

# **Revolução Científica? Reflexões para uma teoria crítica da Ciência dos Dados**

*Max Melquiades da Silva<sup>a</sup> e Simone Cristina Dufloth<sup>b</sup>*

**Resumo:** O trabalho apresenta uma investigação teórico-conceitual que analisa, à luz da noção de paradigma científico na filosofia da ciência, especialmente os estudos epistemológicos pós-popperianos de Thomas Kuhn (1994) e Inre Lakatos (1979), as características do chamado quarto paradigma da ciência (e-Science). O paradigma da e-Science é apresentado por Gray e outros (2009) como um novo padrão intelectual de produção de conhecimento científico baseado no trabalho intensivo com dados tem conseguido consolidar uma ampla comunidade de prática de pesquisa ao redor do mundo. O trabalho evidencia o excesso de pesquisas de estudo de caso, em contraposição a escassos trabalhos teórico-conceituais que explicitem e discutam os fundamentos teóricos e epistemológicos da e-Science.

**Palavras-chave:** e-Science. Big data. Epistemologia.

## **Scientific revolution? Reflections for a critical theory of Data Science**

**Abstract:** The paper presents a theoretical-conceptual research that analyzes, in light of the notion of scientific paradigm in the philosophy of science, especially the post-Popperian epistemological studies of Thomas Kuhn (1994) and Inre Lakatos (1979), the characteristics of the so-called fourth paradigm of science (e-Science). The paradigm of e-Science is presented by Gray et al. (2009) as a new intellectual standard of scientific knowledge production based on data intensive work has managed to consolidate a wide community of research practice around the world. The work evidences the excess of case study research, as opposed to scarce theoretical-conceptual works that explain and discuss the theoretical and epistemological foundations of e-Science.

**Keyword:** e-Science. Big data. Epistemology.

---

a Universidade Federal de Minas Gerais (UFMG). E-mail: [iemax@bol.com.br](mailto:iemax@bol.com.br). ORCID: <https://orcid.org/0000-0003-3372-8615>. Currículo: <http://lattes.cnpq.br/5624247693651969>

b Fundação João Pinheiro. E-mail: [sduf@uol.com.br](mailto:sduf@uol.com.br). ORCID: <https://orcid.org/0000-0002-1963-7365>. Currículo: <http://lattes.cnpq.br/6166815079725368>

## **1 Introdução**

Se a história da ciência no século XX é caracterizada pela emergência de um sem número de recortes temáticos e a consolidação de diversas áreas do conhecimento com o status de ciências (SHAPIN, 1996), o fim do século reservava inovações tecnológicas que permitiriam aumentar a escala e a rapidez das transações informacionais – o que em última instância altera a forma e o ritmo do trabalho científico e a interação entre áreas do conhecimento.

A possibilidade de gerar novos dados de pesquisa em um volume sem precedentes é vista por cientistas de dados como uma realidade análoga à descrita pela lei de Moore ao se referir ao aumento progressivo da capacidade de processamento dos computadores com redução de custos: Em intervalos regulares de tempo o volume de dados cresce e o custo diminui: o desafio se desloca da obtenção para o tratamento e análise desses dados (SAYÃO e SALES, 2015). Nessa linha, Gray aponta para o fato de que a evolução em curso estaria transformando a própria dinâmica da ciência e seus paradigmas: “Quase tudo sobre ciência está mudando por causa do impacto da tecnologia da informação. A ciência experimental, teórica e computacional está sendo afetada pelo dilúvio de dados, e um quarto paradigma da ciência, ‘intensivo em dados’, está emergindo. O objetivo é ter um mundo em que toda a literatura científica esteja on-line, todos os dados científicos estejam on-line e eles interajam uns com os outros” (GRAY, 2009, 29).

Conforme Gray, uma das maiores referências em se tratando de quarto paradigma da ciência, mesmo nas “ciências de poucos dados” (Idem, p. 20) o trabalho do cientista consiste cada vez mais em analisar a informação disponível – com crescente emprego de recursos de software e hardware – em detrimento dos processos tradicionais de coleta. Nesse sentido, uma das mudanças mais significativas na pesquisa contemporânea residiria justamente no surgimento de “um novo padrão intelectual de produção de conhecimento científico – conhecido como o Quarto Paradigma Científico ou e-Science – caracterizado pela produção e uso intensivo de dados e pelo compartilhamento e colaboração baseados em rede de computadores de alto desempenho” (SALES e SAYÃO, 2015).

Mas como se constitui e evolui um paradigma científico? As transformações em curso no sentido de uma crescente ressignificação do dado como unidade de análise e sentido chegam a constituir um novo paradigma? A produção teórico-conceitual no campo da ciência da informação tem acompanhado o progressivo trabalho com big data?

