

## A datificação de um campo de conhecimento: como algoritmos, números e abordagens quantitativas estão mudando a comunicação

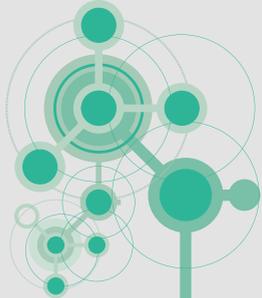
The datification of a knowledge field: how algorithms, numbers and quantitative approaches are changing communication

La datificación de un campo de conocimiento: como los algoritmos, los números y los enfoques cuantitativos están cambiando la comunicación



Márcio Carneiro dos Santos

- Doutor em Tecnologias da Inteligência e Design Digital (TIDD) pela Pontifícia Universidade Católica de São Paulo (PUC-SP)
- Docente na Universidade Federal do Maranhão (UFMA)
- E-mail: [mcszen@gmail.com](mailto:mcszen@gmail.com)



## Resumo

Discute-se o impacto das transformações geradas pelo ecossistema digital dos meios no campo da comunicação, a partir da explosão quantitativa dos emissores, sustentada pela ubiquidade das redes e dispositivos tecnológicos de suporte à produção de conteúdo. Analisam-se, a partir das propostas da *design science* e da abordagem dos métodos digitais, as possibilidades de expansão epistemológica e metodológica, baseadas nas iniciativas interdisciplinares e na incorporação de novas habilidades na formação de profissionais e pesquisadores, para enfrentar a atual situação de excesso de dados e ferramental pouco adaptado para compreendê-la.

PALAVRAS-CHAVE: DADOS • ALGORITMOS • COMUNICAÇÃO.

## Abstract

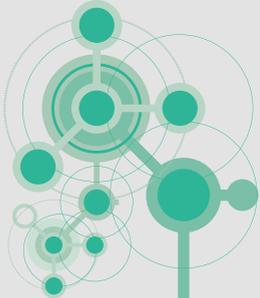
We discuss the impact of the transformations generated by the digital media ecosystem in the field of communication, from the quantitative explosion of the emitters, sustained by the ubiquity of networks and technological devices that support the production of content. Based on the proposals of Design Science and the approach of Digital Methods, we think the possibilities of epistemological and methodological expansion, through interdisciplinary initiatives and the incorporation of new skills in the formation of professionals and researchers, in order to face the current situation of data overload and tools poorly adapted to understand it.

KEYWORDS: DATA • ALGORITHMS • COMMUNICATION.

## Resumen

Discutimos el impacto de las transformaciones generadas por el ecosistema de medios digitales en el campo de la comunicación, a partir de la explosión cuantitativa de los emissores, que se sostiene en la ubicuidad de las redes y dispositivos tecnológicos que apoyan la producción de contenido. Con base en las propuestas de *design science* y el enfoque de métodos digitales, analizamos las posibilidades de expansión epistemológica y metodológica, fundamentadas en iniciativas interdisciplinarias y en la incorporación de nuevas habilidades en la formación de profesionales e investigadores, para enfrentar la situación actual del exceso de datos y herramientas mal adaptadas para entenderla.

PALABRAS CLAVE: DATOS • ALGORITMOS • COMUNICACIÓN.



## INTRODUÇÃO

O campo da comunicação se desenvolveu a partir das escolhas epistemológicas das humanidades e das ciências sociais, incorporando abordagens, métodos e um modo de operação no qual a descrição e, principalmente, a interpretação subjetiva sempre estiveram em primeiro lugar.

Ainda hoje são estranhas para os descendentes dessa tradição científica as funções da predição e da prescrição, formas de trabalho mais orientadas aos projetos, ao enfrentamento de problemas reais e à proposição de soluções, como ocorre no design, nas engenharias, na medicina e nas ciências da terra.

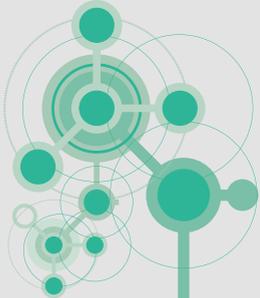
As transformações causadas pela disseminação dos processos digitais no ecossistema informacional contemporâneo trouxeram, entretanto, um conjunto totalmente novo de objetos de estudo e questões da maior importância para a ciência da comunicação. Utilizando apenas um dos diversos tópicos possíveis a discutir, na complexa situação em que vivemos, poderíamos listar perguntas do tipo: como se propagam as notícias falsas e os memes? Que padrões tal difusão segue? Há padrões? Como identificá-los?

Perguntas desse tipo esbarram numa dificuldade inerente a um conjunto de questões semelhantes. O cenário em que tais fenômenos ocorrem é caracterizado por processos que têm como características a velocidade, a variedade e o volume das informações produzidas (González-Bailón, 2013; Lewis; Lima Junior., 2012; Mahrt, Scharkow, 2013; Westlund, 2015). Tal fato inviabiliza determinadas abordagens, que pouco podem contribuir para sua compreensão, simplesmente porque não conseguem sequer identificá-los, registrá-los e, muito menos, buscar neles semelhanças ou divergências para classificá-los.

Amostras pequenas ou tratadas manualmente pouco podem fazer em movimentos de milhares ou milhões de ações humanas, realizadas muitas vezes de forma quase sincrônica, tais como um conjunto de pessoas publicando *tweets* sobre um tema que momentaneamente arrebatou a atenção coletiva (Santos, 2013). Por mais dedicado que seja o pesquisador e por mais horas que dedique à coleta desses dados, tratá-los sem ajuda computacional oferecerá pouco potencial de extrair inferências ou, sequer, de apreender o que efetivamente está acontecendo.

