema, com seus respectivos países, Quartil ao qual pertencem, Índice h e áreas do conhecimento dos periódicos. Ainda, apresentaram-se os países mais produtivos nas temáticas relacionando-os com o desenvolvimento em Ciência, Tecnologia e Informação.

Como procedimento final, mapearam-se, por meio do software *Vosviewer*, as coautorias internacionais dos países envolvidos nas publicações relacionadas ao tema *Big Data*. Para o mapeamento, utilizou-se a totalidade dos dados recuperados na pesquisa e traçou-se a rede de coautorias internacionais, entre os países coautores, procedendo-se à análise.

## 2. APRESENTAÇÃO DOS DADOS

Ao recuperar os 7.103 artigos segundo os critérios descritos anteriormente, analisaram-se as principais temáticas relacionadas a *Big Data*, bem como as principais fontes de publicação periódicas, países que mais publicaram sobre o tema e como estes países relacionam-se em termos de coautorias internacionais.

Para a análise das principais temáticas, recuperaram-se as palavras-chave com maior incidência no conjunto de artigos obtidos pelo termo de busca *Big Data*, assim como apresentado no Quadro 1. Percebe-se que, somadas, estas palavras-chave estão presentes em 5.123 artigos.

O Quadro 1 apresenta nove palavras-chave como as principais relacionadas aos artigos. Dessa maneira, destaca-se que a temática mais incidente nesta coleta de dados é representada pela palavra-chave *Human(s)*. Entretanto, como já descrito

anteriormente, as palavras-chave *Human* (860) e *Humans* (610) foram somadas, visto que apresentam apenas a variação relativa à flexão de número.

### Quadro 1 - Principais temas relacionados (palavras-chaves)

Keywords*	Palavras-Chave	Nº de Artigos
<i>Human(s)</i>	Humano(s)	1470
<i>Data mining</i>	Mineração de dados	893
<i>Data Handling</i>	Manipulação de dados	622
<i>Article</i>	Artigo	607
<i>Digital Storage</i>	Armazenamento Digital	452
<i>Cloud Computing</i>	Processamento em nuvens	405
<i>Information Management</i>	Gerenciamento de Informação	346
<i>Distributed Computer Systems</i>	Sistemas de Computação distribuídos	328
	<b>Total</b>	5.123

Nesse sentido, pode-se dizer que as temáticas mais relevantes desta pesquisa são *Data Mining* (Mineração de dados) presente em 893 artigos, seguidas por *Data Handling* (Manipulação de dados) em 622 artigos. Tanto *Data Mining* quanto *Data Handling* podem ser relacionados a áreas do conhecimento com forte aporte computacional e tecnológico, como Ciência da Computação e Engenharias, corroborando com o que expressa a literatura relacionada a *Big Data*.

Pode-se ainda destacar, neste mesmo segmento, as palavras-chave *Digital Storage* (Armazenamento Digital), *Cloud Computing* (Processamento em nuvens), *Information Management* (Gerenciamento de Informação) e *Distributed Computer Systems* (Sistemas de Computação Distribuída) com 452, 405, 346 e 328 respectivas incidências nos artigos,



É nessa área que atuam os profissionais da informação, principalmente preparando dados para serem processados, como destacado por Coneglian, Gonzalez e Segundo (2017).

Visando identificar as principais fontes de publicação dos artigos analisados e a fim de observar possíveis relações com a temática, foram recuperadas as publicações periódicas que mais publicaram artigos com o termo *Big Data* presente em seus Títulos, Resumos ou Palavras-Chave. Dessa forma, apresenta-se o Quadro 2, composto pelas dez fontes mais produtivas, juntamente com seus respectivos países de procedência, Quartil, Índice h e áreas do conhecimento a que estão associadas.

Observou-se que, das dez publicações periódicas, as duas mais produtivas são oriundas da Venezuela (*Boletín Técnico* e *Revista de La Facultad de Ingeniería*), com respectivos 156 e 120 artigos publicados, efeito ocorrido somente em 2016, não se repetindo em 2017. Porém, mesmo como mais produtivos, estes periódicos estão situados no Quartil de menor relevância e possuem também os menores Índice h, ambos situados no Quartil Q4 e com Índice h, igual a 3 e 6, respectivamente. Representando as áreas de Engenharia Civil, Ciências de Matérias e Ciências Planetárias e da Terra, essas revistas podem ser caracterizadas como *outliers*.

**Quadro 2** - Principais fontes de publicação

Fontes	País	Nº de Artigos	SJR Rankings		Áreas
			Quartil	Índice h	
<i>Boletín Técnico</i>	Venezuela	156	Q4	3	Engenharia Civil; Ciências de Materiais
<i>Revista de La Facultad de Ingeniería</i>	Venezuela	120	Q4	6	Ciências Planetárias e da Terra
<i>IEEE Access</i>	EUA	102	Q1	36	Ciências da Computação; Engenharias

<i>Future Generation Computer Systems</i>	Holanda	65	Q1	85	Ciência da Computação
<i>Agro Food Industry Hi Tech</i>	Itália	60	Q3	16	Engenharia
<i>Neurocomputing</i>	Holanda	54	Q1/Q2	100	Ciência da Computação; Neurociências
<i>Lectures Notes In Computer Science</i>	Alemanha	53	Q2	296	Ciência da Computação; Matemática
<i>Cluster Computing</i>	Holanda	51	Q2	31	Ciência da Computação
<i>International Journal of Control Theory and Applications</i>	Índia	50	Q4	13	Ciência da Computação
<i>Jisuanji Yanjiu Yu Fazhan Computer Research and Development</i>	China	47	Q3	30	Ciência da Computação
<b>Total</b>	-	758	-	-	-

**Fonte:** Construção dos autores

O fato dessas publicações serem oriundas da Venezuela pode suscitar discussões com relação ao papel do país no âmbito do desenvolvimento da temática e possível incentivo a pesquisas relacionadas a Ciência, Tecnologia e Informação, visto que, historicamente, a Venezuela, fazendo parte da América Latina, não está situada entre as grandes produtoras de ciência no mundo como os EUA, a China e a Alemanha. Esta questão apontada carece de maior aprofundamento de estudos.

Em contraponto, os periódicos com Quartil de maior relevância (Q1) são oriundos dos EUA e Holanda: *IEEE Access* (EUA), *Future Generation Computer Systems* (Holanda) e *Neurocomputing* (Holanda). Tais periódicos foram responsáveis

pela produção de 102, 65 e 54 artigos respectivamente. Pode-se também apontar a notoriedade destas publicações ao observar seu Índice h, visto que apresentam respectivamente os valores de 36, 85 e 100 como índice. Com relação às áreas do conhecimento às quais estão associadas estas publicações, nota-se a Ciência da Computação, as Engenharias e as Neurociências.

Ademais, deve-se destacar que o periódico com o maior Índice-H é o alemão *Lectures Notes In Computer Science*, com 296, e é relacionado à Matemática e à Ciência da Computação com Quartil igual a Q2. Dessa maneira, pode-se observar, que mesmo que os periódicos mais produtivos sejam oriundos da Venezuela, os de maior relevância segundo o Quartil e Índice h são provenientes dos EUA, Holanda e Alemanha, países com destaque internacional na produção científica mundial. Além disso, nota-se que todos os periódicos são relacionados a áreas do conhecimento das Ciências Exatas. Nesse contexto, analisaram-se ainda os países mais produtivos com relação ao tema *Big Data*. Os países são descritos na Quadro 3, podendo-se destacar China e EUA como os mais produtivos.

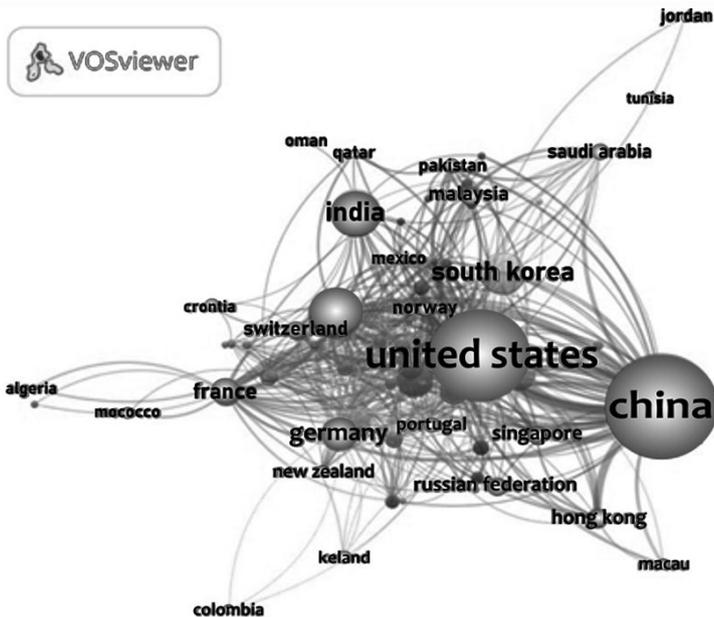
### Quadro 3 – Países mais produtivos

Países	Artigos
China	2.288
EUA	1.759
UK	550
Índia	495
Coreia do Sul	402
Alemanha	289
Áustria	289
Canadá	242
Itália	231
Espanha	214

**Fonte:** Construção dos autores

Ao se destacarem os dez países mais produtivos, a China ganha notoriedade por estar presente em 2.288 artigos relacionados ao tema, seguida dos EUA, com 1.759 artigos. Entretanto, para que se possa ter uma maior dimensão dos países mais produtivos, mapearam-se as coautorias entre estes países, assim como apresentado na Figura 2.

**Figura 2** - Mapeamento das Coautorias Internacionais.



**Fonte:** Elaborada pelos autores por meio do software Vosviewer no formato CSV (Excel).

Os fenômenos relacionados ao *Big Data* estão ligados diretamente à supercomputação e de acordo com o site TOP500list, que divulga duas vezes ao ano o ranking dos computadores mais velozes do mundo, a China aparece em primeiro lugar, tendo 206 dos 500 supercomputadores listados, Estados Unidos com 124, Japão com 36, Alemanha e

França com 21 e 18 supercomputadores respectivamente. Essa hegemonia tecnológica dos Estados Unidos e da China refletiu-se diretamente na produção científica, como é ilustrado na Figura 2, com a indicação dos maiores círculos para os países mais produtivos.

Entretanto, observa-se que existe uma colaboração em coautorias entre os pesquisadores de diferentes países, destacando as coautorias entre Coreia do Sul, Suíça e Estados Unidos com a China, e da França com os Estados Unidos. Pode-se identificar também o domínio Chinês, principalmente com os países asiáticos como Coreia do Sul, Singapura, Macau e Hong Kong. Dos países mais produtivos que apresentam pouca coautoria internacional destacam-se o Reino Unido, Alemanha, Austrália, Canadá, Itália e Espanha.

É possível destacar que o conhecimento dessa produção científica relacionada aos temas que compõem a configuração aqui analisada permite a identificação das características da discussão científica que ocorre no contexto teórico empírico em que é produzida.

## CONSIDERAÇÕES FINAIS

Com a análise da produção de artigos na base *Scopus*, relativa aos anos de 2016 e 2017, foi possível identificar que os temas mais pesquisados envolvem os seres humanos, seja em técnicas de coleta de dados, seja na identificação de padrões de comportamento ou genético.

Examinando o tema sob a dimensão metodológica, observa-se que para realizar essas análises são necessários o desenvolvimento de técnicas de mineração e manipulação de dados, gerenciamento de informações, principalmente na

otimização do armazenamento de grandes massas de dados. Conta-se, para isso, com estudos relacionados a sistemas de computação distribuída e processamento na nuvem. Esse resultado reflete as palavras-chave mais utilizadas na análise dos artigos deste estudo.

Na análise das revistas mais produtivas, descartado o desvio das revistas venezuelanas, sugestivo do caráter geopolítico pelo qual pode ser analisada a América do Sul e seu processo de integração em curso, pode-se observar que a concentração dos estudos é nas áreas de Ciência da Computação e Engenharia. Salienta-se, entretanto, que há uma intensificação no processo de integração regional que tem desencadeado movimentos de natureza político-estratégica em termos de produção científica. Observou-se que os países mais produtivos são a China e os Estados Unidos, refletindo diretamente o investimento realizado por sua política no desenvolvimento de supercomputadores, ou o agrupamento (*cluster*) de computadores, sendo esses países os detentores do maior número de supercomputadores instalados.

Considerando-se que os Estados Unidos e a China são os maiores produtores sobre a temática, esperava-se que a intensidade da colaboração internacional entre pesquisadores destes países fosse mais destacada.

Conclui-se que o *Big Data* é um tema novo na ciência, seu reflexo pode ser observado com a produção de trinta artigos sobre o tema em 2011, para 3.890 artigos produzidos em 2017, e um crescimento de 20,07% em relação a 2016. Por estar concentrada em países mais ricos, a produção em acesso aberto representou apenas 6,1% de toda produção analisada.

Salienta-se, desse modo, o aprofundamento sobre a informação e os artefatos tecnológicos como suporte aos

grandes sistemas de armazenamento e tratamento da informação especializada, em decorrência da forma de fazer ciência e suas transformações ao longo dos anos. É perceptível o círculo de inter-relações que se efetiva entre áreas do conhecimento, autores, instituições, países e configurações temáticas e isso tem sido um desafio aos especialistas em estudos métricos, especialmente pelo potencial de visibilidade que podem oferecer e a consequente complexidade para sua análise e interpretação.

Esse constante e dinâmico processo de influência mútua entre a ciência e as tecnologias utilizadas na produção e comunicação do conhecimento científico tem sido potencializado, por um lado pelos avanços das tecnologias de informação e comunicação, em particular das redes computacionais e, por outro, da possibilidade de compartilhamento dos dados por parte de intermediários que atuam a partir das possibilidades concretas de compartilhar, utilizar e reutilizar dados com os seus pares.

## REFERÊNCIAS

BESSER, H. Longevidade digital. *Acervo*, Rio de Janeiro, v. 23, no 2, p. 57-70, jul/dez 2010 –Request PDF. Disponível em: [https://www.researchgate.net/publication/260301901\\_Longevidade\\_Digital](https://www.researchgate.net/publication/260301901_Longevidade_Digital). Acesso em: 21/08/2018.

BUFREM, L. S. *et al.* Produção internacional sobre ciência orientada a dados: análise dos termos Data Science e E-science na Scopus e na Web of Science. *Informação & Informação*, v. 21, n. 2, p. 40-67, 2016. DOI: 10.5433/1981-8920.2016v21n2p40.

CONEGLIAN, C. S.; GONÇALEZ, P. R. V. A.; SEGUNDO, J. E. S. O profissional da informação na era do Big Data. *Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação*, v. 22, n. 50, 2017.

CONEGLIAN, C. S.; SEGUNDO, J. E. S.; SANT'ANA, R. C. S. G. A. Big Data: fatores potencialmente discriminatórios em análise de dados. *Em Questão*, v. 23, n. 1, 2017.

COORDENAÇÃO APERFEIÇOAMENTO DE PESSOAL DE NÍVEL SUPERIOR (CAPES). *Tabelas de Áreas do Conhecimento*. Brasília: Ministério da Educação, 2014, atualizado em 2018. Disponível em: <http://www.capes.gov.br/avaliacao/instrumentos-de-apoio/tabela-de-areas-do-conhecimento-avaliacao>. Acesso em: 21/08/2018.

EATON, J. P. *et al.* Structural analysis of co-author relationships and author productivity in selected outlets for consumer behavior research. *Journal of Consumer Psychology*, v. 8, n. 1, p. 39-59, 1999.

FURLAN, P. K.; LAURINDO, F. J. B. Agrupamentos epistemológicos de artigos publicados sobre Big Data analytics. *Transinformação*, v. 29, n. 1, p. 91-100, 2017. DOI: 10.1590/2318-08892017000100009.

GLÄNZEL, W. *Bibliometrics as a research field: a course on theory and application of bibliometric indicators*. Bélgica: [s.n.], 2003. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.97.5311&rep=rep1&type=pdf>. Acesso em: 21/08/2018.

KATZ, J. S.; MARTIN, B. R. What is research collaboration? *Research Policy*, Amsterdam, v. 26, p. 1-18, 1997.

ROA-MARTÍNEZ, S.M.; VIDOTTI, S.A.B.; SANTANA, R. C. Estructura propuesta del artículo de datos como publicación científica. *Revista Española de Documentación Científica*, n. 40, v.1, p. e167, 2017. DOI: 10.3989/redc.2017.1.1375.

TORRES-SALINAS, D. Compartir datos (data sharing) en ciencia: contexto de una oportunidad. *Anuario ThinkEPI*, v. 1, pp. 258-261, 2009. Disponível em: <http://www.thinkepi.net/compartir-datos-data-sharing-en-ciencia-el-contexto-de-una-oportunidad>. Acesso em: 21/08/2018.

# INTERSECÇÃO TEMÁTICA DE PROGRAMAS DE PÓS-GRADUAÇÃO BRASILEIROS: CONSIDERAÇÕES SOBRE *BIG DATA*

*Jacquelin Teresa Camperos-Reyes*

*Luiza de Menezes Romanetto*

*Ricardo Cesar Gonçalves Sant'Ana*

*Plácida Leopoldina Ventura Amorim da Costa Santos*

O termo *Big Data* surgiu em meados dos anos 2000, definido inicialmente a partir de 3 V's: volume, variedade e velocidade, desde então tem sido empregado para descrever um fenômeno vinculado à terceira era da informação (NESELLO; FACHINELLI, 2014). Os autores supracitados consideram que, “[...] as contribuições de big data são possíveis de serem realizadas desde que a matéria-prima informacional, gerada pela exploração dos dados, seja transformada em inteligência, por meio do processo de análise” (NESELLO; FACHINELLI, 2014, p.19).

Ao analisar a produção científica nas bases de dados *Web of Science* e *Scopus* no período de 2009-2014, focando *Big Data* e Inteligência Estratégica, Nesello e Fachinelli (2014) identificaram a predominância de trabalhos apresentados em

eventos científicos, o que levou os pesquisadores a sugerir a atualidade do tema para a comunidade científica. Cabe destacar que, ao analisar a produção de livros foi constatada a predominância na área de Ciência da Computação (NESELLO; FACHINELLI, 2014).

[...] a gestão de dados se complexificou com a emergência do Big Data, que aponta para novos horizontes científicos e pragmáticos conduzindo para a exigência de novas habilidades e conhecimento para os profissionais da área, visando o enquadramento no perfil para a gerência dos dados (RODRIGUES; NOBREGA; DIAS, 2017, não paginado).

Neste contexto, a gestão de dados torna-se fundamental para obter informação e transformá-la em conhecimento. Rodrigues, Nóbrega e Dias (2017, não paginado) consideram que a gestão de dados no contexto dos *Big Data* está fundamentada em quatro pilares: técnico, social, político e legal, e caracterizam a gestão de dados como

disciplina que tem como finalidade definir, planejar, implantar e executar: estratégias, procedimentos e práticas necessárias para gerenciar de forma efetiva os recursos de dados e informações das organizações, incluindo planos para sua definição, padronização, organização, proteção e utilização.

Em seguida, a matemática viabiliza inferir probabilidades e correlações em enormes quantidades de dados, o que possibilita realizar previsões (CRAVO, 2015).

Compreende-se que o fenômeno *Big Data* se refere a geração e uso de dados estruturados, semi-estruturados, em que têm sido adotadas diversas técnicas para coletar, integrar,

analisar e tomar decisões, a partir de diversos tipos de dados, para os mais diversos fins. Para tanto, torna-se imprescindível que haja desenvolvimento tecnológico atrelado a uma soma de habilidades e conhecimentos entre os mais diversos setores de atividade na sociedade contemporânea.

Entender este fenômeno recente e de tamanha complexidade requer, em um primeiro momento, analisar como tem sido a atuação de institutos de ensino superior e pesquisa. Reconhecendo que, requer mão de obra qualificada e que, para seu desenvolvimento e aplicação na sociedade, torna-se desejável uma soma de estudos previamente desenvolvidos que apontem para os riscos e fatores positivos e negativos a serem considerados.

Há de se considerar como base as premissas de Borko (1968), em que a Ciência da Informação é responsável por investigar a origem, coleta, organização, armazenamento, recuperação, interpretação, transmissão, transformação e o uso da informação, assim como, reconhecer que no contexto dos *Big Data* a informação é resultado da geração, processamento e análise de dados, ciclo este realizado por diferentes áreas do conhecimento.

Considera-se relevante destacar o papel dos programas de pós-graduação no Brasil, bem como frisar a importância das atividades desenvolvidas pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES, instituição vinculada ao Ministério da Educação, que tem como principal linha de atuação promover:

- 1) avaliação da pós-graduação *stricto sensu*; 2) acesso e divulgação da produção científica; 3) investimentos na formação de recursos de alto nível no país e exterior; 4) promoção da cooperação científica internacional; e, 5) indução e fomento da formação inicial e continuada de professores para a educação

básica nos formatos presencial e a distância (BRASIL, 2008, não paginado).

A CAPES, inicialmente chamada por Campanha Nacional de Aperfeiçoamento de Pessoal de Nível Superior, foi criada por meio do Decreto 29.741 de 1951, com o objetivo direcionado a formação profissional em quantidade e qualidade para atender às necessidades dos setores públicos e privados necessários para o desenvolvimento do país (BRASIL, 2008). O que nos remete à importância dos programas de pós-graduação para a formação profissional nos mais diversos setores de atuação na sociedade contemporânea.

A avaliação de programas adotada pela CAPES foi implementada em 1997, baseada em uma escala numérica de 1 a 7, estabelecendo reavaliação de cursos a cada 4 anos (NOBRE; FREITAS, 2017). O Plano Nacional de Pós-Graduação *Stricto Sensu*, elaborado desde 1981, tem sido desenvolvido com intuito de “definir novas diretrizes, estratégias e metas para dar continuidade e avançar nas propostas para a política de pós-graduação e pesquisa no Brasil” (NOBRE; FREITAS, 2017, p. 29).

No contexto deste estudo, considera-se relevante destacar o papel dos Programas de Pós-Graduação no terceiro PNPG, período 1986-1989, diante do reconhecimento de que

No que se refere à pós-graduação, essa ideia se expressava no fato que não havia um quantitativo de cientistas suficiente para se atingir plena capacitação científica e tecnológica no país, tornando importante, portanto, o progresso da formação de recursos humanos de alto nível, já que a sociedade e o governo buscavam a independência econômica, científica e

tecnológica para o Brasil no século XXI (NOBRE; FREITAS, 2017, p.30).

A Avaliação do Sistema Nacional de Pós-Graduação foi implantada desde 1997 e, representa uma atividade fundamental para manter a qualidade dos cursos de Mestrado e Doutorado no Brasil. Atualmente a avaliação é dividida em 49 áreas, distribuídas em dois níveis: três colégios e nove grandes áreas<sup>1</sup> (BRASIL, 2018).

Com base na classificação de áreas realizada pela CAPES, foram selecionados alguns cursos por Área de Avaliação na Plataforma Sucupira, conforme será descrito na próxima seção. O estudo apresentado neste artigo, foi realizado com o objetivo de identificar, no âmbito de programas de pós-graduação, quais as características, temáticas e disciplinas oferecidas podem conter similaridades pertinentes às necessidades do fenômeno Big Data no Brasil, assim como, qual a contribuição da Ciência da Informação para este fenômeno.

De modo que o problema de pesquisa pode ser assim definido: Qual é o tipo de relação teórica existente entre os programas de Ciência da Informação, Ciência da Computação e Matemática, no Brasil, ao abordar temáticas relacionadas com o Big Data?

---

<sup>1</sup> Divisão de áreas da CAPES: Colégio: Ciências da vida (Grandes áreas: Ciências agrárias; Ciência biológicas; Ciências da saúde); Colégio: Ciências exatas, tecnológicas e multidisciplinar (Grandes áreas: Ciências exatas, tecnológicas e multidisciplinar; Engenharias; Multidisciplinar); Colégio: Humanidades (Grandes áreas: Ciências humanas; Ciências sociais aplicadas; Linguística, letras e artes) (BRASIL, 2018).

## 1 PERCURSO METODOLÓGICO UTILIZADO

Para responder ao problema e atingir o objetivo da pesquisa, objetivo que a encaminha como exploratória de natureza qualitativa, realizou-se, para coleta dos dados, revisão bibliográfica, e, para a análise e interpretação dos dados, utilizou-se da análise de conteúdo, seguindo as etapas:

1. revisão bibliográfica acerca do fenômeno *Big Data*;
2. coleta de dados na plataforma Sucupira<sup>2</sup> buscando identificar os programas de pós-graduação (PPG) alvo da pesquisa;
3. análise das características, temáticas e disciplinas oferecidas em cada programa identificado, a partir de informações descritas nos sites dos PPG.

A revisão bibliográfica realizada na etapa 1 foi realizada com intuito de identificar e caracterizar as tendências relacionadas aos estudos desenvolvidos na temática. Para Cravo (2015), o surgimento dos *Big Data* está diretamente relacionado à habilidade de extrair de grandes bases de dados perspectivas inovadoras ou novas formas de valor econômico, o que gera impacto e pode repercutir em mudanças em mercados, organizações, relações sociais e governos.

Na etapa 2, o procedimento para identificar os programas que iriam compor o *corpus* da pesquisa foi desenvolvido mediante consulta na plataforma Sucupira. Dita plataforma é uma “[...] ferramenta para coletar informações, realizar análises e avaliações, e ser a base de referência do Sistema Nacional de Pós-Graduação (SNPG)”, do Brasil. A Sucupira é mantida pela CAPES (CAPES, 2018).

---

<sup>2</sup> <https://sucupira.capes.gov.br/sucupira/public/index.xhtml>

O site da Sucupira apresenta a opção de menu rotulada como *Cursos avaliados e reconhecidos*, que permite recuperar segundo *Dados quantitativos de programa*, programas classificados segundo a *Nota de Avaliação* obtida por cada PPG. No Brasil, as notas estão no leque de 1 ao 7<sup>3</sup>, tendo então 7 notas não concorrentes possíveis para cada programa. Nesta pesquisa optou-se por identificar programas destacados, avaliados com notas 5, 6 e 7, nas áreas de avaliação *Ciência da Computação, Comunicação e Informação*; e *Matemática / Probabilidade e estatística*.

Para realizar a etapa 3 aplicou-se leitura crítica, em cada site de PPG identificado, na busca de inferências, mediante análise de conteúdo (BARDIN, 2010) das missões, linhas de pesquisa e ementas das disciplinas, focando a referência ao processamento de grandes volumes de dados, mineração de dados, e/ou, Big Data.

A análise de conteúdo possibilita a produção de inferências recorrendo ou não a indicadores quantitativos (BARDIN, 2010), portanto, buscou-se identificar quais as temáticas pertinentes ao fenômeno Big Data têm sido desenvolvidas pelos PPG no Brasil e a contribuição da Ciência da Informação nesse cenário, apresentando uma mensuração dessa abordagem no Brasil.

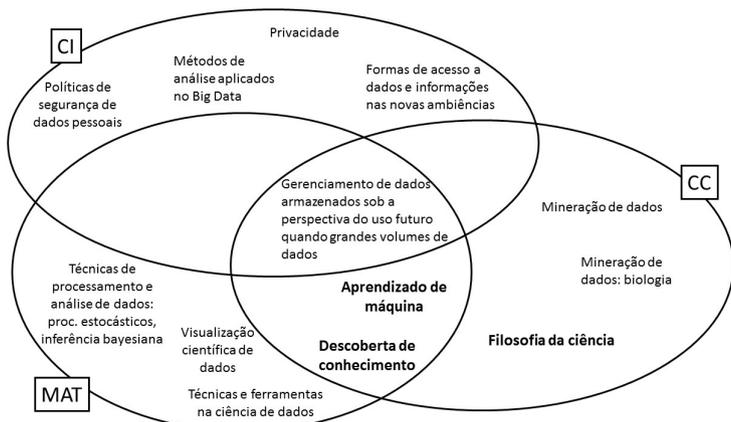
## 2 APRESENTAÇÃO E ANÁLISE DOS RESULTADOS

A figura 1 apresenta a distribuição dos 42 programas identificados na coleta, indicando as notas que ganharam na avaliação vigente na CAPES.

---

<sup>3</sup> Maior detalhamento do sistema de avaliação de programas de pós-graduação no site <http://www.capes.gov.br/acesoainformacao/perguntas-frequentes/avaliacao-da-pos-graduacao/7421-sobre-avaliacao-de-cursos>

**Figura 1.** Distribuição dos programas identificados segundo nota CAPES



Fonte: Plataforma Sucupira (2018)

As áreas de avaliação foram referenciadas destarte: Ciência da Informação (CI); Ciência da Computação (CC); e, Matemática/Probabilidade e Estatística (MAT). Consideraram-se unicamente PPG, mestrado (ME) e doutorado (DT), *strictu sensu*, somando no total 42 programas como *corpus* da pesquisa, sendo 5 da Ciência da Informação, 17 da Ciência da Computação e 20 de Matemática/Probabilidade e estatística.

## 2.1 MISSÕES/APRESENTAÇÕES

Entre os cursos analisados, foi possível identificar que apenas dois apresentam em sua missão/apresentação temáticas relacionadas aos *Big Data*. O Programa de pós-graduação em Engenharia de Sistemas da Universidade Federal do Rio de Janeiro descreve sobre o desenvolvimento de pesquisas em

mineração de dados e abordagem da temática “tratamento da sobrecarga de informação”.

Por outro lado, o Programa de pós-graduação em Gestão e Organização do Conhecimento da Universidade Federal de Minas Gerais indica como prioridade o desenvolvimento de pesquisas relacionadas a “tendência de conexão e interação entre equipamentos via Web, através de vocabulários controlados, popularmente denominada internet das coisas [...] a explosão de fontes de informação e conhecimento heterogêneas, que tem sido denominada de big data, e a necessidade de torná-las úteis” (2018, não paginado).

Os demais PPG analisados não indicavam a temática *Big Data* na missão ou apresentação do curso. Em alguns casos, a informação não foi localizada ou o site estava fora do ar, conforme indicado no quadro 1.

**Quadro 1.** Categorias observadas nas missões/apresentações

<b>Categoria detectada</b>	<b>Ocorrências</b>
Indica explicitamente estudo do fenômeno	2
Não indica relação ao BD	35
Site fora do ar	1
Informação não localizada	4

**Fonte:** Dados da pesquisa

As missões/apresentações que se manifestam explicitamente sobre o estudo do fenômeno, aludem duas nomeações como “Campos de pesquisa”: a) *mineração de dados e tratamento da sobrecarga de informação*, e, b) *explosão de fontes de informação e conhecimento heterogêneas*. A primeira categoria na UFRJ, no PPG em Engenharia de Sistemas e a segunda na UFMG, no PPG em Gestão & Organização do conhecimento.

## 2.2 SÍNTESE DE INFERÊNCIAS NAS LINHAS DE PESQUISA DOS PROGRAMAS

Dos 42 programas explorados, 18 indicam nas suas linhas de pesquisa o estudo do fenômeno Big Data. O quadro 2 apresenta as inferências realizadas logo após da análise dos conteúdos.

**Quadro 2.** Inferências nas linhas de pesquisa

Categoria	Quantidade	Áreas que a abordam
Mineração de dados	9	CI, CC
Aprendizado de máquina	6	CC
Bancos de dados em grande escala	4	CC
Geração de conhecimento	4	CC
Processos estocásticos	4	MAT
Inferência Bayesiana	2	MAT
Novas formas de acesso nos novos ambientes	2	CI
Mineração de dados especificamente em processos biológicos, linguísticos, neuronais	1	CC
Inferências em dados complexos	1	CC
Bioinformática e biologia computacional	1	CC
Foco no fenômeno Big Data como tudo	1	CI
Visualização de dados	1	CI, CC
Metadados	1	CI
Evolução e sistemas sociais	1	MAT
Epidemiologia	1	MAT
Informação não localizada	24	

**Fonte:** dados da pesquisa

Somando as ocorrências por PPG, na Ciência da Informação, três PPG indicam dentro das linhas de pesquisa o estudo do fenômeno *Big Data*, onze na Ciência da Computação e quatro nos PPG de Matemática.

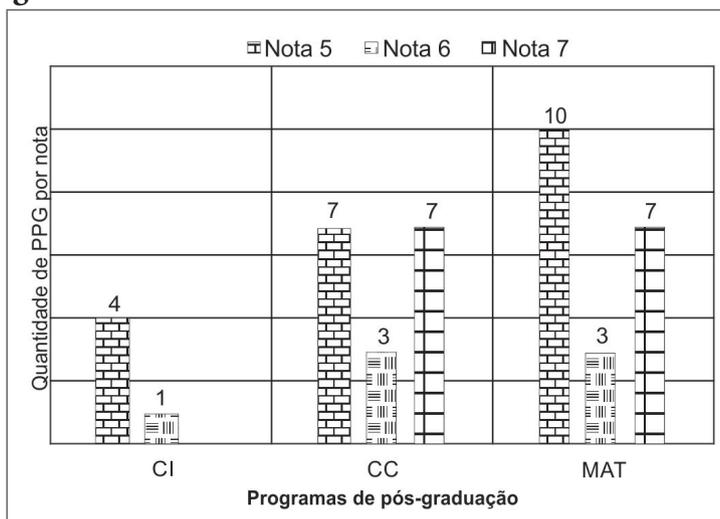
Observou-se um relacionamento mediante a categoria Mineração de dados, a qual é abordada como parte de linhas de pesquisa de um programa da Ciência da Informação e de várias da Ciência da Computação, destacando que, no caso da Ciência da Informação, a UFMG ressalta que a abordagem do estudo proposto do fenômeno *Big Data* é realizada sem a busca da produção de tecnológica, mas sim, do seu uso efetivo (UFMG, 2018).

Cabe assinalar que a única área, entre as observadas, que manifesta nas linhas de pesquisa o estudo dos Metadados é a Ciência da Informação, caso identificado no site da UFMG.

### **2.3 RESULTADO NAS EMENTAS DAS DISCIPLINAS**

A leitura dos conteúdos das ementas das disciplinas oferecidas pelos PPG da amostra possibilitou a criação das inferências representadas na figura 2.

**Figura 2.** Inferências trazem análise do conteúdo de ementas



**Fonte:** dados da pesquisa

A representação gráfica utiliza-se do diagrama de Venn para apresentar cada área avaliada, assinalando em cada uma as inferências. Destaca-se a presença de só uma na intercepção das três áreas: *Gerenciamento de dados armazenados sob a perspectiva do uso futuro quando grandes volumes de dados*.

As categorias *Aprendizado de Máquina* e *Descoberta de conhecimento* encontram-se de modo concomitante entre Matemática e Ciência da Computação, fato que chama a atenção por conter discussões que podem implicar questões das ciências sociais aplicadas, e que, contudo, não foram distinguidas no âmbito da Ciência da Informação.

No caso da Ciência da Computação, cabe destacar uma disciplina oferecida pela USP, de nome “Fundamentos Metodológicos para Inteligência Artificial”, cuja ementa permite inferir ênfase na *Filosofia da Ciência*, viés não considera-

do nas outras áreas da amostra, e pelo impacto que esse programa atribui aos *Big Data* nos métodos científicos tomados como clássicos, assumindo a modelagem de grandes volumes de dados como entradas para o processo científico em si, e não unicamente como técnicas para a avaliação de modelos produtos da pesquisa científica.

A análise corrobora-se por Castro Filho, Faria e Santos (2013) ao indicar que as mudanças inerentes a processos relacionados com ativos informacionais precisam ser observadas pelos profissionais da informação, desenvolvendo habilidades intuitivas de julgamento.

## CONSIDERAÇÕES FINAIS

Em uma sociedade em que a geração e o processamento de dados têm sido realizados de maneira expressiva, influenciando setores de grande importância para o desenvolvimento social e econômico, considera-se relevante identificar quais áreas do conhecimento têm atuado na formação de profissionais capacitados para atuar na era dos *Big Data*.

Os programas de pós-graduação, mantidos por instituições de ensino e pesquisa, representam em nível nacional um importante meio para a formação de mão de obra qualificada para atuar profissionalmente e no desenvolvimento de pesquisa. Com base nos resultados obtidos neste estudo foi possível verificar que os Programas de pós-graduação em Ciência da Informação, Ciência da Computação e Matemática, considerados como referência no Brasil, estão desenvolvendo estudos direcionados à temática.

Cada área do conhecimento apresenta linhas de pesquisa e disciplinas direcionadas a sua área de atuação profes-

sional, entretanto, a Ciência da Informação enquanto ciência social aplicada apresenta maior grau de interdisciplinaridade com estudos de abrangência temática com a Ciência da Computação, envolvendo estudos sobre: mineração de dados, novas formas de acesso, visualização de dados e metadados. Ademais, apresenta disciplinas direcionadas ao estudo de aspectos legais e governança (jurídica, técnica, digital e informacional).

Os Programas de pós-graduação em Matemática apresentam maior incidência de estudos de probabilidade e correlação para o tratamento de grandes quantidades de dados, com destaque para inferência bayesiana e processos estocásticos. Com base nessa constatação, consideramos importante ampliar o escopo da análise realizada neste estudo para as áreas do conhecimento de probabilidade e estatística.

Considera-se que a Ciência da Informação poderia, em princípio, manifestar maior cobertura da temática do fenômeno *Big Data* e suas atuações contextualizadas em grupos sociais, sobretudo pela manifestação do estudo em questões de *Aprendizado de máquina*, que as outras áreas de estudo estão realizando. Acredita-se que deveria existir um maior interesse em observar criticamente a influência que o *Aprendizado de máquina* poderia exercer, cuidando da não desconsideração de fatores sociais que levariam a um uso inadequado da tecnologia em determinados contextos.

Cabe esclarecer que o fenômeno *Big Data* envolve diversas outras áreas do conhecimento, de modo que se considera que o estudo apresentado neste artigo pode ser ampliado por meio de observação em outras áreas do conhecimento, o que permitiria um melhor entendimento sobre seu desenvolvimento, assim como seu nível de complexidade.

## REFERÊNCIAS

- BARDIN, L. *Análise de conteúdo*. Lisboa: Edições 70, 2010.
- BORKO, H. Information science: what is it? *American Documentation*, v. 19, n. 1, p. 3-5, jan. 1968. Disponível em: <http://pt.scribd.com/doc/39023567/Information-Science-What-is-It>. Acesso em: 17/05/2018.
- BRASIL. MINISTÉRIO DA EDUCAÇÃO. *Sobre as áreas de avaliação*. 2018. Disponível em: <http://www.capes.gov.br/avaliacao/sobre-as-areas-de-avaliacao>. Acesso em: 15/04/2018.
- CAPES. *Plataforma Sucupira*. 2018. Disponível em: <http://www.capes.gov.br/avaliacao/plataforma-sucupira>. Acesso em 15/05/2018.
- CASTRO FILHO, C.M.; FARIA, A.C.C.; SANTOS, D.S. O bibliotecário como profissional da informação: o mundo do trabalho, habilidades e competências. CASTRO FILHO, C.M. *Olhares sobre a atuação do profissional da Ciência da Informação*. São Paulo: Todas as Musas, 2013.
- CHEN, M., MAO, S., ZHANG, Y., LEUNG, V. C. *Big data: related technologies, challenges and future prospects*. Heidelberg: Springer, 2014.
- CRAVO, V. O Big Data e os desafios da modernidade: uma regulação necessária? *Revista de Direito Setorial e Regulatório*, Brasília, v. 1, n. 2, p. 243-257, outubro 2015.
- HEKIMA. 2016. *O guia definitivo de Big Data para Iniciantes*. Disponível em: <http://www.bigdatabusiness.com.br/ebook-o-guia-definitivo-de-big-data-para-iniciantes/>. Acesso em: 29/03/2018.
- NESELLO, P.; FACHINELLI, A.C. Big Data: O novo desafio para gestão. *Revista Inteligência Competitiva*, v. 4, n. 1, p. 18-38, jan./mar. 2014. ISSN: 2236-210X
- NOBRE; L.N.; FREITAS, R.R. A evolução da pós-graduação no Brasil: histórico, políticas e avaliação. *Brazilian Journal of Production Engineering (BJPE)*, v.3, n.2, p.18-30, 2017. ISSN: 2447-5580
- RODRIGUES, A. A.; NOBREGA, E.; DIAS, G.A. Desafios da gestão de dados na era do big data: perspectivas profissionais. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 18, 2017. *Anais...* Marília: UNESP, 2017.