## **2 Objetivos e metodologia**

Este trabalho busca analisar, à luz da noção de paradigma científico na filosofia da ciência, as características do chamado quarto paradigma da ciência. O trabalho apresenta uma investigação exploratória através de revisão bibliográfica. No próximo tópico são apresentadas características acerca do chamado quarto paradigma da ciência como descritas por alguns de seus principais expoentes. Na sequência, analisa-se como surge um paradigma científico segundo a visão de Thomas Khun e Inre Lakatos, dois dos mais destacados nomes do pensamento científico pós/popperiano. Por fim, são discutidas questões acerca da base teórica que sustenta as discussões e trabalhos empíricos no campo da e-Science.

### **3 O chamado quarto paradigma da ciência: história e características**

As inovações na pesquisa científica a partir das mudanças no volume de dados e no emprego de estratégias e instrumentos computacionais para sua análise tem sido chamada de e-Science e o primeiro registro histórico dessa nomenclatura é atribuído ao diretor-geral do Escritório de Ciência e Tecnologia do Reino Unido, John Taylor, em um documento de 1999. A utilização do termo, contudo, foi ganhando amplitude no decorrer da década de 2000, com o avanço das tecnologias de computação em nível de software e hardware. O termo, que experimentou algumas variações semânticas, talvez por sua amplitude, faz referência a métodos de obtenção de resultados científicos através da utilização de computação intensiva, paralela ou distribuída, de imensos volumes de dados.

O termo e-Science tem sido mais amplamente utilizado desde então, como “a aplicação da tecnologia da computação para a realização de investigação científica moderna, incluindo a preparação, experimentação, coleta de dados disseminação de resultados, armazenamento em longo prazo e acessibilidade de todos os materiais gerados através do processo científico, tais como modelagem e análise de dados, notebooks de laboratório, conjuntos de dados brutos e ajustados, produção de manuscritos e versões preliminares, pré-impressões, e publicações impressas e ou eletrônicas” (BOHLE, 2013, p. 1 – tradução nossa).

Outros conceitos têm sido empregados para designar a nova área científica: “eScience”, “ciência orientada para dados” (data-driven Science), “ciberinfraestrutura”, “quarto paradigma”, “dos dados ao conhecimento”. Em todos eles está presente a ideia de que cientistas da computação se unem a pesquisadores de diferentes domínios para o desenvolvimento de novos conceitos e teorias a partir de grandes massas de dados, em um trabalho que envolve hardware, software e processos que possibilitem a captura, armazenamento, análise e visualização dos dados.

A noção de um “quarto paradigma” da ciência ganha impulso a partir da publicação da obra homônima, em 2009, e da contribuição de diversos autores motivados pelo trabalho do cientista norte-americano James Nicholas “Jim” Gray. Gray foi um renomado cientista da computação que atuou em laboratórios científicos da IBM e da Microsoft. Entusiasta da noção de e-Science, Gray desapareceu no mar em 28 de janeiro de 2007 – junto com seu barco – poucos dias após proferir a palestra que, transcrita, abre o livro em questão<sup>7</sup>. Nele Gray apresenta sua visão sobre a evolução dos paradigmas científicos. Ele concebia a “ciência intensiva em dados” ou “e-Science” como um “quarto paradigma” da ciência (empírica, teórica, computacional e agora baseada em dados). A ciência empírica ou experimental caracteriza o primeiro paradigma, que cedeu lugar às generalizações teórico-conceituais resultantes do uso de modelos abstratos. O terceiro paradigma é baseado nas simulações assistidas por computadores ou outros equipamentos tecnológicos, enquanto o quarto paradigma é resultante da exploração de dados capturados ou gerados pela simulação.

<sup>7</sup> Gray, um experiente marinheiro, possuía um iate de quarenta pés. Em 28 de janeiro de 2007, ele não retornou de uma curta viagem solo às Ilhas Farallon, a 43 km de San Francisco, para espalhar as cinzas de sua mãe. O tempo estava claro e nenhuma chamada de socorro foi recebida e nem sinais automáticos de emergência detectados. Uma busca de quatro dias pela Guarda Costeira usando aviões, helicópteros e barcos não encontrou nada. Em 16 de fevereiro, esta busca foi suspensa, e uma busca subaquática usando equipamentos sofisticados terminou em 31 de maio.

Em sua análise, Gray (2009) aponta alguns caminhos que caracterizam ou deveriam caracterizar o quarto paradigma:

- A comunicação científica pode deixar de apresentar apenas resultados finais na forma de publicações, mas passa a ter a possibilidade de também partilhar os próprios dados brutos da pesquisa, ampliando a possibilidade de falseamento e novas pesquisas utilizando os mesmos dados;
- As bibliotecas digitais podem experimentar novas formas de distribuição das publicações em diferentes países, idiomas e mais baixos custos em relação à publicação em revistas tradicionais;
- A revisão por pares pode experimentar novos processos ou estruturas digitais e serviços como wikis podem ser criados online para ampliar a discussão sobre os trabalhos após a publicação;
- A internet está se tornando um sistema orientado a objeto, cada vez mais dependente de ontologias, vocabulários controlados e esquemas conceituais que possibilitem dar mais confiabilidade ao significado das coisas e à interação entre diferentes sistemas;
- Bibliotecas de dados digitais podem ser fomentadas, para além de bibliotecas de publicações acabadas.