Foram os jornalistas investigativos e os profissionais no ambiente organizacional que primeiro tiveram que lidar com esse tipo de problema. Dados dos portais públicos de transparência e as histórias que podiam conter escondidas atrás dos números; métricas dos indicadores relacionados à presença nas plataformas de mídias sociais; informações trazidas pelas ferramentas de monitoramento; e a enxurrada de dados das soluções de *analytics*, tudo isso trouxe ao campo não apenas um conjunto novo de problemas, mas uma espécie de reação em cadeia que começou a impactar também a necessidade de novas habilidades desses profissionais, diferentes formas de abordagem, a busca por novos modelos de negócio e – por que não também? – a revisão e readequação teórica e epistemológica de um saber que tinha sido construído num mundo analógico, bem diferente do atual.

Por mais estranho que possa parecer, foi a complexificação do comportamento humano que tornou necessária a inserção dos processos maquímicos nas áreas de conhecimento em que antes tais abordagens nunca foram comuns ou bem-vindas. As humanidades digitais (Lemos, 2002; Moretti, 2007), a linguística computacional (Santos, 2014), o jornalismo automatizado ou guiado por dados (Bradshaw, 2014; Bruns, 2017; Rodrigues, 2009), a matemática aplicada à sociologia (Bonachi; Lu, 2012) e tantas novas formas híbridas de conhecimento desafiam os que se propõem a estudar os processos contemporâneos, hoje sustentados por redes e máquinas binárias, gerando objetos digitais com características e ontologia próprias (Manovich, 2001), incapazes de serem desvelados sem uma reorientação de métodos, ferramentas e técnicas de pesquisa (Moretti, 2007; Santos, 2015; Van Dijck, 2013).

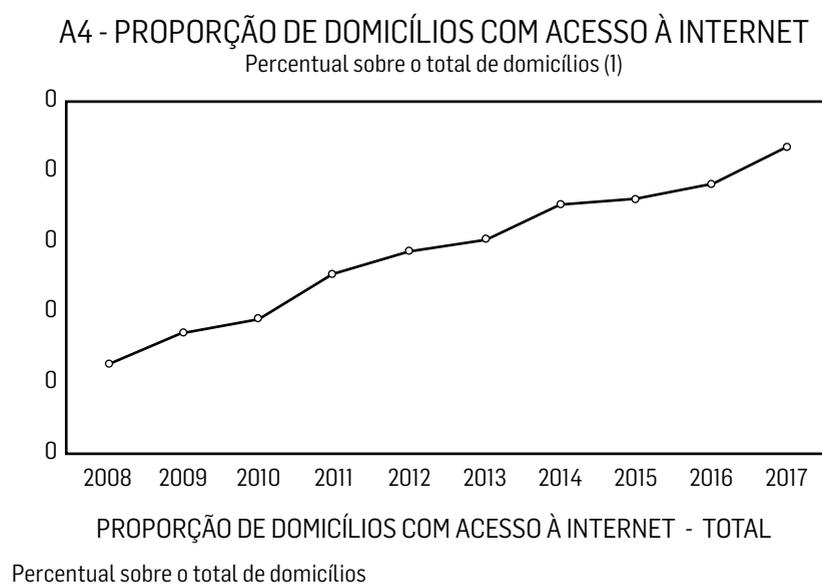


O empoderamento público sustentado pelos meios digitais de comunicação criou uma explosão de emissores, reconfigurando as redes de difusão de informação do mundo analógico, antes concentradas em grandes *hubs* de atenção, como os grandes veículos de mídia e as fontes oficiais.

Como numa cidade onde se abriam diversas novas vias, a circulação de dados ganhou velocidade e volume nunca antes vistos, engendrando processos só passíveis da compreensão a partir das teorias da complexidade, das redes e dos jogos, todas muito estranhas ao pesquisador tradicional da comunicação.

No Brasil os dados disponíveis indicam um sólido movimento ascendente no acesso dos domicílios à internet (Figura 1), sendo o ano de 2015 o marco temporal do momento quando a maioria desses lares passou a ter conexão com a rede (Figura 2).

Figura 1: Série histórica (2008-2017) do crescimento do acesso à internet no Brasil através do percentual de domicílios

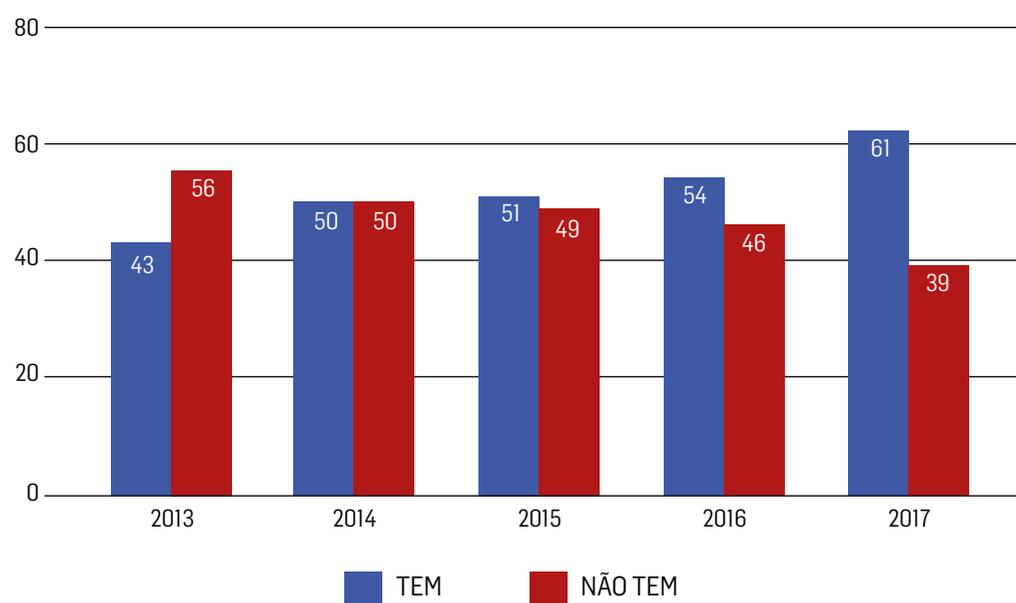


(1) Base: 67.038.766 domicílios. Dados coletados entre Novembro de 2015 e Junho de 2016.  
Fonte: Núcleo de Informação e Coordenação do ponto BR (NIC.br).