**INFORMAÇÃO, TRANSDUÇÃO E  
METADADOS:  
FENÔMENOS DE EMERGÊNCIA?**

# INFORMAÇÃO COMO AÇÃO SIGNIFICATIVA EM PROCESSOS SEMIÓTICOS EMERGENTES BASEADOS EM MULTIAGENTE

*Davidson Bruno da Silva*  
*Tiago Campos Ferreira*  
*Leonardo Lana de Carvalho*

O objetivo deste capítulo é analisar o conceito de informação como uma ação significativa em sistemas complexos adaptativos. A comunicação e a linguagem são vistas como sistemas complexos por Maturana & Varela (1994), cuja principal característica é a auto-organização<sup>1</sup>. Para estes autores, a *comunicação* é literalmente o conjunto de *ações comuns* em um coletivo de agentes. Através destas ações comuns o coletivo permanece como um grupo coeso, mantendo sua identidade. A auto-organização de um sistema coletivo de agentes pode ocorrer, pois o grupo modula as ações de cada agente, ações estas que mantêm a própria dinâmica coletiva. Para Maturana

---

<sup>1</sup> O desenvolvimento da linguagem é entendido como um fenômeno complexo. Para uma exposição da investigação psicológica do desenvolvimento da linguagem indicamos ver Tomasello & Brooks (1999).

& Varela (1994), o domínio linguístico ocorre quando a comunicação se refere a ela mesma. Neste sentido, no momento em que os agentes do grupo passam a comunicar comunicações ocorridas, os agentes agem de modo linguístico. Segundo Thompson (2007/2013), processos informacionais ocorrem na dinâmica comunicativa de sistemas auto-organizados.

Modelos e simulações baseadas em sistemas multiagente (ver FERBER, 2006) são ferramentas poderosas para a investigação de fenômenos sociais e da linguagem em específico (STEELS, 1996; 1998; 2000a; 2000b; 2003; LOULA *et al.*, 2003; LOULA 2004). Steels (2003) rompe com linhas de pesquisa da linguagem baseadas ora no cognitivismo, ora no enativismo radical, o primeiro radicalmente a favor de uma hipótese simbólica da linguagem e o segundo entendendo radicalmente a linguagem como fenômeno comunicacional sem representações. Segundo Carvalho (2008), o autor busca definir, modelar e simular sistemas multiagentes sobre os quais podemos inferir a existência de representações emergentes. Para conceituar as representações emergentes, Luc Steels aproxima-se do conceito peirceano de processos semióticos.

Sob uma vertente enativa semiótica, decidimos então:

1. Replicar resultados de convergência léxica e ampliar os resultados obtidos por Loula *et al.* (2003) e Loula (2004);
2. Para isso, foi concebido e implementado o SIMPS - *Simulador de Processos Semióticos*; e
3. Analisar o conceito de informação como ação significativa, isto é, como ações envolvidas em processos semióticos, cujos significados se encontram nos usos coletivos das palavras e expressões de jogos de linguagem. Apesar de entendermos que nem toda informação é adaptativa, buscamos resultados com informações passíveis de trazer vantagens adaptativas para grupos de agentes.

Na seção 1 apresentaremos os trabalhos relacionados à modelagem e simulação multiagente da linguagem como um sistema complexo. Na seção 2, mostraremos uma defesa do conceito de informação como ação significativa em processos semióticos emergentes. Na seção 3, apresentaremos o SIMPS e a arquitetura de nossos agentes semióticos. Por último, na seção 4, discutiremos os resultados das simulações e a análise dos dados produzidos.

## **1 TRABALHOS RELACIONADOS**

Segundo Axtell (2007), a modelagem e simulação baseada em multiagente de fenômenos sociais chegou na primeira década do século ao fim de seu alvorecer, sendo então hoje um conjunto de conceitos, métodos e técnicas largamente utilizados. Em específico, a modelagem e simulação da linguagem emergente em sistemas complexos baseados em multiagente já possui também uma longa história.

Steels (1996; 1998; 2000a; 2000b; 2003) propõe passar de uma abordagem baseada no enativismo radical (HUTTO & MYIN, 2013; 2017), radicalmente contrário ao conceito de representação, tal como podemos encontrar em Brooks (1991), às funções cognitivas de alto nível através de representações emergentes em sistemas multiagente. Para Steels (2003), interações entre agentes ocorrem envolvendo signos, ou seja, representações emergentes ativas. Steels (1996; 1998; 2000a; 2000b; 2003) observa que o surgimento da comunicação é o resultado da relação entre unidades autônomas que, ao interagirem, constituem um coletivo de agentes auto-organizado, o que corrobora a visão enativa da comunicação e da linguagem (MATURANA & VARELA, 1994). Todavia, o autor conjectura que é possível assimilar

o conceito de representação, entendido como processos semióticos, aos processos comunicacionais auto-organizados em sistemas multiagente. Neste sentido, os agentes de um sistema coletivo auto-organizado devem ser dotados de uma arquitetura que os possibilite uma comunicação baseada em processos semióticos.

Ao contrário de uma abordagem cognitivista, para a qual toda comunicação e inteligência está estritamente baseada em símbolos físicos, a teoria enativa entende que a fonte da inteligência se encontra na auto-organização dos sistemas (VARELA, THOMPSON & ROSCH, 1991). Neste sentido, Steels (2003) entende que processos semióticos precisam estar envolvidos em dinâmicas coletivas auto-organizadas.

Nem todo processo comunicacional ocorre ao nível semiótico (MATURANA & VARELA, 1994). Um exemplo disto pode ser visto no modelo de forrageamento (DROGOUL, 1993; DROGOUL & FRESNEAU, 1998), em que os agentes não são dotados de processos semióticos e se caracterizam como estritamente reativos. Neste modelo, a comunicação ocorre de modo indireto, isto é, as ações dos agentes são moduladas pelo ambiente modificado pelos próprios agentes (LENAY, 1994). As ações dos agentes sobre o ambiente (assim também uns sobre os outros) leva a trajetórias emergentes, também descritas como comportamentos emergentes. Com relação a essas trajetórias emergentes, Steels (2003) diz que se tratam de estruturas que cumprem o papel de organizar a atividade coletiva dos agentes.

Estas trajetórias cumprem um papel de exploração de soluções de um problema específico (que se constitui no próprio ambiente em questão) e de busca de novas soluções neste ambiente. Assim, a comunicação ocorre

como um sistema complexo adaptativo em um equilíbrio dinâmico entre *exploration* e *exploitation*<sup>2</sup> (ver KENNEDY *et al.*, 2001). As trajetórias nas organizações são como um registro das atividades, como uma memória coletiva (ver HALBWACHS, 1950/1997), afirma Steels (2003). Kennedy *et al.* (2001) notam que uma série de meta-heurísticas são baseadas em sistemas complexos adaptativos, exatamente por explorarem esta propriedade dos sistemas complexos, o equilíbrio entre *exploration* e *exploitation*, a fim de atingir um ótimo global.

No desenvolvimento de uma visão enativa semiótica (ver HALL, 2006; CARVALHO, 2008; KULL *et al.*, 2009; STEWART, 2010; JESUS, 2016), a transição de um sistema complexo adaptativo para um sistema cognitivo depende do surgimento de *representações*, mas representações no sentido de processos semióticos emergentes (CARVALHO, 2008). Várias contribuições dedicaram-se a propor sistemas complexos baseados em agentes semióticos, os quais constituem os trabalhos relacionados de nossa pesquisa.

Luc Steels procura assegurar que agentes semióticos possam convergir não apenas sobre um léxico, mas também sobre regras gramaticais, o que abre a possibilidade de surgimento de sentenças (STEELS, 2000a; 2000b; 2007). Na interação humano-computador, o surgimento de uma linguagem comum através de convergência de léxico foi proposta

---

<sup>2</sup> Kennedy *et al.* (2001) relatam que um sistema pode fazer *exploration* e *exploitation* de informações. O sistema informacional realiza *exploration* quando busca novas informações e *exploitation* quando utiliza informações existentes. Como componente do sistema, o agente está envolvido em processos informacionais de *exploitation* e *exploration* em diferentes momentos. Os conceitos evoluíram superando a dicotomia individualista, entendendo os conceitos como processos em equilíbrio em um coletivo de agentes, em sistemas complexos baseados em multi-agente. Em meta-heurísticas, um balanceamento correto entre *exploration* e *exploitation* leva a um método eficiente de solução de problemas de otimização, caracterizando-o como um fenômeno de inteligência coletiva.

por Stuber, Hassas & Mille (2003; 2005). Um experimento digital foi realizado com robôs “*Talking Heads*”, colocando uma população de alguns milhares de agentes para controlar o movimento de câmeras, a fim de capturar imagens de figuras geométricas sobre um quadro branco e convergir sobre um léxico comum. Assim as “cabeças” (câmeras) usariam os mesmos símbolos para fazer referência à quadrados, triângulos, etc. (STEELS, 2003).

Steels (2003) propõe, através deste tipo de jogo de linguagem, que os símbolos que têm sua maior frequência de uso são representações emergentes. Com efeito, os usos dos símbolos são trajetórias emergentes no sistema complexo baseado em agentes. Trata-se também de um processo semiótico porque fazem parte do triângulo semiótico peirceano (Interpretante - Objeto - Signo). É inicializada uma lista de símbolos possíveis, mas o símbolo que toma seu lugar no triângulo semiótico o faz por mecanismos de *feedback* positivo. Todavia, sua perenidade é um caso de criticalidade auto-organizada. Assim, a manutenção do uso de um símbolo em uma relação semiótica depende da complexa dinâmica do coletivo em um ambiente.

Como resultado, Steels (2003) apresenta gráficos mostrando a frequência de uso de certas formas simbólicas em um contexto semiótico. A abscissa mostra o número de jogos de linguagem e a ordenada mostra a frequência de uso de todas as formas simbólicas disponíveis. Ao final, um símbolo é coletivamente “escolhido” para referenciar um objeto (STEELS, 2003).

Loula *et al.* (2003), Loula (2004) e Loula *et al.* (2005) criaram um ambiente para a simulação de ecossistemas que permite a interação cooperativa entre agentes “macacos” para

se proteger de seus três tipos de predadores, o “tigre”, a “cobra” e o “falcão”. Macacos se comunicam emitindo sinais de um para outro. O objetivo da simulação, para além de um modelo predador/presas, foi desenvolver uma arquitetura para o surgimento da comunicação semiótica entre agentes artificiais. O desenvolvimento dessa arquitetura foi inspirado na semiótica de C. S. Peirce, cujos resultados apontam convergência léxica. Neste modelo, cada macaco utiliza uma lista comum de símbolos para informar sobre a ocorrência de determinado evento. Após certo número de interações entre os agentes no ambiente, os macacos passam a utilizar os mesmos símbolos para indicar a presença dos predadores e, assim, conseguem evitá-los melhor. Loula *et al.* (2003), Loula (2004) mostram, deste modo, a eficiência adaptativa do léxico emergente.

Arnellos, Spyrou & Darzentas (2003, 2006) também reproduziram os resultados de L. Steels. Mas estes autores deram uma importante contribuição ao entender a expressão “representação emergente” usada por Steels como um sistema auto-organizado envolvendo relações semióticas. Arnellos *et al.* (2003, 2006) também enfatizam que o significado de *significado* é uma propriedade semiótica. Loula *et al.* (2003), Loula (2004) e outros entendem os usos emergentes dos símbolos como o significado dos mesmos, adotando assim uma visão pragmática da semântica em sistemas complexos baseados em agentes semióticos. Arnellos, Vosinakis, Spyrou & Darzentas (2006) destacam que o surgimento de “representações autônomas” inclui funções em estruturas dinâmicas que fornecem suporte tanto para o surgimento espontâneo de uma sintaxe quanto de um léxico.

Construções fluidas de gramáticas (STEELS, 2011; WELLENS et al, 2013) estão disponíveis como um formalismo de gramática de código aberto que permite abordar a questão de como mudanças dinâmicas da linguagem ocorrem, descobrindo os mecanismos cognitivos e os processos culturais que impulsionam a evolução da linguagem.

## 2 INFORMAÇÃO COMO AÇÃO SIGNIFICATIVA EM PROCESSOS SEMIÓTICOS EMERGENTES

Steels (1998; 2000a; 2000b; 2003) defendeu a passagem do uso de simples agentes reativos, tal como visto em Brooks (1991), à concepção de agentes semióticos. Agentes semióticos são entendidos como um tipo de agentes reativos dotados de uma arquitetura semiótica, não possuindo, necessariamente, capacidades de inferência lógica, tal como em agentes cognitivistas<sup>3</sup> (ver GOMES *et al.*, 2015).

Segundo Andler (1998), Halpin (2006), Carvalho (2008) e outros, o problema do conceito de representação é um problema central e assim incontornável das ciências cognitivas. Ao entender a representação como um sistema simbólico, o cognitivismo inaugura uma abordagem simbólica física da informação, o que ficou conhecido como o paradigma do processamento da informação. Newell & Simon (1976) postulam a hipótese simbólica física da mente, entendendo que qualquer sistema mental (cognitivo, inteligente, etc.) é um *sistema simbólico físico* (SSF). Assim SSF são, para o cognitivismo, necessários e suficientes para todo e qualquer sistema inteligente (inclusive aqui os artificiais e os naturais, como o ser humano). Pylyshyn (1984) descreve que a semântica é uma

---

<sup>3</sup> Para uma revisão da tipologia de agentes, ver Ferber (2007).

propriedade epifenomênica<sup>4</sup> de SSF, não tendo assim nenhuma eficiência causal sobre o processamento das informações. Baseadas em símbolos físicos, as informações somente são causais enquanto entidades físicas instanciadas, ocorrendo no cognitivismo uma separação entre a semântica (significado) de seu suporte físico causalmente eficiente (o símbolo). Como herdeiro deste pensamento, Searle (1980) não soube dissolver este problema e perdido nesse labirinto conceitual, após criticar o epifenomenismo, acabou por concluir que SSF artificiais são vazios semanticamente.

Uma visão da semântica aliada ao pensamento wittgensteiniano entende o significado como uso (Wittgenstein, 1953/2001). Também, para a teoria enativa, nem toda ação cognitiva é baseada em símbolos, mas toda ação simbólica é, antes de tudo, *ação*. A primazia da ação levou Varela, Thompson e Rosch (1991) a chamar a teoria corpórea da cognição de teoria *enativa (em ação)*. Um sistema é cognitivo e inteligente para a teoria enativa se ele existe como uma ação autopoiética. Um conjunto complexo de ações capazes de replicar as estruturas necessárias para a manutenção das ações é a condição necessária e suficiente para toda cognição, inclusive para a cognição simbólica.

Para a teoria enativa, o sentido das ações dos organismos autopoiéticos encontra-se no nicho. Assim os processos informacionais (e não processamento da

<sup>4</sup> O epifenomenalismo entende que propriedades mentais são causadas por eventos físicos, todavia não tendo a capacidade de afetar nenhum evento físico. Em organismos complexos, a ação é vista como um comportamento causado pela contração de músculos sob controle de impulsos neurais, e os impulsos neurais sendo gerados pela estimulação em órgãos sensoriais. Para o epifenomenalismo esta linha causal é suficiente para explicar as ações, sendo os eventos mentais fenômenos causados como adereços pois não desempenham papel causal nesse processo. A teoria enativa critica duramente a visão epifenomenalista da mente ao argumentar que a mente está na ação. O que se chama de propriedades mentais somente existe ao se realizarem como processos corpóreos, sendo que estes processos possuem um campo fenomênico isomórfico (ver CARVALHO, 2018)

informação) (THOMPSON, 2007/2013), ocorrem em sistemas auto-organizados, onde agentes agem sobre um meio, modificando-o e sendo modelados por este mesmo ambiente. É nesta relação dos sistemas auto-organizados que as ações são significativas e informam.

Visando modelar computacionalmente processos semióticos, buscamos gerar um sistema auto-organizado, sem o qual não faz sentido falarmos de semântica e também de informação. As relações de determinação entre agentes de um sistema auto-organizado gera um sistema comunicacional muito eficiente, mesmo na solução de problemas difíceis (ver KENNEDY & EBERHART, 2001). Todavia, entendemos que uma arquitetura de agentes semióticos é necessária para avançar de um sistema multiagente reativo para um sistema multiagente semiótico (ver GOMES *et al.*, 2015).

Neste sentido, entendemos que processos semióticos são processos comunicacionais, mas nem todo processo comunicacional possui uma estrutura semiótica, como dito anteriormente. Ao situá-lo no campo da ciência cognitiva enativa, defendemos que informações ocorrem estritamente em processos semióticos. Assim entendemos que processos informacionais são processos comunicacionais semióticos. Esta é uma tese que difere profundamente da vertente radical do enativismo (HUTTO & MYIN, 2013; 2017), a qual continua defendendo que semiótica é uma teoria representacionista dualista, vestígio do pensamento cartesiano. O enativismo semiótico (ver HALL, 2006; CARVALHO, 2008; KULL *et al.*, 2009; STEWART, 2010; JESUS, 2016), busca assimilar o conceito peirceano de semiótica aos conceitos de enação e autopoiese de Maturana e Varela. Ao fazer isso, entendem a semiótica

como processos, isto é, a semiótica ocorrendo na ação, mais especificamente na comunicação (na auto-organização).

A partir da vertente enativa semiótica, hipotetizamos que a informação é uma ação significativa que ocorre em processos semióticos emergentes, em sistemas complexos, baseados em multiagente. Para que seja uma ação significativa, o critério foi de que os agentes devem usar coletivamente os mesmos símbolos para fazer referência aos mesmos predadores e assim obter vantagens de fuga. Os resultados iniciais que tivemos confirmam esta hipótese, revelam que a convergência do léxico que surge das interações entre os agentes aumenta em muito as chances de sobrevivência de agentes presa. Isto é, envolvidos em um processo semiótico coletivo, os agentes de fato informam uns aos outros.

A análise do conceito de informação proposta aqui pode ser vista como uma compatibilização do conceito de processamento da informação do cognitivismo com o conceito de informação como ação significativa na teoria enativa. Entendemos que esta compatibilização se faz sob uma perspectiva de sistemas complexos da cognição. Para a teoria enativa, a informação existe na comunicação, isto é, nos processos envolvidos na autopoiese de um sistema cognitivo.

Assim defendemos que a informação como ação significativa em processos emergentes semióticos baseados em multiagente ocorre a partir da convergência de um léxico em processos auto-organizados, isto é, na comunicação. O processamento de símbolos por uma estrutura algorítmica é necessária (tese cognitivista) em multiagente, mas não é suficiente para caracterizar a informação, pois é necessário que esteja envolvida em um processo complexo auto-organizado

para que tenha sentido, para que seja uma ação significativa (tese enativista).

Nas seções seguintes, passamos à análise, concepção e implementação de sistemas complexos auto-organizados baseados em multiagente, apontando como fornecem um modo promissor de investigação dos conceitos de informação, linguagem, semiótica e semântica. Enfatizamos a importância da informação (ação significativa) em sua função adaptativa, ao fortalecer o coletivo auto-organizado.

### **3 SIMPS**

O propósito desta seção é analisar, conceber e implementar um léxico que emerge das interações entre agentes semióticos de um sistema multiagente auto-organizado. Para conduzirmos o experimento, partimos da construção de um simulador, originalmente proposto por Loula (2004), em que presas e predadores coexistem em um ambiente virtual e interação entre si através de processos básicos como memória associativa, percepção e foco de atenção. Neste trabalho, propomos avaliar o quanto a comunicação que surge por meio da convergência do léxico é efetiva para a sobrevivência das presas, ou seja, o quão melhor esses agentes conseguem se manter vivos em um ambiente hostil, utilizando a comunicação semiótica que constroem através dos usos dos símbolos. A esse simulador demos o nome de Simulador de Processos Semióticos (SIMPS).

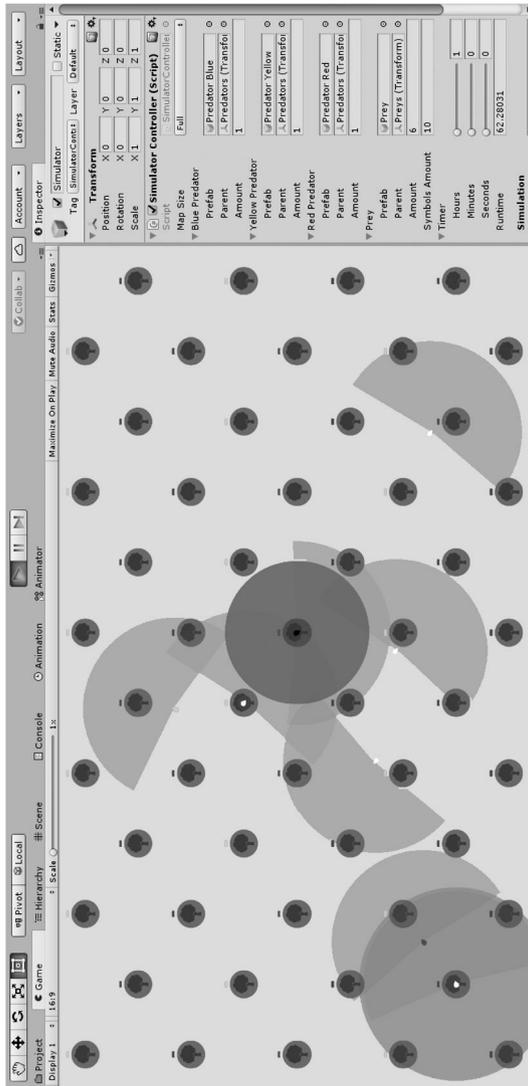


Figura 1: Captura de tela do SIMPS (Fonte: o autor).

SIMPS é inspirado no estudo etológico dos macacos *vervets* apresentado por Loula (2004). Tal como no simula-

dor proposto pelo autor, o SIMPS é composto de criaturas, divididas entre presas e predadores, e objetos estáticos como esconderijos. Os agentes interagem entre si através de sensores como audição e visão e os agentes-presa se comunicam através da emissão de sons (símbolos) de maior relevância para eles para descrever a situação de perigo de um momento específico.

### 3.1 TECNOLOGIAS DE PROGRAMAÇÃO, CONFIGURAÇÕES DE AMBIENTE E INTERFACE

SIMPS foi construído na linguagem de programação C# sobre o motor de desenvolvimento de jogos Unity<sup>5</sup>. Com o SIMPS o usuário consegue executar, pausar e interromper as simulações, além de poder avançá-las *frame a frame* e configurá-las através do *Inspector* do Unity. Entre as opções de configuração disponíveis estão, por exemplo, a duração da simulação, a quantidade de presas e predadores que atuarão no cenário, o tamanho do ambiente, a quantidade de testes e os modos das simulações, que podem ser *com aprendizado*, *sem aprendizado* ou *alternado*.

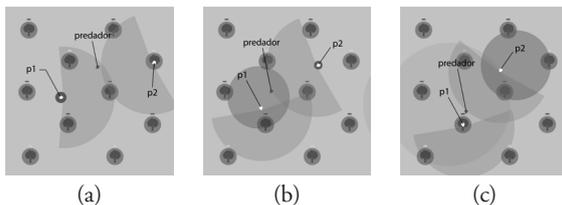
No modo de simulação *com aprendizado*, as presas conseguem associar determinado símbolo que recebem a um predador de sua área visual, o que não ocorre no modo de simulação *sem aprendizado*. De outra forma, no modo *alternado*, pressupõe-se que o usuário deseja realizar várias simulações com os mesmos parâmetros, porém não deseja ter que iniciá-las toda vez. Assim, as simulações são alternadas entre *com aprendizado* e *sem aprendizado*.

---

<sup>5</sup> O Unity foi escolhido por possuir um ecossistema de desenvolvimento bastante simples e poderoso, facilmente customizável e com vasta documentação. Essa ferramenta se integra muito bem às tecnologias e linguagens de programação mais utilizadas atualmente e ainda torna muito simples o processo de criação da parte visual.

O ambiente do SIMPS é parcialmente observável, estocástico, sequencial, dinâmico, contínuo e multiagente e varia entre os tamanhos pequeno, médio e grande, a depender da escolha do usuário. O usuário vê a simulação por meio de uma visão *top-down* de duas dimensões, em que é possível observar as interações com uma câmera vista de cima.

Toda a parte gráfica do SIMPS foi construída especificamente para o projeto. Os agentes são identificados no ambiente por marcadores em forma de agulha. Para identificar as presas, adotamos, como convenção, a cor branca. De outro modo, distinguimos os predadores por meio de três classes de cores: azul, amarelo e vermelho. O mesmo critério de divisão em classes também é atribuído aos esconderijos, de maneira que presas que se escondem em árvores com marcadores vermelhos, por exemplo, estão protegidas dos ataques de predadores vermelhos. A distribuição desses esconderijos foi feita de maneira uniforme e alternada para garantir as mesmas condições para todas as simulações. Os alarmes também são exibidos na interface como uma área circular que se expande e desaparece gradativamente. Eles também podem assumir as cores das classes dos predadores avistados pelas presas. Os raios de visão são exibidos na interface em forma de “leque” para se aproximarem de como ocorre no mundo real.



**Figura 2:** *Storyboard* do aprendizado das presas: (a) uma presa  $p1$  (marcador branco à esquerda) vê um predador (marcador vermelho acima do centro) e dispara um alarme (área circular vermelha em torno de  $p1$ ) ao ambiente para indicar sua aproximação; (b) após verificar o próprio raio de ação, a presa  $p1$  se dirige ao escon-

derijo mais próximo que a protege contra o predador visto. Enquanto isso, outra presa  $p2$  (marcador branco à direita) vê o mesmo predador e também dispara um alarme (área circular vermelha em torno de  $p2$ ) ao ambiente; (c) O alarme emitido por  $p1$  é escutado por  $p2$  e, nesse momento,  $p2$  associa o sinal recebido ao predador que está vendo. (Fonte: o autor).

### 3.2 ORGANIZAÇÃO DO PROJETO

No desenvolvimento do SIMPS, os *scripts* foram separados em *comportamentos*, *controladores*, *enumeradores* e *utilitários*. Os comportamentos são *scripts* que definem como determinado objeto vai agir no ambiente. No caso das presas, por exemplo, os comportamentos atribuídos foram o de explorador, aprendiz, medroso e mortal, enquanto que os predadores receberam os de explorador e caçador.

Os controladores, por outro lado, são responsáveis por gerenciar os *gatilhos* do Unity. O controlador de visão, por exemplo, registra quem está presente no campo de visão e o *controlador de audição* registra informações sobre o último sinal ouvido pelo agente. Também há outro tipo: o *controlador de contato*. Esse último altera o estado e a animação da presa quando ela encontra uma árvore ou é capturada por um predador. Por fim, há também o controlador de fuga, que mantém o registro sobre todos os esconderijos vistos pelas presas em uma área circular de raio igual ao do campo de visão. Os *enumeradores* são como trechos de código que facilitam a organização e o entendimento do projeto por meio de índices definidos pelo programador. Os *utilitários*, por sua vez, são pequenos programas independentes que executam tarefas concorrentes com a simulação. Estes programas são criados, em sua maioria, para gravar os estados da simulação durante sua execução.

### 3.3 COMPORTAMENTO DOS AGENTES

Diferentemente do trabalho de Loula (2004), as presas do SIMPS possuem ambos os papéis de emissoras e aprendizes. Elas também exploram o ambiente, fogem, escondem-se e morrem, assim como os predadores exploram o ambiente e caçam. As presas que morrem não são eliminadas do simulador, porém suas mortes são contabilizadas para efeitos de avaliação da adaptação dos agentes. Também propomos um modelo mais simples em relação ao apresentado por Loula (2004) no que diz respeito aos drivers dos agentes: seus comportamentos são guiados apenas por fadiga, fome e medo.

A exploração dos agentes é alternada entre um período curto de descanso e outro período de transição de uma posição para outra, ambos sendo gerados aleatoriamente em um intervalo de tempo parametrizável. As posições seguintes também são geradas ao acaso até o limite da área de ação do agente, o que também pode ser configurado pelo usuário.

Outra diferença perceptível em relação ao modelo proposto por Loula (2004) é que não há distinção motora entre os predadores. Para este primeiro momento, as simulações foram conduzidas de modo que todos os agentes, independentemente de suas classes e papéis, se movam a uma mesma velocidade. Acreditamos que essa simplificação torna o modelo um pouco mais fácil de ser interpretado, sobretudo quando fizermos a análise da adaptatividade dos agentes-presa, que será apresentada na seção de resultados. Justificando através dos eventos naturais, podemos pensar que as presas possuem predadores com velocidade média semelhante às delas.

No SIMPS, cada presa compartilha do mesmo conjunto de símbolos, que inicialmente possuem probabilidades

de uso distintas para indicar a aproximação de diferentes predadores. Cada presa notifica ao ambiente um evento capturado por ela através da emissão do sinal que possui maior valor para aquele predador específico que está vendo. Se em algum momento dois ou mais símbolos possuem o mesmo valor na tabela de associação símbolo-predador, um deles é escolhido ao acaso para ser emitido. Em outras palavras, se em algum momento a presa utiliza mais de um símbolo para identificar um único predador, ela alerta sobre qualquer um deles. Admitte-se a posição da presa como posição de origem do alarme no instante em que ela o dispara.

Outras presas que estiverem no raio de propagação do sinal certamente irão ouvi-lo. O sucesso do aprendizado ocorre quando a presa associa o sinal que recebeu a um predador que está em sua área visual (aprendizagem por reforço positivo). Caso mais de um predador esteja no campo de visão da presa no momento em que ela receber o sinal, um deles é escolhido ao acaso para participar da associação.

Para efeitos de simplicidade, os predadores possuem somente o sensor de visão, de modo que podem perseguir presas que entrem nessa área de contato. Ao caçarem, eles se utilizam das informações armazenadas pelo *controlador de visão* para definirem os alvos a serem perseguidos. Se uma presa entrar em sua área visual, a posição-alvo do trajeto do predador é alterada para a posição da presa avistada. Ainda que mais de uma presa seja vista, o alvo continuará sendo o primeiro. Se uma presa estiver no campo de visão do predador no momento em que ela entra em um esconderijo, o predador deixa de vê-la e, conseqüentemente, para de persegui-la.

Quando veem algum predador ou reconhecem algum sinal emitido, as presas sentem medo. Há pelo menos duas

ações tomadas por elas quando veem um predador: 1) se houver algum esconderijo dentro de sua área de ação que a protege contra aquele predador, a posição-alvo da presa passa a ser a posição do esconderijo mais próximo e 2) se não houver esconderijo disponível dentro de sua área de ação, a posição-alvo da presa passa a ser alguma posição aleatória contida na área de ação, especificamente do quadrante oposto ao da posição do predador.

Quando um alarme é percebido por uma presa que não está vendo nenhum predador, ela tenta reconhecê-lo comparando-o com os sinais entendidos naquele momento. Se o sinal não for reconhecido, a presa não age. Do contrário, a presa se dirige ao esconderijo mais próximo que a protege contra o predador que ela acredita que faz parte do alerta.

Em seu trabalho, Loula (2004) fixa todos os valores das funções que utiliza em um intervalo de valores reais entre zero (inclusive) e um (inclusive). No modelo proposto por ele, dizemos que um símbolo está em uso quando o seu valor é maior que os dos outros símbolos utilizados para identificar um mesmo predador. Se o autor utilizasse somente o *reforço positivo*, haveria complicações no modelo, uma vez que, à medida em que a simulação ocorre, dois ou mais símbolos podem assumir os valores máximos da função. Isso significa que determinada presa utilizaria mais de um símbolo para identificar um mesmo predador, tornando a linguagem que emergiu pouco efetiva. Para não cair neste problema, Loula (2004) emprega um decremento no valor da probabilidade do uso de um símbolo ao longo das iterações, um processo de *extinção*<sup>6</sup> no caso. Em outras palavras, à medida que o tempo

---

<sup>6</sup> Usamos aqui os conceitos de Skinner (1969) para caracterizar processos como extinção e também reforços e punições, os quais podem ser definidos como positivos ou negativos. A extinção de um comportamento ocorre pela ausência de reforços ao longo do tempo. Qualquer estímulo (entrada) que leve ao aumento da probabilidade de ocorrência de uma resposta (saída) é um reforço. Sendo a punição qualquer estímulo (entrada) que leve

passa e o símbolo não é utilizado, o valor da associação é reduzido, permitindo que outros símbolos assumam o posto ao atingirem valores mais altos.

Embora tenhamos como perspectiva a utilização da extinção na continuidade deste projeto, neste trabalho procedemos de maneira diferente. Seja o valor de *associação símbolo-predador para certo agente* (triângulo peirceano), a função que descreve o aprendizado das presas é uma função recursiva dada por:

$$f(x + 1) = f(x) + 1, \quad x \in \mathbb{N}.$$

Removendo o limite superior do intervalo, podemos nos basear apenas no emprego do *reforço positivo*, tal como pode ser observado acima.

## 4 RESULTADOS DAS SIMULAÇÕES E ANÁLISE DOS DADOS

### 4.1 RESULTADOS DAS SIMULAÇÕES

Ao todo foram realizadas sessenta simulações, das quais trinta eram *com aprendizado* e o restante *sem aprendizado*. Cada simulação foi executada com tempo de duração fixo de sessenta minutos, com a adição de seis presas, um predador do tipo vermelho, um do tipo amarelo e um do tipo azul. O conjunto de símbolos, definido com dez elementos, foi compartilhado entre as presas no processo de vocalização do alarme.

---

à diminuição da probabilidade de ocorrência de uma resposta (saída) de um agente. Um estímulo é dito positivo quando é apresentado e ele é negativo quando retirado das relações. Um exemplo de reforço negativo pode ser obtido ao se retirar a tarefa de lavar as louças, condicionada a tarefas escolares realizadas com apreço.

Mesmo com a simplificação da ideia proposta por Loula (2004), obtivemos o mesmo resultado alcançado por ele em relação à construção da linguagem emergente pela auto-organização. Inicialmente, dividimos os nossos testes de modo igual entre simulações *com aprendizado* e *sem aprendizado*. Porém, graças a essa abordagem, pudemos perceber em nosso experimento um grupo de simulações *com aprendizado* que apresentou resultado pior do que as *sem aprendizado*. Elas possuem em comum a existência de *polissemia*, que ocorre quando uma ou mais presas utilizam o mesmo símbolo para identificar predadores distintos. A essas demos o nome de *simulações com convergência parcial*. Também colocamos como parte deste grupo as simulações em que a convergência não ocorreu dentro do limite de tempo estabelecido. Ao grupo em que as presas utilizam símbolos distintos para identificar cada um dos predadores, demos o nome de *simulações com convergência total*. E por fim, ao grupo em que a capacidade de aprendizado das presas foi desabilitada, demos o nome de *simulações sem convergência*.

**Tabela 1** – Número de simulações conduzidas por tipo de simulação

Tipo de simulação	Frequência Absoluta (F.A.)	Frequência Relativa (F.R.)
Simulações com convergência parcial	15	25%
Simulações com convergência total	15	25%
Simulações sem convergência	30	50%
Total	60	100%

**Fonte:** o autor.

Dentre as trinta simulações com aprendizado, apenas uma não convergiu dentro do limite de tempo estabelecido, tendo sido esta uma convergência parcial.

Visamos analisar a adaptatividade das presas em relação à linguagem construída com base nos usos dos símbolos. Para isso, uma vez que queremos saber se a convergência melhora o desempenho das presas no ambiente, partimos do princípio de contar o número de mortes delas ao longo do tempo para cada tipo de simulação, conforme é mostrado pela Figura 4 e pela Tabela 2 a seguir.

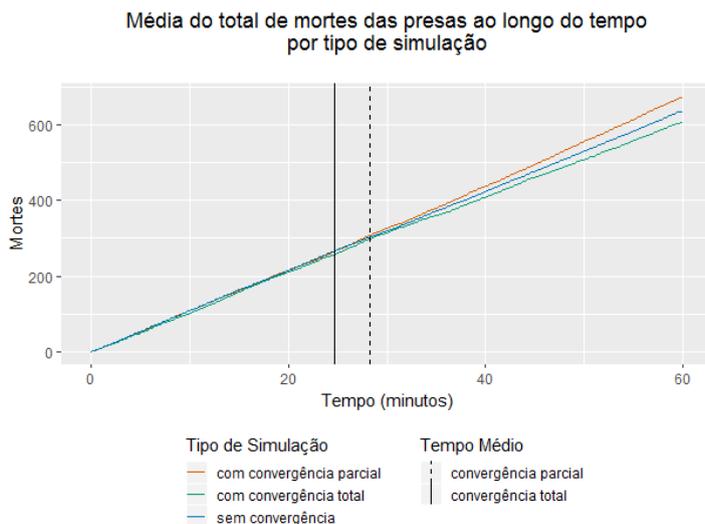
**Tabela 2** - Média de mortes das presas por tipo de simulação ao final das simulações

<b>Tipo de simulação</b>	<b>Média de mortes</b>
Simulações com convergência parcial	674,87
Simulações com convergência total	607,47
Simulações sem convergência	637,77

**Fonte:** o autor.

Observamos então que as simulações parciais apresentam um quadro de informações falsas. Dizemos que estas não trazem vantagens adaptativas por gerarem um número de mortes maior do que o das simulações com convergência léxica desabilitada (com o sistema funcionando estritamente ao nível comunicacional).

**Figura 4** - Média do total de mortes das presas ao longo do tempo por tipo de simulação. Resultados obtidos pelo SIMPS



**Fonte:** o autor

Passemos na subseção seguinte à análise dos dados produzidos.

## 4.2 ANÁLISE DOS DADOS

Como é possível observar na Figura 4, a partir de trinta minutos aproximadamente, momento posterior ao tempo médio de convergência total e parcial das presas, o número de mortes começa a variar entre os tipos de simulação. Pelo desenho do gráfico, somos levados a crer que se aumentarmos o tempo das simulações, mais acentuada será essa variação. De acordo com os resultados da Tabela 2, o número de mortes das simulações com convergência total foi 4,8% menor do que nas simulações sem convergência, até o instante de tempo limite determinado. Por outro lado, as simulações com conver-

gência parcial mostraram-se piores do que as simulações sem convergência em 5,8%. Isso nos mostra que eventos completamente aleatórios podem ser melhores do que eventos com convergência parcial, uma vez que esse último é determinado por erro semântico (polissemia). Além disso, fica evidente que o processo de efetiva comunicação semiótica depende da construção de uma linguagem que é entendida pelos indivíduos, sendo necessário que utilizem símbolos distintos para cada um dos predadores (discriminação de estímulos).

Entendemos que o número de mortes nas *simulações com convergência parcial* aumenta à medida em que o alarme disparado por uma presa  $p1$  para indicar a presença de um predador  $a$  é entendida por outra presa  $p2$  como um indício da presença de um predador  $b$ , uma vez que o símbolo utilizado no processo de comunicação é o mesmo para ambos. Desse modo,  $p2$  é levado a agir erroneamente em resposta ao estímulo interpretado por ele a partir do alerta recebido, o que pode findar em sua própria morte. Caracterizamos esse processo como um erro semântico.

Entendemos como informações no modelo em questão o fluxo de entrada e saída processada pelos agentes semióticos. Este fluxo de ações no sistema multiagente não está limitado ao interior dos agentes, mas se encontra em suas relações envolvidas na auto-organização do coletivo. Com a função semiótica habilitada ou não obtemos, pelas interações dos agentes, um sistema auto-organizado, isto é, comunicacional. Ao habilitarmos os processos semióticos no sistema auto-organizado, ele apresenta usos emergentes dos símbolos (convergência léxica).

Alguns processos informacionais trazem vantagens adaptativas, sobretudo aqueles que discriminam totalmente os

predadores. Nestes, as informações não carregam ambiguidade e trazem vantagens de sobrevivência para os agentes presas, preservando o coletivo das presas. As informações ambíguas são polissemias, no caso, um símbolo faz referência a mais de um predador, o que não permite ao agente presa adotar uma boa estratégia de fuga em função do tipo de predador, levando-o a morte. Ações significativas podem assim ser falsas e, no caso de nossos resultados, eles mostram que informações falsas podem mesmo ser mais prejudiciais aos agentes presa do que um caso de comunicações sem relações semióticas.

## CONCLUSÃO

O tema deste trabalho foi o estudo da informação como ação significativa em processos semióticos emergentes em sistemas complexos baseados em multiagente. Tratou-se de uma pesquisa interdisciplinar na área de Ciências Cognitivas, com foco em psicologia cognitiva e vida artificial. Com a finalidade de caracterizar o conceito de informação como um processo semiótico emergente, concebemos e implementamos um léxico emergente baseado em símbolos que, ao informar uma situação de perigo, podem fornecer vantagens adaptativas a agentes de um sistema multiagente auto-organizado. Nossas simulações corroboram os resultados da literatura, notadamente os obtidos por Steels (1996; 2003), Loula *et al.* (2003), Loula (2004), Arnellos, Spyrou & Darzentas (2003, 2006). A nossa análise foi feita segundo o método da análise conceitual.