O paradigma emergente envolveria colaboração de pesquisadores em computação com pesquisadores de outras áreas do conhecimento, além de métodos computacionais sofisticados para lidar com grandes volumes de dados (Big Data), alçados agora à condição de objetos intelectuais de primeira grandeza. Computadores, bancos de dados e redes não são vistos e utilizados apenas como ferramentas, mas se tornam uma parte fundamental do processo de descoberta de conhecimento e mesmo de interpretação dos dados. Assim, o novo paradigma agrupa, a seu instrumental, métodos e algoritmos para analisar os dados armazenados nessas bases de grande escala, além de protocolos de comunicação padronizados que possibilitem a interação entre todas essas fontes de dados.

No Brasil, Sayão e Sales (2015) são alguns dos mais entusiasmados pesquisadores engajados na reflexão e prospecção de um modelo e uma infraestrutura para e-Science, especialmente no que diz respeito a novas fontes de pesquisa, arranjos para novos serviços em diversos estágios do processamento dos dados de pesquisa e uma oportunidade para a reinvenção das bibliotecas de pesquisa. Para eles, “o digital não é o antagônico do impresso, como o rolo de papiro não é o antagônico do livro e a tecnologia não é algo estranho à biblioteca” (SALES e SAYÃO, 2015, p. 33). Embora neguem um eventual antagonismo, os autores se mostram entusiastas de uma ciência orientada para dados. Para eles, “a convergência de tecnologias de computação, armazenamento on-line, redes de alto desempenho, somados ao desenvolvimento de instrumentos científicos, escalas, dispositivos experimentais e sensores cada vez mais sofisticados e ao uso intensivo de simulações, desloca a pesquisa científica contemporânea na direção de uma ciência orientada por dados, onde o maior problema não é a escassez, mas sim o excesso de dados e a capacidade de interpretar seus padrões ocultos na forma” (Idem, p. 34).

#### **4 O surgimento de novos paradigmas científicos**

Nesta seção, recorreu-se aos estudos epistemológicos pós/popperianos de Thomas Kuhn (1994) e Inre Lakatos (1979), caracterizados por uma estreita conexão com a história da ciência na tentativa de compreensão dos programas de pesquisa científica<sup>8</sup>.

Kuhn se debruçou sobre o que ele chamou de paradigma científico, termo com o qual buscava designar um conjunto de “conquistas científicas universalmente reconhecidas, que por certo período fornecem um modelo de problemas e soluções aceitáveis aos que praticam em certo campo de pesquisas” (KUHN, 1994, p. 23).

Esse conceito culmina na noção de programa de pesquisa. Com efeito, programa de pesquisa é como Kuhn (1994) chama o estudo acadêmico de determinado tema no interior de uma ciência ou conjunto de ciências – no caso de objetos caracterizados pela interdisciplinaridade. Ao redor daquele programa de pesquisa se reúne um grupo de pesquisadores denominado comunidade científica (KUHN, 1994).

Dentre os membros da comunidade científica, Kuhn (1994) afirma que alguns ganham destaque sobre os demais, devido, basicamente, a dois fatores: a ordem cronológica de publicação dos trabalhos – estudos precursores que enfatizam um aspecto particular do programa de pesquisa e que ganham adeptos podem fazer originar uma nova comunidade científica; outro fator é a originalidade, associada ao grau de aceitação dos trabalhos por parte dos demais membros da comunidade. Na teoria kuhniana a comunidade científica está caracterizada pela prática do que ele chama de ciência normal, que é a “pesquisa estavelmente baseada em um ou mais resultados alcançados pela ciência do passado, aos quais uma comunidade científica particular, por certo período de tempo, reconhece a capacidade de constituir o fundamento de sua práxis ulterior” (KUHN, 1994, p. 68).

A ciência normal consiste, então, na tentativa de realização das promessas do paradigma de pesquisa, aprofundando as pesquisas originárias e confrontando as formulações com a realidade, articulando novos conceitos e sugerindo novas ferramentas com as quais provê o paradigma. A teoria de Kuhn sugere que o trabalho do cientista normal órbita ao redor das noções basilares do paradigma e os eventuais problemas teóricos e práticos que encontrar tendem a significar antes uma deficiência do pesquisador do que das construções originárias do paradigma. Isso se deve à noção de ciência normal, que consiste no estado de uma ciência na qual, suas pesquisas e seus resultados são previsíveis e os métodos e resultados das pesquisas corroboram os resultados esperados da pesquisa. Entretanto, podem ocorrer crises, por meio de anomalias sucessivas que não se conformam com os resultados esperados. Se a quantidade desses “problemas” cresce demasiadamente, inicia-se um período de crise do paradigma em que se começa a questionar a validade de suas concepções iniciais: é a fase da ciência extraordinária, em que se colocam os dogmas em questão e cresce o movimento dos pesquisadores ‘reformistas’. Esse movimento pode desencadear uma revolução científica, momento em que um novo paradigma sobressai em relação ao anterior por meio da “conversão” dos cientistas às promessas do novo paradigma. Kuhn (1994) afirma que essa conversão pode se dar por diversas razões: por convencimento, pela

