



UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Campus de Marília



CULTURA
ACADÊMICA
Editora

Intersecção temática de programas de pós-graduação brasileiros:

considerações sobre Big Data

Jacquelin Teresa Camperos-Reyes

Luiza de Menezes Romanetto

Plácida Leopoldina Ventura A. da Costa Santos,

Ricardo César Gonçalves Sant'Ana

Como citar: CAMPEROS-REYES, J. T. *et al.* Intersecção temática de programas de pós-graduação brasileiros: considerações sobre Big Data. *In:* MARTÍNEZ-ÁVILA, D.; SOUZA, E. A.; GONZALEZ, M. E. Q. (org.). **Informação, Conhecimento, Ação Autônoma e Big Data: Continuidade ou Revolução?** Marília: Oficina Universitária; São Paulo: Cultura Acadêmica, 2019. p. 67-82.

DOI: <https://doi.org/10.36311/2019.978-85-7249-055-9.p67-82>



All the contents of this work, except where otherwise noted, is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported.

Todo o conteúdo deste trabalho, exceto quando houver ressalva, é publicado sob a licença Creative Commons Atribuição - Uso Não Comercial - Partilha nos Mesmos Termos 3.0 Não adaptada.

Todo el contenido de esta obra, excepto donde se indique lo contrario, está bajo licencia de la licencia Creative Commons Reconocimiento-NoComercial-CompartirIgual 3.0 Unported.

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 <https://doi.org/10.36311/2019.978-85-7249-055-9.p67-82>

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 áreas¹ (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?

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 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. revisão bibliográfica acerca do fenômeno *Big Data*;
2. coleta de dados na plataforma Sucupira² 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).

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

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

PPG. No Brasil, as notas estão no leque de 1 ao 7³, 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.

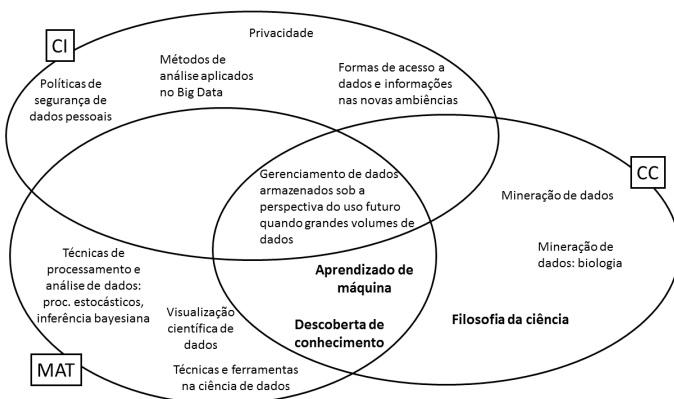
A análise de conteúdo possibilita a produção de inferências recorrendo ou não a indicadores quantitativos (BARDIN, 2010), portanto, buscou-se identificar quais as temáticas pertinentes ao fenômeno Big Data têm sido desenvolvidas pelos PPG no Brasil e a contribuição da Ciência da Informação nesse cenário,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

³ 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

Categoria detectada	Ocorrências
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

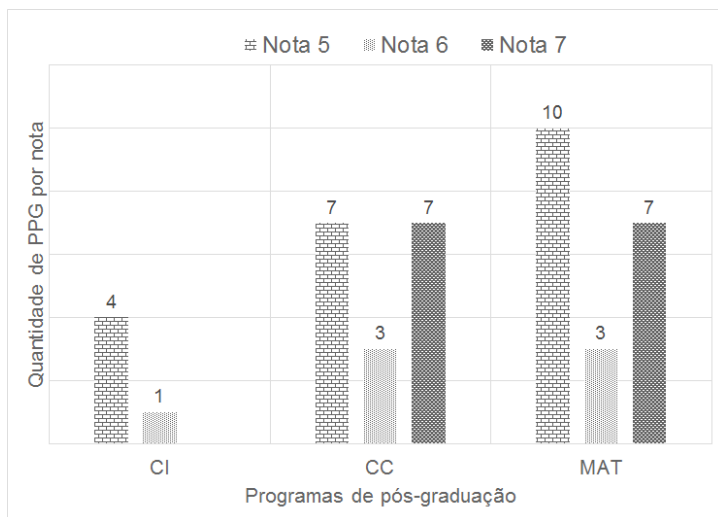
Somando as ocorrências por PPG, na Ciência da Informação, três PPG indicam dentro das linhas de pesquisa o estudo do fenômeno *Big Data*, onze na Ciência da Computação e quatro nos PPG de Matemática.

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 profis-

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?**