Procedemos pela construção de um *Simulador de Processos Semióticos*, o qual chamamos de SIMPS, para representar um cenário real, originalmente proposto por Loula et al. (2003), em que agentes presas (“macacos *vervets*”)

interagem entre si e com o ambiente através de processos básicos como memória associativa, percepção e foco de atenção, possuindo também *drivers* de fadiga, fome e medo. Cada presa compartilha do mesmo conjunto de símbolos que inicialmente possuem probabilidades de uso distintas para cada um deles para indicar a aproximação de diferentes predadores. Cada presa notifica ao ambiente um evento capturado por ela através da emissão do sinal que possui maior probabilidade de uso. Os demais agentes-presa podem perceber este sinal através de seus sensores visual e auditivo. Para cada sinal recebido com sucesso, há um reforço empregado à probabilidade de uso daquele símbolo para fazer referência ao predador que apareceu na área visual.

Os resultados de nossas simulações mostraram que, mesmo com a simplificação do modelo proposto por Loula (2004), obtivemos os mesmos resultados de convergência dos símbolos. Estes resultados foram ainda obtidos a partir de um número maior de simulações do que os de Loula (2004). Também repartimos a população de agentes em três grupos: (1) convergência total, (2) convergência parcial e (3) sem convergência. Com estes três grupos, ao invés de usar somente os grupos (1) e (3), como nos resultados de Loula (2004), pudemos perceber melhor a eficiência adaptativa da convergência de léxico ao termos encontrado um número intermediário de mortes no grupo (3), com relação aos grupos (1) e (2).

Hipotetizamos que a informação é uma ação significativa que ocorre em processos semióticos emergentes. Para que sejam uma ação significativa, o critério foi o de que os agentes deveriam usar coletivamente os mesmos símbolos para fazer referência aos mesmos predadores. A partir dos resultados, percebemos que, para que a informação fosse adaptativa e forne-

cesse às presas vantagens de fuga, foi necessário no modelo que houvesse uma discriminação total dos predadores pelas presas. Neste caso, a eficiência protetiva da informação é maior se um símbolo não é usado para referenciar mais de um predador.

A análise do conceito de informação proposta aqui trata de uma compatibilização do conceito de processamento da informação do cognitivismo com o conceito de informação como ação significativa na teoria enativa. Entendemos que esta compatibilização se faz sob uma perspectiva de sistemas complexos da cognição, da qual a vertente semiótica em teoria enativa se aproxima mais. Para a teoria enativa, a informação existe na comunicação, isto é, nos processos envolvidos na autopoiese de um sistema cognitivo. Assim, defendemos que a informação como ação significativa em processos emergentes semióticos baseados em multiagente ocorre a partir da convergência da linguagem em processos auto-organizados, isto é, na comunicação. Todavia, a informação ocorre em processos semióticos comunicacionais. Com isso, defendemos que processos semióticos são uma condição necessária para processos informacionais, também em sistemas autopoieticos naturais.

Em sistemas artificiais, o processamento em uma estrutura simbólica física é necessária (tese cognitivista) para a arquitetura multiagente, mas não é suficiente para caracterizar a informação, pois é necessário que esta esteja envolvida em um processo auto-organizado complexo para que tenha sentido, para que seja uma ação significativa (tese enativista). Informação não é simplesmente um dado processado por regras lógico-matemáticas, mas antes de tudo uma ação significativa em sistemas auto-organizados semióticos. No mais, enfatizamos a importância da concepção e implementação de

arquiteturas de agentes semióticos a fim de obtermos processos semióticos emergentes.

Concluimos que a análise, concepção e implementação de sistemas complexos auto-organizados baseados em multi-agente fornecem um modo promissor de investigação dos conceitos de informação, linguagem, semiótica e semântica. Enfatizamos a importância da informação em sua função adaptativa, ao fortalecer o coletivo auto-organizado. Como perspectivas, visamos o aprofundamento teórico para lapidar a ainda incipiente teoria enativa semiótica. Também almejamos o aprofundamento prático na implementação de gramáticas emergentes e ampliação do poder computacional dos modelos.

## REFERÊNCIAS

- ANDLER, D. Cálculo e representação: as fontes. In: *Introdução às Ciências Cognitivas* (pp. 25-54). São Leopoldo: Ed. UNISINOS. 1998.
- ARNELLOS, A., SPYROU, T. & DARZENTAS, J. Towards a Framework that Models the Emergence of Meaning Structures in Purposeful Communication Environments. *The 47th Annual conf. of The Int. Society for the Systems Sciences (ISSS)*, 3, 103. 2003.
- ARNELLOS, A., VOSINAKIS, S., SPYROU, T. & DARZENTAS, J. The Emergence of Autonomous Representations. *Journal of Computers*, 1(6), 29-36. 2006.
- AXTELL, R. The End of the Beginning for Multi-agent Systems Social Science. In F. AMBLARD, & D. PHAN (Ed.). *Modélisation et simulation multi-agents: applications pour les Sciences de l'Homme et de la Société* (pp. 115-). Paris: Lavoisier, 2007.
- BROOKS, R. Intelligence Without Representation. *Artificial Intelligence Journal*, 47, 139-159, 1991.

- CARVALHO, L. L. *Représentations émergentes: une approche multi-agents des systèmes complexes adaptatifs en Psychologie Cognitive*. Thèse de Doctorat, Université Lumière Lyon 2, France, 2008.
- CARVALHO, L. L. A realização da mente: crítica da teoria enativa ao conceito de propriedade emergente. In TOLEDO, G. L. *et al. Debates Contemporâneos em Filosofia da Mente*. São Paulo: FiloCzar, 2018.
- DROGOUL, A. *De la simulation multi-agent à la résolution collective de problèmes*. Une étude de l'émergence de structures d'organisation dans les systèmes multi-agents. Thèse de doctorat, Université de Paris 6. 1993.
- DROGOUL, A. & FRESNEAU, D. Métaphore du fourragement et modèle d'exploitation collective de l'espace pour des colonies de robots autonomes mobiles. *JFIADSMASMA'98*. Paris, France: Hermès, 1998.
- FERBER, J. Concepts et méthodologies multi-agents. In F. Amblard, & D. Phan (Ed), *Modélisation et simulation multi-agents: applications pour les Sciences de l'Homme et de la Société* (pp. 23-48). Paris: Lavoisier, 2006.
- GOMES, A., GUDWIN, R., QUEIROZ, J. Meaningful agents: a semiotic approach. In: *International Conference on Integration of Knowledge Intensive Multi-Agent Systems (KIMAS)*, 2005, Waltham. Proceedings KIMAS 2005. Boston: IEEE Press, 2005. p. 399-404.
- HALBWACHS, M. *La mémoire collective*. Paris: Albin Michel, (1950) 1997.
- HALL, W. P. Emergence and growth of knowledge and diversity in hierarchically complex living systems. *Selection, Self-Organization and Diversity CSIRO Centre for Complex Systems Science and ARC Complex Open Systems Network*. Katoomba, NSW, Australia 17-18 May. 2006.
- HALPIN, H. Representations: The Hard Problem for Artificial Life. *Oxford Journals*, 2006. 527-535.
- HUTTO, D. D. & MYIN, E. *Radicalizing enactivism: basic minds without content*. Cambridge: MIT, 2013.
- HUTTO, D. D. & MYIN, E. *Evolving enactivism: basic minds meet content*. Cambridge: MIT, 2017.

KENNEDY, J. F., EBERHART, R. C. & SHI, Y. *Swarm Intelligence*. San Francisco, CA: Morgan Kaufmann, 2001.

KULL, K., DEACON, T., EMMECHE, C., HOFFMEYER, J., STJERNFELT, F. Theses on biosemiotics: Prolegomena to a theoretical biology. *Biological Theory*, 4(2) 2009, 167–173.

JESUS, P. From enactive phenomenology to biosemiotic enactivism. *Adaptive Behavior - Animals, Animats, Software Agents, Robots, Adaptive Systems*, 24(2), 2016, pp. 130-146.

LENAY, C. Organisation émergente dans les populations: biologie, éthologie, systèmes artificiels. *Intellectica*, 19, 1994, 9-17.

LOULA, A. C. *Comunicação Simbólica Entre Criaturas Artificiais: um experimento em Vida Artificial*. Dissertação (Mestrado em Engenharia Elétrica) – Faculdade de Engenharia Elétrica e Computação, Universidade Estadual de Campinas, Campinas, 2004.

LOULA, A., GUDWIN, R. & QUEIROZ, J. Synthetic approach of symbolic creatures. *S.E.E.D. Journal -- Semiotics, Evolution, Energy, and Development*, 3(3), 2003, 125-133.

LOULA, A., GUDWIN, R. & QUEIROZ, J. Symbolic communication in artificial creatures: an experiment in Artificial Life. *Lecture Notes in Computer Science*, 3171, 2004, 336-345.

LOULA, A., GUDWIN, R., EL-HANI, C. N., QUEIROZ, J. The Emergence of Symbol-Based Communication in a Complex System of Artificial Creatures. *5th International Conference Integration of Knowledge Intensive Multi-Agent Systems KIMAS'05: Modeling, Evolution and Engineering*. Waltham, Massachusetts: IEEE, 2005.

MATURANA, H. & MPODOZIS, G. *La dérive des espèces par la voie de la sélection naturelle*. Lyon: Presses Universitaires de Lyon, 1999.

MATURANA, H. & VARELA, F. *L'arbre de la connaissance*. Paris: Addison-Wesley France, (1984) 1994.

NEWELL, A. & SIMON, H. Computer science as empirical inquiry: Symbols and search. *Communications of the Association for Computing Machinery*, 19(3), 1976, 113-126.

- PYLYSHYN, Z. W. *Computation and cognition: toward a foundation for cognitive science*. Cambridge: MIT press, 1984.
- SEARLE, J. R. Minds, brains and programs. *Behavioral and Brain Sciences*, v. 3, 1980, p. 417-424.
- SKINNER, B. F. *Contingencies of reinforcement: A theoretical analysis*. New York: York: Appleton-Century-Crofts, 1969.
- STEELS, L. Emergent Adaptive Lexicons. In: Maes, P. (ed.) *From Animals to Animats 4: Proceedings of the Fourth International Conference On Simulation of Adaptive Behavior*, The MIT Press, Cambridge Ma, 1996.
- STEELS, L. The origins of ontologies and communication conventions in Multi-Agent Systems. *Autonomous Agents and Multi-Agent Systems*, 1(2), 1998, 169-194.
- STEELS, L. Language as a complex adaptive system. *Parallel Problem Solving from Nature -PPSN VI*. Springer, Lecture notes in Computer Science, 2000a.
- STEELS, L. The emergence of grammar in communicating autonomous robotic agents. *ECAI2000* (pp. 764-769). Amsterdam: IOS Press, 2000b.
- STEELS, L. Intelligence with representation. *Philosophical Transactions of the Royal Society, A* 361(1811), 2003, 2381-2395.
- STEELS, L., VAN TRIJP, R. & WELLENS, P. Multi-Level Selection in the Emergence of Language Systematicity. *ECAL07*(pp. 425-434). Springer, 2007.
- STEELS, L., VAN TRIJP, R. How to make construction grammars fluid and robust. In Luc Steels (Ed). *Design Patterns in Fluid Construction Grammar* (pp. 301-330). Amsterdam: John Benjamins, 2011.
- STUBER, A., HASSAS, S. & MILLE, A. Combining multiagents systems and experience reuse for assisting collective task achievement. *Proceedings of ICCBR-2003 Workshop "From structured cases to unstructured problem solving episodes for experience-based assistance*. Trondheim, Norway, 2003.

- STUBER, A., HASSAS, S. & MILLE, A. Language games for meaning negotiation between human and computer agents. *Engineering Societies in the Agents' World (ESAW 2005)*. Kusadasi, Aydin, Turkey: Springer Verlag, 2005.
- STEWART, J. R., GAPENNE, O., & DI PAOLO, E. A. *Enaction: Toward a new paradigm for cognitive science*. MIT Press, 2010.
- THOMPSON, E. *A Mente na Vida: Biologia, Fenomenologia e Ciências da Mente*. Lisboa: Instituto Piaget, 2013. Original publicado em 2007.
- TOMASELLO, M., BROOKS, P. J. Early syntactic development: A Construction Grammar approach In: Barrett, M. (ed.) *The Development of Language* (pp. 161-190). Hove, UK: Psychology Press, 1999.
- VARELA, F., THOMPSON, E., ROSCH, E. *The Embodied Mind*. Cambridge, MA: MIT Press, 1991.
- WELLENS, P., VAN TRIJP, R., STEELS, L., BEULS, K. Fluid Construction Grammar for Historical and Evolutionary Linguistics. *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*. Sofia, Bulgaria, 2013.
- WITTGENSTEIN, L. *Philosophical Investigations*. Malden, MA: Blackwell Publishing, (1953) 2001.

# TRANSDUÇÃO INFORMACIONAL: IMPACTOS DO CONTROLE SOBRE OS DADOS

*Ricardo César Gonçalves Santana*

*[...] a reciprocidade da troca de papéis entre o destinatário e o falante confere à relação dialógica um caráter igualitário [...] o respectivo outro tem de ser levado a sério em seu cerne individualizado e ser reconhecido em sua alteridade radical (Jürgen Habermas, 2014, p. 60).*

O conceito de transdução é muito utilizado em várias áreas, como na Biologia e até mesmo nas engenharias, principalmente relacionado às necessárias transformações de sinais nos processos de comunicação, mas não com a perspectiva apresentada neste texto. Aqui se busca explicitar as camadas de abstração oriundas das diversas transformações a que a informação está sujeita quando utilizada nos suportes digitais (figura 3) e seus impactos na percepção dos envolvidos a partir da opacidade gerada por estas transformações, que vão muito além dos tratamentos mais superficiais de seus conteúdos. Assim, propõe-se o conceito de transdução informacional como

um dos elementos estruturantes para a explicitação das possíveis modificações, e até manipulações, a que estes conteúdos passam a estar suscetíveis. Este conceito pode contribuir com a redução da insciência dos diversos atores envolvidos nos ciclos de vida que os dados estão sujeitos.

Para que se possa estabelecer uma percepção clara sobre os diversos aspectos relacionados à questão do acesso a dados, considerando atores, ferramentas e, ainda, impactos sociais, culturais e econômicos, é necessário entender as estruturas sobre as quais os processos envolvidos se sustentam. Entre essas estruturas, cabe destaque a constituição das camadas de abstração que emergem dos processos de interpretação quando da transdução informacional dos conteúdos registrados nos suportes digitais, altamente estruturados, até ao conteúdo próprio para resposta aos usuários. Essas camadas, sobre as quais os usuários são inscientes (AFFONSO & SANT'ANA, 2018) e das quais derivam transformações que, mesmo não se configurando como violação do que se poderia considerar como integridade lógica e física dos conteúdos envolvidos, podem gerar novas percepções “fatuais” da realidade.

A transdução, considerada como transformação de um tipo de sinal em outro, envolve tipos distintos de energia que propiciam o controle sobre processos ou a realização de medições, como representação do transcurso do conteúdo entre os bits persistidos nos suportes digitais, passando pelas sucessivas agregações semânticas, até os formatos resultantes nas interfaces líquidas, abertas e altamente aderentes às características dos usuários finais amplamente difundidas e ubíquas.

## CICLO DE VIDA DOS DADOS

O processo de transdução envolve, em diferentes momentos, diferentes aspectos, tais como: competências necessárias, conhecimentos sobre o contexto e sobre o usuário, recursos e métodos. A necessidade de identificação da temporalidade das especificidades leva à necessidade de se estabelecer o escopo de tais momentos, de forma a concatená-las, mesmo que dentro de limitações, como por exemplo a concomitância de alguns fatores ou mesmo ações que estariam presentes em momentos distintos (SANT'ANA, 2016).

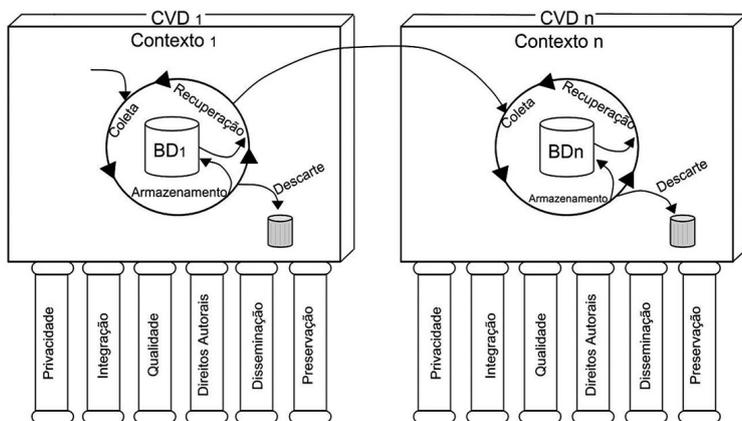
Como base da estruturação destes distintos momentos, pode-se considerar fases em que a concatenação de parte considerável das especificidades justifica tal categorização. Essas fases (Figura 1), consideradas como parte de ciclos específicos de cada detentor e de suas necessidades, inicia-se por um momento em que o objetivo é a obtenção de conteúdos, ou o que se pode denominar de fase de coleta. Nesta fase, coabitam especificidades que levam da definição do escopo das necessidades informacionais, passando pelo planejamento de ações, localização de fontes e culminando no acesso ao conteúdo desejado.

Uma vez consolidado o acesso, pode se configurar a necessidade de acesso futuro a tal conteúdo, o que leva a outro momento do processo em que agora competências e recursos mais próximos do tecnológico ganham protagonismo, encadeando ações de registro em suporte, o que na Ciência da Computação remeteria ao conceito de persistência do conteúdo e pode ser denominada fase de armazenamento, quando se considera o ciclo de vida dos dados.

Uma vez armazenado o conteúdo, não se tem ainda garantia da presença das características que seriam necessárias para estes conteúdos sejam acessados no futuro da forma esperada. Manifesta-se, assim, o momento em que o objetivo está relacionado a fases de coleta de outros ciclos de vida ou mesmo do próprio e que, por conseguinte, deve considerar a disponibilidade de tais conteúdos. Ainda no contexto do ciclo de vida dos dados, pode-se denominar tal fase como sendo a de recuperação.

Uma das consequências do acesso a dados é o potencial de persistência crescente de certos conteúdos. Grande parte dos dados coletados é eliminada já nos momentos iniciais da obtenção, seja por filtro previamente definido, seja por seleção construída em função dos objetivos do acesso ou mesmo por exceder a capacidade de retenção e uso de tais dados, situação que tende a ser cada vez mais comum na fase e coleta, em função da evolução acelerada das formas de captação.

Mesmo os conteúdos que, após esta fase inicial, são destinados à fase de armazenamento podem, em determinado momento, passar por processo de eliminação. Esta fase pode ser definida como a fase de descarte e se, no passado recente, tinha grande foco em função de limitações tecnológicas, volta a ser tema de grande atenção em função das implicações relacionadas à privacidade (BRASIL, 2018).



**Figura 1:** Ciclo de Vida dos Dados (SANT'ANA, 2016)

Estas fases que estruturam o processo de acesso a dados, ao longo do tempo, têm ainda outros fatores que requerem atenção, mas que não se restringem apenas a uma das fases identificadas em função do tempo. São fatores transversais às fases do ciclo de vida dos dados e que podem ser identificados como privacidade, integração, qualidade, direitos autorais, disseminação e preservação.

Ao se considerar a existência de um vasto número de ciclos de vida dos dados, aos quais uma determinada instância encontra-se submetida, pode-se conceber a possibilidade de estabelecimento de conceitos como densidade e gradiente, próprios da concepção de campo (SANT'ANA, 2019), o que leva à possibilidade de se considerar também a densidade relativa ao volume de conteúdos que de um lado são obtidos e por outro são fornecidos pela instância em análise. Emergem, assim, duas dimensões de interação, compostas ambas pela interação da instância com ciclos de vida dos dados, tendo de um lado o conjunto de processos de coleta aos quais ela está

submetida (*output*) e de outro lado o conjunto de processos de recuperação ao qual ela tem acesso (*input*). Em conjunto, estas duas dimensões comporão um campo informacional da instância em análise (Figura 2).

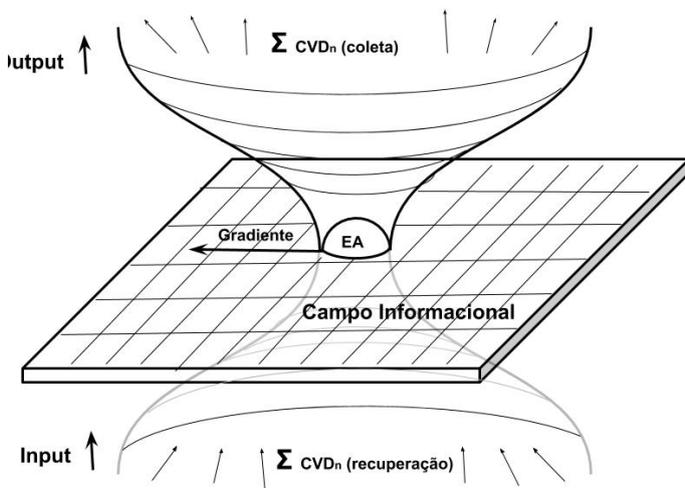


Figura 2: CVDs sobre o elemento alvo compo Campo Informacional

A diferença entre a densidade do que acessamos e do que é acessado, por exemplo, a nosso respeito, deixa clara a crescente assimetria existente entre os usuários (e a sociedade em geral) e os detentores dos dados; o que remete à necessidade de explicitação dos processos e dos atores envolvidos

Assim, a reflexão aqui proposta se ampara em duas dimensões de análise. A primeira, resultante de transformações oriundas do processo de produção sgnica, necessárias e intrínsecas ao uso de suportes tecnológicos digitais e que se sustentam em distintos formatos e energias, para o registro e persistência de conteúdos. A segunda remete à questão dos atores

envolvidos no ciclo de vida dos dados, com especial destaque à desconformidade de agendas entre detentores e usuários.

Em relação ao processo de produção sógnica das instâncias mediadoras, que envolve camadas de código necessárias à transformação de conteúdos, resumindo, ocultando e minimizando a complexidade do real, cabe a explicitação da simplificação resultante desse processo abstrativo. Conforme ressalta Eco:

Escolho então situações descontínuas, discretas, recortadas no *continuum* dos fatos possíveis, e as erijo em *traços pertinentes* aos fins da comunicação que me interessa (ECO, 2001, p. 27).

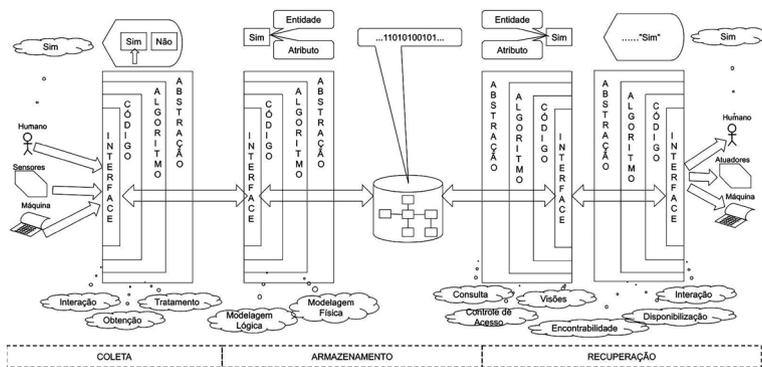
Elementos emergentes são selecionados de um campo produzido pelas interfaces entre elementos maquínicos e entre estes e ambientes ou usuários, exigindo, assim, como resultado destes procedimentos abstrativos e, por consequência, simplificação obrigatória, pela própria natureza contínua e analógica dos conteúdos em uma transdução para conteúdos discretos e digitais. Cabe aqui o seguinte questionamento:

Como pode o real construir-se sob a forma de signos? Em que condições alguns signos que se tornam matemáticos acessam o real e o representam? Que estatuto conceder àqueles que não lhe têm acesso a não ser por esse conjunto de características? (SERRES, 2003, p. 70).

Esse processo de abstração, tanto nas codificações como nas decodificações, exige a participação do agente maquínico como executor das operações, o que por sua vez constitui a obrigatoriedade de sequências finitas e muito bem definidas de passos, algoritmos, para que estas operações possam ser rea-

lizadas, o que remete à segunda dimensão de análise em que se deve considerar o papel dos detentores dos dados. A elaboração de algoritmos facilitadores desse processo requer não só competências técnicas específicas, como, ainda, conhecimento sobre os significados envolvidos com os conteúdos tratados e objetivos desejados, criando uma camada de opacidade sobre as operações modeladas, implementadas e realizadas por parte dos detentores e intermediários ao construir e reconstruir as tessituras das denotações e das conotações.

Assim, apesar da aparente participação democrática dos diversos atores, são os detentores de dados, compondo o que poderia ser considerado como oligopólios, que controlam e direcionam a coleta, armazenamento e recuperação dos dados (SUN *et al*, 2015), sendo que tal opacidade resulta em profunda insciência dos usuários em relação à transdução dos conteúdos.



**Figura 3:** Transdução de conteúdos da coleta à recuperação

Os conteúdos, para que sejam passíveis de interpretação, e, portanto, possam ser considerados como dados, requerem elementos complementares que garantam a semântica mínima e que podemos inferir como a estrutura composta por

entidade, atributo e valor  $\langle e,a,v \rangle$  (SANTOS & SANT'ANA, 2015), tríade esta que vai além de elementos dêiticos e embreadores que sustentam a integração de cada valor em sua respectiva estrutura.

As diversas interfaces envolvidas nesse processo transformativo, ao mesmo tempo em que viabilizam os seus usos e interpretação, necessariamente erguem barreiras à percepção dos usuários sobre as transduções exigidas para que os conteúdos possam ser tratados pelos dispositivos digitais, que, por sua essência, dependem totalmente dos algoritmos disponibilizados pelos detentores.

Quando se analisa as fases de recuperação às quais uma determinada instância tem acesso, observa-se as que são mais perceptíveis e tendem a ser consideradas como retorno das concessões obtidas pelos detentores de dados. Vários fatores podem desviar uma percepção mais clara sobre elas, tais como a ideia preconcebida de que o que se está acessando em um determinado ambiente, como uma rede social, por exemplo, é livre de organização intencional por parte dos interesses daqueles que as mantém. Aliás, esta falsa sensação de liberdade pode até agravar a assimetria informacional, por não suscitar no usuário a motivação por precaução ao interpretar os conteúdos a que tem acesso nestes ambientes.

Já com relação às fases de coleta, a insciência dos usuários acentua-se, chegando mesmo ao extremo de não lhe ser possível, muitas das vezes, nem mesmo ter ciência de que o processo está ocorrendo. Quando se trata de formas mais diretas de obtenção de dados, tais como formulários ou mesmo por meio de registro de interações com mídias sociais por exemplo, os usuários são convencidos de que os dados coletados terão como finalidade sua comodidade e podem até ser alvo de legislações,

como as que preveem garantias em relação aos dados pessoais (BRASIL, 2018). No entanto, muitas das coletas, que são realizadas por meio de dispositivos que podem atuar sem a participação ativa do usuário, podem coletar dados, inicialmente considerados como não identificadores, e que, portanto, ficam livres de controle ou de desconfiças mais diretas. No entanto, quando esses dados são integrados com outros dados, para subsequente tratamento, eles podem gerar o que se denomina de efeito mosaico e levar à identificação e violação de privacidade, que não poderiam ser previstas sem o acesso às camadas mais internas do ciclo de vida dos dados.

Um bom exemplo deste tipo de resultado pode ser observado em pesquisas como a realizada por Keith Chen e Ryne Rohla e publicada na *Science* em junho de 2018 (CHEN & ROHLA, 2018) e que, utilizando informações obtidas sobre a localização de celulares e consideradas como não identificadoras, ou seja, que não traziam elementos de identificação dos donos dos dispositivos, permitiram que se chegassem a conclusões como quantificar o impacto do partidatismo e da propaganda política nos jantares de Ação de Graças em 2016, de mais de 10 milhões de americanos, e ainda: que os democratas encurtaram suas visitas aos lares republicanos entre 20 e 40 minutos; que os republicanos reduziram seu tempo com os anfitriões democratas de 50 a 70 minutos; famílias incompatíveis de áreas com alta exposição política a anúncios gastaram menos tempo juntas. Informações sensíveis, obtidas a partir de processos aos quais os usuários estavam totalmente inscientes não só sobre quais os dados coletados, como também sobre a existência do próprio processo de coleta. Outras compensações podem ser utilizadas na obtenção das concessões para acesso aos dados, tais como: segurança (exemplo: câmeras em ambientes públicos utilizadas como instrumental dos agentes

de segurança) e visibilidade (exemplo: maior volume de acesso em redes sociais) entre outras.

## REFLEXÕES

Deter o controle sobre os processos do ciclo de vida dos dados remete ao domínio sobre os algoritmos envolvidos nas diversas fases do acesso a dados. A simples possibilidade de imbricamento dos interesses desses detentores com os pressupostos objetivos relacionados ao campo informacional deve ser motivo de reflexões e busca por redução das assimetrias entre os detentores e usuários.

A insciência do usuário tende a ser compensada e amenizada por processo constante de construção de senso comum, que tem por objetivo convencer as pessoas sobre uma pressuposta eficiência intrínseca, objetiva, de agentes maquínicos, o que agrava ainda mais a situação.

## REFERÊNCIAS

AFFONSO, E.P.; SANT'ANA, R.C.G. Privacy awareness issues in user data collection by digital libraries. *IFLA Journal*. Volume: 44 número: 3, out de 2018. pp.170-182. Disponível em: <https://journals.sagepub.com/doi/abs/10.1177/0340035218777275>. Acesso em: 10/11/2018.

BRASIL. Lei nº 13.709, de 14 de agosto de 2018. Dispõe sobre a proteção de dados pessoais e altera a Lei nº 12.965, de 23 de abril de 2014 (Marco Civil da Internet). Brasília, 2018. Disponível em: [http://www.planalto.gov.br/ccivil\\_03/\\_ato2015-2018/2018/lei/L13709.htm](http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709.htm). Acesso em: 10/01/2019.

CHEN, K. & ROHLA, R. The Effect of Partisanship and Political Advertising on Close Family Ties. *Science*, vol. 360, nº 6392, pp. 1020-1024. 01 de jun de 2018. Disponível em: <http://science.sciencemag.org/content/360/6392/1020>. Acesso em 10/07/2018.

ECO, U. *A estrutura ausente*: introdução à pesquisa semiológica. São Paulo: Perspectiva, 2007.

HABERMAS, J. *Na esteira da tecnocracia*: pequenos escritos políticos. São Paulo: Editora Unesp, 2014.

SANT'ANA, R.C.G. Ciclo de vida dos dados e o papel da ciência da informação. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 14., 2013, Florianópolis. *Anais...* Florianópolis, 2013. Disponível em: <http://enancib.sites.ufsc.br/index.php/enancib2013/XIVenancib/paper/viewFile/284/319>. Acesso em: 05/05/2014.

SANT'ANA, R.C.G. Ciclo de vida dos dados: uma perspectiva a partir da ciência da informação. *Informação & Informação*, [S.l.], v. 21, n. 2, p. 116–142, dez. 2016. ISSN 1981-8920. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/27940>. Acesso em: 29/12/2016. doi:<http://dx.doi.org/10.5433/1981-8920.2016v21n2p116>.

SANT'ANA, R.C.G. Campo Informacional Resultante da Interação de Ciclos de Vida dos Dados. In: DIAS, G.; FREIRE, B. (Orgs). *Dados Científicos*: perspectivas e desafios. Editora UFPB - João Pessoa. 2019 p.5-19 (prelo)

SANTOS, P.L.V.A.C & SANT'ANA, R.C.G. Dado e Granularidade na perspectiva da Informação e Tecnologia: uma interpretação pela Ciência da Informação. *Ciência da Informação*, Brasília, v. 42, p. 199-209, 2015.

SERRES, M. *Hominescências*: o começo de outra humanidade? Rio de Janeiro: Bertrand Brasil, 2003.

SUN, M.; WEI, Y.; JIANG, J. Big Data: risks and regulatory strategies. *IEEE International Conference on Progress in Informatics and Computing*, 2015. p. 358-362.

## METADADOS DE PRESERVAÇÃO DIGITAL NA ERA DE *BIG DATA*

*Jean Fernandes Brito*  
*Daniel Martínez-Ávila*

Com o avanço das tecnologias, os suportes de informação e dispositivos tornam-se obsoletos cada vez mais rapidamente. Além dos suportes, na era dos grandes volumes de dados, a informação contida neles também apresenta as características dos *Big Data*, consideradas os cinco V's: volume, velocidade, veracidade, variedade e valor. Nesse sentido, as descrições adequadas dos objetos digitais que transmitem os *Big Data* necessitam de uma representação efetiva para garantir a preservação das informações e registros descritivos.

Nesse contexto, além de se tornarem os suportes obsoletos, há um grande volume de dados disponível, que corre o risco de desaparecer ou se distorcer, havendo a necessidade de criar e pensar meios de descrever e preservar os objetos físicos e ainda preservar os metadados de objetos digitais que correspondem aos *Big Data*.

Os *Big Data* representam quantidades maciças de diversos tipos de informação, que são constantemente

coletadas, processadas, misturadas e analisadas estatisticamente na esperança de encontrar correlações que revelem significado (MARTÍNEZ ÁVILA *et al.*, 2018).

Nesse sentido, essas quantidades massivas de dados e informações necessitam ser representadas para garantir a preservação digital e proporcionar contexto a teorias e hipóteses científicas (pragmática). Os metadados de preservação digital têm como propósito gerenciar e administrar a representação e a descrição dos objetos digitais, disponibilizar as informações suficientes para tomada de decisões dos gestores, e manter o acesso ao conteúdo dos objetos em um ambiente tecnológico em longo prazo.

Nesse contexto, esses metadados devem ser criados de maneira adequada para que futuramente “[...] os dados possam ser copiados, ‘re-criados’ e atualizados como migrações periódicas necessárias em novos formatos para seu uso em novas gerações de *hardware* e *software*” (MÁRDERO ARELLANO, 2004, p. 21).

Desse modo, os profissionais da informação que lidam com os metadados de preservação digital devem sempre atualizar o acervo digital e acompanhar as mudanças tecnológicas e informacionais, para, assim, disponibilizar informações atuais do cotidiano para os usuários.

Castro (2012) comenta que a integração semântica dos metadados seria “o processo de utilização de uma representação conceitual dos dados e seus relacionamentos, com o objetivo de eliminar as possíveis heterogeneidades de compreensão e interpretação”.

Diante do exposto, a presente pesquisa tem por objetivo apresentar uma reflexão teórica sobre os metadados de preservação digital à luz do contexto de *Big Data*.

A pesquisa está ancorada em uma abordagem teórica, descritiva e qualitativa. Segundo Bartunek e Seo (2002), a abordagem qualitativa é útil e necessária para identificar e explorar os significados dos fenômenos estudados, possibilitando o desenvolvimento de novas compreensões e reflexões – no caso desse estudo, os metadados de preservação digital no contexto dos *Big Data*.

O trabalho está estruturado da seguinte forma para melhor entendimento: introdução, em que são apresentadas as considerações iniciais e os objetivos; seções: (1) os metadados e a preservação digital, seus conceitos e funções; (2) os metadados de preservação digital e *Big Data*: algumas reflexões; e considerações finais.

## **1 METADADOS DE PRESERVAÇÃO DIGITAL**

Os Metadados têm a função de descrever, identificar, e recuperar os objetos informacionais de acordo com seu conteúdo, contexto, estrutura ou apresentação, para que no futuro o usuário encontre a informação que deseja (ALVES, 2010).

Eles possuem a função de padronização e interoperabilidade dos sistemas de informação, como por exemplo, bibliotecas, mapas, redes sociais e qualquer outro meio, para que elas possam trocar e compartilhar informações. Se os metadados possuem essas funções, pode-se afirmar que estão relacionados à catalogação de recursos informacionais no meio digital, que descreve e representa os documentos.

Os metadados podem ser classificados em várias categorias, entretanto os mais frequentes são: os descritivos, estruturais e administrativos.

**Metadados descritivos:** é a face mais conhecida dos metadados, são eles que descrevem um recurso com o propósito de descoberta e identificação; podem incluir elementos tais como título, autor, resumo, palavras-chave e identificador persistente.

**Metadados estruturais:** são informações que documentam como os recursos complexos, compostos por vários elementos, devem ser recompostos e ordenados. Por exemplo, como as páginas de um livro, digitalizadas separadamente, são vinculadas entre si e ordenadas para formar um capítulo.

**Metadados administrativos:** fornecem informações que apoiam os processos de gestão do ciclo de vida dos recursos informacionais. Incluem, por exemplo, informações sobre como e quando o recurso foi criado e a razão da sua criação. Nessa categoria, estão metadados técnicos que explicitam as especificidades e dependências técnicas do recurso; inclui também os metadados voltados para apoio à gestão dos direitos relacionados ao recurso (SAYÃO, 2010, p. 5, grifo do autor).

Para Campos (2007, p. 30), “[...] os metadados tornam-se especialmente úteis em facilitar a navegação e o planejamento das consultas” de acordo com cada perfil diferente de usuário, para que, quando este for realizar a busca, encontre o que realmente deseja obter.

Os metadados são utilizados, por exemplo, em bibliotecas e arquivos, sendo que, nesses ambientes, possuem funções que vão além da identificação e descrição dos documentos, referindo-se também a várias seções.

Metadados referentes à administração, acesso, preservação e uso das coleções são constantemente empregados em repositórios. Isso é particularmente notável em bibliotecas e arquivos digitais. Nesses ambientes, os metadados não apenas descrevem e identificam um objeto informacional, mas explicitam as condições corretas ou ideais de seu gerenciamento, as relações do objeto com outros na coleção, sua função, utilização, comportamento, contexto de criação e condições de preservação (CAMPOS, 2007, p. 34-35).

Para não correr o risco de perder os materiais digitais deve-se utilizar *software*, *hardware*, e métodos de preservação com qualidade e que armazenem as informações e os dados de maneira adequada para permanecer em longo prazo, utilizando-se padrões de metadados.

A seguir, no Quadro 1, serão apresentados os principais padrões de metadados de preservação digital que foram apresentados na revisão de literatura por Formenton (2015), bem com a sua especificação e apresentação.

**Quadro 1-** Padrões de Metadados de Preservação Digital

Padrão de Metadados	Apresentação
ANSI/NISO Z39.87, <i>Data Dictionary – Technical Metadata for Digital Still Images</i> <sup>32</sup>	Padrão de metadados técnicos para desenvolver, intercambiar e interpretar arquivos de imagem fixa digital, facilitar a interoperabilidade entre sistemas, serviços e software, assim como apoiar a gestão a longo prazo e contínuo acesso a coleções de imagens digitais (BARBEDO <i>et al.</i> 2007; NATIONAL INFORMATION STANDARDS ORGANIZATION 2011)
<i>Dublin Core - DC</i>	Padrão para catalogação e suporte a descoberta de recursos eletrônicos na Web (CAMPOS; SARAMAGO, 2007, 2012; SAYÃO, 2010)

<i>Encoded Archival Description (EAD)</i>	Padrão para codificação de instrumentos arquivísticos de acesso, tais como inventário, índices, dentro outros (ANDRADE, 2007; BARBEDO <i>et al.</i> ; CHAN ZENG, 2006)
<i>Learning Object Metadata (LOM)</i>	Padrão que define um conjunto mínimo de atributos para gerir, localizar e avaliar recursos ou objetos de aprendizagem (CHAN; ZENG, 2006, NATIONAL INFORMATION STANDARDS ORGANIZATION, 2004)
<i>Machine Readable Cataloging (MARC)</i>	Padrão para a representação e a comunicação de registros ou informações bibliográficas entre bibliotecas, através da utilização de computadores (CHAN, ZENG, 2006; GRACIO, 2012; THOMAZ, 2004)
<i>Metadata Encoding and Transmission Standard (METS)</i>	Padrão para codificar metadados descritivos, administrativos e estruturais sobre objetos digitais num repositório (ALMEIDA; CENDÓN; SOUZA 2012, BARBEDO <i>et al.</i> 2007)
<i>Metadata Object Description Schema (MODS)</i>	Padrão de metadados descritivos derivado do esquema bibliográfico MARC 21, cujo enfoque são os recursos eletrônicos e os serviços de biblioteca. (NATIONAL INFORMATION STANDARDS ORGANIZATION, 2004)
<i>ISO/IEC Moving Picture Experts Group (MPEG) Multimedia</i>	Padrões para a descrição e a representação de recursos ou objetos audiovisuais e multimídiaicos digitais tais como MPEG-7 e MPEG-21 (NATIONAL INFORMATION STANDARDS ORGANIZATION, 2004; SAYÃO, 2010)
<i>PREservation Metadata Implementation Strategies (PREMIS)</i>	Padrão que define um conjunto de elementos básicos para codificar, armazenar, gerir e intercambiar metadados de preservação no contexto de um sistema de repositório para a preservação digital (ARELLANO, 2008; CAMPOS; SARAMAGO, 2007, SAYÃO, 2010)
<i>Record Export for Art and Cultural Heritage (REACH)</i>	Padrão para a descrição de objetos no âmbito de coleções de museus (GRACIO, 2012)

Fonte: (FORMENTON, 2015, p. 62)

Para entender o processo de preservação digital é necessário compreender primeiro o que é um objeto digital, sendo assim, “Um objeto digital é aquele que foi criado em computador, podendo ser original ou uma versão depois de haver sido convertido (ou digitalizado).” (MÁRDERO ARELLANO, 2004, p. 16). Para Hedstrom (1998, p. 190, tradução nossa), a preservação digital é “[...] o planejamento, alocação de recursos e aplicação de métodos e tecnologias para assegurar que a informação digital de valor contínuo permaneça acessível e utilizável”.