---

<sup>8</sup> Esta seção condensa parte da análise epistemológica desenvolvida por ocasião da dissertação de mestrado de um dos autores deste artigo (SILVA, 2008).

sedução que as novas ideias proporcionam (caráter emocional) por familiaridade, nacionalidade e até mesmo por razões estéticas.

A explicação oferecida por Kuhn ao progresso da ciência evidencia um elemento ‘problemático’ nesta. A ciência de que o autor fala já não é aquela catedrática da ‘objetividade científica’, da imparcialidade metodológica e das certezas conceituais: é uma ciência que pressupõe escolhas às vezes arbitrárias, cujos pesquisadores não podem ignorar assaltos da dimensão emocional e precisam endossar os dogmas que a sustentam, sob pena de ‘perder seu chão’, as bases teóricas de sua produção.

A mesma postura crítica quanto aos pressupostos metodológicos frequentemente empregados na atividade científica é encontrada na obra de Imre Lakatos (LAKATOS, 1979), que também estuda a temática dos programas de pesquisa e apresenta pontos de convergência e de divergência com o trabalho de Kuhn. Sobre as teorias científicas, Lakatos (1979) argumenta que é componente essencial ao progresso da ciência a concorrência entre programas de pesquisa antagônicos. A partir do estudo da crítica à teoria de Kuhn, Lakatos deu um importante passo na filosofia da ciência ao formular a tese segundo a qual os sistemas teóricos são compostos por um núcleo forte portador de características que o definem como programa, circundado por um cinturão de suporte, constituído de hipóteses e suposições auxiliares que poderiam ser refutadas sem prejuízo para o programa como um todo.

Quando submetido a testes, o núcleo forte da teoria que possui o papel de “pressuposto básico” não é afetado, de modo que é o cinturão, e não a teoria central em si, que passa pelo processo de averiguação. Desta forma, uma anomalia em qualquer teste afeta unicamente o cinturão de proteção e jamais a teoria central (SILVA, 2008). Esse núcleo – por decisão metodológica dos pesquisadores – é considerado não falseável, isto é, geralmente não se contestam no interior do programa as ideias contidas no seu núcleo fundamental porque sua destruição implicaria problemas para o programa como um todo. A teoria de suporte que acaba sendo refutada pode ser substituída por outra, mas para Lakatos o caso mais comum é o de reforçar a cinta de proteção do núcleo com teorias ad hoc, ou seja, teorias auxiliares não possuem propósito algum a não ser o de sustentar o núcleo forte.

Embora sedutora, a teoria de Lakatos não escapou de importantes críticas. Ao extremo, pode-se cogitar que a priori nenhum programa de pesquisa poderia ser refutado, uma vez que seu núcleo não-falseável poderia ser mínimo a ponto de conter somente princípios amplamente aceitos, e suas teorias de suporte – que efetivamente desenvolvem as ideias do programa – podem ser pragmaticamente descartadas conforme tenham dificuldade de resistir a novas críticas. Lakatos admitia a existência dessa possibilidade mas asseverava que, para um programa científico, o acúmulo de teorias ad hoc é desconfortável. Com efeito, este acúmulo leva a comunidade científica a aderir a determinada teoria em detrimento de outra. A nova teoria nem sequer precisa responder a todas as questões que a anterior respondia, bastando que ela responda um número satisfatório de problemas (dentre esses, alguns que a anterior não conseguia responder) e não ter um número grande de fundamentações ad hoc (SILVA, 2008).

## 5 Considerações finais

As breves considerações sobre como Khum e Lakatos entendem a formação de um paradigma científico possibilitam lançar luzes sobre o atual momento histórico. Primeiramente, independentemente da designação que demos é inegável o fenômeno do trabalho com grandes dados na produção científica e seus resultados, cujos exemplos mais notórios são o projeto do genoma humano e os aceleradores de partículas que produzem uma infinidade de dados impossíveis de se processar e analisar sem um suporte computacional. O surgimento de uma comunidade de pesquisa trabalhando com grandes dados é, portanto, um fenômeno do mundo da práxis que convive com uma ainda incipiente produção teórico-conceitual que a fundamenta. Uma rápida pesquisa em portais de periódicos demonstra que há muito mais pesquisa sobre casos de implementação de e-Science que estudos predominantemente acerca de sua epistemologia, seus fundamentos conceituais ou os aspectos éticos e políticos sem os quais o pensamento científico não se completa – permanece no campo da técnica.