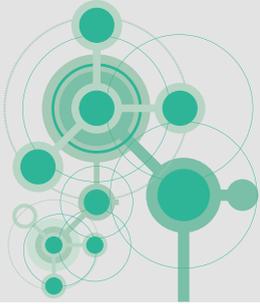
Fonte: Cetic.br, 2019.

Figura 2: Gráfico demonstrando que em 2014/2015 houve o ponto de inflexão quando a maioria dos domicílios brasileiros passou a ter acesso à internet

Percentual de Domicílios com Acesso à Internet - Média Brasil. Fonte: Cetic.br - Tic Domicílios (2013-2017)



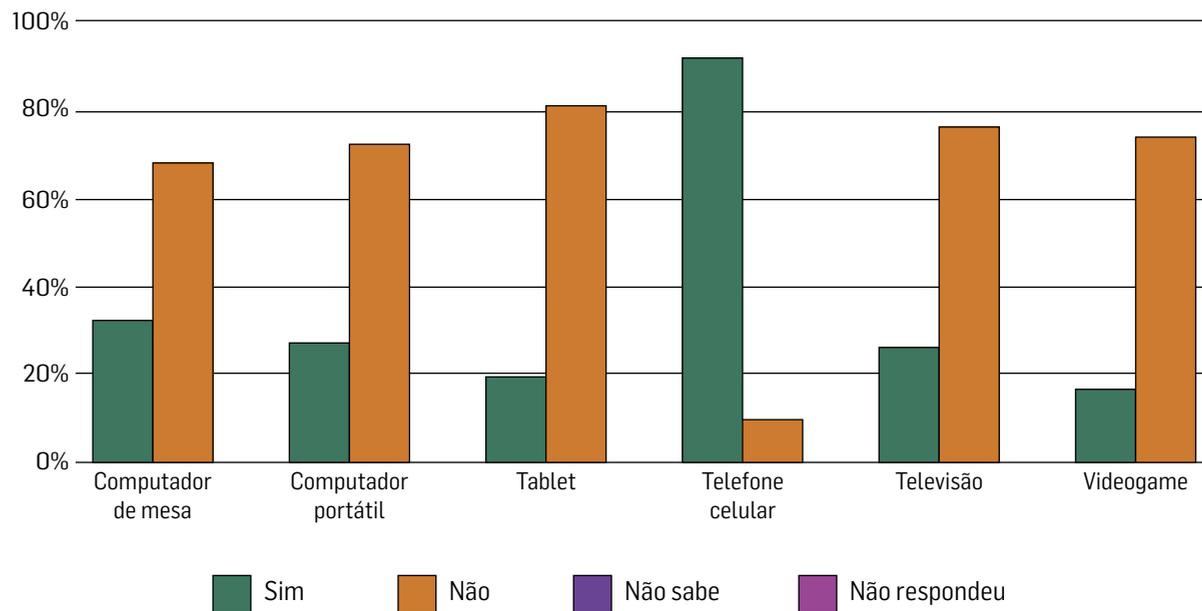
Fonte: Elaborado pelo autor com dados do TIC Domicílios – Cetic.br, 2019.



Outro aspecto importante é a utilização dos dispositivos móveis como principal tecnologia de acesso não só entre adultos, mas também entre crianças e adolescentes (Figura 3).

Figura 3: Distribuição das tecnologias de acesso entre jovens e adolescentes indicando forte predominância do celular na pesquisa, TIC Kids Online 2017

A1 - CRIANÇAS E ADOLESCENTES, POR DISPOSITIVOS UTILIZADOS PARA ACESSAR A INTERNET  
Total de usuários de internet de 9 a 17 anos (I)



Total de usuários de Internet de 9 a 17 anos

(I) Fonte: CGI.br, NIC.br, Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação (Cetic.br), Pesquisa sobre o Uso da Internet por Crianças e Adolescentes no Brasil - TIC Kids Online Brasil 2017.

Fonte: Núcleo de Informação e Coordenação do Ponto BR (NIC.br).

Fonte: Cetic.br, 2019.

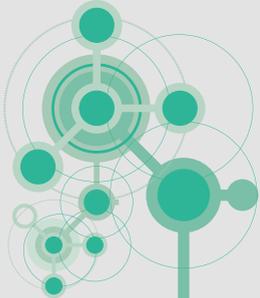
Assim, pensar em inovação na comunicação só é possível a partir de uma atualização de modelos teóricos e ferramentas para coleta e análise de dados, numa atividade empreendedora, aberta a novos paradigmas e à busca de soluções viáveis para nossos problemas atuais. De acordo com a etimologia, “empreender” está relacionado ao termo francês *entrepreneur*, palavra cujo sentido tem a ver com correr riscos e começar algo novo, ou, ainda, à raiz latina *imprehendere*, que comporta a ideia de apreender, captar a essência, compreender.

Assim, inovação e empreendedorismo na comunicação do ambiente digital devem traduzir-se numa atitude, antes de tudo, de humildade, de reconhecimento das transformações pelas quais temos passado e, por estarmos ainda num processo de transição, pelas que ainda estão por vir. O apego a feudos de conhecimento que têm se mostrado inadequados para lidar com a situação atual só atrasa o desenvolvimento científico, qualquer que seja o campo em questão.