A preservação digital busca armazenar as informações e os dados de uma instituição para que permaneça sempre em acesso e uso contínuo, preserva também a memória e a cultura de uma determinada comunidade. “A preservação digital compreende os mecanismos que permitem o armazenamento em repositórios de dados digitais que garantiriam a perenidade dos seus conteúdos” (MÁRDERO ARELLANO, 2004, p. 17).

## **2 OS METADADOS DE PRESERVAÇÃO DIGITAL NA ERA DOS *BIG DATA*: ALGUMAS REFLEXÕES**

De modo a elucidar as reflexões dos metadados de preservação digital à luz dos *Big Data*, apresenta-se uma conceituação desse termo:

O Big Data é um termo derivado dos avanços recentes relativos à massificação da utilização de recursos tecnológicos e da farta produção de dados. Em suma, é um conceito que caracteriza volumosos conjuntos de dados heterogêneos, os quais não são passíveis de processamento por soluções computacionais tradicionais considerando dinamismo e sua complexidade (RAUTENBERG; CARMO, 2019, p 57).

Rautenberg e Carmo (2019), ao apresentarem e discutirem esse termo, dialogam com aportes teóricos da Ciência da Informação, no que diz respeito à otimização do acesso e à forma como os metadados se comportam nesse processo de representação da informação.

Os autores supracitados apresentam, ainda, alguns questionamentos sobre a aplicação dos dados e metadados, ancorando sua reflexão à luz dos metadados de preservação digital. São questionamentos dos autores: Como armazenar os dados e metadados em ecossistemas de *Big Data*? Como organizar e catalogar dados e metadados nesses ecossistemas? Ao colocar esses questionamentos diante do grande volume de dados e ancorá-los à luz dos metadados de preservação digital e de seu processo de armazenamento, deve-se pensar numa política dos padrões de metadados de preservação digital que discuta esse grande volume de dados.

Grácio (2012, p. 75) caracteriza os aspectos da preservação digital em três grupos: organizacional, legal e técnico que se subdividem:

**O Organizacional** em: missão, visão, objetivos institucionais; equipe multidisciplinar; responsabilidades; recursos financeiros e atos administrativos.

**O Legal** em: legislação vigente e direitos autorais;

**O Técnico** em: modelos, padrões, iniciativas; infraestrutura tecnológica; repositórios institucionais; seleção, descarte; estratégias de preservação; autenticidade; metadados e suporte digital.

A figura a seguir exemplifica melhor a relação entre esses grupos.

**Figura 1:** Aspectos da Preservação digital



**Fonte:** Grácio (2012, p. 76).

A prática de preservação digital é necessária em todas as unidades de informação para que os usuários tenham acesso *on-line* e para que a informação esteja sempre preservada e ativa, pois os documentos ou conteúdos que estão no suporte físico (as informações impressas, livros, periódicos, teses, dissertações, entre outros) com o passar do tempo podem se desintegrar ou perder, nunca mais sendo recuperadas, lembrando que, o usuário não pode ter acesso ao documento impresso a todo o momento e nem poderá pesquisar em qualquer lugar que esteja (ARAKAKI *et al.*, 2018).

## CONSIDERAÇÕES FINAIS

Nas discussões sobre *Big Data* no contexto documental, apresentamos um novo V que caracteriza os *Big Data* e a informação em ambientes digitais: a *volatilidade*. Neste sentido, consideramos os metadados de preservação digital como

uma ferramenta que facilitaria a preservação e melhor contextualização dos dados.

Os metadados de preservação digital assumem um papel primordial nas descrições e representações dos dados e da informação possibilitando a interoperabilidade entre sistemas. As políticas de preservação digital no contexto dos *Big Data* possuem grande importância global, pois têm o papel de garantir o armazenamento e a descrição não ambígua de informações, permitindo também uma melhor significação contextual dos *Big Data* em relação ao conhecimento acumulado.

## REFERÊNCIAS

ALMEIDA, M.B.; CENDÓN, B.V. & SOUZA, R.R. Metodologia para implantação de programas de preservação de documentos digitais a longo prazo. *Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação*, Florianópolis, v. 17, n. 34, p. 103-130, ago. 2012. ISSN 1518-2924. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2012v17n34p103>. Acesso em: 27/05/2019.

ALVES, R. C. V. *Metadados como elementos do processo de catalogação*. 2010. 132 f. Tese (Doutorado em Ciência da Informação)–Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Faculdade de Filosofia e Ciências, Marília, 2010.

ANDRADE, R. S. Aspectos introdutórios da representação de informação arquivística: a Norma Brasileira de Descrição Arquivística (Nobrade), a Descrição Arquivística Codificada (EAD-DTD) e o projeto Archives Hub. *Ponto de Acesso*, Salvador, v. 1, n. 2, p. 70-100, jul./dez. 2007. Disponível em: <http://www.brapci.inf.br/index.php/article/view/0000005065/7cd7f7403d73975ed2619d2d8310720/> . Acesso em: 23/05/2019.

ARAKAKI; F. et al. Web Semântica e Preservação digital: o padrão de metadados PREMIS na proposta do Liked Data. *Informação e Tecnologia*. v. 2, 2018. Disponível em: <http://www.periodicos.ufpb.br/ojs/index.php/itec/article/view/38118>. Acesso em: 22/05/2019.

- ARELLANO, M. Á. M. *Crerios para a preservação digital da informação científica*. 2008. 354 f. Tese (Doutorado em Ciência da Informação) – Departamento de Ciência da Informação e Documentação, Universidade de Brasília, Brasília, 2008. Disponível em: [http://repositorio.unb.br/bitstream/10482/1518/1/2008\\_MiguelAngelMarderoArellano.pdf](http://repositorio.unb.br/bitstream/10482/1518/1/2008_MiguelAngelMarderoArellano.pdf). Acesso em: 11/01/2015
- BARBEDO, F. *et al.* RODA: repositório de objectos digitais autênticos. *Actas dos Congressos Nacionais de Bibliotecários, Arquivistas e Documentalistas*, Lisboa, n. 9, 2007. Disponível em: <https://www.bad.pt/publicacoes/index.php/congressosbad/article/view/535>. Acesso em: 13/05/ 2019.
- BARTUNEK, J. M. & SEO, M. Qualitative research can add new meanings to quantitative research. *Journal of Organizational Behavior*, v. 23, n.2, 2002.
- CAMPOS, L. F. de B. Metadados digitais: revisão bibliográfica da evolução e tendências por meio de categorias funcionais. *Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação*, Florianópolis, v. 12, n. 23, p. 16-46, 1. sem. 2007. Disponível em: <http://www.brapci.ufpr.br/download.php?dd=11673>. Acesso em: 26/05/2017.
- CAMPOS, F. M. G.; SARAMAGO, M. L. Preservação digital de longo prazo em instituições patrimoniais: reutilização e adaptação de metadados. *Actas dos Congressos Nacionais de Bibliotecários, Arquivistas e Documentalistas*, Lisboa, n. 9, p. 1-7, 2007. Disponível em: <https://www.bad.pt/publicacoes/index.php/congressosbad/article/view/540>. Acesso em: 11/05/2019.
- CASTRO, F. F. de. *Elementos de interoperabilidade na catalogação descritiva: configurações contemporâneas para a modelagem de ambientes informacionais digitais*. 2012. 202 f. Tese (Doutorado em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2012.
- CHAN, L. M.; ZENG, M. L. Metadata interoperability and standardization: a study of methodology part i: achieving interoperability at the schema level. *D-Lib Magazine*, v. 12, n. 6, june 2006. Disponível em: <http://www.dlib.org/dlib/june06/chan/06chan.html>. Acesso em: 23/05/2019.

FOMERTON, Danilo. *Identificação de padrões de metadados de preservação digital*. Dissertação de Mestrado em Tecnologia, Informação e Sociedade: Universidade Federal de São Carlos - Ufscar – São Carlos, 2015.

GRÁCIO, J. C. A. *Preservação digital na gestão da informação: um modelo processual para as instituições de ensino superior*. São Paulo, SP: Cultura Acadêmica, 2012. Programa de Pós-Graduação em Ciência da Informação, Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília. Disponível em: <http://www.HYPERLINK> “[http://www.culturaacademica.com.br/\\_img/arquivos/Preservacao\\_digital\\_na\\_gestao\\_da\\_inform%20acao-WEB\\_v2.pdf](http://www.culturaacademica.com.br/_img/arquivos/Preservacao_digital_na_gestao_da_inform%20acao-WEB_v2.pdf)” w.culturaacademica.com.br/\_img/arquivos/Preservacao\_digital\_na\_gestao\_da\_inform HYPERLINK “[http://www.culturaacademica.com.br/\\_img/arquivos/Preservacao\\_digital\\_na\\_gestao\\_da\\_inform%20acao-WEB\\_v2.pdf](http://www.culturaacademica.com.br/_img/arquivos/Preservacao_digital_na_gestao_da_inform%20acao-WEB_v2.pdf)” acaoHYPERLINK “[http://www.culturaacademica.com.br/\\_img/arquivos/Preservacao\\_digital\\_na\\_gestao\\_da\\_inform%20acao-WEB\\_v2.pdf](http://www.culturaacademica.com.br/_img/arquivos/Preservacao_digital_na_gestao_da_inform%20acao-WEB_v2.pdf)”-WEB\_v2.pdf. Acesso em: 26/05/2019.

HEDSTROM, M. Digital preservation: a time bomb for digital libraries. *Computers and the Humanities*, Michigan, U.S.A, n. 31, p. 189–202, 1998. Disponível em: <http://deepblue.lib.umich.edu/bitstream/handle/2027.42/42573?sequence=1>. Acesso em: 24/05/2019.

MÁRDERO ARELLANO, M. Á. Preservação de documentos digitais. *Ciência da Informação*, Brasília, DF, v. 33, n. 2, p. 15-27, maio/ago. 2004. Disponível em: <http://www.brapci.ufpr.br/download.php?dd0=9026>. Acesso em: 24/05/2017.

MARTÍNEZ ÁVILA, D. *et al.* CLASSIFYING THE LOD CLOUD. *Brazilian Journal of Information Science: research trends*, v. 12, n. 4, p. 06-10, 12 dez. 2018. Disponível em: <http://www2.marilia.unesp.br/revistas/index.php/bjis/article/view/8328>. Acesso em: 29/05/2019

NATIONAL INFORMATION STANDARD ORGANIZATION. *Understanding Metadata*. Bethesda, MD: NISO Press, 2004. Disponível em: <https://www.niso.org/publications/understanding-metadata-2017>. Acesso em: 12/11/2018.

RAUTENBERG, S.; CARMO, P. R. BIG DATA E CIÊNCIA DE DADOS. *Brazilian Journal of Information Science: research trends*, v. 13, n. 1, p. 56-67, 29 mar. 2019. Disponível em: <http://www2.marilia.unesp.br/revistas/index.php/bjis/article/view/8315>. Acesso em: 29/05/2015.

SAYÃO, L. F. Uma outra face dos metadados: informações para a gestão da preservação digital. *Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação*, Florianópolis, v. 15, n. 30, p. 1-31, 2010. Disponível em: <http://www.brapci.ufpr.br/download.php?dd0=12621>. Acesso em: 18/05/2019.

THOMAZ, K. P. *A preservação de documentos eletrônicos de caráter arquivístico: novos desafios, velhos problemas*. 2004. 388 f. Tese (Doutorado em Ciência da Informação) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2004. Disponível em: <http://www.bibliotecadigital.ufmg.br/dspace/handle/1843/VALA-68ZRKF>. Acesso em: 11/01/2019.

# CURADORIA DIGITAL E REPRESENTAÇÃO DE ACERVO MUSEOLÓGICO NO ATOM: UM ESTUDO DA CUSTOMIZAÇÃO DO *SOFTWARE*

*Nandia Leticia Freitas Rodrigues*

*Lucinéia da Silva Batista*

*Lais Alpi Landim*

*Maria José Vicentin Jorente*

Os avanços nas Tecnologias de Informação e Comunicação (TIC) de meados dos anos 1990 têm condicionado uma complexa transformação na cultura ocidental, impactando significativamente o comportamento dos sujeitos nas formas de buscar, adquirir e produzir conhecimento (JORENTE, 2017). Na era dos *Big Data* – caracterizada pelo grande volume de dados variados, produzidos em alta velocidade, com expressiva variabilidade e passíveis de atribuição de valor aos dados no ambiente *Web* – a acumulação e produção exponencial de conhecimento social a partir dos sistemas computacionais complexos evidencia a urgência em se desenvolver e disponibilizar representações das informações estruturadas. Assim, o pleno acesso nos ambientes *Web* é facilitado, de modo a atri-

morar o compartilhamento da informação. Nesse contexto, a descrição de acervos musealizados é essencial no processo de tratamento e gerenciamento documental.

O conceito de Big Data pode ser compreendido como conjuntos de dados tão complexos e de volume tão extensos que aplicações tradicionais de processamento de dados são inadequadas para o seu tratamento. Suas características o distinguem de esforços clássicos de coleta de dados, principalmente quanto a tudo que é possibilitado pelo Big Data. Os dados, nesse contexto, advêm da reunião de múltiplas fontes, que já reuniram outros conjuntos de dados para diferentes propósitos (BATES *et al.*, 2018).

A natureza dos dados no contexto do Big Data é caracterizada pelos desafios de volume, veracidade, variedade, velocidade e valor – os 5 Vs. Nesse contexto, alguns dos maiores desafios encontram-se na ausência de estruturação dos dados – a atribuição de metadados de descrição que possibilitam o seu processamento por sistemas automatizados (JAMES, 2014).

No escopo da Ciência da Informação (CI), discutem-se estratégias e métodos para o armazenamento, a preservação, os processos de digitalização de acervos e as novas formas de organizar e disponibilizar informação em sistemas digitais: as questões que giram em torno da preservação e do acesso são desafios constantes para a CI. A Museologia encontra, então, novas perspectivas no contexto da sua conformação, bem como a possibilidade de adoção dos espaços híbridos para divulgação e compartilhamento dos acervos musealizados. É premente reinventar a comunicação museológica, sobretudo a partir da adesão às inúmeras possibilidades de inovação advindas dos avanços das TIC. Na era dos dispositivos tecnológicos, é essencial discutir questões emer-

gentes, tais como as demandas na laboração de novos procedimentos específicos para tratamentos, curadoria, preservação e descrição acervos nos ambientes digitais, para pensar os rumos da Museologia contemporânea.

O ciberespaço modifica a forma como os museus lidam com seus acervos – o acesso à informação, à obra de arte e ao objeto musealizado transcende o espaço físico do museu (BEIGUELMAN, 2017). A atuação tradicional dos museus, centrada na territorialidade e nos acervos físicos, não é mais suficiente para alcançar com total eficiência o seu público-alvo que, embora muito heterogêneo, encontra-se cada vez mais inserido num contexto comum, isto é, nos ambientes digitais *online*. Neste sentido, considerando que a relação entre o museu, o acervo e o público sofre interferências significativas da cultura digital, é necessário adotar estratégias e metodologias que viabilizem a representação dos conteúdos informacionais das dos museus também nos ambientes digitais. A disponibilização da imagem, seguida da descrição textual dos objetos, é uma eficiente estratégia de hibridização nos ambientes digitais de museus na busca pela aproximação do público através dos ambientes digitais.

A descrição de objetos museais e a sua disponibilização no ambiente digital possibilita a representação das informações e facilita o acesso. Porém, esta é uma atividade que demanda a adoção de uma série de procedimentos, normas, estruturas, padrões e metadados adequados, e de um *software* específico para realização do tratamento descritivo do objeto. A adoção adequada destes elementos influencia no sucesso da representação da informação, na sua recuperação, interoperabilidade e importação de dados entre sistemas. Agrega-se,

assim, valor à informação digital, o que a torna mais adequada ao contexto do Big Data.

Trataremos, neste estudo, das possibilidades da utilização do *software* de descrição arquivística AtoM na representação da informação museológica por meio da descrição. Acreditamos que o AtoM possa ser aplicado nesse processo de tratamento descritivo, visto que se trata de um sistema *open source* que possibilita convergências de normas, interoperabilidade, descrições multiníveis, multilíngues e que possui *design* simples, eficiente e eficaz para a representação da informação. Considerando a heterogeneidade dos acervos de museu e a sua semelhança com os acervos arquivísticos, acreditamos que os campos descritivos do AtoM atendam à necessidade de representação da informação dos museus.

## 1 MUSEUS NA ERA DIGITAL

Os museus, tradicionalmente, são os “guardiões de nossa herança cultural” (DEKKER, 2014, p. 66). Enquanto instituições culturais, os museus são promissores espaços de informação, cultura e lazer, atuantes no processo educacional, formador de conhecimento sociocultural e no exercício de direitos do cidadão (ROCHA, 2017).

Segundo definição do Conselho Internacional de Museu (ICOM):

O museu é uma instituição permanente sem fins lucrativos, ao serviço da sociedade e do seu desenvolvimento, aberta ao público, que adquire, conserva, investiga, comunica e expõe o patrimônio material e imaterial da humanidade e do seu meio envolvente com fins de educação, estudo e deleite (ICOM, 2015, [online]).

As instituições museológicas – por meio de narrativas expositivas, atividades e ações culturais voltadas aos sujeitos que com elas interagem – desempenham uma função mediadora entre a informação histórica e cultural e a sociedade (LOUREIRO; SOUZA; SAMPAIO, 2007). A ideia tradicional de museu modificou-se com o passar do tempo, bem como suas práticas museológicas, seus discursos e seus públicos. Os museus tornaram-se espaços informacionais progressivamente híbridos, o que possibilita que as informações museológicas sejam produzidas e compartilhadas tanto no ambiente físico, quanto em ambiente dígito/virtual. Assim, a articulação e a coexistência entre o conteúdo informacional físico e o digital no âmbito do museu reinventam as formas de comunicação cultural (CASTELLS, 2011).

O acesso aos conteúdos informacionais dos acervos dos museus na contemporaneidade vai além dos muros das instituições. Há pouco menos de 50 anos, por exemplo, somente havia duas formas possíveis de se ter contato com uma obra de museu: ou a pessoa se deslocava até o museu físico, ou visualizava a representação da obra por meio das fotografias em livros, revistas, e manuais (CARMO; FOFONCA, 2012).

Além das fronteiras geográficas, havia também o contexto das fronteiras socioeconômicas e culturais que restringia o acesso a uma parcela da população. A interação com as obras era dificultada. No entanto, com a popularização da Internet, e nela a *Web* em seu estágio de *Web 2.0*, modificou-se essa realidade: não é mais impreterível o deslocamento até o Museu do Louvre em Paris para contemplar um simulacro da obra “Mona Lisa”, de Leonardo da Vinci. Basta o acesso a um dispositivo móvel conectado à Internet para acessar o ambiente digital do museu e, a partir de poucos cliques, a representação da obra.

Certamente, a experiência sensorial e cognitiva na contemplação de uma obra no museu *in loco* é bem distinta da contemplação de uma obra no ambiente digital. Suprir as necessidades informacionais do internauta por intermédio de uma interface de interação com eficiência é um grande desafio contemporâneo da área. Salienta-se que não se trata de um anseio de substituir a visita do museu físico pela visita virtual, mas sim de uma estratégia de aproximação do público com o patrimônio, por meio do acesso remoto aos ambientes, já que a maioria dos indivíduos dispõe de aparatos portáteis de conexão a eles, e encontra-se inserida na cultura digital.

Ademais, para além das polêmicas no tocante à experiência subjetiva do indivíduo em contato com o objeto musealizado, independentemente da ambiência, é importante analisarmos o impacto das experiências dos museus virtuais na sociedade; principalmente se considerarmos que em países como o Brasil ainda há “restrição do acesso a instituições culturais a uma parcela da sociedade, coexistindo uma barreira entre a sociedade e o patrimônio” (ROCHA, 2017, p. 2). Nota-se no acesso *online* uma possibilidade de visita e conhecimento de acervos de museus locais e globais, especialmente àquelas pessoas que não frequentam os museus, seja pela barreira histórico-social, pela falta de adequação de dias e horários acessíveis à população que trabalha, pelos valores dos ingressos, por questões de acessibilidade, ou pela falta de divulgação, dentre outros.

Assim, a ampla utilização dos recursos tecnológicos desencadeia um processo de desmaterialização e desterritorialização dos acervos, estabelecendo novas relações entre o público e o museu (LUPO, 2016). “[...] As grandes redes de informação e comunicação, com seus fluxos invisíveis, imateriais,

as quais formam territórios abstratos que escapam às antigas territorialidades” (MATTELART, 1999, p. 166).

No intuito de propiciar uma experiência satisfatória e privilegiada do internauta no acesso às representações imagéticas (estáticas ou dinâmicas) dos acervos de museu no ambiente digital, é importante que, além da disponibilização das imagens dos objetos, sejam disponibilizadas também para o acesso, tanto do internauta quanto de sistemas, descrições ricas em detalhes de cada item do acervo, complementando o compartilhamento da mensagem pretendida com a sua representação. Para isso, é desejável que estas informações sejam disponibilizadas de forma estruturada.

Embora a área da museologia tenha desenvolvido recentemente *softwares* para a catalogação de acervos, tais como o *Collective Access* (internacional), o Donato/SIMBA e a Tainacan (nacionais) – *softwares* livres para descrição de objetos museais –, não há, entre eles, as características encontradas no AtoM, um *software* criado com o suporte do *International Council on Archives (ICA)*, a partir de normas e padrões para representação informacional, caracterizado por ser gratuito, livre, *open source*, com acesso via *Web*, multilíngue, multi-repositório, customizável, beta perpétuo, interoperável, com importação e exportação em Dublin Core, EAD, SKOS, EAC. Assim, acredita-se que este sistema proporcione uma descrição mais completa, além de facilitar a interoperabilidade com outros sistemas.

A utilização do AtoM nos museus proporciona a convergência da representação informacional e imagética dos objetos no ambiente digital, pois é interoperável com *softwares* como o *Archivematica*, para a preservação e acesso à informação, o que torna o acesso efetivo, eficiente e eficaz.

## 2 DESCRIÇÃO DE ACERVOS DE MUSEUS NA *WEB 2.0*

O termo descrição refere-se a um conjunto de procedimentos técnicos, utilizados com o intuito de construir a representação de um documento ou objeto a partir da identificação dos principais elementos representativos – metadados – do item a ser descrito. Trata-se de uma técnica muito utilizada no campo da Arquivologia, por meio da qual se realiza a descrição documentária com a finalidade de criar instrumentos de pesquisa que possibilitem a identificação, localização, recuperação e a utilização dos documentos (BELLOTO, 2006). No campo da Museologia, a descrição do objeto musealizado é uma atividade que faz parte da catalogação.

Os metadados – dados ou informações sobre dados – são descritores representativos, capturados de um determinado recurso informacional para a construção da descrição das informações semânticas e sintáticas de um item, elementos essenciais para o entendimento do objeto representado em um sistema de informação e para a sua posterior recuperação (LIMA, SANTOS, SANTARÉM SEGUNDO, 2016). Embora o termo metadado seja “[...] usado frequentemente para referir-se às informações legíveis por máquinas [...] é usado **também** para se referir a qualquer esquema de descrição de recurso, aplicado a qualquer tipo de objeto, sendo ele digital ou não” (LIMA, SANTOS, SANTARÉM SEGUNDO, 2016, p. 52, grifo nosso). A atribuição de metadados é essencial para favorecer a recuperação de dados digitais no contexto do Big Data.

A atividade de descrição de objetos de museus tem sido tradicionalmente realizada sobre suportes de fichas catalográficas ou fichas de registro. A descrição das obras de museu, realizadas em tais fichas, feitas manualmente e em papel, segundo Magalhães (2014), produz uma “documentação historicizada”

sobre as obras, construída a partir de um procedimento meramente técnico em que se registram as informações mínimas e com os dados técnicos “permanentes” da obra. Com o advento da computação, as fichas em papel, na grande maioria das instituições museológicas, foram substituídas por formatos digitais e, além disso, uma gama significativa de informações adicionais sobre o objeto passou a ser inserida nas atuais fichas catalográficas durante o processo de descrição, visando diversas funcionalidades dos sistemas eletrônicos de informação e funções derivadas das novas concepções da Curadoria Digital (CD).

Segundo orientações de Costa (2006), no guia de consulta “Princípios Básicos da Museologia”, os elementos a serem descritos são: nome da instituição, número de registro, categoria, nome do objeto, título, autor, época/data/período, estilo, procedência, origem, material/técnica, fabricação, dimensões, inscrições, marcas, assinatura, forma de aquisição, valor, estado de conservação, restauração, descrição do objeto, dados bibliográficos, referências bibliográficas, exposições, observações, e fotografias dos objetos.

Na Museologia, muito além de elencar meramente as referências técnicas e básicas das obras, no processo de descrição é necessário identificar e evidenciar na representação o máximo de informações possível referentes à obra, atentando-se às formas, estilo, e detalhes físicos. No processo de análise para construir a descrição, deve-se “decompor o objeto” e analisá-lo de forma minuciosa, considerando sua tridimensionalidade com vistas a construir sua representação fidedigna.

A adoção de normas e padrões de metadados no processo de descrição de objetos de museu é, portanto, fundamental para construir uma representação uniforme do acervo no setor documental dos museus. Nesse cenário, a escolha do *software* de descrição é igualmente importante, haja vista que

os campos do sistema escolhido devem ser compatíveis às normas de descrição para a inserção dos metadados.

## 2.1 A ADAPTAÇÃO DO *SOFTWARE* AtoM AO SETOR DOCUMENTAL DE MUSEUS

O AtoM é um sistema de representação da informação documental via descrição arquivística, criado para promover um ambiente de acesso com base em normas arquivísticas internacionais. Sua sigla “AtoM” (*Acess to Memory*) descende do ICA-AtoM – *International Council on Archives* (ICA). É um sistema voltado à documentação arquivística que propicia a descrição de acervos memoriais. Um dos seus objetivos é possibilitar a representação da informação. É um sistema de descrição que viabiliza o acesso e o compartilhamento da informação, a padronização das representações, a descrição padronizada dos objetos, além de ser interoperável entre sistemas, *open source*, livre, gratuito, em beta perpétuo e multilíngue.

Os modelos descritivos do AtoM baseiam-se nas normas arquivísticas ISAD(G), ISAAR(CPF) e outras, e possui estrutura multinível e hierárquica. Pelas características semelhantes aos arquivos que os setores documentais dos museus possuem, consideramos a possibilidade de aplicação do AtoM para a descrição e representação dos acervos musealizados e sua disponibilização em rede. Os campos de descrição do AtoM incluem:

**Zona de identificação:** identificador, título, data, nível de descrição; **Zona de contexto:** nome de produtor, entidade detentora, história do arquivo, fonte imediata de aquisição ou transferência; **Zona de conteúdo e estrutura:** âmbito e conteúdo, avaliação, seleção, eliminação, ingressos adicionais, sistema de organização; **Zona de condição de acesso**

**e utilização:** condições de acesso, condições de reprodução, idioma do material, *script* do material, notas ao idioma e *script*, características físicas e requisitos técnicos e instrumento de descrição; **Zona de documentação associada:** existência e localização de originais, existência e localização de cópias, unidades de descrição relacionadas, notas de publicação; **Zona de notas:** notas; **Ponto de acesso:** notas, assuntos, locais, gênero e assuntos; [autores]; **Zona de controle de descrição:** identificador da descrição, identificador da instituição, regras e convenções, estatuto, nível de detalhe, data de criação, revisão e eliminação, línguas e escritas, *script(s)*, fontes e notas do arquivista; **Zona de administração:** língua original, esquema padrão de exibição (BATISTA, 2018, p. 105, adaptado).

### 3 A REPRESENTAÇÃO DE ACERVOS EM AMBIENTES DIGITAIS E AS POSSIBILIDADES DO ATOM NESSES AMBIENTES

Analizamos, nesse estudo, ambientes digitais de quatro museus brasileiros, custodiadores de tipos diferentes de acervos, a fim de compreender como os acervos dessas instituições vêm sendo representados em tais ambientes. Traçamos, assim, um cenário laboratorial simulado para as possibilidades da utilização *software* AtoM utilizado na representação dos acervos musealizados em ambientes digitais. Exploramos nesse intuito os ambientes digitais do Museu da Imagem e do Som (MIS), do Museu Afro Brasil (MAB), do Museu de Arte de São Paulo (MASP) e do Museu do Futebol (MF).

Na fase exploratória de nosso estudo nos ambientes digitais dos citados museus, ao abrirmos a seção “Acervo” ou “Acervo digital”, encontramos as representações imagéticas e textuais dos acervos disponibilizadas e identificamos quais os elementos descritores utilizados para descrever os acervos em

cada um dos ambientes eleitos para nossa exploração laboratorial. Realizamos, posteriormente, uma comparação entre os elementos descritivos utilizados pelos museus para representar seus acervos e os campos descritivos – equivalentes – do AtoM, como pode ser observado no quadro 1:

**Quadro 1** – Museus e descritores de acervo

MUSEUS	DESCRITORES	Descritores equivalentes no AtoM (modelo ISAD (G))	Observações
MIS (objeto tridimensional)	Título	Título	
	Tipo	Âmbito e conteúdo – zona de conteúdo e estrutura	
	Número do item	Código de referência – área de identificação	
	Número de registro	Identificadores alternativos – área de identificação	
	Modelo	Âmbito e conteúdo - zona de conteúdo e estrutura	
	Coleção pertencente	Nível de descrição – área de identificação	
	Marca do objeto	Âmbito e conteúdo - zona de conteúdo e estrutura	
	Fabricante	Âmbito e conteúdo - zona de conteúdo e estrutura	Pode ser utilizado também o Ponto de acesso de autor, especificando em notas de âmbito (modelo descritivo de assunto)
	Local de fabricação	Ponto de acesso/Locais	
	Data de fabricação	Datas	
	Descrição dos componentes	Âmbito e conteúdo	
	Descrição das dimensões físicas	Dimensão e suporte – área de identificação	
	Descritores (palavras-chave)	Ponto de acesso/ assunto	
	Uso e acesso	Zona de condições de uso e acesso	
	Informações complementares	Zona de notas	

Informação, Conhecimento, Ação Autônoma e Big Data

MAB (objeto tridimensional)	Acervo	Entidade detentora – zona de contexto	
	Coleção	Nível de descrição – área de identificação	
	Número de registro	Identificadores al- ternativos – área de identificação	
MAB (objeto tridimensional)	Autoria/ grupo cultural	Nome do produtor – zona de contexto	
	Título/ denomi- nação	Título – área de iden- tificação	
	Data	Datas – área de iden- tificação	
	Técnica	Dimensão e suporte – área de identificação	
	Dimensões	Dimensão e suporte – área de identificação	
	Origem	Ponto de acesso/Locais	Pode ser usado o cam- po âmbito e conteúdo
	Categoria/ Tipologia	Nível de descrição – área de identificação	
	Função	Âmbito e conteúdo - zona de conteúdo e estrutura	
	Informações de contexto	Notas – Zona de notas	
	Palavras-chave	Ponto de acesso/ assunto	
	Estado de conservação.	Características físicas e requisitos técnicos – zona de condições de acesso e utilização	Também pode ser uti- lizada a zona de notas

MASP (objeto tridimensional)	Autor	Nome do produtor – zona de contexto	
	Dados biográficos	História – área de descrição	(modelo ISAAR-CPF)
	Título	Título	
	Data da obra	Datas – área de identificação	
	Técnica, dimensões	Dimensão e suporte – área de identificação	
	Aquisição	Fonte imediata de aquisição ou transferência – zona de contexto	
	Designação	Ponto de acesso de gênero	
	Número de inventário	Identificadores alternativos – área de identificação	
	Créditos da fotografia	Notas do arquivista – zona de controle da descrição	
MF (descrição de acervo fotográfico).	Título	Título – área de identificação	
	Tipo	Dimensão e suporte – área de identificação	
	Local de produção	Ponto de acesso/locais	
	Data	Datas – área de identificação	
	Suporte e cromia	Âmbito e conteúdo – zona de conteúdo e estrutura	Também pode ser utilizado o campo dimensão e suporte – área de identificação.

**Fonte:** Autoras, 2018.

A partir da análise proposta, destacamos a ausência de padronização para a representação das informações entre as quatro instituições representadas em seus ambientes digitais. Isso acarreta uma indesejável heterogeneidade na qualidade informativa entre os ambientes explorados, de modo que algumas delas são pouco detalhadas sobre os elementos visuais característicos do objeto, tais como a descrição física, as técnicas utilizadas, dentre outros. Conseqüentemente, áreas essenciais

de descrição como título, data, autoria, dimensões e número de registro são elementos básicos presentes em 3 das 4 descrições analisadas; Por outro lado, com relação à disponibilização das imagens dos objetos, embora os quatro museus cumpram esse requisito, a qualidade na pixelização das imagens é variável: verificamos uma notável diferença de qualidade entre as apresentações das representações.

O MAB disponibiliza a informação sobre a instituição detentora em suas descrições; promove a descrição exaustiva e rica em detalhes de cada peça, com sua contextualização histórica, grupo cultural a qual o objeto pertencia originalmente, sua função utilitária anteriormente a musealização, e palavras-chave; também disponibiliza mais de uma imagem de cada peça. O MASP, por outro lado, disponibiliza uma imagem grande de cada peça descrita, em alta resolução e passível de *download* e, além de representação por imagem e descrição textual, conta com o recurso de descrição em áudio de algumas imagens, trazendo informações sobre o autor, contextualização e técnicas utilizadas na criação da obra. Os outros ambientes digitais analisados apresentam-se mais pobremente.

No que diz respeito a nossa proposta de utilização do *software* AtoM em ambientes museais digitais, constatamos, a partir do estudo realizado, que ele possui significativas possibilidades de aplicação para a representação e recuperação da informação dos objetos musealizados, a partir dos vários níveis descritivos existentes no sistema e suas muitas funcionalidades, respondendo de forma satisfatória às necessidades descritivas de acervos musealizados compostos por objetos tridimensionais e heterogêneos.

Além de atender com eficiência e eficácia a todos os elementos descritivos utilizados para a representação dos obje-

tos nas instituições museológicas analisadas, observamos que o sistema amplia a variedade de campos descritivos utilizados atualmente pelas instituições, tais como: nome de produtor (autor), fonte imediata de aquisição ou transferência, unidades de descrição relacionadas, condições de acesso e uso, dentre outros. A plena utilização desses campos, a partir do processo de descrição minuciosa dos objetos, enriquece a descrição museológica, possibilitando uma representação da informação musealizada mais semântica e adequada a um futuro emergente de interação e interoperatividade entre as instituições dos sistemas memoriais no contexto do Big Data.

Se a descrição dos objetos de museu é fundamental para a representação, recuperação, uso e compartilhamento da informação dos seus acervos, sua disponibilização nos sistemas digitais, sobretudo na *Web 2.0*, potencializa o alcance e a resolução das necessidades informacionais dos cidadãos que buscam por informações de uma maneira global e integrada.

## CONSIDERAÇÕES

A adoção do AtoM para a descrição de acervo musealizado pode solucionar o problema da heterogeneidade pontuado em nosso empreendimento, considerando a padronização dos metadados, a exaustividade das descrições e a compatibilidade e possível adequação dos elementos da diversidade compreendida nos objetos de museus aos também diversos campos descritivos do AtoM. O *software* AtoM possibilita o registro informacional de objetos de museu de forma eficiente, podendo representar acervos de maneira satisfatória no ambiente digital. Além disso, apresenta muitas funcionalidades que o tornam flexível na sua adaptação ao contexto museológico para responder a possíveis emergências.

Evidenciou-se, no entanto, que para a realização da descrição, o profissional em tela necessita de competências em leitura visual do objeto, oriunda da capacitação em leitura imagética e também da percepção, tempo, técnica e competência para perceber ser necessária extensiva pesquisa sobre a obra e o seu contexto histórico. A heterogeneidade das descrições demonstra, portanto, que para o profissional da informação, não basta o conhecimento das técnicas de descrição ou de *softwares* que cumpram um papel operacional no acesso aos acervos custodiados.

Frente aos desafios da representação da informação museológica nos ambientes digitais tendo como objetivo propiciar a sua recuperação adequada no contexto do Big Data, a Curadoria Digital destes ambientes deve envolver a escolha de *softwares* de descrição e organização de metadados como o AtoM. O AtoM favorece o enriquecimento dos dados digitais a partir da atribuição de metadados padronizados que tornam possível o compartilhamento, o uso e o reuso da informação em ambientes digitais de museus de forma mais adequada ao novo paradigma do Big Data.

## REFERÊNCIAS

BATES, D. W. *et al.* Why policymakers should care about “big data” in healthcare. *Health Policy and Technology*, [s. l.], 2018. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S2211883718300996>. Acesso em: 10/06/2018.

BATISTA, L. S. *O Redesign do sistema Access to Memory (AtoM) para a curadoria digital de acervos museológicos heterogêneos*. 2018. 175 f. Dissertação (Mestrado em Ciência da Informação) – Universidade Estadual Paulista, Faculdade de Filosofia e Ciências, Marília, 2018.

- BEIGUELMAN, G. Museus do inacabado para memórias efêmeras: notas sobre a conservação de obras de net art. *Museologia & Interdisciplinaridade*, v. 6, n. 12, jul./dez. 2017.
- BELLOTTO, H. L. *Arquivos permanentes: tratamento documental*. 4. ed. Rio de Janeiro: FGV, 2006.
- CARMO, A. C. R.; FOFONCA, E. Arte na sociedade híbrida: a experiência dos museus virtuais. In: SIMPÓSIO NACIONAL DE HISTORIA CULTURAL: escritas da história: ver-sentir-narrar, 6., 2012, Teresina. *Anais...* Teresina, UFPI: GT Nacional de História Cultural, 2012.
- CASTELLS, M. Os museus na era da informação, conectores culturais de tempo e espaço. Tradução de Claudia Storino. *MUSAS revista brasileira de museu e museologia*, ano VII, n. 5, 2011, p. 8-21. Brasília: Instituto Brasileiro de Museus, 2011.
- COSTA, E. P. (Org.). *Princípios básicos da museologia*. Curitiba: Coordenação do Sistema Estadual de Museus; Secretaria de Estado da Cultura, 2006.
- DEKKER, A. Os desafios da conservação da net art: a aceitação da perda da subjetividade. In: BEIGUELMAN, G.; MAGALHÃES, A. G. *Futuros possíveis: arte, museus e arquivos digitais*. São Paulo: Peirópolis: Edusp, 2014.
- ICOM. Definição de Museu. 2015. Disponível em: <http://icom-portugal.org/2015/03/19/definicao-museu/>. Acesso em: 31/08/2018.
- JAMES, R. Out of the box: Big data needs the information profession - the importance of validation. *Business Information Review*, 31(2), 2014, 118–121.
- JORENTE, M. J. V. Relações sistêmicas entre a Teoria da Complexidade, o Design da Informação e a Ciência da Informação na Pós-Modernidade. In: ALVES, M. A.; GRÁCIO, M. C. C.; MARTÍNEZ-ÁVILA, D. (Org.). *Informação, conhecimento e modelos*. Campinas: UNICAMP, Centro de Lógica, Epistemologia e História da Ciência; Cultura Acadêmica, v. 78, 2017, p. 181-198.
- LIMA, F. R. B.; SANTOS, P. L. V. C.; SANTARÉM SEGUNDO, J. E. Padrão de metadados no domínio museológico. *Perspectivas em ciência da informação*, v. 21, n. 3, 2016.