Por sua vez, tentativas de refutação da e-Science não são raras, sendo Martin Frické um de seus maiores críticos. Frické (2015) é céptico em relação à possibilidade de o big data produzir uma lógica de descobertas científicas universais ou ferramentas de inferência e refutação que torne a e-Science paradigmática. Para ele, a lógica indutiva por trás do trabalho com dados é falível e contaminada pela teoria: “Os instrumentos são construídos ou adotados à luz do que sabemos ou das teorias que temos”, uma vez que “dados não falam (...); é necessário uma enorme quantidade de conhecimento básico ou suposições ou pesquisa prévia” (FRICKÉ, 2015, pp. 5-6).

Tal como vem se configurando, o fenômeno da e-Science, a despeito da resistência sofrida, vem diluindo ainda mais algumas tênues fronteiras entre várias ciências – queremos destacar as ciências da informação e da computação – sem chegar a constituir apenas uma disciplina autônoma (e-Science), mas uma concepção de ciência e de fazer científico distinta do que se via até então. Talvez não valha a pena entrar em uma discussão se é adequado ou não chamar esse movimento de um novo paradigma científico. O mundo e a natureza fornecem ‘resistência’ a nossas conceituações arbitrárias sob a forma de anomalias, situações em que fica claro que algo está errado com as estruturas atribuídas ao mundo por nossas construções conceituais aleatórias. Nesse sentido, o tempo e a produção teórico-conceitual da e-Science em diálogo com a tradição científica poderão atestar seu alcance e falsear seus pressupostos, embora seja forçoso reconhecer a urgência de mais fundamentos (núcleo fundamental e teorias *ad hoc*) para que se possa mais confortavelmente chamá-la de um novo paradigma científico. A esse respeito, embora sem mencionar especificamente a e-Science, refletia Boaventura Santos há 30 anos: “Depois da euforia cientista do século XIX e da consequente aversão à reflexão filosófica, bem simbolizada pelo positivismo, chegamos a finais do século XX possuídos pelo desejo quase desesperado de complementarmos o conhecimento das coisas com o conhecimento do conhecimento das coisas” (SANTOS, 1988, p. 57). Precisamos conhecer mais como a e-Science quer conhecer o mundo.

## Referências

- BOHLE, Shannon. (2013, 12 June). What is E-science and How Should it be Managed? **Scientific and Medical Libraries. Scilogs. Nature and Spektrum der Wissenschaft.**

Disponível em: [http://www.scilogs.com/scientific\\_and\\_medical\\_libraries/what-is-e-science-and-how-should-it-be-managed/](http://www.scilogs.com/scientific_and_medical_libraries/what-is-e-science-and-how-should-it-be-managed/). Acesso em setembro de 2018.

FRICKÉ, Martin. Big data and its epistemology. **Journal of the Association for Information Science and Technology**. 66, 4, p. 651-661 11 p., 2015.

GRAY, Jim. **Jim Gray on eScience**: A Transformed Scientific Method. Edited by Tony Hey, Stewart Tansley, and Kristin Tolle. In: Hey, T., Tansley, S. and Tolle, K. (Ed.). **The Fourth Paradigm: dataintensive scientific discovery**. Redmond, WA, USA: Microsoft Research, 2009.

KUHN, Thomas. **A estrutura das revoluções científicas**. São Paulo: Perspectiva, 1994.

LAKATOS, Imre; MUSGRAVE, Alan (Org.). **A crítica e o desenvolvimento do conhecimento**: Quarto volume das atas do colóquio internacional sobre Filosofia da Ciência, realizado em Londres em 1965. Trad.: Octavio Mendes Cajado. São Paulo: Cultrix e EDUSP, 1979.

SALES, Luana Farias; SAYÃO, Luís Fernando. Há Futuro para as Bibliotecas de Pesquisa no Ambiente de eScience? **Informação & Tecnologia**: Marília/João Pessoa, 2(1): 30-52, jan./jul., 2015.

SANTOS, Boaventura de Sousa. Um discurso sobre as ciências na transição para uma ciência pós-moderna. **Estudos avançados**. Vol.2 no.2 São Paulo May/Aug. 1988.