## TÓPICOS DE RECONFIGURAÇÃO

Diante do amplo espectro de problemas a serem enfrentados e das limitações deste texto, optamos por organizar a ação empreendedora e científica que entendemos necessária a partir de grandes linhas de atuação relacionadas a essa temática:

- Ampliação epistêmica: incorporação das atividades ou tentativas de predição e prescrição nas iniciativas de pesquisa ligadas às classes de problemas do mundo real; incentivo à pesquisa aplicada; e orientação a abordagens guiadas por dados presentes em propostas como a da *design science*;



- b. Extensão metodológica: experimentação e teste de abordagens como a dos métodos digitais, baseadas na premissa de uma ontologia específica dos entes binários que impacta suas formas de apreensão, como, por exemplo, a metodologia de análise de redes sociais (ARS);
- c. Interdisciplinaridade: abertura à conexão e operação de grupos de pesquisadores oriundos de outras tradições acadêmicas para o enfrentamento de problemas que também se constituem a partir da hibridização de temáticas de complexidade aprofundada;
- d. Revisão teórica: adaptação e proposição de constructos, modelos e explicações com possibilidade de operação no atual ecossistema informacional contemporâneo;
- e. Diversificação de habilidades: abertura à aproximação de abordagens quantitativas mais consistentes, que incluam análise e visualização de dados, estatística e até, em alguns casos, matemática e linguagens de programação, como Python;
- f. Atualização nos programas de formação de profissionais e pesquisadores: o que implica revisão (ou renovação de propostas) de disciplinas, ementas, programas e, eventualmente, áreas de concentração e linhas de pesquisa, no caso dos programas de pós-graduação.

Ressalte-se que, de forma alguma, a inserção dessas mudanças deverá acontecer de forma acrítica ou contemplativa, a partir de uma visão simplista de que a aproximação com dados, algoritmos e ferramentas computacionais por si só nos salvará a todos. Hoje, no Brasil, já existe, ainda que de forma embrionária, um conjunto de pesquisadores que têm explorado, por exemplo, o viés e os erros que a utilização de ferramentas de inteligência artificial (especificamente *machine learning*) têm gerado em processos que acabam descambando para invasão da privacidade, distorção dos fatos e até práticas como racismo e xenofobia, como demonstram Amadeu (2019) e Silva (2019).

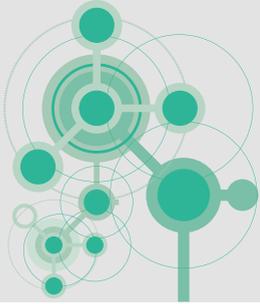
Entretanto, é justamente para que a tradição crítica do campo da comunicação possa continuar a ser exercida no ambiente digital, que se faz necessária uma aproximação direta de tais temas ou, pelo menos, de forma indireta, através do trabalho em equipes interdisciplinares, capazes de operar mais efetivamente com tais questões. Afastar-se dessas temáticas, isso sim, poderá nos colocar em posição de fragilidade científica e até irrelevância, num movimento de distanciamento dos processos essenciais que guiam as transformações no ambiente digital.

Importante também lembrar que a identificação das transformações causadas pelas mudanças estruturais nos sistemas informacionais de essência digital/binária vem sendo feita por vários teóricos da comunicação e de outras áreas, como Castells (1999), Chwe (2000), Feenberg (2002), Lemos (2002), Santaella (2003) e Vilches (2003), entre outros, há pelo menos vinte anos. Tal fato nos leva a crer que, depois de todo esse tempo, estudos simplesmente constatatórios ou repetitivos dessas proposições têm pouco a agregar à expansão dos respectivos campos científicos. Se tais fenômenos existem e já foram apontados, é importante agora estudar como se dão, que padrões seguem, que forças os movem e que impactos poderão gerar no futuro próximo. A profusão de estudos de caso, tão comuns na nossa área, indica um baixo potencial de inferências e proposições de mais largo alcance, principalmente no desenvolvimento de novas teorias e revisão das anteriores.

## DESIGN SCIENCE E MÉTODOS DIGITAIS

Os itens a e b da lista elaborada no item anterior podem ser encaminhados, entre outras possibilidades, mediante a aproximação com dois tipos de abordagens ainda pouco conhecidas no ambiente nacional da pesquisa em comunicação. Temos trabalhado no Laboratório de Convergência de Mídias (Labcom)<sup>1</sup> na difusão dessas possibilidades teórico-metodológicas, não só em

<sup>1</sup> Disponível em: [www.labcomdata.com.br](http://www.labcomdata.com.br).



termos do desenvolvimento de projetos orientados à pesquisa aplicada, mas também a partir de reflexões já documentadas anteriormente que resumiremos a seguir para fins de detalhamento das proposições que ora defendemos.

### *Design science (DS)*

O termo *science of design*, que posteriormente passou a ser *design science* (DS), foi introduzido pelo economista e filósofo Herbert Simon numa obra considerada seminal para esse campo, *The sciences of the artificial* (As ciências do artificial), publicada pela primeira vez em 1969. Nela o autor começa a esboçar um novo paradigma epistemológico, que hoje se caracteriza pela orientação à solução de problemas, seja através da criação de novos artefatos (conceito que vamos detalhar adiante), seja pela melhoria das soluções existentes. Focada inicialmente nos campos da engenharia e dos sistemas de informação, a DS viu sua utilização expandir-se para gestão e educação, bem como pelas ciências sociais aplicadas de forma geral, oferecendo um caminho alternativo para pesquisadores que desejam ir além das fases de descrição e análise de objetos de pesquisa dados previamente.

O caráter prescritivo e propositivo dessa vertente procura integrar projetos que, mantendo o rigor dos métodos científicos tradicionais, buscam também a relevância social de seus achados na implementação de melhorias objetivas a problemas de determinada classe.