- LOUREIRO, J. M. M.; SOUZA, D. M. V. de; SAMPAIO, A. C. O. Museus de história natural e a construção da 'nação'. In: ENCONTRO DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 8., 2007, Salvador. *Anais...* Salvador: ENANCIB, 2007. P. Não paginado.
- LUPO, B. Tecnologia, materialidade e espacialidade no museu contemporâneo. VENTURELLI, S; ROCHA, C. (Org.). In: ENCONTRO INTERNACIONAL DE ARTE E TECNOLOGIA, 15., 2016, Brasília. *Anais...* Brasília: Universidade de Brasília 2016.
- MAGALHÃES, A. G. Considerações para uma análise histórico-crítica da catalogação de acervos artísticos. *Futuros possíveis: arte, museus e arquivos digitais= Possible futures: art, museums and digital archives. 1ed. São Paulo: EDUSP/ED. PEIRÓPOLIS LTDA, 2014, 1: 34-44.*
- MATTELART; A. e M. *História das teorias da comunicação*. Tradução de Luiz Paulo Rouanet. São Paulo: Loyola, 1999.
- MUCHACHO, R. Museus virtuais: A importância da usabilidade na mediação entre o público e o objecto museológico. In: CONGRESSO DA ASSOCIAÇÃO PORTUGUESA DE CIÊNCIAS DA COMUNICAÇÃO (SOPCOM), 4., 2005, Aveiro, Portugal. *Livro de Actas...* Portugal: FLUP, 2005.
- ROCHA, B. F. R. Conhecendo os museus virtuais e cibermuseus: aplicativo 'Fala Sério'. *Revista Museologia & Interdisciplinaridade*, v. 6, n. 11, p. 269-291, 2017.

**IMPACTOS DE *BIG DATA* NA  
CIÊNCIA E NA INDÚSTRIA 4.0**

REDES SOCIAIS DIGITAIS E A ESFERA  
PÚBLICA:  
“FAKE NEWS” E A MANIPULAÇÃO DA  
OPINIÃO COLETIVA

*Magaly Prado*

*A imagem pública substitui o espaço público com a colonização publicitária – o espaço onde se dava a comunicação social, as avenidas, as praças públicas foram substituídos pela teleimagem das máquinas de visão, capazes de ver e de perceber em nosso lugar.*

(VIRILIO)

Ao considerar o tema das *fake news*, excluindo a parte histórica de que elas sempre existiram, priorizaremos neste capítulo o suceder dos tempos na trajetória da internet, com a eclosão das redes sociais e dos mensageiros instantâneos, da atroz “recentidade” que põe em risco a credibilidade do jornalismo – já padecendo da perda de audiência para o espaço numérico. A nossa proposta é discutir a popularização do termo – tornado politicóide por Donald Trump, quando passa a

usá-lo para rebater, atacar e insultar a imprensa *mainstream*<sup>1</sup>, com a qual não concorda, após ganhar as eleições de 2016 nos EUA – tidas (não oficialmente) como manipuladas por ações de marketing<sup>2</sup> político ou a serviço de interesses particulares. Apesar do estratagema de campanha para persuadir os britânicos nas negociações do plebiscito do Brexit, para a saída do bloco europeu, o ressoo maior foi com a raposice de Trump.

Sobre Trump, Latour (2018) vilipendia “um presidente que inventa os fatos para se adequar ao seu humor e busca a credibilidade de qualquer um que o contradiga, parece representar a culminação dessa podridão epistêmica”, em entrevista ao *The New York Times Magazine*.

Na tentativa de embaralhar a noticiabilidade, *fake news* alcançam espalhamento rápido com a força do crescimento da republicação pelo público<sup>3</sup>, que acata *fake news* como verdade, impulsionando, assim, os algoritmos<sup>4</sup> das redes, que passam a intensificar a visibilidade da desordem da informação de maneira progressiva e, conseqüentemente, a viralizar de forma espaventosa, confluindo no revigoramento dos filtros-bolha de informação e na cultura com a polarização de opiniões,

---

<sup>1</sup> *New York Times*, *Washington Post*, *Wall Street Journal*, CNN e a BBC. “Trump os despreza. Cunhou o termo *fake news* para referir-se às suas matérias exclusivas e os declarou ‘inimigos do povo’” (El País).

<sup>2</sup> “O marketing e a propaganda usam os algoritmos de IA para identificar os hábitos e as preferências dos consumidores e produzir campanhas mais assertivas e segmentadas” (KAUFMAN, 2018, p.45).

<sup>3</sup> “Trump sabe que as pessoas que votaram nele são desconfiadas e preferem obter suas informações da [rede de TV] Fox e de outras fontes da direita. Atacando a grande mídia, se reconecta com sua base. A polarização nutre a desinformação. Seus eleitores querem acreditar que é digno e capaz. E quando se publicam informações que sugerem que não é, rejeitam aceitá-las e preferem considerar que são *fake news*”, afirma Shanto Iyengar, professor de Ciência Política de Stanford (El País).

<sup>4</sup> Os algoritmos acoplados aos *Big Data* dão azo a uma situação na qual “um número finito de componentes gera uma diversidade praticamente ilimitada de combinações” (DELEUZE, 2004 *apud* CASTRO, 2018).

sobretudo políticas, culminando na destilação desenfreada de ódio de cada polo.

Outrora, porém, existia um otimismo generalizado devido ao surgimento da internet. Como conta Parisier: “a grande esperança era de se estabelecer um meio no qual comunidades, e até mesmo países inteiros, pudessem cooperar para estabelecer uma cultura através da comunicação” (PARISIER, 2011, p. 164). Mas, em vez disso, “a personalização levou a algo diferente: uma esfera pública manipulada e organizada por fórmulas, segregada pelo design e contrária ao diálogo”. A opinião pública coletiva, influenciada pelo conteúdo fragmentado das mídias sociais, deslegitima a imprensa e enaltece a desinformação noticiosa ou propagandística.

Antes de mais nada, a expressão *fake news* não pode ser traduzida ao pé da letra, porque ela abrange diversas categorias; dentre elas: notícias fraudulentas ou frágeis; informação falsa (em geral com fontes forjadas), manipulada, adulterada ou fabricada (com a intenção de enganar); desinformação (criada para prejudicar) ou má informação (sem apuração ou mal apurada ou mesmo usando a verdade, muitas vezes fora de contexto, para causar danos; notícias antigas requentadas; sensacionalismo (próprio dos tabloides); mentiras, maquiagens, boatos, fatos alternativos etc. – todas ameaçam a qualidade do jornalismo e, por conseguinte, a formação da opinião coletiva.

Problemas gerados por *fake news* fazem parte do que podemos entender por cibercultura, que traz à tona muitos outros problemas, como o da dominação e do controle, da exploração e da vigilância, das informações duvidosas, do isolamento e da sobrecarga de informações (que nem sempre geram conhecimento), afora atitudes que imitam a vida real, como, por exemplo, a pornografia, agora online, e a visibili-

dade maior que a pedofilia ganhou. Ainda assim, há outras ações, mais inofensivas, como a avalanche de *spams*, que tanto incomodam as pessoas (PRADO, 2011).

De que maneira as *fake news* se propagam estrondosamente e exercem uma ascendência sobre o público<sup>5</sup>, manipulando sua maneira de pensar, é a incógnita por trás dos algoritmos complexos que as inflamam? A velocidade da ação algorítmica a nos trazer ilações é fato. “Seus benefícios são inegáveis, e os indivíduos e a sociedade os reconhecem. Em paralelo, contudo, proliferam impactos negativos a serem compreendidos e equacionados”, alerta Kaufman (2018, p. 46). “Dentre eles, destacam-se o viés nos processos de decisão automatizados, a invasão da privacidade e as novas formas de controle, e a personalização dos acessos e das pesquisas online” (ibidem).

Como os algoritmos funcionam, separando quem é quem, para direcionar conteúdo personalizado? De acordo com a análise de comportamento, por exemplo, o usuário entra com informações pessoais e a máquina devolve resultados a partir dos dados obtidos, através de um processo de cognição computacional, em que a máquina interpreta os signos dos internautas. A partir desses dados, o software cruza informações e oferece diversos caminhos para se navegar (PRADO, 2009) pela emoção dos perfilados. Temos assim, usuários com acesso ao que rodeia seus amigos, as comunidades traçando perfis de comportamento, reconhecendo signos semelhantes, combinados, recombinaados, estimulando direcionamentos e, ainda, mostrando dados de representação icônica mais ínti-

---

<sup>5</sup> “A transformação de qualquer classe de grupos em públicos explica-se por uma necessidade crescente de sociabilidade, que torna necessário que os associados se disponham em comunicação regular mediante uma corrente contínua de informações e excitações [reações discursivas] comuns. Esta transformação é, por conseguinte inevitável. [...]” (TARDE, 1986, p. 56).

mos, como fotos e vídeos em situações particulares. O processamento de dados, conforme ressalta Setzer (2015),

[...] limita-se exclusivamente a manipulações estruturais dos mesmos, e é feito por meio de programas. Estes são sempre funções matemáticas, e portanto também são “dados”. Exemplos dessas manipulações nos casos de textos são a formatação, a ordenação, a comparação com outros textos, estatísticas de palavras empregadas e seu entorno etc.

A dadosfera é objeto de estudo desta pesquisadora desde 2005; porém, aqui, o que se tentará sondar é de onde vêm algumas das vertentes do algoritmo de redes sociais. Assim, o nosso intuito é partir da inteligência artificial, entre outras ramificações. Apesar da teoria ou metodologia Ator-Rede<sup>6</sup> (ANT: *actor network theory*), de Law, Callon e Latour<sup>7</sup>, e outros que a estudam desde meados da década de 1980, ser anterior ao que se denomina *Big Data*, será o lume desta breve reflexão.

De seus estudos de laboratórios, Latour (2018) tinha visto “como um item aparentemente fraco e isolado – um instrumento científico, um pedaço de papel, uma fotografia, uma cultura bacteriana – poderia adquirir enorme poder por causa da complicada rede de outros itens, conhecidos como atores, que foram mobilizados em torno dele”. Em suas palavras:

Quanto mais socialmente “conectado” fosse um fato (quanto mais pessoas e coisas envolvidas em sua produção), mais efetivamente ele poderia refutar suas alternativas menos plausíveis. A ciência era “social” não apenas porque era executada por pessoas (isso,

---

<sup>6</sup> A ANT vem sendo apropriada de diversos modos pelas ciências humanas e sociais, especialmente por pesquisas em que a tecnologia e a ciência ocupam lugares centrais (BRUNO, 2011).

<sup>7</sup> Em: [https://www.sagepub.com/sites/default/files/upm-binaries/5222\\_Ritzer\\_\\_Entries\\_\\_beginning\\_with\\_A\\_\\_\[1\].pdf](https://www.sagepub.com/sites/default/files/upm-binaries/5222_Ritzer__Entries__beginning_with_A__[1].pdf).

ele pensou, era um mal-entendido reducionista da palavra “social”); em vez disso, a ciência era social porque reunia uma multidão de entidades humanas e não-humanas e aproveitava seu poder coletivo para agir e transformar o mundo (LATOURET, 2018).

Cientistas políticos, enquanto isso, mostraram que os chamados indivíduos “irracionais”, especialmente aqueles que são altamente educados, em alguns casos, na verdade, agarram suas opiniões com mais força quando confrontados com fatos que os contradizem. Em vez de acusar os defensores de Trump e os negacionistas do clima de irracionalidade, Latour argumenta que é insustentável falar sobre fatos científicos como se a correção deles fosse persuasiva.

Com o intuito de estabelecer uma base histórica para a reflexão que se segue, cabe indicar o exemplo de uma máquina teórica denominada “Memex”. Essa máquina foi criada em 1945 por Vannevar Bush<sup>8</sup> para melhorar a memória humana, ao permitir o armazenamento e recuperação de documentos ligados por associações; sendo considerado por muitos como o padrinho da era da internet (XIMENES *apud* LEÃO, 1999). Esta vinculação associativa era semelhante ao que é conhecido como hipertexto – termo cunhado na década de 1960 por Ted Nelson, reconhecedor, inclusive, de sua dívida para com Bush, tanto que o credita como sua principal influência. “Bush estava certo”, escreveu Nelson, como citado no livro *From Memex to Hypertext: Vannevar Bush and the Mind’s Machine*, de Nyce e Kahn (1991).

A ideia inovadora de Bush para automatizar a memória causou influência na ciência nos Estados Unidos. “Seu trabalho de criar uma relação entre o governo e a comunidade

---

<sup>8</sup> *As We May Think*, publicado no The Atlantic Magazine.

científica durante a Segunda Guerra Mundial mudou a maneira como a pesquisa científica é feita nos EUA e promoveu o ambiente em que a internet<sup>9</sup> foi criada”. Bush queria incitar as pessoas a “pensar sobre a ciência no mundo moderno”, enfatizando a oportunidade para sua aplicação, pois reiterava que a automatização era um campo “negligenciado pela ciência”.

A propósito, o significado de automatização, cujo efeito é de automação, é de um sistema em que os processos operacionais são controlados e executados por meio de dispositivos mecânicos ou eletrônicos substituindo o trabalho humano<sup>10</sup>. Wardle (2017) traz a definição para a atualidade ao dizer que “**automação** é o processo de projetar uma ‘máquina’ para completar uma tarefa com pouca ou nenhuma direção humana. Realiza tarefas que seriam demoradas para os humanos completarem e as transformam em tarefas que são concluídas rapidamente e quase sem esforço”. A pesquisadora dá um exemplo: “É possível automatizar o processo de envio de um *tweet*, portanto, um ser humano não precisa clicar ativamente em ‘publicar’. Processos de automação são a espinha dorsal das técnicas usadas para efetivamente ‘fabricar’ a amplificação da desinformação” (WARDLE, 2017). O que explica uma das

---

<sup>9</sup> No âmbito da formação da consciência pública, Maria Lucia Becker (2010: s/p) reforça que “as ações online ocorrem no sentido de tornar a internet uma fonte alternativa de informações, por meio de veículos de comunicação independentes, blogs, sites de organizações, listas de e-mails, fóruns de discussão e comunidades de relacionamento” (PRADO, 2015, p. 5).

<sup>10</sup> Com registro de 2007, em programas como o Musicoverly, o usuário pode ouvir músicas de sua preferência, selecionando as listas do esquema. A interface trabalha com cores que correspondem a estilos musicais em uma ideia de iconicidade e combinação de botões deslizantes, imputando uma espécie de conexãoismo. É possível eleger músicas, conforme o estado de espírito do usuário. Por exemplo, se está alegre, vai preferir ouvir músicas mais animadas. Trata-se de um *mashup* instantâneo. Traz alternativas de músicas tristes, pesadas, suaves etc. Com as diferenças de humor e a mudança de um estado para outro, a máquina, à base de probabilidade, oferece músicas de acordo com as alterações de humor do usuário. E quando ele altera, o sistema altera as possibilidades de audição, em uma maneira de deslanchar uma reação automática do sistema, como uma autocorreção (PRADO, 2009).

formas do desvario das chamadas *fake news*, especialmente na turbulência dos recentes anos eleitorais no Brasil e no exterior.

É conveniente descrever o funcionamento do Memex, no qual informações armazenadas no microfilme poderiam ser encontradas e/ou recuperadas e projetadas em uma tela no intuito de nos remeter às ações dos dias de hoje. O propósito era que a máquina pudesse estender os poderes da mente humana para formar memórias através de associações. O usuário do Memex seria capaz de fazer links entre documentos e chamadas por trilhas associativas (LEÃO, 1999).

No cenário atual, o funcionamento do Memex já não se aplica e a prática da comunicação jornalística precisa ser repensada com urgência. Para tanto, deve-se considerar a emergência da multiplicidade de ferramentas, as *affordances*<sup>11</sup> para produção do conteúdo, outros procedimentos recentes de captura, apresentação de dados e novos modelos de pensamento e sociabilidade. “A imensa quantidade de informações na Web criou necessidades de pesquisa em sistemas de busca e compreensão de informações, que por sua vez geraram inúmeros avanços na compreensão automática de textos”, diz Cozman (2018). Finalmente, “a Web abriu mercado para empreendedores interessados em sistemas automáticos de compreensão de texto, de busca de imagens, de detecção de spam”. Outro fenômeno ocorrido há 20 anos, é a emergência de técnicas de *Big Data*, aqui provisoriamente caracterizado em termos de grande escala de dados que pode ser estruturada, organizada em bancos e relacionada com outros dados ou conjunto de dados trazendo resultados agregados, tanto positivos quanto negativos.

---

<sup>11</sup> *Affordance*: característica de um ambiente, ou objeto que informa, ao humano que interage com ele, a forma de abordagem ou utilização (HOLANDA, 2014).

Grande parte das ferramentas em rede são automatizadas, sendo a maioria a partir da codificação concebida pelo humano. É notório que muitas delas, paulatinamente, venham facilitando as tarefas dos humanos – através de sistemas de auto-organização, análise preditiva, reconhecimento de padrões etc. –, tanto que muitos já não trabalham sem esses recursos. A Teoria Ator-Rede<sup>12</sup>, “como o próprio nome diz, é feita simultaneamente de atores e redes”, conforme ressaltam Santaella e Lemos (2010, p. 38).

Fernanda Bruno (2011) resalta que “uma das afirmações mais contundentes da teoria ator-rede, especialmente reiterada por Latour, é a de que o social não existe”. Uma das muitas de suas provocações é a de que boa parte da sociologia teria se poupado do trabalho essencial de explicar como se constrói ‘o social’, transformando-o numa espécie de grande estrutura ou substância que tudo explica. O “social” não é o que explica, diz a teoria ator-rede, mas o que merece ser explicado. E explicar, nesse caso, diz Bruno, “é também construir o próprio social, ou um mundo comum, ou seja, trata-se de um ‘programa’ a um só tempo cognitivo e político”.

## PONTOS DE VISTA DA VISUALIZAÇÃO DE DADOS

Com a rede, a possibilidade de analisar particularidades *a priori* – e com elas chegar a outras particularidades – é mais interessante do que imaginar, *a posteriori*, um todo (porque estamos acostumados a isso), ficar preso a ele, e tentando dar conta das experiências de conhecimento das pesquisas para ter a sua confirmação. Resumidamente, a partir da reflexão sobre

<sup>12</sup> Em 2011, a Teoria Ator-Rede foi aplicada em produtos experimentais para a tese de doutorado desta pesquisadora, que no estudo associam-se humanos – estudantes de jornalismo (atores ou actantes) – e não-humanos (dispositivos móveis), deixando reportagens de radiojornalismo dispostas em ambiente digital.

o pensamento de Gabriel Tarde (1843-1904)<sup>13</sup>, LATOUR *et al.* (2011) aproveitam a oportunidade dada pela “súbita proliferação de bancos de dados digitais para visitar a antiga teoria social, proposta por Tarde antes da disponibilidade de grandes ferramentas estatísticas e antes do entrincheiramento de muitas teorias sociais dentro do paradigma 2-LS” (*idem*). No sentido de dramatizar o contraste, os autores afirmam que “há mais complexidade nos elementos do que nos agregados”, ou, sendo mais provocativos, que “o todo é sempre menor do que a soma das partes”. Eles chamam “essa hipótese de ‘ponto-de-vista de um nível’ (1-LS) em contraste com o ‘ponto-de-vista de dois níveis’ (2-LS). Tal hipótese tem a chance de seguir apenas se ocorrer uma diferença empírica no tratamento dos dados (*idem*). Em suas palavras:

Esse é o porquê de tentarmos demonstrar dois pontos: a) algumas das novas técnicas digitais e, em particular, algumas das ferramentas oferecidas pela análise de rede podem permitir o rastreamento e a visualização do fenômeno social de uma forma que faz o 1-LS ligeiramente mais direto que a alternativa 2-LS; b) agora pode ser possível explicar características de ordem social mais duradoura, aprendendo a navegar através de ‘mônadas’ sobrepostas, em vez de alternar entre os dois níveis de individual e de agregado (LATOUR *et al.*, 2012).

Esses dados digitais de que falam os autores são considerados como “os vestígios digitais deixados por atores de dentro de banco de dados recentemente disponíveis” e que poderiam, ressaltam eles, “modificar a própria posição dessas questões clássicas de ordem social” (*idem*).

---

<sup>13</sup> Ver artigo: *The whole is always smaller than its parts - How digital navigation may modify social theory.*

O argumento dos autores é o de que, quando era difícil, praticamente impossível, e bastante complicado, ou simplesmente lento, “navegar através da massa de informações sobre itens específicos, fazia sentido tratar dados sobre conexões sociais com a definição de dois níveis: um para o elemento, outro para os agregados” (idem).

Porém, uma vez que temos a experiência de seguir “os indivíduos através de suas conexões (que é frequentemente o caso com os perfis), poderia ser mais gratificante começar a navegar pelos conjuntos de dados, sem fazer distinção entre o nível do componente individual e o da estrutura agregada” (idem). Sendo assim, os autores afirmam que “torna-se possível dar alguma credibilidade à estranha ideia de ‘mônadas’, de Tarde”. Eles afirmam que esse tipo de prática de navegação “tornou-se possível somente agora pelas bases de dados disponíveis digitalmente, que tal prática poderia modificar a teoria social se pudessemos visualizar esse novo tipo de exploração de uma forma coerente” (LATOURET *et al.*, 2012). Nota-se que Latour *et al.* tratam do assunto antes do surgimento dos *Big Data*.

Importante lembrar que em relação à discussão de se voltar ao todo para entender as partes, a definição das “Mônadas”, de Tarde, nas palavras dos autores, “dissolve[m] o dilema e redefine[m] a noção do todo, realocando-a como entidades sobrepostas que herdaram uma das outras” (idem). A questão é polêmica e eles estão conscientes de que essas bases de dados estão cheias de defeitos, e que elas mesmas “incorporam uma definição bastante crua da sociedade; que são marcadas por fortes assimetrias de poder e, acima de tudo, que elas marcam apenas um momento passageiro de cruzamento na rastreabilidade das conexões sociais”. Além disso, eles estão “dolorosamente conscientes das limitações estreitas que lhes são colo-

cadadas pela análise de rede<sup>14</sup> e pelas limitações das ferramentas de visualização disponíveis hoje”. Mas eles frisam um ponto com o qual é possível pactuar: “seria uma pena perder esta oportunidade de explorar uma alternativa tão poderosa, capaz de fornecer uma outra maneira de abordar as ciências sociais de forma empírica e quantitativa, sem perder a sua necessária ênfase nos detalhes” (idem).

Byung-Chul Han (2018) profere opinião adversa: “*Big data* torna o pensamento supérfluo porque se tudo é contável, tudo é o igual. Estamos em plena datação: o homem não é mais soberano de si mesmo, mas é o resultado de uma operação algorítmica que o domina sem que ele perceba isso”.

Vemos isso na China com a concessão de vistos de acordo com os dados manipulados pelo Estado ou na técnica de reconhecimento facial. A revolta seria parar de compartilhar dados ou de estar nas redes sociais? [...] Precisamos de uma carta digital que recupere a dignidade humana e pense em uma renda básica para as profissões que devorarão novas tecnologias (HAN, 2018).

Quanto à influência das novas tecnologias nas profissões, ao ser questionado sobre o impacto de mudanças na estrutura da informação, Steve Coll<sup>15</sup> afirma que “ele é mais positivo e, possivelmente, de efeito mais duradouro, pois ele tem a ver

---

<sup>14</sup> “A junção dos conhecimentos de sociologia com as teorias dos grafos deu origem a análise de redes sociais, uma ciência complexa, não exata, mais inclinada à heurísticas do que a soluções determinísticas, mas que ainda assim, segue padrões matemáticos conhecidos. (...) Utilizando as teorias e algoritmos de redes complexas é possível analisar uma rede social e extrair diversas métricas desta rede. Como por exemplo, o número de comunidades, ou grupos existentes nessas redes. Quais são os vértices (nós) mais importantes da rede. E a partir desse conhecimento direcionar esforços para os pontos mais importantes da rede, otimizando recursos e tempo” (GABARDO, 2015).

<sup>15</sup> Diretor da *Columbia Journalism School*, da *Columbia University*. Em entrevista aos jornalistas Ricardo Gandour e Jorge Tarquini para o Observatório da Imprensa.

com a estrutura que a informação passou a ter com o advento dos *Big Data* e o uso de softwares baseados em algoritmos, capazes de processar imensos volumes de dados”. Disse, ainda, que isso impactou diretamente o jornalismo investigativo:

Em projetos como Panama Papers<sup>16</sup> – mudando a forma como o jornalismo é feito, com aumento da colaboração entre as organizações de mídia. Houve investigações que foram feitas por até 60 veículos em conjunto, todos guardando os mesmos segredos e publicando ao mesmo tempo. Isso era algo impensável dez anos atrás. Outro aspecto são os jornalistas colaborando com cientistas da computação e de dados para o bem público. Há diversos casos de reportagens premiadas feitas a partir da análise de uma grande quantidade de dados (COLL, 2018).

Sobre acreditar ou não que programação, tecnologia de dados e outras áreas relacionadas deveriam fazer parte do ensino atual nas escolas de jornalismo, Coll respondeu que sim: “Porque, na era do *Big Data*, é difícil para o jornalista levar adiante sua função democrática e constitucional de informar, se ele não souber examinar da maneira correta os dados e algoritmos que estão sendo utilizados e seus impactos na sociedade”. Ele reforça que é preciso “literalmente olhar por baixo dos códigos e ver como eles são feitos. Os tribunais americanos estão usando softwares de algoritmo para determinar sentenças baseadas em indicações de como alguns criminosos irão reagir, se podem ou não retornar ao crime”. Ressalta, ainda, “que os processos de engenharia utilizados nos códigos das chamadas ‘*fake news*’ podem influenciar a opinião pública. Há várias questões muito importantes de interesse público, que estão sendo desenvolvidas com o suporte de engenheiros e especialistas em dados”, finaliza

---

<sup>16</sup> <https://www.icij.org/investigations/panama-papers/>.

Coll, afirmando que “isso muda significativamente as habilidades necessárias para o exercício do jornalismo hoje”.

Temos um panorama midiático no qual mais técnicas serão adotadas na produção de notícias: “análises algorítmicas de base de dados, visualização de dados, solicitações de conteúdo por parte de amadores, produção automatizada de narrativas, criação de narrativas baseadas em dados entre outros”, como afirmam Anderson, Bell & Shirky (2013). Eles entendem que em 2020, os exploradores de dados, criadores de visualizações de informação, ou designer de experiência interativa, terão em mãos um conjunto muito mais refinado de ferramentas do que qualquer um existente na atualidade.

Faz-se necessário debater também a interatividade, jamais ocorrida de forma tão intensa, e a entrada do jornalismo aberto gerado pelo usuário (ou jornalismo colaborativo, participativo etc.) como estratégia de acompanhamento de vanguardas, mesmo que desconfiando do que chega desse material de não-jornalistas e, com isso, aumentando o grau de checagem, fazendo paralelos com as novas funções e manifestações que surgem com as diferentes etapas da web (1.0<sup>17</sup>, 2.0, 3.0)<sup>18</sup> e quantas numerações mais possam aparecer como marcos), como a personalização (quando os sites monitoram o rastro do usuário), os filtros baseados em palavras-chave (*tags*) e ranqueamento mostrando preferências.

Uma visão distinta é defendida por Anderson (2018), editor da revista *Wired*, que entende que grandes quantidades de dados e matemática aplicada substituem todas as outras

<sup>17</sup> No início da web, a chamada web 1.0, as pesquisas da fase pré-Google eram realizadas em sites como o Lycos, o Altavista e o Yahoo.

<sup>18</sup> Estes indicadores apontam para o que se chama de web 3.0, que conta com sistemas inteligentes de busca por contexto, que ao mesmo tempo em que facilitam a procura, permitem a maior produção e segmentação do conteúdo, além da disponibilização deste para o público de interesse.

ferramentas que podem ser usadas; ele questiona: “Quem sabe por que as pessoas fazem o que fazem?” e como o fazem. Anderson é taxativo ao afirmar que “podemos rastreá-lo e medi-lo com uma fidelidade sem precedentes. Com dados suficientes, os números falam por si”.

A nova disponibilidade de enormes quantidades de dados, juntamente com as ferramentas estatísticas para processar esses números, oferece uma maneira totalmente nova de entender o mundo. A correlação substitui a causalção, e a ciência pode avançar mesmo sem modelos coerentes, teorias unificadas ou, na verdade, qualquer explicação mecanicista (ANDERSON, 2018).

Apesar de polemica, a afirmação de Anderson parece encontrar ecos, por exemplo em Castro (2018), ao afirmar que: “muitas relações que enlaçam os traços digitais dos usuários não estão dadas com antecedência, mas precisam ser descobertas, desafio a que se propõe a analítica dos dados, peça-chave da governança algorítmica. O grande trunfo desta é o escopo do Big Data”, ao qual ela se acopla,

no tocante ao volume, velocidade e variabilidade – a fórmula dos 3 Vs, introduzida por Laney (2001). “O prospecto alardeado do *Big Data* é trazer à luz e descobrir o que nunca foi observado”, alega Reigeluth (2014, p. 248), “abandonando os rígidos métodos hipotético-dedutivos de raciocínio, e abraçando novas ferramentas indutivas, que se apoiam em vastas extensões de dados aráveis” (CASTRO, 2018).

Com a finalidade de evocar a presença de autores que se debruçam sobre temas quentes, similares ao que é aqui estudado, é proveitoso dar réplicas às falas de Beate

Josephi. Ela acaba por pressionar as fronteiras do jornalismo, ao invés de estabelecê-las, ao dizer que “uma nova estética precisa ser explorada e assumida, de forma a oferecer uma aparente impressão de participação não-mediada nos eventos. Não mediar, ou mesmo não mostrar que se está mediando, é crucial nesse momento, pois as pessoas estão cansadas de serem mediadas” (JOSEPHI, 2016, p. 13). A partir do surgimento da internet e a certeza de que todos podem participar, cai retumbantemente o monopólio da informação de quem quer que seja e – não é preciso ir atrás dos números (mas, eles existem para comprovar) – esse movimento não tem volta.

Trata-se de um envolvimento que deixa de ser latente e emerge, pois, se audiência vem de audição e de prestar atenção a quem vos fala, ela também quer ser ouvida, chamar a atenção para sua participação. “Isso implica não só a interatividade em relação à palavra escrita, mas também uma expectativa em ‘estar presente’ em eventos e mostrá-los da maneira como ‘eles aconteceram’” (JOSEPHI, 2016). Muito disso ocorre porque a imprensa *mainstream* nem sempre cobre os eventos como realmente acontecem, já que não estamos falando apenas do foco em que se dá, ou o recorte escolhido, (até porque, não dá mesmo para mostrar tudo, o tempo todo, no espaço limitado da mídia tradicional) e, sim, da manipulação como distorção da verdade que sempre existiu, mas que na era digital é mais aparente e o incremento dado nos é apresentado de modo sistemático.

Com a análise do comportamento dos perfis nas redes sociais, seja de humanos e não humanos (*bots, ciborgues*), procuramos entender o veloz processo de manipulação da opinião na esfera pública conectada. De forma contínua e multiplicadora, a utilização de *fake news* e similares ocorre com a intenção de influenciar diretamente um certo segmento ignorante e/ou mal intencionado da sociedade.

Concluiremos provisoriamente nossa reflexão relembrando Gilbert Simondon (1924-1989), mais conhecido pela sua filosofia da técnica do que pela sua teoria da individuação, da qual sua reflexão sobre os objetos técnicos é decorrente – entendendo que é possível levar este estudo a patamares que não dissociam humanos, coisas e máquinas, ou seja, internautas, jornalistas, não jornalistas, softwares automáticos, robôs, *bots* etc.. Afinal, uma das características marcantes do pensamento de Simondon está, conforme ressalta JUNK (2017), na noção científica de informação, entendida como entidade imaterial, possuidora de propriedades organizacionais, com estrutura matemática, e que reúne, em um mesmo grupo, seres vivos em geral, seres humanos em particular e seres artificiais.

## REFERÊNCIAS

- ANDERSON, C. The end of theory: The data deluge makes the scientific method obsolete. *In: Wired Magazine*, 23 June 2008.
- ANDERSON, C.W.; BELL, E.; SHIRKY, C. Jornalismo pós-industrial - Adaptação aos novos tempos. *In: Revista de Jornalismo ESPM*, São Paulo, nº 5, p. 32-89, 2013.
- BRUNO, F. (coord.); SANTAELLA, L.; FELINTO, E. *In: V SIMPÓSIO NACIONAL ABCIBER*. 2011. *Ator-Rede e Cibercultura: diálogos I*. Disponível em: <http://dispositivodevisibilidade.blogspot.com.br/2011/11/teoria-ator-rede-e-cibercultura.html>. Acesso em: 02/10/2017.
- BUSH, V. *As we may think*. 1994. Disponível em: <http://ccat.sas.upenn.edu/~jod/texts/vannevar.bush.html>. Acesso em: 23/06/2018.
- CASTRO, J. C. L. Redes sociais como modelo de governança algorítmica. *In: revista Matrizes*. V.12 - Nº 2, maio/ago. São Paulo. Brasil. 2018. Disponível em: <http://www.revistas.usp.br/matrizes/article/view/140890/147048>. Acesso em: 12/09/2018.

COLL, S. Os desafios da reportagem em tempos de algoritmos e antagonismo político. In: *Observatório da Imprensa*. 2018. Disponível em: <http://observatoriodaimprensa.com.br/edicao-brasileira-da-columbia-journalism-review/os-desafios-da-reportagem-em-tempos-de-algoritmos-e-antagonismo-politico/>. Acesso em: 26/08/2018.

COZMAN, F. A Inevitável Vitória da Inteligência Artificial. 2017. In: *Estadão.com*. Disponível em: <http://cultura.estadao.com.br/blogs/estado-da-arte/a-inevitavel-vitoria-da-inteligencia-artificial/>. Acesso em: 25/03/2018.

El País. *Trump e as 'fake news' atacam de novo*. 2017. Por Jan Martínez Ahrens. Disponível em: [https://brasil.elpais.com/brasil/2017/05/06/internacional/1494087975\\_053461.html](https://brasil.elpais.com/brasil/2017/05/06/internacional/1494087975_053461.html). Acesso em: 21/01/2018.

GABARDO, A. C. *Análise de redes sociais, um texto introdutório*. 2015. Disponível em: <http://imasters.com.br/midia-e-marketing-digital/redes-sociais/analise-de-redes-sociais-um-texto-introdutorio/?tracce=1519021197&source=single>. Acesso em: 21/11/2018.

HAN, B-C. Agora você explora a si mesmo e acredita que está se realizando – diz filósofo coreano. In: *Pensarcontemporaneo.com.br*. Disponível em: <https://www.pensarcontemporaneo.com/agora-voce-explora-si-mesmo-e-acredita-que-esta-se-realizando-diz-filosofo-coreano/>. Acesso em: 28/08/2018.

HOLANDA, A. *Traduzindo o Jornalismo para Tablets com a Teoria Ator-rede*. 2014. (Tese). Faculdade de Comunicação Social. Universidade Federal da Bahia (UFBA), Salvador, 2014.

JOSEPHI, B. Profissionalismo jornalístico na era digital. In: *Brazilian journalism research* - 2016. Volume 12 - Número 3 - SBPjor / Associação Brasileira de Pesquisadores em Jornalismo. 2016. Disponível em: <https://bjr.sbpjor.org.br/bjr/article/view/958/851>. Acesso em: 12/04/2018.

JUNK, I. Simondon: uma perspectiva ontoepistemológica para a contemporaneidade. In: *Transobjeto*. 2017. Disponível em: <https://transobjeto.wordpress.com/2017/04/15/simondon-uma-perspectiva/>. Acesso em: 25/06/2017.

KAUFMAN, D. O protagonismo dos algoritmos de Inteligência Artificial: observações sobre a sociedade de dados. In: revista *TECCOGHS* edição 17. 2018. Disponível em: [http://www4.pucsp.br/pos/tidd/teccogs/artigos/2018/edicao\\_17/teccogs17\\_artigo02.pdf](http://www4.pucsp.br/pos/tidd/teccogs/artigos/2018/edicao_17/teccogs17_artigo02.pdf). Acesso em: 11/09/2018.

LATOURE; JENSEN; BOULLIER; GRAUWIN; VENTURINI. "The Whole is Always Smaller Than Its Parts" — How Digital Navigation May Modify Social Theory. *In: British Journal of Sociology*, v. 63, n. 4, p. 591-615, 2011. Tradução de Flávia Gonsales e Beatriz Redko para a Parágrafo. Jul./Dez. 2015. V. 2, N. 3, 2015. Disponível em: <http://revistaseletronicas.fiamfaam.br/index.php/recicofi/artic/view/329/336>. Acesso em: 22/03/2017.

LATOURE, B. Bruno Latour, the Post-Truth Philosopher, Mounts a Defense of Science - He spent decades deconstructing the ways that scientists claim their authority. Can his ideas help them regain that authority today? Por Ava Kofman. *In: The New York Times Magazine*. 2018. Disponível em: <https://www.nytimes.com/2018/10/25/magazine/bruno-latour-post-truth-philosopher-science.html>. Acesso em: 14/05/2019.

LEÃO, L. *Labirinto da hipermídia - arquitetura e navegação no ciberespaço*. São Paulo: Iluminuras, 1999.

LEMO, A. *Cidades invisíveis, cidades rastreadas*. 2011. Disponível em: <https://andrelemos.info/cidades-invisiveis-cidades-rastreadas/>. Acesso em: 21/04/2019.

NYCE, J. M.; KAHN, P. *From Memex to Hypertext: Vannevar Bush and the Mind's Machine*. London: Academic Press, 1991.

PARISER, E. *The Filter Bubble: What the Internet Is Hiding from You*. New York: Penguin Press, 2011.

PRADO, M. *Ciberativismo e Noticiário: da mídia torpedista as redes sociais*. Rio de Janeiro: Alta Books, 2015.

PRADO, M. *Webjornalismo*. Rio de Janeiro: LTC/GEN, 2011.

PRADO, M. *Audiocast nooradio: redes colaborativas de conhecimento*. 2009. Dissertação Mestrado em Comunicação e Semiótica. Pontifícia Universidade Católica de São Paulo, São Paulo.

SANTAELLA, L; LEMOS, R. *Redes sociais digitais: a cognição conectiva do twitter*. São Paulo: Paulus, 2010.

SETZER, V. W. *Dado, Informação, Conhecimento e Competência*. Departamento de Ciência da Computação, Universidade de São Paulo. 25 maio 2015. Disponível em: [www.ime.usp.br/~vwsetzer/dado-info.html](http://www.ime.usp.br/~vwsetzer/dado-info.html). Acesso em: 13/11/2017.

TARDE, G. *Les lois de l'imitation*. Paris: Kimé Éditeur, 1895.

WARDLE, C. & DERAKHSHAN, H. *Information Disorder: Toward an interdisciplinary framework for research and policy making*. Council of Europe. 2017. In: *RM*. Disponível em: <https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-research/168076277c>. Acesso em: 11/02/2019.

# O *BIG DATA* NO DESENVOLVIMENTO DA INDÚSTRIA 4.0: NOVAS PERSPECTIVAS PARA O EMPREENDEDORISMO ACADÊMICO

*Selma Leticia Capinzaiki Ottonicar*  
*Gisele Rodrigues Atayde*  
*Luis Antonio de Santa-Eulalia*

A quarta Revolução Industrial, também conhecida como Indústria 4.0, tem trazido mudanças para as organizações contemporâneas. A principal transformação é a produção massiva de dados e informações disseminadas por objetos, pessoas e elementos biológicos (SCHWAB, 2016). Esse novo contexto foi reconhecido a partir de 2013 na Alemanha e desde então vem influenciando pesquisas científicas e a indústria manufatureira em diversos países. As empresas sofrerão mudanças significativas em sua estrutura e nos modos de produção, modificando, inclusive, a vida em sociedade. A Indústria 4.0 (I4.0) traz novas oportunidades de negócios e novas formas de compreensão a respeito da qualidade de vida da população.