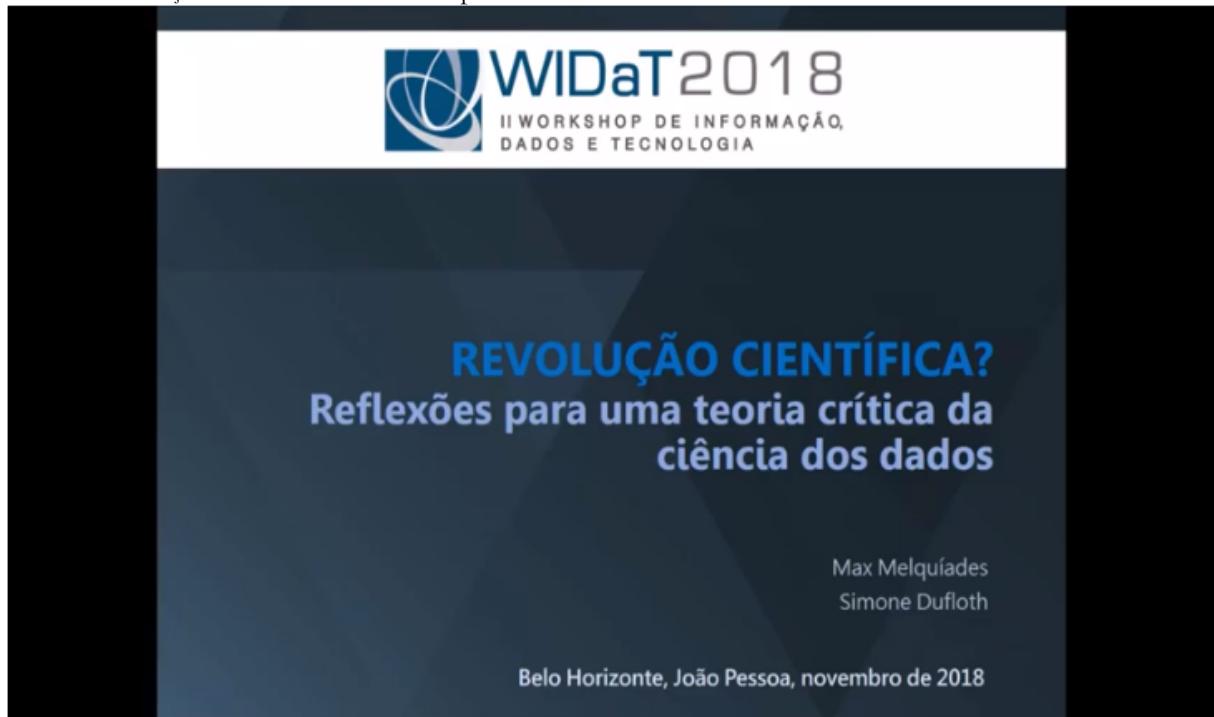
<http://dx.doi.org/10.1590/S0103-40141988000200007>. Acesso em setembro de 2018.

SHAPIN, Steven. **The Scientific Revolution**. Chicago: The University of Chicago Press, 1996.

SILVA, Max Melquíades da. A gestão do conhecimento entre as propostas teórico-conceituais e a efetividade de sua implementação na administração pública: problemas e possibilidades. 2008. 175 f. **Dissertação** (Mestrado em Administração Pública) - Fundação João Pinheiro, Belo Horizonte, 2008.

## Vídeo da apresentação

Título: Revolução Científica? Reflexões para uma teoria crítica da Ciência dos Dados.



Disponível em: [http://dadosabertos.info/enhanced\\_publications/idt/video.php?id=2](http://dadosabertos.info/enhanced_publications/idt/video.php?id=2)

## **Transcrição da apresentação**

Olá participantes do WIDAT 2018, meu nome é Max e eu quero apresentar esse trabalho que a gente submeteu para o evento: Revolução científica? Reflexões para uma teoria crítica da ciência dos dados. Um trabalho meu e da professora Simone Duflot que deixo aqui as apresentações, professora Simone é Doutora em Ciência da Informação pela UFMG e eu Max sou doutorando em gestão e organização do conhecimento também pela UFMG e ficam aqui os nossos contatos.

Bem o trabalho vai seguir essa organização e o trabalho escrito segue essa organização em que a gente também a utilizou aqui para apresentação online. Vamos lá então, a gente sempre gosta de começar a falar desse tema falando sobre essas recentes inovações tecnológicas que tem permitido uma mudança significativa na forma de trabalho da ciência como escala de dados nunca antes vista.

Alguns teóricos têm feito paralelo com a lei de Moore, que trazia essa premissa de que a cada ano e meio aproximadamente, a velocidade de processamento das informações dos dados duplicaria com uma razoável redução também dos custos, nenhuma redução proporcional dos custos de produção de processadores e dos transistores que se utilizavam na época e agora a gente vive uma época também com um aumento significativo no volume de dados produzidos com redução dos custos de acesso a esses dados, de tal forma que o desafio nesse momento se desloca no tratamento desses dados em detrimento da sua obtenção que hoje se torna algo mais simples de se fazer em comparação com períodos passados, de tal forma que alguns teóricos têm falado da emergência de um quarto paradigma da ciência, que é um paradigma focado em uma lógica computacional de trabalho com grandes volumes de dados.

Então a partir disso a gente propôs essas questões de pesquisa como evoluir um paradigma científico? As transformações em curso no sentido de uma crescente ressignificação do dado como uma unidade de análise chega a constituir um novo paradigma? A produção teórico-conceitual no campo da ciência da informação tem acompanhado esse progressivo trabalho com Big Data?