A DS propõe uma espécie de extensão, direcionada à solução de problemas reais e à melhoria de artefatos existentes. Em Gibbons (1994), Le Moigne (1994), March e Smith (1995), Romme (2003), Walls, Wyidmeyer e Sawy (1992) encontramos diversas referências a essa visão. No Brasil o trabalho de Dresch, Lacerda e Antunes Júnior (2015) é uma referência sobre o tema.

O direcionamento da DS para critérios objetivos na solução de problemas pode ser entendido a partir de um de seus conceitos fundamentais, o de “validade pragmática”, que “busca assegurar a utilidade da solução proposta para o problema. Considera: custo/benefício da solução, particularidades do ambiente em que será aplicada e as reais necessidades dos interessados na solução” (DRESCH; LACERDA; ANTUNES JÚNIOR, 2015, p.59).

### *Artefatos e classes de problemas*

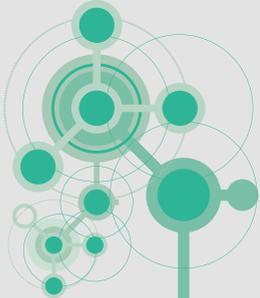
Outro conceito fundamental para a DS é o de “artefato”. Design Science é a “ciência que procura consolidar conhecimento sobre o projeto e desenvolvimento de soluções para melhorar sistemas existentes, resolver problemas e criar novos artefatos” (DRESCH; LACERDA; ANTUNES JÚNIOR, 2015, p.59). O conceito de artefato pode ser entendido como o produto final do percurso proposto pela DS e por isso algo que está associado ao contexto específico do problema a ser resolvido. O artefato (Figura 4), criado pelo homem, representa um intermediador entre um conjunto do conhecimento estabelecido em determinada área e as condições específicas que envolvem o problema que o artefato deverá resolver.

Figura 4: Escala de tangibilidade dos artefatos na DS



Fonte: Elaborado pelo autor.

Os artefatos podem ser divididos em categorias, sendo uma das classificações mais aceitas a de March e Smith (1995), que propõe quatro tipos: constructos, modelos, métodos e instanciações.



**Constructos:** os componentes mais básicos no desenvolvimento da DS. São elementos conceituais cujo objetivo é estabelecer um conjunto de definições utilizadas na solução de problemas, constituindo uma espécie de vocabulário sobre determinado campo em que tal problema está inserido. São os conceitos sobre os quais a solução opera e que o pesquisador utilizará para evoluir do puramente abstrato para o tangível (Figura 4) e aplicado à determinada situação.

**Modelos:** são descrições sobre determinado sistema que estabelecem relações entre os constructos previamente definidos. São uma espécie de representação da realidade que procura descrevê-la, mesmo que através de simplificações, mas que tem o objetivo de apreender sua lógica de operação interna para utilização como solução.

**Métodos:** são conjuntos de procedimentos e ações orientados para o desempenho de determinada tarefa ou solução de dado problema. Os métodos podem estar relacionados a modelos previamente estabelecidos, sendo um passo a mais na escala entre abstração e tangibilidade da solução que propomos anteriormente.

**Instanciações:** o conceito de instância ou instanciar, bastante conhecido entre programadores e cientistas da computação, talvez represente na DS o nível mais tangível da solução criada no contexto prévio que a inspirou, ou seja, o artefato em operação no ambiente que gerou a necessidade da solução. As instanciações nos permitem também avaliar algo importante dentro da proposta da DS: sua efetividade em relação ao problema proposto ou às melhorias pretendidas no sistema existente.

Um quinto tipo de artefato é admitido por alguns autores que se referem a ele usando termos como “regras tecnológicas”, “regras de projeto” ou, mais comumente, *design propositions*.

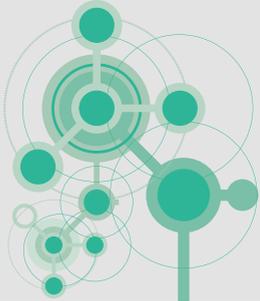
*Design propositions:* essas proposições de design ou do projeto seriam contribuições teóricas que podem ser feitas a partir da aplicação dos princípios da DS diante de um tipo específico de problemas ou, em seus próprios termos, operando sobre uma classe de problemas.

O termo “classe de problemas”, que temos utilizado, também faz parte dos conceitos importantes da DS. Conjuntos de problemas práticos ou teóricos que têm já estabelecido um conjunto de soluções ou artefatos a eles ligados constituem uma classe de problemas. Como exemplo da comunicação e das ciências sociais, poderíamos citar a necessidade geral de coletar dados em repositórios na internet, que poderíamos nomear como coleta de dados digitais. Seja para a produção de uma matéria jornalística, seja para um plano de gestão ou a definição de uma política pública sobre determinado tema, com os processos de digitalização e o crescimento do uso de bases de dados, a necessidade de conseguir tais informações, acessando seus repositórios disponíveis na rede – tais como portais de transparência, por exemplo – caracteriza uma classe de problemas. Nela operam artefatos como os métodos de *scraping* (raspagem) e extração automatizada, bem como as instanciações disponíveis exemplificadas pelos algoritmos em determinada linguagem de programação, que atuam para resolver tais problemas. Nesse último exemplo, os códigos poderiam não só ser classificados como instanciações mas também como métodos, já que executam sequências de comandos para realizar suas funções.

## MÉTODOS DIGITAIS

Rogers (2013) afirma que, mesmo portando métodos tradicionais para o emprego em pesquisas ligadas ao digital, podemos, em algumas situações, estar utilizando um ferramental inadequado.