A Indústria 4.0 possui várias tecnologias que a complementam, dentre elas estão os recursos de *Big Data*. Os *Big*

*Data* podem ser entendidos como o armazenamento de dados provenientes de sistemas e da web a fim de que possam ser utilizados no futuro. O desafio das pessoas é saber como fazer o uso de tais dados para tomar decisão, inovar, construir conhecimento, empreender e atuar perante as transformações da estrutura social.

O acesso a esses dados é um instrumento fundamental para estimular a criatividade, aprendizagem e inovação no contexto universitário. A universidade é um ambiente de construção de conhecimento teórico e prático. Por isso, algumas dessas instituições têm se tornado ambientes que propulsionam a inovação e o empreendedorismo dos indivíduos. Os discentes, principalmente das áreas de Engenharia e Ciências Sociais aplicadas lidam com a aplicação prática em sala de aula. Desse modo, o empreendedorismo acadêmico contribui para que possam criar novos negócios e realizar parcerias.

Nesse contexto, os profissionais necessitam saber como usar a informação para tomar decisões eficazes. Com isso, defende-se que essas novas transformações podem impulsionar o empreendedorismo entre alunos universitários à medida que vêm, nas lacunas do contexto, novas oportunidades de negócios.

Assim, o problema de pesquisa é: como a Indústria 4.0 e os *Big Data* podem ser utilizados para estimular o empreendedorismo acadêmico? O objetivo do artigo é demonstrar como os *Big Data* no contexto da Indústria 4.0 contribui com novas perspectivas para o empreendedorismo acadêmico. Para atingir os objetivos propostos utilizou-se revisão bibliográfica sobre os temas *Big Data*, Indústria 4.0 e empreendedorismo acadêmico. Essa revisão teórica possibilitou a construção de um quadro conceitual inter-relacionando as temáticas e,

posteriormente, a construção de um modelo, demonstrando a relevância dos temas para a área da Gestão Empresarial e da Ciência da Informação.

## 1 INDÚSTRIA 4.0

A terceira Revolução Industrial iniciou-se em meados de 1970, devido à revolução digital. Houve o aumento no uso de semicondutores, automatização e robotização da produção, além do avanço nas comunicações e a internet (COSTA, 2017). A I4.0 complexifica tais tecnologias, ao possibilitar que elas estejam conectadas, produzam dados e se comuniquem entre si de maneira inteligente.

A nova Revolução Industrial refere-se ao que estão sofrendo as empresas manufatureiras por meio das tecnologias inteligentes, conectadas e da Internet das Coisas (IoT). Suas fontes de informação são as pessoas, os objetos e os elementos biológicos (SCHWAB, 2016; DUTTON, 2014; ALMADA-LOBO, 2015). A I4.0 traz novos desafios e oportunidades para as organizações, principalmente devido à modificação na estrutura e nas formas de se realizar o trabalho. Amorim (2017, p. 248) explica que:

O termo “Indústria 4.0” surgiu a partir de um projeto do governo alemão<sup>27</sup> que visava o desenvolvimento das tecnologias voltadas para as indústrias, objetivando, sobretudo, aumentar a competitividade, através de “fábricas inteligentes”. Tais unidades de produção “inteligentes” seriam proporcionadas através da conexão de máquinas, sistemas e ativos criando redes inteligentes ao longo do processo produtivo controladas autonomamente, ou seja, com intervenção humana ínfima, consoante o tratamento das informações produzidas pela cadeia produtiva e pela demanda de produção.

A I4.0 vem trazendo novos impactos socioeconômicos que influenciarão os países em escala global com o estímulo à criação de novos negócios, o aumento da tecnologia de informação com foco na segurança dos sistemas, modificações no mercado de trabalho por meio das demandas de novas profissões e a extinção de outros cargos (AMORIM, 2017). Por isso, os países necessitam estar atentos a tais novas mudanças e desenvolver políticas públicas que contribuam com a economia e a vida profissional dos cidadãos.

Costa (2017) explica que as consequências dessa revolução envolvem formas complexas de inovação. Há a combinação de diferentes tecnologias que vêm modificando as organizações completamente. Os gestores precisam refletir sobre novas formas de produção, desenvolvimento de processos, criação de novos produtos e serviços e ajustar o marketing e a distribuição da empresa.

Além disso, a I4.0 propicia modificações nas demandas dos consumidores, produtos inteligentes e produtivos, parcerias e inovação colaborativa, modificações no modelo operacional e a transformação para um novo modelo digital (SCHWAB, 2016). A I4.0 é baseada em três principais pilares tecnológicos que deverão ser adotados pelas organizações nos próximos anos. Esses pilares são a Internet das Coisas (IoT), os sistemas *ciber* físicos e o Big Data (COSTA, 2017).

As fábricas da I4.0 são baseadas em máquinas, produtos e processos conectados. Tais objetos são capazes de analisar dados e tomar decisões programadas. Existe a sintonia entre o ser humano e a máquina voltados ao aumento da produtividade, qualidade e melhora dos lucros. As pessoas e a tecnologia avançada trabalham juntos (SALTIÉL; NUNES, 2017).

Essa automação e conexão eliminará postos de trabalho, principalmente aqueles que exigem esforços físicos ou condições perigosas. Nesse sentido há a valorização dos indivíduos, pois as empresas precisarão de funcionários que atuem na gestão e controle dessas tecnologias inteligentes (GORECKY *et al.*, 2014; JO, 2010; LEE; JO, 2007; NUNES; MENEZES, 2014; RUSSWURM, 2014).

Entretanto, a I4.0 tem como consequência o aumento do desemprego da população, pois vários cargos deixarão de existir e outros serão substituídos pelas tecnologias e robotização, como é o caso da linha de produção. Nesse contexto, será cada vez mais necessário saber como acessar, avaliar e usar a informação para tomar decisão, resolver problemas e aprender. Concomitantemente, novas demandas profissionais surgirão, como é o caso daqueles que saibam atuar com a análise de dados, mercado, inteligência competitiva, entre outros.

## 2 *BIG DATA*

Os *Big Data* envolvem quantidade massiva de dados armazenados em sistemas de informação complexos (LETOUZÉ, 2012; GOLDMAN *et al.*, 2012). Os sistemas tradicionais não conseguem suportar tamanha quantidade de dados (GOLDMAN *et al.*, 2012). Os *Big Data* é uma tecnologia fundamental para a I4.0 e a tendência é que as empresas utilizem cada vez mais esses sistemas. Isso acontece porque não só as pessoas serão produtoras de dados e informação, como também os objetos e elementos biológicos.

Costa (2017) explica que os *Big Data* contribuem para armazenar a quantidade massiva de dados gerados pela Internet das Coisas (IoT), que são produzidos em tempo real. Essa tecnologia traz desafios para as empresas de tecnologia da informação como onde guardar os dados de maneira segura e que sejam

facilmente acessados, além disto, como processar tais dados a fim de que sejam transformados em informação voltada a tomada de decisão inteligente?

Essa quantidade massiva de dados exige que sejam trabalhados e atribuídos significados, pois os dados são números, palavras e sinais simples não codificados. Eles podem ser verificados, mas, para tanto, precisam de análise, interpretação e contextualização, tornando-se uma informação. As informações são elementos fundamentais para que os indivíduos desenvolvam suas hipóteses e consigam prever o futuro (COSTA, 2017).

A análise desses dados disponíveis em sistemas, nas nuvens, nas redes de negócios e na internet contribuem para perceber as tendências do mercado e com isso, os gestores desenvolvem novas oportunidades de negócios. O desafio dos *Big Data* é de propiciar o armazenamento de informações de maneira segura, respeitando a privacidade e a ética que envolve o uso da informação organizacional e pessoal.

Pimenta (2013, p. 6) explica que: “Não obstante, torna-se evidente o potencial político do controle do *Big Data* quando compreendemos que seu advento representa um momento, ou uma experiência de “concretização” de uma mudança tecnológica”. Assim, os *Big Data* possibilitam que os “traços” das pessoas sejam monitorados pela internet, pois os sistemas identificam seus gostos e preferências. São deixados rastros de gostos políticos, relações comerciais e privados de modo que os sistemas coletam esses dados por meio das tecnologias de vigilância e monitoramento (PIMENTA, 2013).

“Há também grandes expectativas sobre o impacto nas relações políticas, seja pelo uso de ferramentas de BD pelos políticos ou pelo novo modelo de organização e engajamen-

to da sociedade através das redes sociais” (LUVIZAN; MEIRELLES; DINIZ, 2014, p. 4). Assim, os *Big Data* contribuem em todas as áreas da sociedade, além dos negócios, podendo ser utilizados na transparência do uso dos recursos públicos, bem como no exercício da cidadania.

Os *Big Data* podem ser compreendidos em duas vertentes diferentes. A primeira os identifica como uma tecnologia que armazena quantidade massiva de dados e informação. Já a segunda vertente compreende os *Big Data* como sendo a quantidade dos dados presentes na internet (RIBEIRO, 2014). Breitman (2014) defende que dessa quantidade enorme de dados, apenas 1% é realmente analisado, por isso, Ribeiro (2014) aponta que essa é uma oportunidade de atuação para os profissionais da informação.

O SEBRAE (2018) explica que os dados coletados por meio dos dispositivos e sensores são aproveitados em processos inteligentes, por meio de algoritmos de mineração de dados (*Data Mining*). Esses algoritmos computacionais contribuem para que novas informações sejam desenvolvidas. Assim as máquinas determinam predições sobre os defeitos, matéria-prima, energia elétrica, estrutura da fábrica a fim de se ajustar às demandas dos consumidores. Essas tecnologias contribuem para que os gestores tomem decisões inteligentes sobre a produção, o mercado e construam aprendizagem durante o processo.

### **3 O EMPREENDEDORISMO ACADÊMICO COMO MECANISMO PARA FORMAÇÃO DE NOVAS COMPETÊNCIAS PARA A I4.0**

Atualmente as universidades estão incorporando o desenvolvimento econômico e social como missão, com isto, vão além do ensino e pesquisa, abrangendo a criação, com-

partilhamento e comercialização de conhecimento, ou seja, direcionando esforços para o empreendedorismo acadêmico (SHANE, 2004; WRIGHT, 2007; ETZKOWITZ, 2008).

Além da exigência e exigente responsabilidade no retorno dos recursos recebidos à sociedade, a universidade contemporânea exerce novas funções. Segundo Thorp e Goldstein (2010), há alguns pontos que precisam ser considerados. Primeiramente, os problemas atuais são grandes e complexos, e para atacá-los, deve haver recursos sem precedentes e abordagens não tradicionais que vão além das disciplinas curriculares.

Em segundo lugar, as ferramentas ricas em informações e a substituição de indivíduos em atividades simples por máquinas, levará o ensino para uma mentalidade empreendedora que explore novas formas de trabalho e novos modelos de negócios. Em terceiro lugar, os estudantes possuem uma nova visão sobre conhecimento e educação, e valorizam colher resultados ao longo do processo. E por último, as fontes de recursos para o financiamento da ciência estão diminuindo levando a novos modelos de parcerias para desenvolvimento de projetos (THORP; GOLDSTEIN, 2010).

Tais pontos levam a crer que esse novo modelo se relaciona ao conceito de Universidade Empreendedora, que segundo Clark (1998) é a universidade que busca frequentemente inovar sua estratégia e adequar sua estrutura para responder rapidamente às novas demandas da academia.

Diante dessas novas diretrizes no ambiente acadêmico, Etkowitz (2013) ressalta a importância da capacitação empreendedora coletiva e institucional, ou seja, a universidade necessita atuar como uma instituição empreendedora em suas atividades.

Ainda sobre capacitação no âmbito da universidade, é fundamental desenvolver competência empreendedora. Rasmussen e Wright (2015) estabelecem dois argumentos sobre a importância de direcionar esforços dentro da universidade para criação de negócios e estabelecer uma competência empreendedora. O primeiro está relacionado ao impacto econômico direto gerado por novos empreendimentos tecnológicos resultando em empregos, receitas e competitividade internacional. Já o segundo argumento é que tais empresas promovem a transferência tecnológica que converte o conhecimento em aplicação na sociedade.

No contexto da I4.0 e *Big Data*, é plausível que somente as instituições acadêmicas serão capazes de lidar com o capital humano e com a escala de tempo necessária para perceber o potencial das invenções descontínuas. Além disso, a possibilidade de conectar ciência, tecnologia e inovação sem cair nas armadilhas da não traduzibilidade do mercado e o conhecimento tácito só poderão ser realizadas dentro dos laboratórios e centros acadêmicos (ETZKOWITZ; VIALE, 2010).

#### **4 METODOLOGIA**

Essa pesquisa foi desenvolvida mediante os princípios teóricos de Marconi e Lakatos (2013), com o uso da pesquisa bibliográfica (GIL, 2012; FONSECA, 2002). O artigo foi desenvolvido e estruturado com base nos conceitos de autores da área da Ciência da Informação e Gestão Empresarial. Esses conceitos estavam presentes em artigos científicos, dissertações, teses e livros.

As buscas foram realizadas nas seguintes bases de dados: *Scientific Electronic Library Online* (SCIELO), Base de Dados Referencial de Artigos de Periódicos em Ciência da Informação (BRAPCI), Portal de Periódicos da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), SCOPUS e *Library and Information Science Abstracts* (LISA). Utilizou com base os termos *Big Data*, Indústria 4.0 e Empreendedorismo Acadêmico.

Essa revisão da literatura contribuiu para a construção de um quadro teórico-conceitual sobre os três principais temas do artigo, *Big Data*, Indústria 4.0 e Empreendedorismo Acadêmico (GIL, 2012; FONSECA, 2002; MARCONI; LAKATOS, 2013). No final, desenvolveu-se um modelo teórico entre tais temáticas para demonstrar como sua inter-relação pode contribuir com a aprendizagem dos indivíduos envolvidos com as instituições de ensino, como é o caso das universidades.

## 5 RESULTADOS E DISCUSSÕES

A I4.0 possui diversas tecnologias que a compõem no ambiente organizacional, dentre elas encontram-se os *Big Data*. Essa tecnologia contribui para armazenar uma grande quantidade de dados produzidas pelas pessoas e objetos. Essa tecnologia e o advento da I4.0 trazem mudanças para as universidades e podem ser aproveitadas para desenvolver novas formas de aprendizagem.

Assim, algumas universidades têm se preocupado com o empreendedorismo dos alunos, professores e funcionários que atuam na instituição. Para empreender, os acadêmicos precisam acessar informações que são disponibilizadas pelos *Big Data*. Es-

sas informações podem ser importantes para a geração de novas ideias de negócios a serem criados pelos indivíduos. Essas relações estão detalhadas no quadro 1, conforme segue.

**Quadro 1** – Contribuições do Big Data, Indústria 4.0 e do Empreendedorismo Acadêmico

<b>Big Data</b>	<b>Indústria 4.0</b>	<b>Empreendedorismo Acadêmico</b>
Armazenamento de um grande conjunto de dados	Produção massiva de dados e informações por meio de máquinas inteligentes e conectadas (TISLUK; MOSCONI; CHAMBERLAND-TREMBLAY, 2015)	A Universidade fornece estrutura e apoio para alunos, docentes e pesquisadores testarem e iniciarem novos empreendimentos. Dessa forma, a Universidade pode ser um laboratório para teste de novos modelos de negócios integrando <i>Big Data</i> e I4.0.
Os dados necessitam ser analisados para que se agregue valor ao seu conteúdo.	É necessário que o profissional saiba como acessar, avaliar e interpretar as informações.	A Universidade pode auxiliar na capacitação e formação de atitudes e habilidades empreendedoras que serão essenciais para o manuseio de dados e informações.
Dados estão presentes em tecnologias remotas, inteligentes e conectadas. Está em processo de transição.	Dados e informações são produzidos, armazenados e compartilhados pelas tecnologias inteligentes e conectadas.	Conectar especialistas de áreas distintas para o desenvolvimento de tecnologias inteligentes é uma estratégia considerável para identificação de oportunidades de mercado e criação de negócios baseados em dados e informações. Networking e interdisciplinaridade no ambiente acadêmico podem apoiar estes negócios.
Dados podem ser acessados por meio da internet.	Dados e informações podem ser acessados por meio do uso das tecnologias inteligentes.	A integração entre os <i>Big Data</i> e I4.0 na universidade podem ser estimuladas por meio de projetos em parcerias com o ambiente corporativo.

**Fonte:** Criado pelos autores – 2018

A quarta Revolução Industrial tende a modificar os modos de produção, a estrutura organizacional e a economia dos países. Além disso, a produção massiva de dados estimula o desenvolvimento de negócios que atuam nessa área e o aumento da procura por novas profissões, como é o caso dos analistas de negócios, de inteligência competitiva e organizacional. Esses profissionais necessitam ser competentes para interpretar as informações. Essas demandas contribuem com o empreendedorismo acadêmico, principalmente nos cursos de Gestão Empresarial, Engenharias, Biblioteconomia e Arquivologia.

As tecnologias associadas aos *Big Data* possibilitam o armazenamento de informação advindos das tecnologias da I4.0. Tais tecnologias podem ser utilizadas na universidade a fim de possibilitar a pesquisa e desenvolvimento pelos estudantes. Esses dados e informações contribuem para estimular o empreendedorismo, à medida que demonstram as oportunidades do mercado.

Tanto no contexto dos *Big Data* quanto da I4.0 as tecnologias estão conectadas e compartilham dados entre elas. Essa quantidade de dados propicia um contexto interdisciplinar, no qual estudantes de várias áreas podem realizar estudos e desenvolver novas formas de negócios e soluções criativas para o mercado.

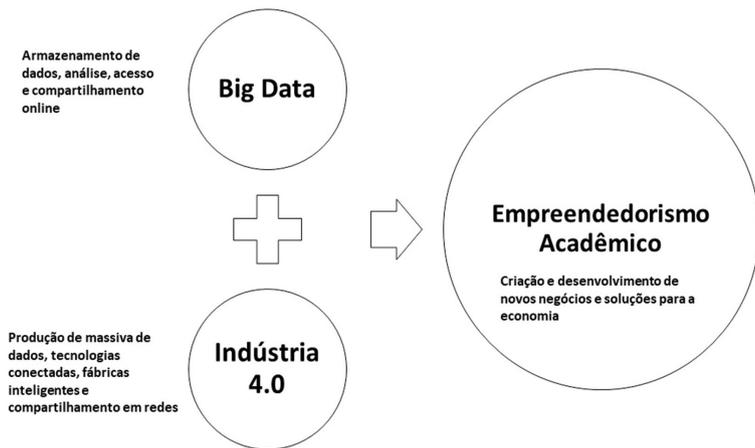
Os dados armazenados precisam de análise eficaz, por isso, os estudantes podem aprender como acessar, avaliar e usar os dados para construir inteligência. Ao criar informações inteligentes é possível aprender sobre as oportunidades do mercado. A universidade tem papel central como espaço de treinamento e capacitação dos estudantes.

Os *Big Data* facilitam o acesso à internet e a I4.0 estimula o acesso a tecnologias conectadas, inclusive as redes

que englobam a *Web*. Esses dados podem ser acessados pelos estudantes. Ressalta-se a relevância em se possibilitar a parceria de pesquisa entre universidade e empresa. Esse tipo de projeto contribui para o crescimento de ambas intuições no que tange aprendizagem, inovação e competitividade.

A partir de tais considerações elaboradas por meio do quadro 1 foi possível desenvolver o modelo teórico-conceitual abaixo que inter-relaciona as temáticas dos *Big Data*, I4.0 e o empreendedorismo acadêmico.

**Figura 1** – Influência do Big Data e I4.0 no Empreendedorismo Acadêmico



**Fonte:** Os autores - 2018

A figura 1 demonstra a relação existente entre *Big Data* e I4.0 como propulsores de novos negócios acadêmicos. O contexto da I4.0 vem trazendo mudanças para a economia dos países desenvolvidos e, conseqüentemente, trará resultados e demandas para os demais países. Essas mudanças in-

cluem transformações no modo de produção das indústrias e no modo de vida da população.

Já os *Big Data* podem ser entendidos como uma tecnologia aplicada na sociedade e seus resultados têm contribuído com a competitividade das organizações, empresas e instituições educacionais. Essa tecnologia faz parte do arcabouço tecnológico da I4.0 e, portanto, demonstra as novas tendências da sociedade.

Os empreendedores acadêmicos precisam estar atentos a essas novas mudanças, a fim de observar o contexto para identificar os problemas e criar soluções viáveis para as pessoas. Por isso, precisam aprender como manusear e o funcionamento das tecnologias da I4.0, aplicando em suas empresas e gerando projetos inovadores.

A utilização de cases para desenvolvimento de produtos que utilizem tecnologias conectadas a dados pode ser aplicada em projetos interdisciplinares, ou seja, abrangendo diversas disciplinas com conteúdos complementares. As áreas que podem utilizar esse modelos são: gestão empresarial, engenharias, biblioteconomia, arquivologia, economia, ciências da computação, entre outras.

## CONSIDERAÇÕES FINAIS

Apesar da quarta Revolução Industrial ter como consequência o desaparecimento de profissões, principalmente aquelas que estão no chão de fábrica e serão substituídas pelas máquinas, há novas demandas profissionais e de negócios que estão cada vez mais relacionados com a busca, interpretação, análise, organização e disseminação dos dados e informações.

A Indústria 4.0 também contribui para que as máquinas inteligentes e a IoT produzam informações e dados complexos. Estes dados armazenados, com recursos dos *Big Data*, precisam ser avaliados e interpretados para a tomada de decisão eficaz. Essa nova necessidade traz oportunidades de negócios para os empreendedores acadêmicos, pois aprendem a partir do acesso e uso de dados e informações.

As informações armazenadas pelas tecnologias dos *Big Data* são também produzidas no contexto da I4.0, principalmente pelos objetos, pessoas e elementos biológicos. Essa quantidade de dados pode ser utilizada nas universidades para estimular o empreendedorismo acadêmico. Os docentes e discentes podem utilizá-los para realizar análises de mercado a fim de identificar as oportunidades e as necessidades que emergem em um contexto crescente.

Isso será possível por meio de informações, anteriormente disponíveis apenas para grandes instituições e agora nas mãos de praticamente todos, dando aos indivíduos e pequenos grupos poder e influência anteriormente reservados a poucos. Essa disseminação de conhecimento tem o potencial inovador, criando uma oportunidade sem precedentes para qualquer pessoa com características empreendedoras. E se o conhecimento é a energia que gera o motor de inovação da academia, essa energia necessita ser livre e disponível para grande parte da população.

## REFERÊNCIAS

ALMADA-LOBO, F. The Industry 4.0 revolution and the future of Manufacturing Execution Systems (MES). *Journal of Innovation Management*, n. 3, vol. 4, p. 16-21, 2015.

- AMORIM, E. A “Indústria 4.0” e a sustentabilidade do modelo de financiamento do Regime Geral da Segurança Social. *Cadernos de Direito Actual*, v. extraordinário, n. 5, p.243-254, 2017.
- BRIETMAN, K. Big Data Overview. Palestra Apresentada no 1o. EMC Summer School on Big Data. EMC/NCE/UFRJ. Rio de Janeiro. 2013.
- CLARK, Burton R. *Creating Entrepreneurial Universities: Organizational Pathways of Transformation*. Issues in Higher Education. Elsevier Science Regional Sales, 665 Avenue of the Americas, New York, NY 10010 (paperback: ISBN-0-08-0433545; hardcover: ISBN-0-08-0433421, \$27)., 1998.
- COSTA, C. Indústria 4.0: o futuro da indústria nacional, *POSGERE*, v. 1, n. 4, set., p. 5-14, 2017. ISSN 2526-4982
- DUTTON, W. H. “Putting things to work: social and policy challenges for the Internet of things”, *info*, Vol. 16 Issue: 3, p.1-21, 2014. <https://doi.org/10.1108/info-09-2013-0047>.
- ETZKOWITZ, Henry. Anatomy of the entrepreneurial university. *Social Science Information*, v. 52, n. 3, p. 486-511, 2013.
- ETZKOWITZ, Henry. *The triple helix: university-industry-government innovation in action*. Routledge, 2008.
- ETZKOWITZ, Henry; VIALE, Riccardo. Polyvalent knowledge and the entrepreneurial university: A third academic revolution? *Critical Sociology*, v. 36, n. 4, p. 595-609, 2010.
- FONSECA, J. J. S. *Metodologia da pesquisa científica*. Fortaleza: UEC, 2002. Apostila.
- GIL, A. C. *Métodos e técnicas de pesquisa social*. 6.ed. São Paulo: Atlas, 2012.
- GOLDMAN, A., et al. Apache Hadoop: Conceitos Teóricos e Práticos, Evolução e Novas Possibilidades. *XXXI JORNADAS DE ATUALIZAÇÕES EM INFORMÁTICA*, 2012.
- GOECKY, D. et al. Human-machine-interaction in the industry 4.0 era. Proceedings - 2014 12th IEEE International Conference on Industrial Informatics, INDIN 2014, p. 289–294, 2014.

- JO, H. J. *The Hyundai Way: The Evolution of a Production Model. Global Asia*, 2010.
- LEE, B.-H.; JO, H.-J. The mutation of the Toyota Production System: adapting the TPS at Hyundai Motor Company. *International Journal of Production Research*, v. 45, n. 16, p. 3665–3679, 15 ago. 2007.
- LETOUZÉ, E. *Big Data for Development: Challenges & Opportunities*. UN Global Pulse, 2012.
- LUVIZAN, S. L.; MEIRELLES, F. S.; DINIZ, E. H. Big Data: publication evolution and research opportunities. *11th International Conference on Information Systems and Technology Management – CONTECSI* May, 28 to 30, 2014 - São Paulo, Brazil.
- MARCONI, M. de A.; LAKATOS, E. M. *Técnicas de pesquisa*. São Paulo: Atlas, 2013.
- NUNES, F. L. DE; MENEZES, F. M. Sistema Hyundai de Produção e Sistema Toyota de Produção: Suas Interações e Diferenças. *Revista Acadêmica São Marcos*, v. 4, n. 2, p. 101–120, 2014.
- PIMENTA, R. M. Big Data e controle da informação na era digital: tecnogênese de uma memória a serviço do mercado e do estado. *Tendências da Pesquisa Brasileira em Ciência da Informação*, v. 6, n. 2, jul./dez. 2013.
- RASMUSSEN, Einar; WRIGHT, Mike. How can universities facilitate academic spin-offs? An entrepreneurial competency perspective. *The Journal of Technology Transfer*, v. 40, n. 5, p. 782-799, 2015.
- RIBEIRO, C. J. S. Big Data: os novos desafios para o profissional da informação. *Informação & Tecnologia (ITEC)*, João Pessoa/Marília, v. 1, n. 1, p. 96-105, jan./jun., 2014.
- RUSSWURM, S. Industry 4.0 - from vision to reality. *Background Information*, p. 1, 2014.
- SALTIEL, R. M. F.; NUNES, F. L. A Indústria 4.0 e o sistema Hyundai de produção: suas interações e diferenças. *Anais do V Simpósio de Engenharia de Produção – SIMEP*, 2017 - ISSN: 2318-9258.
- SCHWAB, K. *The fourth industrial revolution*. Crown Business, New York, 2016.

SHANE, S. A. *Academic entrepreneurship: University spinoffs and wealth creation*. Edward Elgar Publishing, 2004.

TISLUCK, G.; MOSCONI, E.; CHAMBERLAND-TREMBLAY, D. Divergence, convergence et confusion au sein des processus de l'intelligence d'affaire et de l'intelligence compétitive: une analyse des implications pour les chercheurs et les praticiens. In: Association des Sciences Administratives au Canada (ASAC), 2015. *Proceedings* [...]. Montreal, Québec, Canada.

THORP, H.; GOLDSTEIN, B. The entrepreneurial university. *Retrieved January*, v. 19, p. 2016, 2010.

WRIGHT, M. *Academic entrepreneurship in Europe*. London, UK. Edward Elgar Publishing, 2007.

# ALGORITMOS GENÉTICOS E APRENDIZAGEM: QUEM, DE FATO, APRENDE?

*Monica Aiub*

Considerando *Big Data*, tal como a definição do *Oxford English Dictionary*, uma quantidade de dados tão grande que sua manipulação exige desafios capazes de levar a mudanças nas formas tradicionais de análise de dados, a questão que se coloca não se refere à quantidade de dados, mas como analisá-los. A forma tradicional de análise e interpretação dos dados, diante de uma grande quantidade, torna-se pouco efetiva. Para otimizar o processo de análise e interpretação de dados, atingindo mais rapidez e precisão, são utilizados métodos automatizados, especialmente estatísticos e de aprendizado de máquina (*machine learning*). A principal tarefa destes métodos é a mineração de dados (*data mining*), que ocorre a partir de observação de padrões ou estabelecimento de relações entre dados. Algoritmos genéticos – métodos de otimização e busca introduzidos por John Holland (1975) e popularizados por David Goldberg (1988) – têm se apresentado como uma

possível resposta, entre outras, para a análise dessa grande e complexa quantidade de dados.

Os algoritmos genéticos são baseados na evolução biológica, mais especificamente, na genética populacional. Inspirados na seleção natural, funcionam a partir do modelo de sistemas adaptativos, buscando as melhores soluções ou as soluções aceitáveis para os problemas através de conhecimentos acumulados sobre os mesmos. Ao invés da codificação de um parâmetro de solução, trabalham com conjuntos de parâmetros; não consideram indivíduos, mas populações; utilizam-se de informações sobre custos e recompensas, e de regras probabilísticas e não determinísticas. Ou seja, tentam replicar os processos da evolução biológica, não apenas apresentando um passo a passo determinístico, mas atuando a partir de cálculos de probabilidade. Consideram, portanto, os elementos históricos referentes a origem do problema e formas anteriormente testadas para solução como fontes de conhecimento que integram o cálculo probabilístico. Tentam, com isso, emular o processo de evolução das espécies.

Seus principais componentes são nomeados numa linguagem composta por termos também utilizados em Biologia, tais como: cromossomo, que corresponde à sequência de códigos que representa uma possível solução. Um cromossomo, como em Biologia, é formado por genes, que aqui correspondem a cada parâmetro codificado no cromossomo. Outro termo é população, que se refere ao conjunto de soluções do problema. A população pode ser escolhida aleatória ou heurísticamente. São também utilizados termos como: avaliação ou função de aptidão, que diz respeito à medida do grau de aptidão de um indivíduo (qualidade quanto à solução ótima);

e seleção, que consiste em copiar boas soluções e eliminar soluções de baixa aptidão.

Na linguagem dos algoritmos genéticos, incluem-se, ainda: operadores genéticos – que visam transformar a população através de sucessivas gerações, objetivando resultado satisfatório; critérios de parada – estabelecidos por resultados ou por tempo previamente determinado; e parâmetros de controle – que correspondem a tamanho da população, taxa de cruzamento, taxa de mutação, número de gerações.

Algumas pesquisas apresentam algoritmos capazes de, por exemplo, reconhecer rostos (Universidade de Hong Kong); diagnosticar depressão (Universidade de Vermont e Universidade de Harvard); identificar sarcasmo (Universidade Hebraica); determinar personalidade (Universidade de Stanford e Universidade de Cambridge), resumir textos (empresa Salesforce); identificar usuário bêbado (Universidade de Rochester); produzir arte criativa (Universidade de Rutgers); identificar orientação sexual por análise facial (Universidade de Stanford); decodificar pensamentos (Universidade da Califórnia), entre outras atividades anteriormente consideradas exclusividade humana. Afirmam os textos de divulgação destas pesquisas que os resultados obtidos com o uso dos algoritmos são mais precisos que a análise humana.

## **1 OBSERVANDO ALGUMAS PESQUISAS QUE UTILIZAM ALGORITMOS GENÉTICOS**

Wu Youyou, Michal Kosinski e David Stillwell (2015), numa parceria entre as universidades de Cambridge e Stanford, desenvolveram um estudo que compara a precisão dos julgamentos de personalidade humanos com aqueles baseados

em algoritmos genéticos. Eles trabalharam com uma amostra de 86.222 voluntários, que preencheram um questionário com cem itens. A conclusão da pesquisa indica que as predições computacionais baseadas em algoritmos genéticos são mais precisas do que as predições feitas pelos amigos de facebook dos participantes.

Os pesquisadores utilizaram o Questionário de personalidade de cinco fatores (IPIP), medindo as características – abertura, consciência, extroversão, amabilidade e neuroticismo ou instabilidade emocional – pelo método lexical. Com 70.520 participantes, foram observados os *likes* no facebook. A associação de padrões de gosto (*likes* no facebook) e traços de personalidade geraram resultados mais precisos que a avaliação feita por amigos de facebook. Para a avaliação humana, foram utilizados dez itens propostos a partir do Questionário de personalidade de cinco fatores. Os critérios para avaliação, em ambos os casos, foram a autoavaliação do voluntário, o cruzamento entre avaliações de humanos (neste caso foram considerados 14.410 voluntários) e a validade externa.

Afirmam os autores que as ferramentas de avaliação de personalidade automatizadas, sendo mais precisas e baratas, podem afetar a sociedade de várias maneiras: marketing adaptado aos usuários; trabalho de recrutadores facilitado; produtos e serviços mais adequados a seus usuários e, para a ciência, coleta de dados de personalidade sem sobrecarregar os participantes com longos questionários. O mais curioso é a afirmação sobre, no futuro, as pessoas poderem abandonar seus próprios julgamentos e confiar mais em computadores para tomar decisões importantes na vida, como escolher atividades, carreira ou até mesmo parceiros para um relacionamento afetivo (YOUYOU; KOSINSKI; STILLWELL, 2015, p. 1039).

Andrew G. Reece e Christopher M. Danforth (2016), das universidades de Harvard e Vermont, pesquisaram o instagram de 166 pessoas e utilizaram algoritmos genéticos para identificar marcadores de depressão. Foram analisadas 43.950 fotos, sendo observados: características das fotos, como cores, saturação, brilho; metadados, como comentários e *likes*; frequência de *posts*; e presença de pessoas nas imagens, identificada com detecção algorítmica de faces. Suas conclusões indicam que pessoas com depressão tendem a postar fotos mais escuras, em tons de azul ou cinza; indicam uma relação inversa entre o número de comentários nas fotos e os índices de depressão, assim como uma relação direta entre a frequência de *posts* e a depressão; maior incidência de fotos de rosto (*selfies*) entre os depressivos. Os resultados, segundo os pesquisadores, superam o sucesso da média dos diagnósticos médicos. Os pesquisadores apontam novos caminhos para detecção precoce de transtornos mentais e afirmam que o programa possibilita diagnosticar com maior precisão e rapidez que os médicos.

Ahmed Elgammal, Bingchen Liu, Mohamed Elhoseiny, Marian Mazzone (2017), do Laboratório de Arte e IA da Universidade de Rutgers, inspirados na obra de Colin Martindale (1990) sobre a necessidade constante de inovação na arte, produziram um software que aprende observando obras de arte e é capaz de produzir obras de arte criativa. As obras criadas pelo programa foram apresentadas à avaliação humana e foi impossível a distinção entre as obras criadas por artistas e as criadas pelo software, para 53% dos voluntários. Os pesquisadores criaram dois algoritmos em competição: um algoritmo gerativo, que gerou aleatoriamente imagens com padrões e formas distintas; e um algoritmo discriminativo, cuja função era avaliar e selecionar as obras adequadas ao padrão de referência. Quando o algoritmo discriminativo passou a iden-

tificar como adequadas ao padrão as obras “criadas” pelo algoritmo gerativo, iniciou-se novo passo no projeto. O algoritmo discriminativo foi alimentado com diferentes estilos de arte, a fim de criar um padrão de identificação do que é comumente aceito como arte. O algoritmo gerativo tentava criar conforme os padrões e o algoritmo discriminativo classificava. A competição entre os algoritmos funcionava segundo o seguinte critério: se o algoritmo discriminativo não conseguisse identificar o estilo da obra criada, o algoritmo gerativo ganhava pontos; se o algoritmo gerativo criasse obras facilmente classificadas, perdia pontos. Na segunda fase da pesquisa, as obras foram avaliadas por humanos.

## **2 ALGORITMOS GENÉTICOS NA APRENDIZAGEM**

O desenvolvimento e a disseminação do ensino à distância trouxeram a necessidade de softwares educacionais que pudessem atender a demandas desta modalidade de ensino. A fim de abarcar o grande campo de probabilidades presente na área educacional, algoritmos genéticos também são utilizados na construção destes softwares. A ideia inicial é que fossem instrumentos para auxiliar os estudantes no desenvolvimento de seu processo de aprendizagem.

Algoritmos genéticos utilizados em cursos de educação à distância auxiliariam no processo de aprendizagem ao indicarem objetos de aprendizagem mais adequados a cada estudante, recomendando objetos sensíveis ao contexto (JÁCOME JÚNIOR; MENDES NETO; SILVA, 2012). A educação à distância, muitas vezes, dada a sua abrangência – principalmente a partir da educação móvel (utilizada em dispositivos sem fio) –, para facilitar o acesso aos conteúdos educacionais, acaba por apresentar conteúdos inadequados

aos usuários. A fim de solucionar este problema, são utilizados algoritmos genéticos, para que sejam oferecidos aos estudantes objetos de aprendizagem mais adequados.

Num primeiro momento, esta poderia parecer uma forma interessantíssima para lidar com uma das questões fundamentais em educação: a necessidade de adequação dos conteúdos ao contexto e à linguagem do estudante. Contudo, contexto, neste caso, não se refere essencialmente à realidade do estudante, que necessitaria de pesquisas muito mais detalhadas, mas à localidade onde ele se encontra e outros elementos, tais como as condições do equipamento que ele utiliza. Para cada localidade há um repositório de objetos de aprendizagem que serão oferecidos ao estudante. Assim, além de escolher seus objetos de aprendizagem, o estudante terá um ambiente virtual criado para ele, a partir dos dados coletados no repositório, considerando-se outros usuários com características similares e padrões de escolhas anteriores. Isto possibilitaria, segundo os pesquisadores (JÁCOME JÚNIOR; MENDES NETO; SILVA, 2012), a criação de um micromundo virtual e o desenvolvimento de autonomia, criatividade, atenção e memória do usuário.

Hoje há plataformas, softwares e jogos educacionais que não são trabalhados apenas nos cursos de educação à distância, mas são utilizados em cursos presenciais e apresentados como um diferencial pedagógico em muitas escolas. Podemos citar como exemplo, entre tantos outros, a plataforma *Létrus*, que consiste num programa que objetiva melhorar a capacidade de escrita dos estudantes, visando a nota de redação no ENEM. Com base em “milhares de textos já escritos”, ou seja, um banco de dados com as redações com excelente pontuação, o programa começa por identificar o perfil de escrita de

cada estudante, oferecendo orientações “personalizadas” para melhorar a performance do aluno. Este programa é apresentado como diferencial em várias escolas particulares e utilizado na terceira série do Ensino Médio em algumas escolas da Rede Pública Estadual de São Paulo. Assim como ele, vários programas já foram incorporados ao cotidiano educacional, os professores também interagem com o programa, que “aprende” e “incorpora” novos elementos.

É preciso, contudo, considerar que, ainda que os algoritmos genéticos trabalhem com probabilidades e que possam “aprender” com o usuário do software, suas bases iniciais estabelecem limites de atuação, traçados pelo programa. Tais limites, se comparados às possibilidades humanas, são, ainda, muito restritos. Porém, a despeito disso, alguns softwares e aplicativos educacionais são usados mais como formas de adestramento do que como processos múltiplos e variados de aprendizagem. Ou seja, muitas vezes, ao invés de desenvolver a habilidade reflexiva e investigativa, tais programas são utilizados de modo a gerar padrões específicos de associação de dados, tornando tais padrões, através da repetição, hábitos incorporados dos estudantes.

Se ao invés de desenvolver suas habilidades de escrita o estudante aprender um formato padrão para escrever seus textos, e apenas reproduzi-lo, impedindo-se de explorar a criatividade na escrita, muito brevemente haverá programas capazes de “redigir textos” melhor do que muitos humanos. Se inserirmos aqui os dados acerca das pesquisas citadas na sessão anterior, especialmente o programa capaz de gerar arte criativa, poderíamos nos questionar: estaríamos gerando programas capazes de criar e nos surpreender, ao mesmo tempo em que

nos “adestramos” para responder conforme o esperado pelos programas utilizados para nos avaliar?

Até mesmo as interações humanas que se dão por estas interfaces são estabelecidas a partir dos padrões previamente determinados por elas, isto é, ao interagir com outros usuários de um aplicativo, programa ou jogo, há limites e possibilidades estabelecidos pelo software para esta interação. Além disso, cada vez que um aplicativo é utilizado, novos dados dos usuários são coletados, incluindo-se as formas de associação de dados; tais dados são considerados “capital” da empresa proprietária do aplicativo.