A partir dessas questões a gente se propôs a este objetivo de analisar à luz da noção de paradigma científico na filosofia da ciência, as características desse paradigma da e-Science e fizemos isso em um trabalho eminentemente de revisão bibliográfica a gente se fundamenta muito na obra desses dois teóricos pós/popperianos, o Thomas Kuhn e Imre Lakatos, dois teóricos que abordam questão da evolução o progresso da ciência.

Então primeiramente a gente fala dessa noção do quarto paradigma, eles surgem na primeira década do ano de 2000 do século XXI e um dos trabalhos mais significativos nesse sentido e o trabalho de Gray, que consegue uma ciência intensiva de dados efetivamente como um novo paradigma, sendo primeiro um paradigma da ciência empírica depois uma ciência teórica, e ciência computacional e agora já seria uma ciência baseada em dados, resultante dessa exploração de dados que são capturados da natureza ou gerados por simulações por experimentos sempre com suporte computacional.

Esse paradigma na perspectiva de Grey seria caracterizado entre outros por esses elementos uma comunicação científica que já não se restringe a comunicar resultados de publicações

finalizadas mas também dados de pesquisa ainda em construção, em dados primários e tal forma que outros pesquisadores possam também a partir desses dados fazer suas simulações, testar hipóteses, tentar refutar ou não, os resultados podem fazer um trabalho na tentativa de falseamento desses dados. Bibliotecas digitais que podem experimentar novas formas de publicações inclusive dados brutos. A revisão por pares passaria a ter essa possibilidade não apenas de descer uma palavra final para dizer que um elemento é tem um artigo, uma publicação pode efetivamente ganhar a comunidade científica, pode ser digamos é referendado, oficializada, publicada, mas você teria possibilidade de construções coletivas, você tem a possibilidade de utilizar pelas digitais ter uma construção permanente de um texto, ou seja, não é porque já recebi uma aprovação dos pares que esse texto está pronto, tá bom, e outros leitores, outros pares também podem continuar colaborando inclusive através de wikis para aperfeiçoar esse trabalho. A internet passa estar cada vez mais controlada por ontologias, vocabulários controlados e outras formas de trabalho propostas ou providas digamos assim pela ciência da informação e a colaboração entre pesquisadores também passa a ter muito mais possibilidades.

Nessa mesma linha no Brasil, Luís Fernando Sayão e Sales são dois dos teóricos mais conceituados nesse sentido de uma reflexão a cerca da e-Science, eles estão sobretudo nessa perspectiva de uma pensada, uma infraestrutura para e-Science e novas possibilidades para bibliotecas de pesquisa.

A partir dessa caracterização muito sucinta, muito breve, a gente traz mais elementos no trabalho escrito, a gente quis pensar o que os filósofos da ciência pensam de um paradigma científico? Até para a gente analisar que ponto que e-Science pode ser caracterizada efetivamente como o quarto paradigma.

Então a gente trouxe que as noções do Tomas Kuhn, com ele traz vários conceitos importantes para a gente compreender a evolução dos paradigmas científicos e a primeira não são Justamente a noção de paradigma que é um mapa, um roteiro através do qual os cientistas dentro de um determinado campo atuam na elaboração de propostas para solução de problemas daquele campo de pesquisa. A noção de enigma, problemas que o paradigma permite resolver e fazer avançar aquele campo do conhecimento, anormalidades problemas persistentes, meio que se recusam a ser resolvidos pelo conteúdo já provido pelo paradigma nesses elementos que desafiam as regras existentes naquele momento.

Isso introduz também de crise que é justamente essa desconfiança no paradigma, parte justamente de um excesso de anormalidades de elementos que aquele paradigma não dá conta de resolver, a noção de revolução científica que reside nessa ideia de uma ruptura e nisso toma com marca uma ruptura também com pensamento anterior a ele sobre todo o pensamento de Karl Popper, que enxergava possibilidade de uma evolução gradual na ciência.

O Thomas Kuhn não fala de rupturas abruptas, efetivamente de revoluções, ele entende que na ciência o progresso se dá dessa forma, através de revoluções realmente as marcantes e alguns exemplos clássicos aí para balizar essa perspectiva nessa disputa entre geocentrismo x heliocentrismo, então passa-se de uma visão em que a Terra era o centro do mundo, do universo, para uma visão em que a gente tem a verdade o sol como centro de um sistema solar de uma galáxia e vários outros cosmos, várias outras estrelas e planetas orbitando aquele corpo que seria

o sol. Também nessa perspectiva do criacionismo e do evolucionismo, nessa passagem de uma perspectiva de que um ser criou tudo para uma perspectiva já darwinista de que na verdade houve um processo de evolução que durou milhões de anos para você chegar, para as coisas chegarem até o grau status que estão hoje, a gente tem uma ruptura brusca na forma de concepção do mundo e da forma como as coisas evoluíram conforme cada perspectiva.