Por exemplo, varredura e extração de dados, inteligência coletiva e classificações baseadas em redes sociais, ainda que de diferentes gêneros e espécies, são todas técnicas baseadas na internet para coleta e organização de dados. *Page Rank* algoritmos similares são

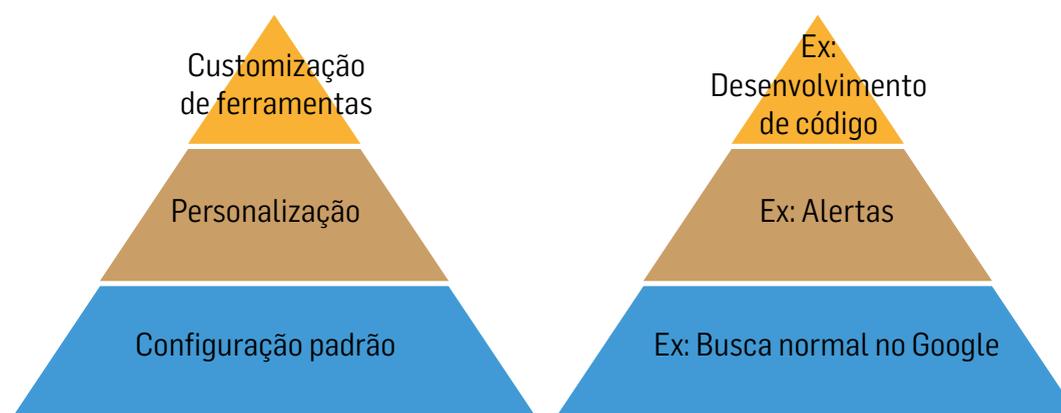


meios de ordenação e classificação. Nuvens de palavras e outras formas comuns de visualização explicitam relevância e ressonância. Como poderíamos aprender com eles e outros métodos on-line para recriá-los? O propósito não seria tanto contribuir para o refinamento e construção de um motor de buscas melhor, uma tarefa que deve ser deixada para a Ciência da Computação e áreas afins. Ao invés disso o propósito seria utilizá-los e entender como eles tratam *hiperlinks*, *hits*, *likes*, *tags*, *datestamp* e outros objetos nativamente digitais. Pensando nesses mecanismos e nos objetos com os quais eles conseguem lidar, os métodos digitais, como uma prática de pesquisa, contribuem para o desenvolvimento de uma metodologia do próprio meio. (ROGERS, 2013, tradução nossa)

A proposta de Rogers vai ao encontro do percurso que ora propomos, partindo de uma visão do mundo contemporâneo onde o digital apresenta uma centralidade crescente, composto por entes com características específicas e, por isso, demandando também uma adequação ou extensão metodológica capaz de colaborar com pesquisas cujos objetos de alguma forma têm essa característica.

Desse modo, definimos métodos digitais como o conjunto de ferramentas, processos e abordagens de pesquisa que consideram a ontologia dos objetos digitais e as estruturas de redes por onde circulam, utilizando-se de recursos computacionais intensivos para coleta e análise de dados. Tais soluções oferecem uma espécie de escala de utilização (Figura 5).

Figura 5: Representação da escala de utilização dos métodos digitais



Tal escala parte da utilização de ferramentas e técnicas já existentes em sua configuração padrão num nível inicial; passa por ajustes a fim de personalizá-las para atender a nossas necessidades específicas, num nível médio; ou, ainda, num nível mais alto, apresenta soluções baseadas em programação e desenvolvimento de código.

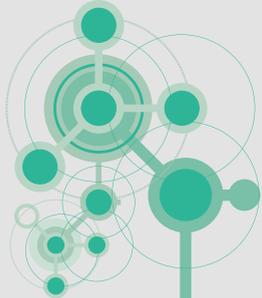
Nas pirâmides da Figura 5, exemplificamos a escala numa situação de coleta de dados que utiliza a busca do Google, inicialmente com sua interface normal, depois a partir de uma solução com maior poder de personalização, como os alertas<sup>2</sup>, e por fim através de um código específico para coletar e armazenar esses dados.

Em termos gerais, a abordagem que propomos resume-se às seguintes etapas:

Etapa 1: identificar a estrutura que contém os dados de que precisamos. Algumas possibilidades apresentam-se com mais frequência:

- a. Bases de dados que permitem consultas amigáveis via preenchimento de formulários ou procedimentos simples. Exemplo: portais de transparência governamentais em que é possível requisitar dados sobre determinado tema e período.

<sup>2</sup> Disponível em: <https://www.google.com/alerts>.



- b. Interfaces de programação de aplicações (*application programming interfaces* – APIs) que exigem requisições estruturadas no formato que estabelecem, ou seja, respeitando sua sintaxe própria. Exemplo: APIs do *Twitter* e do *Facebook* que precisam ou de uma aplicação específica para solicitar conteúdo, como os aplicativos que as acessam em nossos celulares, ou de um código customizado que consiga estabelecer tal diálogo e coletar as informações que a API entrega a partir de cada tipo de requisição.
- c. Conteúdo disponível em páginas de internet que podem ser extraídos diretamente via técnicas de *scraping* (raspagem de dados), como textos de matérias em portais jornalísticos ou tabelas e informações gerais publicadas, tais como previsão do tempo, cotação do dólar e resultados de competições esportivas.
- d. Informações protegidas em ambientes fechados, acessadas apenas por usuários cadastrados e que contam com mecanismos de proteção como encriptação de dados e outros. Tais ambientes eventualmente podem ser acessados por técnicas de *hacking* que estão além do escopo deste texto.

Etapa 2: formatar a consulta ou requisição de dados alinhada ao tipo de repositório em que se encontram, de acordo com as opções anteriormente descritas.

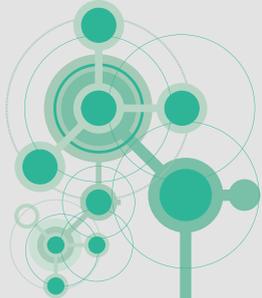
Etapa 3: analisar os dados coletados a partir do processamento possível partindo do que foi efetivamente conseguido.