Se considerarmos as recentes mudanças em nossa Lei de Diretrizes e Bases da Educação Nacional (LDBN), permitindo que parte da carga horária seja à distância, inclusive na Educação Básica; as discussões no Senado acerca de educação domiciliar (*homeschooling*); a 4ª Revolução Industrial, com a aposta da paulatina substituição de profissionais (tais como professores, médicos, advogados, engenheiros etc.) por softwares, que rumos estaríamos traçando para a educação? Seriam tais rumos uma revolução educacional ou a continuidade de um processo de precarização na educação?

### **3 UMA LEITURA A PARTIR DE PEIRCE E EDELMAN**

A Teoria da Seleção dos Grupos Neurais ou Darwinismo Neural, de Gerald Edelman (1987), também inspirada na teoria evolutiva de Darwin, é baseada em três princípios. O primeiro é a seleção no desenvolvimento, ou seja, um processo que conduz à formação da neuroanatomia. Esta, apesar de característica da espécie, possui um enorme grau de variação, no qual as informações transmitidas por

herança genética constituirão, a partir do processo de desenvolvimento do feto, um repertório primário de categorização das experiências. Assim, o desenvolvimento neuronal do feto gerará, mesmo em caso de gêmeos univitelinos, cérebros diferentes ao nascer.

No segundo princípio, as conexões sinápticas são fortalecidas ou enfraquecidas por processos bioquímicos. O conjunto de variantes dos circuitos neuronais é chamado de repertório secundário. O terceiro princípio “sugere que os mapas cerebrais interagem através de um processo chamado reentrada” (EDELMAN, 1995, p. 128).

Tenciono provar aqui que a seleção somática, atuando sobre as cartografias globais, com novos tipos de cartografias juntando-se aos anteriores ao longo da evolução, é um poderoso meio de aquisição de novas funções tais como memórias especializadas e capacidades conceituais (EDELMAN, 1995, p. 148).

Poderíamos considerar a proposta de Edelman como um elemento favorável ao desenvolvimento de algoritmos genéticos que adquirissem, a partir da construção de um repertório secundário, novas funções, tais como “memórias especializadas e capacidades conceituais”. Para tal, precisaríamos criar padrões como se houvesse uma natureza regular, ordenada. O próprio Edelman afirma que o padrão da natureza é a variação, a diversidade. O estabelecimento de padrões e repetições é próprio da mente humana. Observamos padrões naturais porque eles se destacam em meio à diversidade, esta é observada como “a música nunca ouvida porque sempre ouvida”.

O pensamento em termos de populações considera que a variabilidade não é um erro, mas sim, como diz o grande evolucionista Ernst Mayr, um fenômeno

real. A variabilidade individual de uma população é a fonte de diversidade, sobre a qual atua a seleção natural para produzir diferentes tipos de organismos (EDELMAN, 1995, p. 111).

Peirce também considera a variação da natureza:

A natureza não é regular. Nenhuma desordem é menos ordenada do que a combinação existente. É verdade que as leis particulares e as regularidades são incontáveis; mas ninguém pensa sobre as irregularidades que são infinitamente mais frequentes. Cada fato verdadeiro sobre qualquer coisa no universo está relacionado a qualquer outro fato verdadeiro. Mas a grande maioria destas relações são fortuitas e irregulares (PEIRCE, CP, 5.342).<sup>1</sup>

A constatação da variabilidade, que segundo Edelman é responsável pelo desenvolvimento, na espécie, de memória e pensamento conceitual, exige um mesmo processo de variação e irregularidade para que um algoritmo pudesse obter resultados tão ou mais precisos que os resultados humanos. Isto porque, ainda que o algoritmo trabalhe com uma quantidade de dados extremamente grande, os padrões de relação observados poderão ser fortuitos, ou seja, apenas coincidências observadas, mas que não apresentam uma relação de fato; ou modificados pela ação do acaso ou por interações diversas. Poderíamos argumentar que não temos condições de processar e relacionar tantos dados quanto um programa é capaz de fazer e, por ter acesso a mais dados, o resultado obtido pelo programa seria mais preciso.

---

<sup>1</sup> A obra de Peirce está citada conforme a convenção: CP (*Collected Papers*). Os números que seguem as letras correspondem, respectivamente, a volume e parágrafo.

Porém, os padrões de relação estabelecidos por algoritmos genéticos são pautados em estatística, ou seja, a partir do método indutivo. O que se observa a partir do método indutivo não pode ser aplicado a cada indivíduo da espécie pesquisada. Assim, não é possível considerar as relações estabelecidas por algoritmos genéticos como processos de categorização universais. Como esclarece Queiroz:

Se lidamos com métodos indutivos, então não podemos garantir universalidade na aplicação das categorias. Metodologicamente, o problema é óbvio. Não é possível, por observação, construir uma lista na qual qualquer coisa que se possa pensar ou experimentar tenha lugar (QUEIROZ, 2004, p. 34).

As formas humanas de categorização dos dados variam de acordo com nossos repertórios primário e secundário e com a reentrada. Ou seja, a cada situação vivida, a cada relação estabelecida com o mundo ou com o outro, os próprios padrões que utilizamos para categorizar nossas experiências podem ser modificados, é o que chamamos plasticidade.

Peirce afirmou que “A mais plástica de todas as coisas é a mente humana e, depois dela, vem o mundo orgânico [...]. Ora, a tendência generalizante é a grande lei da mente, a lei da associação, a lei de adquirir hábitos” (CP, 7.515).

A aprendizagem é, segundo Peirce, a realização da plasticidade, ou seja, a possibilidade que temos de modificar nossos hábitos diante das necessidades colocadas por nossas relações com o mundo. Neste caso, não apenas os dados do mundo são diversos, mas a mente humana modifica seus hábitos de categorização e associação dos dados a partir das experiências, ela é capaz de aprender com as experiências e os erros.

Os algoritmos genéticos, em sua construção, utilizam regras probabilísticas. A partir delas, categorizam e relacionam dados, tais como cores e saturação de uma foto e depressão – conforme pesquisa descrita na sessão anterior.

A própria probabilidade é uma ideia essencialmente imprecisa, exigindo, no seu uso toda a precaução do pragmatismo, no qual sua origem indutiva deve ser firmemente mantida como se fosse uma bússola pela qual devemos guiar com segurança nosso barco neste oceano de probabilidades. A indução poderia ser definida, em termos precisos, como sendo a inferência virtual de uma probabilidade, se é que a probabilidade pode ser definida sem a ideia da indução (CP, 2.101).

Neste “oceano de probabilidades” que é o real, não podemos esquecer o caráter impreciso da indução. Ainda que apontássemos as relações observadas por dados estatísticos como um processo abduutivo, faltaria a tal processo a capacidade de interagir com a rede de signos e a diversidade de significados que podem ser atribuídos; faltaria, ainda, a interação e movimentação constante entre os métodos dedutivo, indutivo e abduutivo.

#### **4 ALGUNS QUESTIONAMENTOS**

A partir destas observações, surgem alguns questionamentos. Não temos respostas para eles, mas colocá-los poderá nos provocar a pensar acerca do tipo de vida e de sociedade que estamos construindo.

Que tipo de aprendizagem possui um estudante que aprende, com seu professor do cursinho, como escrever uma redação segundo os critérios do algoritmo utilizado para a

correção da mesma? E o estudante que aprende com programas como o Létrus? Se considerarmos aprendizagem a realização da plasticidade, estaríamos moldando nossas mentes, como próteses cognitivas a partir das quais conduziremos nossa existência? Seriam os algoritmos genéticos estímulos ao desenvolvimento de nossas habilidades cognitivas ou a replicação de padrões e modos de vida úteis à manutenção da ordem vigente?

O estudante de música que aprende a “compor” a partir do aplicativo “compositor” sem conhecer harmonia musical, teoria ou acústica, sem compreender os significados das relações entre a música e a sociedade será um compositor ou um programa o superará na qualidade das composições?

Correríamos o risco, ao criarmos máquinas capazes de escrever livros, compor e executar músicas, criar obras de arte, ou seja, atividades consideradas exclusividade humana, de nos tornarmos dependentes delas e, conseqüentemente, dependentes dos detentores da propriedade de tais máquinas?

Se nossos saberes e capacidades criativas podem se tornar “propriedade” de uma empresa, que tipo de sociedade estamos criando? E nela, quais os limites de nossas ações e de nossa criatividade?

Poderíamos, de fato, considerar que no uso de aplicativos e/ou jogos, construídos a partir de algoritmos genéticos há aprendizagem? Se há, quem, de fato, aprende? O usuário, o aplicativo ou a empresa proprietária do aplicativo?

Se o programa for uma ferramenta, um instrumento de apoio para auxiliar estudantes, professores, pesquisadores, profissionais etc., ele poderá ser muito útil, principalmente porque nos oferecerá acesso a uma quantidade maior

de dados. Contudo, esses dados já serão “filtrados” segundo critérios previamente estabelecidos e, conseqüentemente, o resultado ficará limitado a este “filtro”. Se pensarmos numa educação desenvolvida exclusiva ou majoritariamente tendo programas como interface, perderemos a perspectiva da diversidade, da variabilidade que o encontro e o convívio propiciam.

Se vistos como instrumentos, *Big Data*, algoritmos genéticos, mineração de dados, programas etc. – ou seja, o desenvolvimento de tecnologias e inteligência artificial não têm o poder de revolucionar, pois apenas nos auxiliarão na continuidade de nossos processos.

Se vistos como suficientes para desempenhar atividades humanas melhor que os humanos, também não serão revolucionários, pois apenas intensificarão o processo de precarização da existência.

## REFERÊNCIAS

AMARAL, F. *Aprenda mineração de dados*. Rio de Janeiro: Alta Books, 2016.

EDELMAN G. M. *Biologia da Consciência*. Lisboa: Instituto Piaget, 1995.

EDELMAN G. M. *Neural Darwinism: The Theory of Neuronal Group Selection*. New York: Basic Books, 1987.

ELGAMMAL, A.; LIU, B.; ELHOSEINY, M.; MAZZONE, M. CAN: Creative Adversarial Networks Generating “Art” by Learning About Styles and Deviating from Style Norms. In *arXiv:1706.07068v1* [cs.AI] 21 Jun 2017. Disponível em: <https://arxiv.org/pdf/1706.07068.pdf>. Acesso em 15/04/2018.

GOLDBERG, D.E. *Genetic Algorithms in Search, Optimization, and Learning*. Boston: Addison-Wesley, 1989.

HOLLAND, J. H. *Adaptation in Natural and Artificial Systems*. Ann Arbor: University of Michigan Press, 1975.

JÁCOME JÚNIOR, L.; MENDES NETO, F. M.; SILVA, L. C. N. Uma abordagem baseada em algoritmo genético para recomendação de objetos de aprendizagem sensível ao contexto do estudante. *Anais do 23o Simpósio Brasileiro de Informática na Educação (SBIE 2012)*, ISSN 2316-6533. Rio de Janeiro, 26-30 de Novembro de 2012. Disponível em: <http://www.br-ie.org/pub/index.php/sbie/article/viewFile/1673/1435>. Acesso em 15/04/2018.

MARTINDALE, C. *The clockwork muse: The predictability of artistic change*. New York, NY, US: Basic Books, 1990.

MEISTER, I. *A terceira margem: O conhecimento nas redes sociais*. São Paulo: Mackenzie, 2014.

OXFORD ENGLISH DICTIONARY [Internet]. Oxford: Oxford University Press; 2015. Big Data; [cited 2015 Apr 17]; [1 paragraph]. Available from: <http://www.oed.com/view/Entry/18833>.

PEIRCE, C.S. *Collected Papers of Charles Sanders Peirce*. (CP). CD-ROM past masters. Charlotterville: InteleX Corporation, 1992.

QUEIROZ, J. *Semiose segundo Peirce*. São Paulo: EDUC/PUC, 2004.

REECE, A. G.; DANFORTH, C. Instagram photos reveal predictive markers of depression. In *EPJ Data Science*, 6, 2016. Disponível em: <https://www.researchgate.net/publication/306186650/download>. Acesso em 15/04/2018.

YOUYOU, W.; KOSINSKI, M.; STILLWELL, D. Computer-based personality judgments are more accurate than those made by humans. *PNAS* January 27, 2015 112 (4) 1036-1040. Disponível em: <http://www.pnas.org/content/112/4/1036>. Acesso em 15/04/2018.

**REGIME DE INFORMAÇÃO E ANÁLISE DO  
DISCURSO: APROXIMAÇÕES TEÓRICAS E  
CONCEITUAIS NA ERA  
*BIG DATA***

*Mariana da Silva Caprioli  
Larissa de Mello Lima  
João Batista Ernesto de Moraes*

Frohmann (1995) sinaliza para o fato de como a informação é materializada através do meio institucional. Ele introduz o conceito de “Regime de Informação”, explicando que este possui uma estrutura polêmica na medida em que é formada por grupos sociais, artefatos científicos, tecnológicos e discursos. Denomina essa estrutura de híbrida porque carrega olhar discursivo e social simultaneamente.

Este viés é importante para ser explorado uma vez que a Ciência da Informação também convive com uma multiplicidade de agentes de caráter híbrido, responsáveis por fomentar a área em contexto nacional. Exemplos desses agentes podem ser identificados a partir da Unesco, da Fundação

Getúlio Vargas e do então Instituto Brasileiro de Bibliografia e Documentação (IBBD, atual IBICT), ou seja, a partir de um respaldo institucional, há a caracterização de instâncias de validação científica propriamente discursivas e institucionais que impulsionaram o nascimento da Ciência da Informação no Brasil, na década de 1970. Interligar a perspectiva institucional com a discursiva é ater-se ao conceito de “Regime de Informação” apresentado por Frohmann.

A partir de Frohmann (1995), González de Gómez (2012) pontua que a criação do então IBBB, atual IBICT, indicou o surgimento de um novo regime de informação no Brasil, uma vez que sua criação foi oriunda de uma parceria da Unesco com a Fundação Getúlio Vargas, por intermédio de seu representante no Brasil, Paulo Carneiro, o qual indicou Lydia de Queiroz Sambaquy, juntamente com Jannice Monte-Mór, para percorrerem as principais bibliotecas e centros de documentação da Europa e Estados Unidos durante quase um ano, viagem essa, que tinha a finalidade de fomentar o estabelecimento de centros bibliográficos nacionais, beneficiando-se da política da Unesco.

Tal ação contou com o apoio da Fundação Getúlio Vargas e serviu para ampliar as ideias de Lygia de Queiroz Sambaquy sobre a estruturação que um órgão que fosse dedicado a trabalhos bibliográficos deveria possuir para impulsionar atividades científicas e tecnológicas do país. Tal criação confirma a teoria de regime de informação.

Um regime de informação seria o modo informacional dominante em recursos, uma formação social, o qual define quem são os sujeitos, as organizações, as regras e as autoridades informacionais e quais os meios e os preferenciais de informação, os padrões de excelência e os modelos de sua organização, interação

e distribuição, enquanto vigentes em certo tempo, lugar e circunstância (GÓMEZ, 2012, p 15).

O novo regime de informação que surgia, caracterizado pela centralidade do fenômeno informacional, apenas viria a se desenvolver plenamente ao longo das décadas seguintes, mas nesse momento já definia suas diretrizes, pautado no surgimento do IBBD e nas novas posturas profissionais.

Paralelo a esse regime de informação, surgia, com o advento da Web 2.0, o fenômeno dos *Big Data*:

Em 1980, surgiram os conceitos de Data Warehouse, até chegar à ideia de Big Data, o que ocorreria em 1990. A diferença de Data Warehouse e Big Data é a velocidade com que os dados precisam ser disponibilizados, uma vez que, em um projeto de Data Warehouse, o processo de ETL torna-se mais lento até que as informações estejam disponíveis. Segundo Taurion (2013), o Big Data está muito além de um grande Data Warehouse. Entretanto, pode-se dizer que o Big Data se relaciona com o Data Warehouse (CALDAS; SILVA, 2016).

A Análise do Discurso de matriz francesa, preocupada com o contexto em que tal discurso está sendo construído, com os atores e entidades envolvidos, relaciona-se com o Regime de Informação, uma vez que:

[...] pode-se concluir que o regime de informação seria aquele que reúne atores e artefatos em práticas de informação que giram em torno de um interesse comum e em cujas relações se legitimam políticas explícitas ou implícitas que direcionam e ordenam essas práticas. Representa um conceito estrutural e estruturante para a Ciência da Informação porque não se prende à informação em si mesma, mas ao

real objeto de estudo da disciplina, ou seja, ao conjunto de relações que a informação estabelece ao redor dos múltiplos interesses de atores e pela adoção específica de artefatos direcionados pelas práticas informacionais desses sujeitos (MAGNANI; PINHEIRO, 2011, p.16).

## **1 FORMAÇÕES DISCURSIVAS E IDEOLÓGICAS E O REGIME DE INFORMAÇÃO**

Pode-se observar o regime de informação de duas formas: o autor, inserido na sociedade, com suas vivências e formações, escreve seu discurso dentro de suas próprias leis de vigência; o poder estando apenas em suas mãos, mas ao mesmo tempo atrelado à sua formação discursiva e ideológica, sendo esses, então, uma forma de controle, de onde o discurso sai “condicionado”. É na Formação Discursiva que se constitui o sentido e a identificação do sujeito. É onde todo sujeito se reconhece, por meio de sua relação consigo mesmo e com outros sujeitos, é onde adquire identidade (ORLANDI, 2008). É o que pode ou não ser dito em uma determinada Formação Ideológica, dessa forma, relacionando ambas.

Então pode-se dizer que a Formação Discursiva é aquilo que, em uma Formação Ideológica dada, articula o que pode ser dito, levando em consideração o contexto em que os sujeitos estão inseridos. Orlandi (2009) dizia que as palavras (ou discursos, aqui) não têm sentido nelas mesmas, mas sim derivam de suas formações discursivas, onde estão inseridas no contexto, ou seja, têm seu suporte em uma formação ideológica.

## 2 METODOLOGIA

A metodologia empregada utiliza o aporte da Análise do Discurso de matriz francesa que apresenta uma definição extremamente estruturada. Segundo Orlandi (1999, p.15):

A Análise de Discurso, como seu próprio nome indica, não trata da língua, não trata da gramática, embora todas essas coisas lhe interessem. O discurso é assim palavra em movimento, prática de linguagem: com o estudo do discurso observa-se o homem falando.

De tal forma, a autora leva a observar que a Análise do Discurso não trabalha com a língua como um sistema abstrato, mas com está no mundo.

Dessa forma, Brandão (2004) procurou entender a Análise do Discurso de matriz Francesa em seu ponto de surgimento, quando suas bases iniciais se debruçavam sobre os discursos políticos com posições bem definidas, como os debates entre direita e esquerda, e era definida como “o estudo linguístico das condições de produção de um enunciado”.

A partir desse procedimento, passou-se a entender que a Análise do Discurso de matriz francesa, não se limita a estudos linguísticos, tratando-se de analisar somente a parte gramatical da língua, como também leva em conta os aspectos externos da língua como os elementos históricos, sociais, culturais, ideológicos que permeiam a produção de um discurso e que, por tal motivo, nele se refletem. Tais elementos externos são partes essenciais de uma abordagem discursiva.

Tem-se então:

Um conceito fundamental para a AD é, dessa forma, o de condições de produção, que pode ser definido como o conjunto dos elementos que cerca a

produção de um discurso: o contexto histórico-social, os interlocutores, o lugar de onde falam, a imagem que fazem de si, do outro e do assunto de que estão tratando. Todos esses aspectos devem ser levados em conta quando procuramos entender o sentido de um discurso (BRANDÃO, 2004, p. 6).

O discurso está em constante influência, como visto, de elementos que o cercam, e dessa forma, observa-se que o contexto institucional é de grande relevância no momento da análise de um discurso. Pontuando acerca dessa perspectiva, Pêcheux (1997) nos lembra que se ater ao contexto é levar em conta uma sequência linguística flexível na medida em que:

É impossível analisar um discurso como um texto, isto é, como uma sequência linguística fechada sobre si mesma, mas [que] é necessário referi-lo ao conjunto de discursos possíveis a partir de um estado definido de condições de produção (PÊCHEUX, 1997, p.74).

Estar sofrendo constante influência do contexto em que se está inserido leva à definição de Formação Discursiva e Ideológica, tão presentes na A.D. e que permitem a melhor compreensão acerca do tema. Dessa forma,

Chamaremos, então, formação discursiva aquilo que, em uma formação ideológica dada, isto é, a partir de uma posição dada em uma conjuntura dada, determinada pelo estado da luta de classes, determina “o que pode e o que deve ser dito” (articulado sob a forma de uma alocução, de um sermão, de um panfleto, de uma exposição, de um programa, etc.) (PÊCHEUX 1997, p. 160)

Baseado em um determinado contexto, pautado em determinada Formação Ideológica, a Formação Discursiva determina o que pode ou não ser dito, levando a concluir que um discurso construído dependerá do contexto inserido, ou seja, de quem o fala, como o fala, em que época o fala, entre outros para expressar o que se deseja. Interligar o olhar discursivo com o viés institucional que consolida as práticas científicas é se ater ao conceito de Regime de Informação proposto por Frohmann (1995). Dessa forma, o regime de informação trata de um modo informacional em uma organização social, o qual define quem são, neste contexto, as organizações e em qual contexto estão inseridas.

Assim, observa-se que tanto a Análise do Discurso de matriz francesa, quanto o Regime de Informação tratam do discurso impresso ou digital em um determinado contexto. A primeira importando-se com os elementos que permeiam a formação do discurso, sendo o âmbito histórico-social ou qual a instituição de onde fala esse discurso, ou seus interlocutores, o assunto de que se está tratando; enquanto o Regime de Informação trata-se de um modo informacional em uma organização social, ao qual define quem são, nesta situação, as organizações e em qual contexto estão inseridos.

### **3 RESULTADOS E DISCUSSÕES**

No que tange ao delineamento do conceito de Regime de Informação, Carvalho (2009, p. 214) nos explica que:

Frohmann (1995) introduz o conceito de regime de informação – qualquer sistema ou rede que permite o fluxo de informação, através de estruturas específicas, de canais e produtores a consumidores, e este conceito passa a ser entendido como um elemento passível de

ser utilizado como ponto de partida para a elaboração de políticas nacionais de informação. A partir de vários regimes de informação, pode-se pensar na discussão de uma política nacional de informação.

A partir desta sinalização, em linhas gerais, dos aspectos teóricos que formam um regime de informação, é possível entender esse enquanto uma rede formada por atores que interagem em uma estruturação hierárquica de poder, essa estrutura atual de grande intercâmbio de dados em plataformas diversas. Tal rede delimitada a partir da estrutura hierárquica de poder é formada através do contorno institucional, possuindo viés discursivo na medida em que as instituições legitimam determinadas instâncias enunciativas em detrimento de outras. A validação faz parte de um jogo complexo em que decisões repercutem em escalas relacionadas às instâncias de poder.

Dessa forma, trazendo tal configuração para o contexto da Ciência da Informação em nível nacional, é possível dizer que o conceito de Regime de Informação, de Frohmann (1995), relaciona dinâmicas de poder propriamente discursivas, a partir do aporte institucional que começou a ser delineado na década de 70, com a criação do IBBD pela fundação Getúlio Vargas, por incentivo da Unesco.

O IBBD pode ser caracterizado então como um ator fundamental para a área em contexto nacional. A partir dele, pode-se dizer que a estruturação hierárquica de poder da área em contexto nacional começa a ser delineada. Neste processo é importante situar dois veículos de divulgação científica que nascem a partir do mesmo: o periódico *Ciência da Informação*, considerado um dos periódicos mais importantes da área, foi

criado em 1972 pelo IBBD, que, no mesmo ano, também criou a *Revista da Escola de Biblioteconomia, da UFMG*.

Essa iniciativa por parte do órgão IBBD pode ser entendida como reflexo da criação do primeiro curso de mestrado em Ciência da Informação, em 1970, ou seja, a partir do momento em que são formados pesquisadores na área, passam a ser necessários veículos para divulgar as pesquisas científicas da mesma. É possível dizer que a conjuntura histórica da Ciência da Informação no Brasil possui alicerces estruturados sob o viés institucional. Denominamos por instâncias de validação científica da Ciência da Informação no Brasil as instituições que caracterizam e até mesmo determinam a história da área no país.

É possível notar a relação hierárquica que há entre as instâncias de validação científica, uma vez que tal relação não é aleatória, representando uma série de decisões que foram tomadas e repercutiram em posições na esfera de institucionalização da área. Sendo assim, é possível dizer que tais órgãos funcionam como atores que formam um Regime de Informação da área em contexto nacional. Estas características são discursivas e refletem o “jogo” que se remete à ação ideológica e às instituições que legitimam a atuação dessa ideologia, na medida em que o discurso é uma instância que registra essa atuação.

Frohmann (1995, p. 5) ainda definiu regime de informação como qualquer sistema ou rede mais ou menos estável, onde a informação flui por determinados produtores específicos, via estruturas organizacionais específicas, para consumidores específicos, para ele, “descrever um regime de informação significa mapear os processos agonísticos que resultam em tentativas de estabilização e conflitos entre grupos sociais,

interesses, discursos e até artefatos científicos e tecnológicos” (FROHMAN, 1995, p.5).

*Big Data*, por sua vez, segundo a definição de Boyd e Crawford, trata-se de

um fenômeno cultural, tecnológico e acadêmico baseado na interação de três fatores: (1) Tecnologia: maximização da precisão dos algoritmos e do poder de computação para reunir, analisar, relacionar e comparar grandes conjuntos de dados; (2) Análise: processamento de grandes conjuntos de dados para identificar padrões para atender às necessidades de ordem econômica, social, técnica e legal; e (3) Mitologia: a ampla crença de que grandes conjuntos de dados possibilitam uma forma mais avançada de inteligência e conhecimento que podem gerar insights até então impossíveis de se alcançar, de forma objetiva e confiável (BOYD; CRAWFORD, 2012, p. 2, tradução de LOTT; CIANCONI, 2018).

Sendo o fator (2) talvez, de interesse para o Regime de Informação aqui apresentado, uma vez que as formações discursivas e ideológicas podem auxiliar no tratamento da massiva quantidade de dados na Era de *Big Data*, pois opõem-se a formas tradicionais de análise e processamento ao fazer o analista se deparar com elementos de intercomunicação e interoperabilidade de dados na esfera institucional da Ciência da Informação.

Os conceitos e aplicabilidades da Análise do discurso e do Regime de Informação ao contexto dessa grande quantidade de dados permitem reunir fontes heterogêneas, explorar e fornecer significado a diferentes conjuntos, facilitando a aplicação do processamento semântico (CONEGLIAN; SEGUNDO; SANTANA, 2017), ou seja, a Análise do Dis-

curso conjuntamente com o Regime de Informação extraem volume, velocidade, variedade e valor da avalanche de dados no momento em que mudam a forma de fazer a análise desses, legitimando determinadas instâncias enunciativas em detrimento de outras e, assim, colaborando para que o fluxo de dados seja mais robusto.

## CONCLUSÕES

É possível pontuar que o conceito de Regime de Informação de Frohmann (1995) relaciona a perspectiva institucional com a discursiva. Ela torna possível a compreensão das instâncias de validação científica da Ciência da Informação no Brasil, situadas enquanto atores que relacionam dinâmicas de poder da área no país, podendo ser uma alternativa teórica para trabalhar a massiva quantidade de dados.

Fica claro que a informação científica também se materializa através do viés institucional. Através da caracterização do IBBD, por meio da iniciativa da Unesco em parceria com Fundação Getúlio Vargas, e a partir dos primeiros periódicos fundacionais da área (*Ciência da Informação* e *Revista da Escola de Biblioteconomia da UFMG*), torna-se viável afirmar que esses se firmam enquanto agentes de caráter híbrido que fomentam a área no Brasil.

A relação hierárquica entre esses órgãos pode ser dita como pautada em dinâmicas de poder na medida em há a caracterização de instâncias de validação científica propriamente institucionais e discursivas que deram suporte para o nascimento da Ciência da Informação no Brasil na década de 1970.

Buscou-se interligar a perspectiva institucional com a discursiva ao fazer um paralelo entre o conceito de “Regime de

Informação”, apresentado por Frohmann (1995), com o contexto brasileiro da Ciência da Informação à luz da matriz teórica da Análise do Discurso de origem francesa. Entendeu-se que tanto a Análise do discurso de matriz francesa, quanto o regime de informação tratam do discurso impresso, atendo-se a determinado contexto.

Conclui-se que o conceito de Regime de Informação de Frohmann (1995) interliga a perspectiva institucional com a discursiva ao tornar possível a compreensão das instâncias de validação científica da Ciência da Informação no Brasil, situadas enquanto atores que tornam possível a relação entre instâncias de poder na área.

Assim, Formação Discursiva e Formação Ideológica, como visto anteriormente, possuem uma relação com Regime de Informação, se observarmos esse primeiro contexto; o discurso do autor, presente nos livros, representado em uma unidade de informação seguindo as regras de representação documental; o profissional que representa o item se observa controlado por tais regras, evidenciando o Regime de Informação que se encontra presente na instituição.

## REFERÊNCIAS

- BOYD, D.; CRAWFORD, K. Critical questions for Big Data. *Information, Communication & Society*, v. 15, n. 5, p. 662-679, 2012.
- BRANDÃO, N. H. H. *Introdução a análise do discurso*. 2.ed. Campinas: Ed. Da Unicamp, 2004.
- CALDAS, M.S.; SILVA, E.C.C. Fundamentos e aplicação do Big Data: como tratar informações em uma sociedade de yottabytes. *Bibliotecas Universitárias: pesquisas, experiências e perspectivas*, Belo Horizonte, v. 3, n.1, p. 65-83, jan. /jun. 2016

CARVALHO, A, M, A. Novos enfoques no campo da Ciência da informação: uma discussão sobre a aplicabilidade do conceito de regime de informação em arranjos produtivos locais. *Perspectivas em Ciência da Informação*, v.14, n. especial, 2009.

CONEGLIAN, C. S.; SEGUNDO, J. E. S.; SANTANA, R. C. G. Big data: fatores potencialmente discriminatórios em análise de dados. *Em Questão*, v. 23, n. 1, p. 62-86, 2017. DOI: 10.19132/1808-5245231.62-86. Disponível em: [http://www.scielo.br/readcube/epdf.php?doi=10.1590/S1413-99362009000400015&pid=S141399362009000400015&pdf\\_path=pci/v14nspe/a15v14nspe.pdf](http://www.scielo.br/readcube/epdf.php?doi=10.1590/S1413-99362009000400015&pid=S141399362009000400015&pdf_path=pci/v14nspe/a15v14nspe.pdf). Acesso em: 11/04/2019

FROHMANN, B. Taking information policy beyond information science: applying the actor network theory. *23<sup>rd</sup> Annual Conference: Canadian Association form Information*, 1995.

GÓMEZ, M. N. G. Regime de informação: construção de um conceito. *Informação & Sociedade: Estudos*, João Pessoa, v.22, n.3, 2012.

LOTT, Y. M.; CIANCONI, R. de B. Vigilância e privacidade, no contexto do big data e dados pessoais: análise da produção da Ciência da Informação no Brasil. *Perspectivas em Ciência da Informação*, [s.l.], v. 23, n. 4, p.117-132, out/dez, 2018. Disponível em: <http://www.scielo.br/pdf/pci/v23n4/1413-9936-pci-23-04-00117.pdf>. Acesso em: 10/04/2019.

MAGNANI, M. C. B.; PINHEIRO, M. M. K. “Regime” e Informação”: a aproximação de dois conceitos e suas aplicações na Ciência da Informação. *Liinc Em Revista*, v.7, n.2, 2011, p. 593-610.

ORLANDI, E. P. *A linguagem e seu funcionamento*. São Paulo: Brasiliense, 1984.

ORLANDI, E.P. *Análise de discurso: princípios e procedimentos*. Campinas: Pontes, 1999.

ORLANDI, E. P. *Discurso e leitura*. 8. ed. São Paulo: Cortez, 2008

ORLANDI, E. P. *Análise de Discurso: princípios e procedimentos*. Campinas, SP: Pontes, 2009.

PÊCHEUX, M. *Semântica e discurso: uma crítica à afirmação do óbvio*. Tradução de Eni Pulcinelli Orlandi *et al.* Campinas: UNICAMP, 1997.

# **HOMENAGEM A MARIANA BROENS**

## CONVERSA COM LAURO FREDERICO BARBOSA DA SILVEIRA<sup>1</sup>

É um prazer, uma honra muito grande, estar aqui com o professor Lauro Frederico Barbosa da Silveira, que vai nos falar a respeito de sua amizade e parceria com a professora Mariana Claudia Broens, homenageada neste livro.

O professor Lauro é Graduado em Filosofia pela Universidade de São Paulo (1969) e doutor em Filosofia pela Pontifícia Universidade Católica de São Paulo (1974). Tem experiência na área de Direito, com ênfase em Semiótica Jurídica. Ele também pesquisa semiótica em Medicina e em Psicanálise, com ênfase nas relações médico-paciente e analista-analisando. Atua, principalmente, nos seguintes temas: semiótica, interpretante, signo, hábito.

### **PROFESSOR LAURO, QUAIS AS MELHORES LEMBRANÇAS QUE O SENHOR TEM DA PROFESSORA MARIANA?**

Primeiro, é uma alegria muito grande estar com você aqui. Eu gostaria de agradecer muito o convite, especialmente porque é uma homenagem a uma colega por

---

<sup>1</sup> Entrevista realizada e transcrita por Edna Alves de Souza, em 09 de maio de 2019.

<https://doi.org/10.36311/2019.978-85-7249-055-9.p235-249>

quem tenho muita admiração, respeito e uma profunda amizade. A vinda da professora à UNESP foi um dos bons encontros, que permitiu que o departamento de filosofia e a pós-graduação se mantivessem estáveis, com todas as dificuldades que temos vivido. Acho que isso é um aspecto muito importante. Estava vendo nesses cartazes dos primeiros eventos realizados no departamento, dentre eles o congresso em homenagem ao professor Antonio Trajano Menezes Arruda, e outros que não estão mais presentes entre nós... Mas a Mariana, sem dúvida alguma, é uma contribuição muito grande, dado o seu conhecimento, o rigor de seu pensamento, honestidade e, sobretudo, seu empenho afetivo com o departamento, conosco, comigo também, muito grande. É uma grande amizade mesmo! Ela é uma pessoa maravilhosa.

### **CONTE UM POUQUINHO DA HISTÓRIA DO DEPARTAMENTO DE FILOSOFIA DA UNESP: O QUE O SENHOR SE LEMBRA DE SEU INÍCIO E DA CHEGADA DA MARIANA?**

Olha, o início, eu vou tentar me lembrar... Primeiro, estávamos todos nós no prédio da Avenida Vicente Ferreira, ainda. Quando eu cheguei lá na Vicente Ferreira, o professor Ubaldo Martini Puppi já estava; a Ligia Fraga Silveira, minha mulher, também já estava lecionando lá. Nem havia acesso até o *Campus* Universitário; ainda não tinha o viaduto de entrada ao *Campus*. Foi quando houve um acidente, em que morreu um aluno [ao tentar atravessar a rodovia para chegar ao *Campus*], que se construiu o viaduto, para aumentar a segurança de quem vem para cá. Infelizmente, no Brasil isso é muito frequente: depois que acontece uma desgraça é que se acaba tomando um investimento público. E foi um

investimento importante, porque foi um investimento de se fazer, de se estudar filosofia, onde quer que a gente estivesse. Foi na gestão da professora Mariana que mudamos da Vicente Ferreira para o *Campus* novo.

**UMA DAS PREOCUPAÇÕES DA PROFESSORA MARIANA É COM O TEMA DO ANTROPOCENTRISMO, AINDA DOMINANTE EM NÓSSA SOCIEDADE. O SENHOR ACHA QUE O ANTROPOCENTRISMO EXPRESSA UMA VISÃO SIMPLISTA SOBRE A NATUREZA DO PENSAMENTO?**

O pensamento não é propriedade privada. Quando se diz ‘autoria de pensamento’, ‘autoria’ quer dizer que a gente propõe, mas o pensamento não é da gente. Essa hipótese é do Peirce; é ele quem diz: o pensamento não está em nós; nós estamos em pensamento. Eu acho bastante razoável entender que o pensamento é uma vibração de todos nós. Por que conseguimos, facilmente, nos encontrar com as pessoas certas? Por que há um campo, vamos chamar de energético (mas no sentido bem amplo da palavra), psicossomático; um campo, antes de tudo, espiritual; não digo ‘espiritual’ como se estivesse pensando em alguma coisa transcendente, mas nessa capacidade que a gente tem de conviver com as coisas e, antes de nós as interpretarmos, elas nos interpretam.

O pensamento não é sequer uma exclusividade do ser humano; ele permeia, nos modos mais diversos, toda a realidade. Ele é o que dá à trama da realidade algum sentido e que nos permite ao longo do tempo entender, de alguma maneira, a cada momento, o significado de viver, o significado de ser. E aí todo nosso trabalho que vem sendo feito; não só com os clássicos, mas em toda a história da filosofia. Quando nos dedicamos à filosofia, buscamos e sugerimos respostas a

questões de várias naturezas ao longo do tempo. Considere, por exemplo, um fragmento de um pré-socrático e veja o que está escrito lá. É impressionante! A discussão de Heráclito com Parmênides e daí por diante: é um compartilhar de algo muito profundo. Nós tivemos oportunidade dentro do pensamento greco-ocidental: Pré-socráticos, Platão Aristóteles... Mas, o mesmo se dá nas mais diversas tradições. Na tradição bíblica, por exemplo, veja o livro da Sabedoria: o que há de reflexão, o que há de expressão, diríamos assim, do que é uma liberação no sentido poético, que a gente entende como *poiesis*! É fazer de um pensamento não a procura de uma outra coisa, mas, sobretudo, aquilo que ele pode nos *constituir*. Não é incrível? Cada um, a seu jeito, compartilha. E não é só exclusivamente o ser humano quem compartilha. Se você tem um cachorrinho dentro de casa, não há um afeto? E é recíproco. Chegando até em situações emocionantes em que falece o dono e o cachorro entra em um processo de luto. Não posso me esquecer, de jeito nenhum: aos oito anos de idade, eu morava em São Paulo, mas a gente vinha para o interior, pois meu pai sempre trazia a gente. Uma vez, eu estava montado a cavalo, escorreguei da sela e caí no chão. O cavalo parou, foi assim com o focinho, sei lá, me agradar, me apoiar.

**A REALIDADE É MUITO COMPLEXA. PARA A PROFESSORA MARIANA, ASSUMIR QUE A AFETIVIDADE E MESMO O PENSAMENTO, A INTELIGÊNCIA, É, NECESSARIAMENTE, EXCLUSIVIDADE HUMANA, CONSISTE EM UMA POSTURA EQUIVOCADA, EM MUITOS ASPECTOS. O SENHOR CONCORDA COM A PROFESSORA?**

Sem dúvida! A afetividade, o pensamento, a inteligência não constituem exclusividade humana. Nós compartilhamos de uma totalidade complexa. Quando dizemos

que uma flor murcha e morre, *morre*. Quando vemos uma sementinha se desenvolvendo em uma planta, é vida. Se nós mudamos um pouco a escala e nos vemos na escala metabólica, é um compartilhar total disso tudo. E nós não somos um acréscimo; somos uma expressão disso; o ser humano não está no centro do universo, mas é uma de suas partes expressivas!

**A PROFESSORA MARIANA HOJE É COORDENADORA DO PROGRAMA DE PÓS-GRADUAÇÃO EM FILOSOFIA DA UNESP DE MARÍLIA E DEFENSORA AGUERRIDA, SOBRETUDO, DA ÁREA DE CONCENTRAÇÃO EM FILOSOFIA DA MENTE, EPISTEMOLOGIA E LÓGICA DESSE PROGRAMA, QUE É BEM INOVADORA EM TERMOS DO MODO DE SE FAZER FILOSOFIA NO BRASIL. O QUE O SENHOR TEM A DIZER SOBRE O MESTRADO, SUA IMPLANTAÇÃO E MANUTENÇÃO, BEM COMO O PAPEL DA PROFESSORA, NESSE CONTEXTO?**

Foi um desdobramento e um avanço, muito feliz, de toda uma experiência de pensamento que carregava comigo até mesmo antes de eu vir para a Faculdade de Filosofia. A partir dessa experiência, o mestrado surgiu com a colaboração das professoras Maria Eunice Quilici Gonzalez, Carmen Beatriz Milidoni, Ligia Fraga Silveira e do professor Cosme Damião. Estava me lembrando também da Revista Transformação: o registro de um pensamento de longa data de uma revista que está aí a quantos números! Inclusive, com aquela parte de entrevistas, dos diversos artigos. É uma contribuição que nós recebemos e tentamos devolver da melhor maneira possível. Espero que a gente consiga fazer sempre isso.