O Inre Lakatos ele traz de novidade em relação de partida de vários elementos convergentes em relação a teoria do Kuhn, mas a gente quis destacar programa de pesquisa, então ele percebe a ciência como organizar em torno de vários programas de pesquisa que oferecem regras metodológicas e o corpus teórico também que banaliza o trabalho do cientista e ele entende que o crescimento da ciência acontece por meio da competição entre programas de pesquisa rivais e essa noção de que todo o programa de pesquisa possui núcleo rígido, núcleo duro, tem várias traduções diferentes para esse conceito, mas sempre um núcleo duro que consiste nas ideias fundamentais daquele programa de pesquisas, essa é uma ação conjunto de construções que não pode ser refutado por que se não implicaria numa crise para todo esse programa de pesquisa e entorno deste núcleo rígido, dessas ideias fundamentais que constituem um programa de pesquisa, estaremos no cinturão protetor composto por teorias, hipóteses auxiliares que podem ser refutadas que podem ser falseadas sem maiores problemas para o programa como um todo, inclusive ideias ou teorias que vão surgindo ao longo do tempo até para dar conta de novos problemas que vão sendo identificados por aquele programa de pesquisa.

Parte dessa breve revisão de literatura a gente traz para discussão, algumas ideias e pensamentos no sentido de enxergar o mundo da prática, então a gente percebe que esse fenômeno do trabalho com grandes dados ou e-Science como um fenômeno do mundo da práxis ele é inegável e é de fato tem conseguido aglutinar uma grande comunidade de pesquisa em torno desse conceito, desse paradigma, tanto inclusive que esse evento que a gente tá participando é uma comunidade de pesquisa que tem já realizado ao longo dos últimos anos não só pesquisas de cunho teórico quanto prático, o que evidencia que esse fenômeno precisa ser discutido.

O que acontece que a gente identifica muito mais estudo de caso, muito mais estudos práticos do que essas formulações teóricas conceituais sobre as bases epistemológicas da e-Science enquanto paradigma científico.

Qual que é o núcleo duro da e-Science? Quais são as suas teorias de suporte? A gente pode caracterizar aquelas ideias propostas pelo Gray como sendo o núcleo duro, elas constituem um fundamento sólido e se algumas dessas ideias começam a ser refutadas no futuro próximo que a gente tem uma crise para esse programa de pesquisa ou para esse paradigma ou na verdade parte daquelas características seriam proposições auxiliares de suporte que podem ser refutadas sem maiores problemas, falta um pouco dessas discussões.

Tentativas de refutação ou críticas só isso já existem, Martin Frické um dos expoentes críticos é muito céptico com relação a ideia, a possibilidade de um big data produzir uma lógica de descobertas científicas que possa ser amplamente aplicada, diversas ciências que eles, três paradigmas mencionados anteriormente, no paradigma teórico-empírico, teórico e o computacional, a gente tem uma a utilização ampla nas mais diversas ciências. A e-Science também teria essa possibilidade ou essa característica de universalidade? Algo a se pensar.

Martin Frické também é bastante céítico com relação à possibilidade dessa forma de e-Science produzir conhecimento através da análise de muitos dados que consiste numa forma sobretudo de métodos indutivo de produção de conhecimento. Frické é muito céítico com relação às possibilidades disso produzir conhecimento novo porque não entende que os nossos pré-conceitos, pré-juízos na hora de analisar dados empíricos ele já implica em uma visão de mundo que pode distorcer a nossa análise. Então a gente entende que embora seja forçoso reconhecer a urgência desses fundamentos teórico-conceituais para e-Science para que ela possa ser mais confortavelmente chamado de um paradigma científico no campo da prática ou práxis, ela efetivamente já é uma realidade.

Então a gente queria finalizar com essa frase de Boaventura Sousa Santos que dizia assim lá em 1988, final do século 20, ele dizia depois da euforia cientista do século XIX e da consequência aversão a reflexão filosófica bem simbolizada pelo positivismo, chegamos ao final do século 20, possuídos pelo desejo quase desesperado de complementarmos o conhecimento das coisas com conhecimento do conhecimento das coisas.

Aqui ele está falando de epistemologia, tá falando de não apenas fazer mas pensar o que fazemos como fazemos, qual que é a lógica subjacente ao trabalho dos cientistas no caso do cientista de dados, então a gente entende que é preciso conhecer mais sobre como a e-Science quer conhecer o mundo, pensar mais nessas bases epistemológicas.

Queremos agradecer e deixar os nossos dados para contato e convidá-los a fazer a leitura do texto completo e a continuar essa discussão aí nos diversos fóruns possíveis.

Um abraço.

## **Slides da Apresentação**

Título: Revolução Científica? Reflexões para uma teoria crítica da Ciência dos Dados.



Disponível em: [http://dadosabertos.info/enhanced\\_publications/idt/presentation.php?id=2](http://dadosabertos.info/enhanced_publications/idt/presentation.php?id=2)