Podemos combinar então a escala de utilização da Figura 3 com as diversas formas de estruturação de dados mais comuns (Quadro 1). A partir do cruzamento das quatro formas mais comuns de repositórios estruturados em bases de dados on-line, listamos algumas possibilidades de aplicação dos métodos digitais em seus três níveis.

Quadro 1: Matriz de possibilidades de coleta via métodos digitais em função da forma e local dos dados e dos níveis de aplicação

Estrutura dos dados	Nível inicial: ferramentas padrão	Nível médio: ferramentas com personalização	Nível alto: desenvolvimento de código
a) Bases de dados e repositórios acessíveis	Solicitação de dados através da própria interface da base de dados, recebendo o resultado no formato padrão de entrega. Ex.: acesso à base Sidra do IBGE e download do arquivo no formato do Excel ou em CSV.	Utilização de filtros e recursos de análise e visualização oferecidos pela plataforma, alterando a forma de entrega do resultado de acordo com as opções oferecidas. Ex.: uso das funções avançadas da Sidra e geração de gráfico.	Código para automatizar o acesso ao banco de dados fazendo requisições sucessivas, customizadas, coletando e salvando os registros em outro tipo de estrutura ou formato de dados. Ex.: Python com módulos Splinter ou Selenium.
b) Servidores com acesso via API específica	Acesso via aplicação oficial da plataforma ou através de sua página web padrão. Ex.: uso do aplicativo do Facebook no celular ou acesso à página <a href="http://www.facebook.com">www.facebook.com</a> .	Acesso através de aplicativos de terceiros que também se conectam ao servidor da plataforma, mas oferecem funcionalidades adicionais. Ex.: Node XL.	Código para acessar diretamente a API da plataforma coletando todas as informações disponibilizadas por ela e também fazendo requisições sucessivas capazes de coletar volumes maiores de dados.

Continua...



Quadro 1. Continuação

Estrutura dos dados	Nível inicial: ferramentas padrão	Nível médio: ferramentas com personalização	Nível alto: desenvolvimento de código
c) Conteúdo em páginas web HTML	Busca do Google, acesso manual e eventual coleta via Ctrl+C e Ctrl+V.	Utilização de ferramentas específicas para <i>scrapping</i> . Ex.: Portia, Web Scraper	Desenvolvimento de código para coleta e análise. Ex.: Python com módulo Beautiful Soup
d) Dados protegidos mediante acesso logado	Acesso via solicitação de cadastro e log in normal.	Ferramentas de <i>hacking</i> geral tipo <i>brute force</i> ou engenharia social.	Desenvolvimento de códigos de invasão tipo <i>worm</i> ou <i>trojan</i> .

Fonte: Elaborado pelo autor.

## CONSIDERAÇÕES FINAIS

A necessidade de uma reconfiguração das práticas de pesquisa no campo da Comunicação nos parece um importante tópico ao tratarmos de objetos e problemas de pesquisa situados no contexto do ecossistema informacional contemporâneo, constituído a partir de entes binários, hiperconectados em redes cada vez mais complexas.

As características de volume, variedade e velocidade impactam diretamente a efetividade de abordagens não orientadas a lidar com sua presença, interferindo em resultados, diminuindo as possibilidades de inferências mais sólidas e, inclusive, inviabilizando a apreensão de fenômenos cada vez mais importantes na lista dos pesquisadores da Comunicação.

As propostas da *Design Science* e dos Métodos Digitais, aqui citados apenas como exemplos, constituem um caminho possível (mas não único) capaz de acrescentar novos horizontes em termos epistêmicos e metodológicos aos pesquisadores de um campo cada vez mais desafiado pela produção avassaladora de dados e conteúdo, gerados não apenas por fontes oficiais e grandes veículos de comunicação, mas também por uma explosão de emissores empoderados por dispositivos móveis e infraestrutura de internet em expansão no país.

Não se sugere aqui, obviamente, o abandono de todo um conjunto de conhecimentos consolidados que fez o campo nacional da Comunicação constituir-se e crescer ao longo de décadas. Pelo contrário, entendemos que uma expansão que reflita a centralidade atual de redes e entes digitais, é necessária, a partir do que já foi feito, sem, contudo, bloquear as novas possibilidades que a interdisciplinaridade e a composição com equipes de pesquisadores oriundas de outras origens acadêmicas têm a oferecer em muitas iniciativas de pesquisa que estão diante dos cenários descritos neste texto.

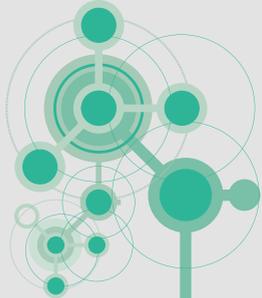
Não se trata de tarefa fácil, mas através das linhas ora apontadas – ampliação epistêmica; extensão metodológica; interdisciplinaridade; revisão teórica; diversificação de habilidades; e atualização nos programas de formação de profissionais e pesquisadores – entendemos que é possível dar o primeiro passo.

Inovação e empreendedorismo em comunicação não devem ser pensados apenas como tópicos externos ligados a mercados e consumo. É preciso usá-los como metáforas ou guias para um movimento interno de reconfiguração e revisão de práticas de pesquisa urgente, necessário e, ao nosso modo de ver, inevitável.