## CONSIDERANDO QUE O SENHOR É UMA PRESENÇA CENTRAL NA HISTÓRIA DO DEPARTAMENTO DE FILOSOFIA, O QUE O SENHOR SE LEMBRA DO INÍCIO DELE, COMO FOI A SUA IMPLANTAÇÃO?

Houve colaboração de toda uma parte que vinha da fenomenologia de tradição europeia, que vinha do Tomismo, inclusive, tanto do professor Puppi como meu. Eu fui dominicano; fiz meu mestrado em Tomas de Aquino: *providência e governo do mundo*; depois fui fazendo outras coisas.

O pensamento *acontece*. Há uma efetivação de um encontro de questões, de propostas, até no sentido, que eu acho maravilhoso, da palavra erro. Erro é caminhar. Errante é aquele que caminha. Uma pessoa absolutamente convicta consigo mesma é estática. Errar é um benefício. No seio do diálogo, o erro é exatamente o que permite que nós construamos nossos conceitos e não nos deixemos prender pelos nossos preconceitos. Isso para mim é absolutamente fundamental. Isso marca, dentro dos meus limites, a minha vida.

Aonde você for, sempre procure admirar. Você sempre vai encontrar o admirável. Ou o admirável acaba te encontrando. Tudo é limitado. É limitado, mas é infinito; pode crescer indefinidamente e em todas as esferas: esfera cultural, esfera de transmissão do código genético... Trazemos conosco toda uma tradição. Para mim, isso é de extrema importância. Isso no departamento foi sendo desenvolvido realmente, com algumas pessoas que foram saindo, outras que foram entrando, sempre em uma procura de avançar o conhecimento, de resolver possíveis impasses. E agora estamos aí...

## **COMO O SENHOR VÊ A TRAJETÓRIA DA MARIANA NESSE PROCESSO?**

Cada um de nós tem tantas coisas que a gente vai descobrindo aos poucos... Mas, pelo que eu conheço da Mariana, primeiro, acho que ela contribuiu e teve ocasião de encontrar um campo afetivo muito importante de acolhimento no departamento; ela foi e está sendo capaz de contribuir profundamente para com ele. Isso é muito importante porque é isso que permite ampliar o conhecimento. Por outro lado, ou em outra dimensão, o afeto é um afeto lúcido. E da parte dela isso me parece muito claro. Há realmente um empenho, que não é um empenho particular, interesseiro, mas é realmente um empenho para que ela e o departamento se desenvolvam. Para que ela se desenvolva, esse empenho precisa ser compartilhado por todos nós; supõe enfrentar as dificuldades, que não foram e nem são, certamente, pequenas. E a Mariana, efetivamente, tem uma história para isso: ela vem da Argentina, é formada em Filosofia e em Direito... Seria interessante ouvi-la também a esse respeito.

## **O SENHOR SE LEMBRA DE QUANDO CONHECEU A MARIANA? FOI NA UNESP?**

Quando eu a conheci, eu estava na UNESP, na Vicente Ferreira antes de nossa mudança para cá, quando ela se propôs e foi aceita no departamento. Eu estou a mais tempo do que ela, certamente. Mas, como estava dizendo, e isso é verdadeiro: foi nesse campo afetivo, que não é um campo boboca. Ao contrário, é um campo lúcido, crítico, no bom sentido, de tentar fazer com que a verdade predomine, capaz de enfrentar as adversidades, algumas extremamente sérias.

**A EXISTÊNCIA DESSE “CAMPO AFETIVO”, COMO O PROFESSOR O DENOMINA, COM UM TIPO DE AFETIVIDADE MUITO PECULIAR, É RECONHECIDA POR AQUELES QUE, COMO EU, CONVIVERAM OU MANTÊM, DE ALGUM MODO, RELAÇÕES ESTREITAS COM O DEPARTAMENTO. COMO SE DEU A CONSTITUIÇÃO DESSE CAMPO AFETIVO? FOI ALGO FÁCIL DE SE CONSTRUIR?**

Não foi fácil, de modo algum. Quando se leva a sério a construção de um campo afetivo, nunca é fácil. Ninguém ganha isso de presente. Se estiver tudo muito fácil, desconfie. Está valendo a lei da inércia. Se existe uma emulação, uma exigência, de que a gente cresça em conhecimento, cresça nas nossas relações, vai ter sempre obstáculos a serem vencidos, dificuldades a serem enfrentadas. Na medida do possível, eu acho que isso foi possível de acontecer no departamento. O tempo vai permitindo que a gente veja com maior clareza. Não foi nada fácil! Mas, foi possível que certas pessoas fossem se unindo afetivamente e acho que isso foi uma coisa boa.

**SE, POR UM LADO, SÓ O TEMPO PERMITE QUE A GENTE VEJA COM MAIOR CLAREZA DETERMINADAS SITUAÇÕES, POR OUTRO, HÁ CERTAS CONFIGURAÇÕES QUE PARECEM RECLAMAR POR MEDIDAS PRÁTICAS URGENTES, MESMO NO ESCURO. COMO O SENHOR VÊ O CENÁRIO ATUAL DO BRASIL, MAIS ESPECIFICAMENTE, NO QUE DIZ RESPEITO À PESQUISA CIENTÍFICA? HOJE, COMO ESTÃO AS COISAS?**

O cenário atual não é nada fácil de entender. Sob alguns aspectos, é uma situação mais favorável, do que, por exemplo, a de todo o período, por nós vivido, da ditadura, com repressão, com mortes, com situações realmente trágicas.

Então, acho que o acesso a mais informação foi sendo possível: as bibliotecas estão aí, o acesso via internet ampliou profundamente o nosso campo de pesquisa. Agora, tudo isso é um *convite* para que a gente interaja, senão as interações ficam sem sentido. Não só às vezes sem sentido, mas às vezes até em prejuízo do próprio avanço do conhecimento. No entanto, existe hoje um campo possível, que vem sendo feito, de interagir com a realidade, de elaborar mais o pensamento, de publicar mais – dificuldades que havia antes.

**HOJE SE FALA QUE VIVEMOS A ERA DA DATIFICAÇÃO, EM QUE TUDO SÃO DADOS, E QUE OS DADOS NÃO SÃO OBJETIVOS. O SENHOR VÊ NESTE NOVO CENÁRIO DE IMPORTÂNCIA, SENÃO DE DEPENDÊNCIA, DOS DADOS, DAS TECNOLOGIAS DE INFORMAÇÃO E DE COMUNICAÇÃO, UM LUGAR PARA UM REALISMO?**

A cada momento em que se vive, se vive tudo isso. Esta configuração é a configuração que nós, neste momento, estamos vivendo. Mas, se você observar, em momentos anteriores ou em lugares diferentes, o pensamento só se desenvolve se ele for desafiado. E no desafio você tem a possibilidade de que se interaja com ele e o risco de se deixar ser dominado pela, por assim dizer, banalidade, pela força bruta.

Chegou o jornal agora... Você vê? Volta-se a se defender que todo mundo possa andar armado... O que é isso, gente?

**ACREDITO QUE SEJA UM DOS SINAIS DO QUE ESTÁ SENDO CHAMADO DE ERA DA ‘PÓS-VERDADE’, ‘PÓS-REALIDADE’. NESTE MOMENTO, DE “DESAFIO DO PENSAMENTO”, QUANDO VEMOS NOTÍCIAS COMO ESSA QUE O SENHOR ACABA DE MENCIONAR, TEMEMOS QUE A BALANÇA ESTEJA SE INCLINANDO PARA A TENDÊNCIA À FORÇA BRUTA E À BANALIZAÇÃO. ESTAMOS DIANTE DA BANALIZAÇÃO DE PRINCÍPIOS MUITOS CAROS À FILOSOFIA, À HUMANIDADE.**

Sim: à humanidade, à dignidade do ser humano. Eu acho que isso é alguma coisa que nós temos que enfrentar... Eu acho que há uma contribuição nossa, da área em que trabalhamos – a filosofia –, de refletirmos sobre isso, de trazer contribuições, porque a tendência à força bruta não só é muito premente, como está sempre presente.

**A MARIANA É UMA PROFESSORA MUITO MILITANTE, MUITO PREOCUPADA COM AS QUESTÕES ÉTICAS, COM AS QUESTÕES SOCIAIS, ENTÃO, ESSE CENÁRIO, QUE SE DESENHA HOJE, É PREOCUPANTE PARA ELA. QUAL O PAPEL, A CONTRIBUIÇÃO, DO REALISMO PEIRCEANO PARA AS DISCUSSÕES CORRENTES DOS RUMOS TOMADOS PELA NOSSA SOCIEDADE NESTA ERA DA DATIFICAÇÃO?**

Antes de tudo, não fugir da *realidade*. O que não é fácil. Em segundo lugar, ter a capacidade poética de, em conjunto, construirmos uma expressão, de uma maneira profundamente falível, mas que permita uma consciência dessa realidade. Esse papel é a maneira de você assumir a ética sem assumir o moralismo. A coisa mais fácil é o moralismo, porque ele enquadra e reitera o poder violento.

**O EXEMPLO QUE O SENHOR DEU A POUCO DA PROPOSTA DE LIBERAÇÃO DO PORTE DE ARMAS É BASTANTE ELUCIDATIVO NESSE SENTIDO: TRATA-SE DE UMA IDEIA BASEADA NO MORALISMO.**

Certamente há uma questão ou até mais de uma. Mas a resposta já é a da violência, da força bruta. Como se isso resolvesse alguma coisa! Eu fico muito impressionado. Esse pensamento, vamos chamar de *reacionário*, carrega a oportunidade de ter os meios de comunicação com ele e tudo o mais. Esta reflexão que estamos fazendo, é a reflexão de uma minoria, que, se possível, é *calada*.

Soluções fáceis são todas muito perigosas. Lembro-me agora de Lewis Carroll... Disse a Rainha de Copas: “Cortem-lhe a cabeça”. Em vez de resolver um problema, *elimina*. Não vem de agora isso. Não vem de agora...

Uma coisa importante, inclusive, é manter o diálogo; entender as pessoas; respeitar as pessoas; não deixar que os nossos preconceitos predominem. Se a gente tiver um pouco de sensibilidade para perceber, quanta sabedoria está nas pessoas que às vezes estão exercendo funções extremamente humildes?

**SIM, MUITA. COMO FAZ FALTA ÀS VEZES OUVIR ESSAS PESSOAS!**

Não é verdade, meu bem? Eu sinto: se a gente não tem esse cuidado, porque não é uma coisa rígida, a gente é tomada por preconceitos.

**E DAÍ A IMPORTÂNCIA DE SE OUVIR AS MAIS DIVERSAS E DIFERENTES VOZES E, EM ESPECIAL, DAQUELAS PESSOAS COM EXPERIÊNCIA. EU SOU UMA ADMIRADORA DO SENHOR. O PROFESSOR NÃO TEM IDEIA DO QUANTO EU O CONSIDERO UMA PESSOA SÁBIA, SERENA. POR ISSO, EU GOSTARIA QUE O SENHOR DEIXASSE UMA MENSAGEM DE ESPERANÇA, UMA MENSAGEM PARA OS INICIANTE NA ARTE DA FILOSOFIA, UMA MENSAGEM PARA A GENTE DA SUA EXPERIÊNCIA DE VIDA, COMO PROFESSOR, COMO FILÓSOFO, COMO PESSOA...**

Você que é um encanto! Espero que eu corresponda a isso...

Existem mais aspectos admiráveis na realidade do que aqueles que nós podemos dominar. Nós compartilhamos desse pensamento, mas não somos donos dele. Não somos donos do pensamento. O pensamento está em nós e nós estamos em pensamento. Ele não é privilégio daqueles que são chamados “intelectuais”. Mais importante do que isso, a dimensão que não pode se perder, é a dimensão da busca da verdade, a dimensão do afeto, a dimensão do respeito, da sensibilidade da gente, porque os preconceitos são insidiosos, e bem mais do que parecem ser. Eu sempre mantenho uma atenção, talvez ao longo do tempo, com minha idade, atualmente, eu até tenha mais coisas das quais eu certamente me iludia, mas isso não é para entrar em um estado de desespero, ao contrário, é para exatamente *se tornar mais simples*. Eu acho que isso é tão gostoso como experiência! Às vezes, estou fazendo compras no supermercado e vejo uma mãe com uma criancinha: é uma coisa tão linda! Aprender, se preocupar também, mas ter uma sensibilidade por essas pessoas, inclusive aquelas que vêm pedir esmolas. O mundo está preconceituoso e muitos estão, assim, em busca de privilégios. Olha, que não sejamos tomados por

privilégios! Mas que avancemos no conhecimento, que avancemos nas questões, sempre guardado que: pensamento que cresce é o pensamento que assume a sensibilidade; que, sobretudo, é gerado num contexto de afetividade. Estava lembrando agora, falando com você, que eu estava lendo Bion [Wilfred Ruprecht Bion], quando ele fala da criança com sua mãe, a busca do seio. A boa mãe é aquela que acolhe no seu seio a criancinha que está nascendo. Essa desenvolve dimensões da personalidade que, ainda bem, ela não tem sequer consciência, não precisa ter consciência. Ter consciência já é um processo que, não digo malévolos, mas tem o perigo de ser restritivo, como se só aquilo que nós temos consciência fosse importante. Ao contrário, ela só tem sentido, em um contexto muito mais amplo de todas as relações, mais complexo, que não nos cabe dominar. Cabe-nos sensibilizar para viver. Se você domina, você restringe. Entram em jogo outras dimensões que iludem. Essa sensibilidade é um mistério: aquilo que está escondido. Mistério, em grego, é aquilo que está escondido.

**ESSA NÃO É A IDEIA DA CIÊNCIA MODERNA: ENTENDER PARA CONTROLAR, PARA DOMINAR, A NATUREZA, E QUE, DE CERTA FORMA, DESEBOCOU NESSA ERA DA DATIFICAÇÃO?**

Exatamente. Ela restringe o acesso à expressão da totalidade das qualidades que a natureza tem, e despreza um oceano de possibilidades para ficar lá tentando no vermelhinho, vamos dizer assim.

## **E O QUE ISSO GERA? QUAIS AS CONSEQUÊNCIAS PRÁTICAS DESSE MODO DE PROCEDER?**

Isso só pode dar coisas monstruosas! Mas, ter essa sensibilidade constante, acho que é a coisa mais importante. Aí, nela, tudo vai podendo ser feito. Como é que podem surgir as grandes hipóteses, realmente, as grandes hipóteses? Porque as pessoas que as propuseram estavam liberadas das cascas de suas crenças. Li em um texto que Einstein estava em uma estação ferroviária e viu um trem andando, então, se perguntou: e se esse trem andar na velocidade da luz? Daí eu me perguntei: mas essa não é uma ideia fugal? Não, espere um pouco: talvez ele reconheça no passado um momento importantíssimo em que a hipótese de uma teoria da relatividade pudesse surgir.

**É MUITO BOM PODER CONVERSAR COM O SENHOR, COMPARTILHAR ESSAS REFLEXÕES. MAS, PROFESSOR, PARA FECHAR ESTA ENTREVISTA, DEIXE UMA MENSAGEM ESPECÍFICA PARA A PROFESSORA MARIANA. O QUE O SENHOR DIRIA, AGORA, NESSE MOMENTO ESPECIAL DA CARREIRA DELA?**

Antes de mais nada, Mariana, que você esteja sempre conosco, nessa relação afetiva, que dá sentido às nossas investigações teóricas e daí por diante. Antes de tudo isso, que esse pensamento do qual compartilhamos, você pode ter certeza, Mariana, traz para nós uma contribuição importantíssima. Que a gente sempre possa continuar fazendo isso.

Você sabe, Mariana, todo encanto que tenho por você; a felicidade que foi tê-la no departamento, que foi tê-la conhecido, e espero, sim, que você possa sempre contar comigo

Eu estou emocionado!

É tão gostoso falar com vocês, viu querida. Eu só tenho a agradecer. Que coisa boa! Foi uma alegria estar com vocês. Vamos continuar sempre assim...

**EU QUE LHE AGRADEÇO, PROFESSOR. A ALEGRIA É RECÍPROCA. SIM, VAMOS CONTINUAR, SEMPRE CONVERSANDO, REFLETINDO, TENTANDO ENTENDER O PEIRCE...**

Errando enquanto caminhando! Errante é aquele que caminha. Lembrar sempre que nunca devemos achar que está tudo bem. Não está. Esse pensamento nos ajuda a estar sempre buscando. Ter sempre um oceano de possibilidades. Nós conhecemos muito poucas coisas, como aquela criança na beira do oceano catando umas areinhas. E tomara que continuemos caminhando. O quanto, não nos cabe saber. Se pudéssemos saber, já seria falso.

**PEQUENAS PALAVRAS PARA TANTO AFETO  
E ADMIRAÇÃO**

**Mariana Cláudia Broens** é uma destacada intelectual, ativista, querida amiga e professora de muitos. Precisaríamos de um novo e grande livro para conter os depoimentos em agradecimento a seu trabalho e sua amizade. Ainda assim, seriam pequenas palavras diante de tanto afeto e admiração cultivados nos encontros da vida. Aqui registramos alguns depoimentos que representam o agradecimento, a amizade e o carinho de todos nós.

## **MARIANA BROENS – QUERIDA AMIGA**

Já de longa data tenho o privilégio de desfrutar da amizade de Mariana Broens. Temos feito muitas coisas juntos, partilhando intensa e contínua convivência acadêmica, mesmo sem sermos colegas em uma mesma instituição. Meu testemunho sobre Mariana nutre-se não apenas de sua seriedade profissional, seu senso de justiça, sua transparência de conduta e sua atuação docente estimulante, mas, do mesmo modo, de seu caráter profundamente humano, comprometido com a realidade social de nosso país e atenta a tudo o que a história de agora faz desfilarmos pela nossa experiência, convocando-nos a nos posicionar em face dela.

Penso que somente uma afinidade de valores e sentimento de mundo, muitas vezes não ditos, guardados em silêncio na interioridade, sustentam o contínuo prazer dos amigos estarem juntos, garantindo a continuidade incólume de um relacionamento que não se deixa fraturar por divergências contingentes. Com respeito a Mariana, com alegria, reconheço esse contínuo prazer de uma convivência que comunga trabalho e afinidade humana.

Rendo à minha querida Mariana o meu mais terno afeto, esperando que nossa sólida amizade permaneça prosperando por longo tempo ainda.

*Ivo Ibri* – 17 de maio 2019

Mariana, falar sobre a Mariana!

Uma das pessoas mais incríveis que conheço e com quem convivo.

Bem formada, competente, responsável, dedicada, excelente professora e acadêmica admirável.

Mas tudo isso é pouco, Mariana é muito mais.

Mariana é exemplar, vive o que pensa, crê e defende.

Em nossa convivência pessoal e acadêmica, reconheço, por sua expressão facial, quando não está de acordo ou algo a desagrada, pois Mariana, mesmo com esforço, não consegue dissimular ou disfarçar.

Mariana é a personificação, em ação, do que pensa e crê.

Uma grande profissional, uma grande mestra, mas, sobretudo, uma inestimável amiga, um exemplo de comportamento ético a ser seguido.

Obrigada, Mariana, é um privilégio estar com você nesta nossa caminhada!

*Ítala Maria Loffredo D'Ottaviano*

É um privilégio muito grande poder trabalhar ao lado da professora Mariana. Ela é uma pessoa que ensina com dedicação e paciência e me auxilia muito nas dificuldades profissionais, principalmente em relação ao microcomputador. Ela age com simplicidade e está sempre disposta a ajudar em qualquer hora, em qualquer momento. É aquela pessoa que, quando tudo está bem difícil, abre um sorriso e diz: calma, você vai conseguir!

Enfim, ela é uma grande professora e uma grande amiga, que me deu a chance de olhar a vida sob um novo horizonte, a partir do momento que comecei a trabalhar no Departamento de Filosofia.

Parabéns Mariana e que todo seu esforço seja recompensado com muitas alegrias!

*Edna Bonini de Souza*

À professora Mariana Broens,

A primeira vez que assisti uma palestra da professora Mariana foi em 2014 no Encontro Nacional de Filosofia da Universidade Federal de Ouro Preto. Como membro da organização eu havia trabalhado bastante durante o dia e a noite seguiria para casa, mas resolvi conferir se tudo estava correndo bem com a palestra da noite intitulada “Uma introdução à Filosofia da Mente”. Comecei a assistir e me esqueci completamente do cansaço permanecendo totalmente atenta. É provável que a professora Mariana não saiba o quanto modificou a minha vida a partir daquele instante. Até então, meu caminho pela Filosofia estava em processo de construção e não havia especificamente nenhuma área a qual eu estivesse inclinada a me especializar. A partir disso passei a ler os poucos livros do meu departamento sobre Ciência Cognitiva e Filosofia da Mente, aumentando meu entusiasmo e curiosidade intelectual.

Procurei participar de eventos sobre Ética e Filosofia da Mente e pude assistir novamente a Mariana e não restou dúvidas: precisava estudar com essa professora! Ao terminar a graduação ingressei no mestrado em Filosofia da Mente da Unesp e tive o privilégio de ser orientada por aquela que acendeu essa inquietação filosófica tão especial. O que eu não fazia ideia é do quanto aprenderia com ela em poucos anos de convivência. Além da orientação, que proporcionou um fecundo desenvolvimento filosófico, aprendi lições sobre ética, moralidade, justiça, empatia e respeito por todas as formas de vida. Obrigada professora Mariana Broens por todas as valiosas lições que levarei para sempre com grande carinho, respeito, admiração intelectual e pessoal.

*Barbara Linda Tavares*  
13 de Agosto de 2019.

A professora Mariana é uma pessoa extraordinária e muito querida. Em sua dinamicidade cotidiana ela abraça seus alunos e as pessoas a sua volta e mesmo sem prometer, garante que tudo dará certo e nos mostra que as coisas podem ser ainda melhores do que esperamos. Sua humildade, seu carinho, seus atos justos e éticos nos mostram como somos afortunados por tê-la em nossas vidas. Desejo que seja feliz, professora. Espero que a injustiça do mundo não embruteça sua humanidade. E enquanto orientanda e aluna, espero que a deixemos tão orgulhosa quanto você nos deixa. Muito obrigada pelos seus ensinamentos.

*Camila da Cruz Silva*

No atual contexto em que vivemos, ter a possibilidade de conviver com pessoas como a Professora Mariana é um grande privilégio e razão de uma imensa alegria. Sua disposição e honestidade com as investigações filosóficas inspiram a todos que estão ao seu redor. Como sua orientanda, sou especialmente grata por toda sua contribuição acadêmica e, para além disso, existencial. Com certeza levarei seus ensinamentos por toda a minha vida.

*Emanuelly Nakaryn Rodrigues*

Além de ser uma excelente intelectual em Filosofia da Mente, a Profa. Mariana é também uma pessoa muito humana, inteligente e criativa que faz seus orientandos se sentirem bastante responsáveis pelos impactos do trabalho que realizam. Eu, particularmente, não teria continuado a estudar filosofia se não fosse por sua dedicação inspiradora e por ela ter acreditado em mim primeiro. Como seu orientando, na maioria das vezes, senti-me parte de um projeto científico muito maior, podendo ousar ser eu mesmo enquanto pesquisador iniciante. Sou imensamente grato pelas excelentes orientações da professora Mariana e, apesar de estar sempre muito ocupada devido à sua generosidade, espero que seja reconhecida o quanto antes como merece, por sempre deixar pessoas razoáveis sempre muito confortáveis a seu redor.

*Felipe Eleutério*

No cenário atual brasileiro há um profundo descaso com a área da educação, algo que é especialmente desanimador para nós filósofos(os). Mas, apesar de estarmos vivendo tempos difíceis e obscurantistas, tenho aprendido com a professora Mariana a não desistir de meus estudos e lutar sempre pelo bem e o justo. Dessa forma, muito obrigada, professora, por ser mais que uma orientadora para a vida universitária, mas também uma orientadora para a vida. Mais que respeito, tenho um profundo carinho e admiração pela senhora, que o tempo jamais será capaz de apagar. Enquanto eu estiver viva me lembrarei de seus ensinamentos e contarei para todos ao

meu redor o quanto a senhora foi importante em minha vida. Meus sinceros e profundos agradecimentos, professora.

*Júlia Rodrigues de Lima*

Professora Mariana, é muito gratificante estar em um campus como o da UNESP de Marília pela proximidade que isso possibilita com uma grande professora como a senhora, poder conviver, cursar disciplinas, tutoria, ser diretamente orientado e poder dizer que tive a honra de tal convívio com tantos aprendizados. Através da senhora tive contato com temas e abordagens que enaltecem e dão direcionamento ao profissional que desejo me tornar, inclusive as suas críticas. Expresso aqui meus mais sinceros agradecimentos, obrigado por tudo.

*Rafael Otavio Ribeiro de Mattos*

## SOBRE OS ORGANIZADORES

## **DANIEL MARTÍNEZ-ÁVILA**

*dmartinezavila@gmail.com*

Daniel Martínez-Ávila é professor do Departamento de Ciência da Informação da Universidade Estadual Paulista (UNESP), Marília. Ele obteve o título de Doutor pela Universidad Carlos III de Madrid, em 2012, sob a supervisão de Rosa San Segundo e de Hope Olson, da University of Wisconsin-Milwaukee. Foi professor assistente no Departamento de Biblioteconomia e Documentação da Universidad Carlos III de Madrid. Em 2013 assumiu o cargo de pesquisador e instrutor adjunto na School of Information Studies at University of Wisconsin-Milwaukee. Em 2014, tornou-se professor titular do Curso de Pós-Graduação em Ciência da Informação da UNESP; seu principal interesse de pesquisa é em organização do conhecimento, estudos de gênero, *Big Data*, epistemologia e avaliação de pesquisas. Atualmente, é membro da International Society for Knowledge Organization e Scientific Advisory Council. Ele também atua como Coordenador Internacional na Satija Research Foundation for Library and Information Science, da Índia, e no Instituto de Estudos de Gênero da Universidad Carlos III de Madrid, na Espanha.

## **EDNA ALVES DE SOUZA**

*souzaednaalves@gmail.com*

Edna Alves de Souza é pesquisadora de pós-doutorado pelo Programa de Pós-Graduação em Filosofia da Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP), *Campus* de Marília, na área de concentração em Filosofia da Mente, Epistemologia e Lógica, com o apoio financeiro da CAPES e sob supervisão da Dra. Maria Eunice Quilici Gonzalez;

participa do Projeto “Understanding opinion and language dynamics using massive data”, que tem apoio financeiro da FAPESP. Ela obteve o título de Doutora em Filosofia pela Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo (USP), em 2014; seu principal interesse de pesquisa é sobre metodologia científica, realismo científico, racionalidade, informação, complexidade e *Big Data*; é membro do Grupo Acadêmico de Estudos Cognitivos (UNESP) desde 1999 e do Grupo Interdisciplinar do Centro de Lógica, Epistemologia e História da Ciência (UNICAMP); possui artigos e capítulos de livro publicados nas Áreas de Filosofia da Ciência e Filosofia da Informação.

## **MARIA EUNICE QUILICI GONZALEZ**

*eunice.gonzalez@unesp.br*

Maria Eunice Quilici Gonzalez é pesquisadora PQ1 do CNPq, professora livre-docente do Departamento de Filosofia da UNESP; sua tese de doutorado: “A cognitive approach to visual perception” foi concluída em 1989, na Universidade de Essex, Inglaterra. Ela é membro fundadora da Sociedade Brasileira de Ciência Cognitiva e do Grupo Acadêmico de Estudos Cognitivos (UNESP) e desde 1989 participa do grupo CLE auto-organização, UNICAMP. Tem experiência na área de filosofia, com ênfase em Epistemologia, Filosofia da Informação e Tecnologia, Ciência Cognitiva e Filosofia da Mente, desenvolvendo pesquisa sobre os seguintes temas: autonomia e informação, auto-organização, sistemas complexos e *Big Data*. Por quatro décadas, ela vem ajudando na construção de uma concepção epistemológico-interdisciplinar da relação entre informação e ação, reunindo a filosofia, as ciências sociais, biológicas e naturais. Atualmente ela é

também membro da Complex Systems Society e da Ecological Psychology Society, e coordena o Projeto “Understanding opinion and language dynamics using massive data”, com apoio financeiro da FAPESP.

## **SOBRE OS AUTORES**

**DANIEL MARTÍNEZ-ÁVILA** é Professor do Departamento de Ciência da Informação da Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP, *Campus* de Marília). E-mail: [dmartinezavila@gmail.com](mailto:dmartinezavila@gmail.com)

**DAVIDSON BRUNO DA SILVA**, graduando em Sistemas de Informação pela Universidade Federal dos Vales do Jequitinhonha e Mucuri. E-mail: [davidsonbruno@outlook.com](mailto:davidsonbruno@outlook.com)

**EDNA ALVES DE SOUZA**, pesquisadora de pós-doutorado pelo Programa de Pós-Graduação em Filosofia da Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP, *Campus* de Marília). E-mail: [souzaednaalves@gmail.com](mailto:souzaednaalves@gmail.com)

**ELY FRANCINA TANNURI DE OLIVEIRA** é Livre Docente na área de Estudos Bibliométricos; doutora e mestre em Educação pela Universidade Estadual Paulista (UNESP); graduada em Pedagogia pela (UNESP) e em Matemática pela Faculdade de Filosofia, Ciências e Letras (Osvaldo Cruz – SP); atua no programa de Pós-Graduação em Ciência da Informação da (UNESP), *Campus* de Marília; é bolsista pesquisadora CNPq (2). E-mail: [ely.tannuri@unesp.br](mailto:ely.tannuri@unesp.br)

**GISELE RODRIGUES ATAYDE** é Gestora da Sevna Startups, Ribeirão Preto; Mestre em Engenharia de Produção pela USP, Bauru; Bacharel em Administração pela Faculdade FGP. E-mail: [giatayde@usp.br](mailto:giatayde@usp.br)

**JACQUELIN TERESA CAMPEROS-REYES** é Doutoranda e Mestre em Ciência da Informação pela Universidad Estadual Paulista (UNESP); Engenheira de Sistemas pela Universidad Francisco de Paula Santander (UFPS), Colômbia. E-mail: [jtcamperos@hotmail.com](mailto:jtcamperos@hotmail.com)

**JEAN FERNANDES BRITO** é Doutorando em Ciência da Informação na Área de concentração: Informação, Tecnologia e Conhecimento pela Universidade Estadual Paulista (UNESP); possui o título de mestre em Ciência da Informação na Área de concentração: Gestão da Informação pela Universidade Federal de Santa Catarina (UFSC); Bacharel em Biblioteconomia pela UNESP. E-mail: [jjeanfernandes@gmail.com](mailto:jjeanfernandes@gmail.com)

**JOÃO BATISTA ERNESTO DE MORAES** é Professor Associado do Departamento de Ciência da Informação, da Faculdade de Filosofia e Ciências (UNESP), Marília. E-mail: [joao.batista@unesp.br](mailto:joao.batista@unesp.br)

**LAÍS ALPI LANDIM** é Mestre e Doutoranda em Ciência da Informação pelo Programa de Pós-Graduação em Ciência da Informação (PPGCI) da UNESP, *Campus* de Marília; graduação em Ciências Sociais pela mesma universidade; coordenadora discente do Laboratório de Pesquisa em Design e Recuperação da Informação. E-mail: [laisalpi@gmail.com](mailto:laisalpi@gmail.com)

**LARISSA DE MELLO LIMA** é Mestre em Ciência da Informação e Doutoranda junto ao Programa de Pós-Graduação em Ciência da Informação da Faculdade Filosofia e Ciências (UNESP), Marília. E-mail: larissalima.unesp@gmail.com

**LAURO FREDERICO BARBOSA DA SILVEIRA**, Graduado em Filosofia pela Universidade de São Paulo (1969) e doutor em Filosofia pela Pontifícia Universidade Católica de São Paulo (1974). É professor do Programa de Pós-Graduação em Filosofia da UNESP. E-mail: lfbsilv@terra.com.br

**LEILAH SANTIAGO BUFREM** é doutora em Ciências da Comunicação (USP) e Mestre em Educação (UFPR), atuando nos Programas de Pós-Graduação em Ciência da Informação das Universidades Federais da Paraíba e de Pernambuco; bolsista pesquisadora do CNPq (1). E-mail: santiagobufrem@gmail.com

**LEONARDO LANA DE CARVALHO**, doutor em Psicologia (Cognitiva) pela Universidade de Lyon II, professor adjunto do Departamento de Computação da Universidade Federal dos Vales do Jequitinhonha e Mucuri. E-mail: leonardolana.carvalho@ufvjm.edu.br

**LUCINEIA DA SILVA BATISTA** é Mestre pelo programa de Pós-Graduação em Ciência da Informação da Faculdade de Filosofia e Ciência (UNESP), Marília; graduada em Biblioteconomia e Arquivologia; colaboradora voluntária no Laboratório de Pesquisa em Design e Recuperação da Informação. E-mail: lucineia.bat@gmail.com

**LUIS ANTONIO DE SANTA-EULALIA** é Professor da École de Gestion da Université de Sherbrooke (UdeS); Doutor em Engenharia de Produção pela Université Laval, Quebec; Mestre em Engenharia de Produção pela USP; Engenheiro de Produção pela UFSCar. E-mail: Luis.Antonio.De.Santa-Eulalia@usherbrooke.ca

**LUIZA DE MENEZES ROMANETTO** é Doutoranda em Ciência da Informação pela Universidade Estadual Paulista (UNESP); Mestre em Ciência da Informação pela UNESP; possui Bacharelado em Biblioteconomia e Ciência da Informação pela Universidade Federal de São Carlos (UFSCar). E-mail: luizaromanetto@hotmail.com

**MAGALY PRADO**, pós-doutoranda no Departamento de Informação e Cultura, da Escola de Comunicações e Artes, da Universidade de São Paulo; doutora em Comunicação e Semiótica e mestre em Tecnologias da Inteligência e Design Digital pela Pontifícia Universidade Católica de São Paulo; Jornalista e pós-graduada em Comunicação Jornalística pela Faculdade Cásper Líbero. E-mail: magalyprado@gmail.com

**MARIA EUNICE QUILICI GONZALEZ**, Professora Livre-Docente do Departamento de Filosofia da Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP, *Campus* de Marília). E-mail: eunice.gonzalez@unesp.br

**MARIA JOSÉ VICENTINI JORENTE** é Livre Docente em Cultura Digital e Informação Pós Custodiada em Redes de Colaboração; Doutora pelo Programa de Pós-Graduação em Ciência da Informação da Universidade Estadual Paulista Júlio de Mesquita Filho (UNESP); Especialista em Design de Produto; Licenciada em Artes pela Fundação Armando Álvares Penteado (FAAP) e em Letras pela Universidade de São Paulo (USP); Professora Adjunta em Regime de Dedicção Integral à Docência e à Pesquisa da UNESP, *Campus* de Marília, Departamento de Ciência da Informação. E-mail: [mj.jorente@unesp.br](mailto:mj.jorente@unesp.br)

**MARIANA DA SILVA CAPRIOLI** é Bibliotecária da Biblioteca Municipal “João Mesquita Valença” – Marília – SP; Mestre em Ciência da Informação e Doutoranda junto ao Programa de Pós-Graduação em Ciência da Informação da Faculdade Filosofia e Ciências da UNESP, Marília. E-mail: [mariana.caprioli@gmail.com](mailto:mariana.caprioli@gmail.com)

**MONICA AIUB**, Graduada em Filosofia (UNISANTOS) e Música (UNESP), com mestrado (UFSCAR) e doutorado (PUC-SP) em Filosofia, atua como orientadora filosófica, pesquisadora e professora no Espaço Monica Aiub: Filosofia, Arte e Cultura, em São Paulo-SP. E-mail: [monicaaiub@monicaaiub.com.br](mailto:monicaaiub@monicaaiub.com.br)

**NANDIA LETICIA FREITAS RODRIGUES** é Mestranda em Ciência da Informação e graduada em Biblioteconomia pela Universidade Estadual Paulista Júlio de Mesquita Filho (UNESP); colaboradora do Laboratório de Pesquisa em Design e Recuperação da Informação. E-mail: [nandiarodrigues@gmail.com](mailto:nandiarodrigues@gmail.com)

**PLÁCIDA LEOPOLDINA VENTURA A. DA COSTA SANTOS** é Livre-docente em Catalogação pela Universidade Estadual Paulista (UNESP); Docente do Departamento de Ciência da Informação da UNESP. E-mail: [placidasantos@gmail.com](mailto:placidasantos@gmail.com)

**RAFAEL GUTIERRES CASTANHA** é Mestre em Ciência da Informação pela Universidade Estadual Paulista (UNESP), *Campus* de Marília e graduado em Licenciatura em Matemática, pela UNESP, *Campus* de Presidente Prudente. Docente da Faculdade Católica Paulista (FACAP); doutorando em Ciência da Informação pela UNESP, *Campus* de Marília. E-mail: [r.castanha@gmail.com](mailto:r.castanha@gmail.com)

**RENE FAUSTINO GABRIEL JUNIOR** graduação em Biblioteconomia e Documentação pela Pontifícia Universidade Católica do Paraná, mestrado em Ciência, Gestão e Tecnologia da Informação pela Universidade Federal do Paraná e doutorado em Ciência da Informação pela Universidade Estadual Paulista Júlio de Mesquita Filho. É professor adjunto da Universidade Federal do Rio Grande do Sul e do Programa de Pós-Graduação em Ciência da Informação da mesma universidade. E-mail: [rene.gabriel@ufrgs.br](mailto:rene.gabriel@ufrgs.br)

**RICARDO CÉSAR GONÇALVES SANT'ANA** é Professor Associado da Universidade Estadual Paulista (UNESP), *Campus* de Tupã; Professor do Programa de Pós-Graduação em Ciência da Informação da UNESP, *Campus* de Marília; Graduado em Matemática e Pedagogia; Mestrado em Ciência da Informação; Doutorado em Ciência da Informação; Livre-Docente em Sistemas de Informações Gerenciais pela UNESP. E-mail: ricardo.santana@unesp.br

**SELMA LETICIA CAPINZAIKI OTTONICAR** é Doutoranda em Ciência da Informação pela UNESP, *Campus* de Marília; Mestre em Ciência da Informação pela UNESP, *Campus* de Marília, e Tecnóloga em Gestão Empresarial, FATEC. E-mail: selma.leticia@hotmail.com

**TIAGO CAMPOS FERREIRA**, graduando em Sistemas de Informação pela Universidade Federal dos Vales do Jequitinhonha e Mucuri. E-mail: camposferreiratiago@gmail.com

## **SOBRE O LIVRO**

### *Capa*

Profa. Maria José Vicentini Jorente  
Isabelle Ribeiro O. C. Lima

### *Diagramação*

Gláucio Rogério de Moraes

### *Assessoria Técnica*

Renato Geraldi

### *Oficina Universitária*

Laboratório Editorial  
labeditorial.marilia@unesp.br

### *Formato*

14 x 21 cm

### *Tipologia*

Adobe Garamond Pro

### *Papel*

cana-de-açúcar 75g/m<sup>2</sup>, branco (miolo)  
papel fotográfico glossy 230g/m<sup>2</sup> (capa)

### *Acabamento*

Colado

### *Tiragem*

300

---

Impressão e acabamento

***FiloCzar***

---

2019

uma perspectiva interdisciplinar, falibilista - sempre aberta à detecção e correção de erros - tendo como leme a interrogação "Big Data: reforma ou revolução?".

Uma das hipóteses discutidas nesta obra é a de que no passado, os colonizadores enviavam exércitos e religiosos para conquistar territórios e apoderar-se de recursos ambientais considerados preciosos; hoje, com recursos avançados de TIC, além dos exércitos, os novos colonizadores financiam técnicos qualificados para minerar e comercializar dados de fontes variadas, incluindo aquela de usuários de redes sociais. As consequências, positivas e negativas, dessa complexa atividade clamam por pensadores que, como Mariana Claudia Broens, homenageada neste livro, ajudam a ampliar o campo de reflexões sobre o presente e o futuro das novas gerações influenciadas pelo acelerado avanço da tecnologia.

Informação, Conhecimento, Ação Autônoma e

# Big Data

continuidade ou revolução?



ISBN 978-85-7249-054-2



9 788572 149054 2