## REFERÊNCIAS

- AMADEU, Sérgio Amadeu da. *Democracia e os códigos invisíveis: como os algoritmos estão modulando comportamentos e escolhas políticas*. São Paulo: Edições Sesc, 2019. (Coleção Democracia Digital)
- BONACICH, Phillip; LU, Phillip. *Introduction to mathematical sociology*. Princeton: Princeton University Press, 2012.
- BRADSHAW, Paul. *Scraping for journalists*. Victoria: Leanpub, 2014.
- BRUNS, Axel. Faster than the speed of print: reconciling big data social media analysis and academic scholarship. *First Monday*, Chicago, v.18, n.10, 2013. Disponível em <https://bit.ly/2lYnryg>. Acessado em: 23 jul. 2017.
- CASTELLS, Manuel. *A sociedade em rede*. São Paulo: Paz e Terra, 1999.
- CHWE, Michael S. Communication and coordination in social networks. *The Review of Economic Studies*, Oxford, v.67, n.1, p.128-156, 2000. doi: 10.1111/1467-937X.00118.
- CODDINGTON, Mark. Clarifying journalism's quantitative turn. *Digital Journalism*, Abingdon, v.3, n.3, p.331-348, 2015. doi: 10.1080/21670811.2014.976400.
- DRESCH, Aline; LACERDA, Daniel; ANTUNES JÚNIOR, José Antonio do Valle. *Design science research: método de pesquisa para avanço da ciência e tecnologia*. Porto Alegre: Bookman, 2015.
- FEENBERG, Andrew. *Transforming technology: a critical theory revisited*. New York: Oxford University Press, 2002.
- GIBBONS, Michael *et al.* *The new production of knowledge: the dynamics of science and research in contemporary societies*. London: Sage, 1994.
- GONZÁLEZ-BAILÓN, Sandra. Social Science in the era of Big Data. *Policy and Internet*, Hoboken, v.5, n.2, 2013. doi: 10.1002/1944-2866.POI328.
- LE MOIGNE, Jean-Louis. *Le constructivisme: tome 1: des fondements*. Paris: ESF, 1994.
- LEMOS, André. *Cibercultura: tecnologia e vida social na cultura contemporânea*. 4. ed. Porto Alegre: Sulina, 2002.
- LEWIS, Seth; WESTLUND, Oscar. Big data and journalism: epistemology, expertise, economics and ethics. *Digital Journalism*, Abingdon, v.3, n.3, 2015, p.447-466. doi: 10.1080/21670811.2014.976418.
- LIMA JUNIOR, Walter. Big data, jornalismo computacional e data jornalismo: estrutura, pensamento e prática profissional na Web de dados. *Estudos em Comunicação*, Covilhã, n.12, p.207-222, 2012. Disponível em: <https://bit.ly/2lTuqIW>. Acesso em: 21 jul. 2017.
- MAHRT, Merja; SCHARROW, Michael. The value of big data in digital media research. *Journal of Broadcasting & Electronic Media*, Abingdon, v.57, n.1, 2013. doi: 10.1080/08838151.2012.761700.



MARCH, Salvatore; SMITH, Gerald. Design and natural science research on information technology. *Decision Support Systems*, v.15, p.251-266, 1995. doi: 10.1016/0167-9236(94)00041-2.

MANOVICH, Lev. *The language of new media*. Cambridge, MA: MIT Press, 2001.

MORETTI, Franco. *Graphs, maps, trees: abstract models for literary history*. New York, Verso, 2007.

RODRIGUES, Adriana Alves. *Infografia Interativa em base de dados no jornalismo digital*. 130f. Dissertação (Mestrado em Comunicação) – Universidade Federal da Bahia, Salvador, 2009.

ROGERS, Richard. *Digital methods*. Cambridge, MA: MIT Press, 2013. *E-book*.

ROMME, A. Georges. Making a difference: organization as design. *Organization Science*, Catonsville, v.14, n.5, p.558-573, 2003.

SANTAELLA, Lucia. *Culturas e artes do pós-humano: da cultura das mídias à cibercultura*. São Paulo: Paulus, 2003.

SANTOS, Márcio Carneiro. Conversando com uma API: um estudo exploratório sobre TV social a partir da relação entre o Twitter e a programação da televisão. *Revista Geminis*, São Carlos, v.4, n.1, p.89-107, 2013. Disponível em: <https://bit.ly/2lTxjck>. Acesso em: 20 abr. 2013.

SANTOS, Márcio Carneiro. Textos gerados por software: surge um novo gênero jornalístico. *In: CONGRESSO BRASILEIRO DE CIÊNCIAS DA COMUNICAÇÃO*, 37., 2014, Foz do Iguaçu. *Anais[...]*. Foz do Iguaçu: Intercom, 2014. p.1-15. Disponível em: <https://bit.ly/2kOhJIT>. Acesso em: 26 jan. 2014.

SANTOS, Márcio Carneiro. Métodos Digitais: a internet e as redes como instrumentos de pesquisa. *In: CONGRESSO BRASILEIRO DE CIÊNCIAS DA COMUNICAÇÃO*, 38., 2015, Rio de Janeiro. *Anais[...]*. Rio de Janeiro: Intercom, 2015. p.1-17. Disponível em: <https://bit.ly/2lYMtxc>. Acesso em: 29 jul. 2016.

SILVA, Tarcízio. Linha do tempo do racismo algorítmico. *Blog do Tarcízio Silva*, [s.l.], 2019. Disponível em: <https://bit.ly/2kQHe2t>. Acesso em: 20 ago. 2019.

SIMON, Herbert. *The sciences of the artificial*. 3. ed. Cambridge, MA: MIT Press, 1996.

VAN AKEN, Joan Ernest. *The research design for design science research in management*. Eindhoven: [s.n.], 2011.

VAN DIJCK, José. *The culture of connectivity: a critical history of social media*. New York: Oxford University Press, 2013.

VILCHES, Lorenzo. *A migração digital*. Rio de Janeiro: Editora PUC-Rio; São Paulo: Loyola, 2003.

WALLS, Joseph; WYIDMEYER, George; SAWY, Omar. Building an information system design theory for vigilant EIS. *Information Systems Research*, Catonsville, v.3, n.1, p.36-60, 1992.