

AS CIÊNCIAS SOCIAIS NA ERA DO ZETTABYTE

Ronaldo Baltar¹
Cláudia Siqueira Baltar²

O volume de informações disponíveis para a pesquisa tem crescido rapidamente nas últimas décadas. De acordo com o relatório intitulado “*The Zettabyte Era—Trends and Analysis*”, publicado pela empresa de telecomunicações norte-americana Cisco Systems, em 2013³, estamos começando, nesta segunda década do século XXI, a era do “zettabyte”. Um byte é uma unidade de informação digital, que representa, por exemplo, uma letra, um sinal de pontuação ou um número em um texto processado em um computador. A capacidade de armazenamento de informações em computadores é medida em múltiplos de mil bytes. Um disquete de 3½ polegadas, muito comum nos anos 80 e 90 do século XX, tinha capacidade para armazenar 1,44 Megabytes ou 1.448.448 bytes, o que seria suficiente para arquivar uma tese com aproximadamente 500 páginas (sem imagens).

No final da primeira década do século XXI, computadores de uso pessoal eram comumente vendidos com um disco rígido de 500 Gigabytes. Um único computador com esta configuração tinha a capacidade para armazenar aproximadamente 500 mil disquetes de 3½ polegadas. Esse volume de informações disponíveis para armazenamento em dispositivos comuns de informática foi rapidamente superado pelos dispositivos com um Terabyte (1.000 Gigabytes ou 1.000⁴ bytes).

Com o crescente incremento na capacidade de armazenamento de informação, aliada ao aumento na velocidade de processamento, contando ainda com a ampliação

¹ Docente do Departamento e do Programa de Pós Graduação em Ciências Sociais da Universidade Estadual de Londrina, Brasil. baltar@uel.br

² Docente do Departamento e do Programa de Pós Graduação em Ciências Sociais da Universidade Estadual de Londrina, Brasil. cbaltar@uel.br

³ Disponível em: <http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/VNI_Hyperconnectivity_WP.html>

nos canais de comunicação entre os dispositivos digitais, cada vez mais diferentes atividades humanas passaram a ter um registro digital em tempo real. De acordo com o relatório da Cisco Systems, estima-se que em 2015 a troca anual de informações que transitarão pela Internet em todo o globo irá superar um Zettabyte (1.000⁷ bytes ou 1 bilhão de Terabytes). Espera-se que entre 2012 e 2017, os próximos cinco anos, haverá um aumento a cada ano de 23% nas informações trafegadas pela rede. Aproximadamente metade deste volume de informações será originada de dispositivos móveis e não de microcomputadores. Calcula-se que a capacidade física de armazenamento de dados somada de todos os computadores e dispositivos digitais em 2013 já tenha alcançado 2 Zettabytes.

Como as Ciências Sociais estão se preparando para os desafios conceituais e metodológicos da era Zettabyte? Essa é uma questão de fundo proposta para reflexão neste Dossiê de Mediações - Revista de Ciências Sociais.

O volume de informações armazenado nos dispositivos digitais e que trafegam pela Internet pode ser visto como um composto de registros de atividades sociais, econômicas, políticas e culturais. Além de dados quantitativos, também circulam em grande parte vídeos, falas, documentos, imagens e todo um conjunto que representa a variedade de expressões cotidianas da atividade humana. O relatório da Cisco Systems estima que em 2017 transitará pela Internet 13,8 Pentabytes de informação a cada cinco minutos. Ou seja, o equivalente a todos os filmes produzidos até hoje circulará globalmente a cada cinco minutos, correspondendo a 73% de todo o tráfego da Internet daqui a cinco anos. Todo esse volume de dados contém informações relevantes para a pesquisa em Ciências Sociais, disponíveis em uma escala inimaginável no século XIX e início do século XX.

Além do conhecimento sobre os processos sociais em tempo real, os dados também têm sido utilizados cada vez mais para a composição, monitoramento de políticas públicas, principalmente por meio de indicadores diversos e cada vez mais especializados em determinados aspectos da vida social. Informações que antes estavam acessíveis a poucos departamentos, institutos e empresas de pesquisa, estão agora mais acessíveis e ao alcance de movimentos sociais, organizações não governamentais, sindicatos e cidadãos.

O acesso e o tratamento analítico dessas informações têm impulsionado uma renovação e ampliação dos métodos nas Ciências Sociais em vários países da União Europeia, Estados Unidos, Japão e outros. Essas mudanças têm se processado por duas vias: uma retomada da formação básica em procedimentos científicos próprios, desenvolvidos pelas Ciências Sociais e Humanas (sociologia, antropologia, ciência política, economia, demografia); e também pelo diálogo aberto com outras áreas do conhecimento, como a computação, física, estatística, biologia e psicologia, entre outras.

No Brasil e em grande parte da América Latina ainda nos encontramos com alguma defasagem em relação a estes debates interdisciplinares que têm movimentado as Ciências Sociais em outros países. Também temos estado aquém da capacitação necessária para o uso de ferramentas de informática e manipulação de bases de dados em pesquisa social, sejam em termos de pesquisa quantitativa, qualitativa, documental ou qualquer outro tipo.

A formação do cientista social no Brasil ainda trata com certo desdém a pesquisa, sobretudo quando se trata de análise quantitativa. Apesar de vários centros estarem estimulando uma formação científica mais aprimorada em termos de análise de dados, tanto em curso de graduação, como na pós-graduação, ainda é comum se observar uma referência pejorativa dissimulada a trabalhos de Ciências Sociais que estejam fundamentados em análise de dados. Não é raro que se associem estudos quantitativos com positivismo, conservadorismo e falta de reflexão crítica. É frequente entre estudantes de pós-graduação de Ciências Sociais, que se arriscam à análise de dados (quase sempre contra a vontade ou sob a desconfiança dos orientadores), usar mais linhas para justificar que não se trata de um estudo "meramente quantitativo", do que analisar de fato os dados disponíveis no trabalho. Mesmo assim, ainda se vêem orientadores, pareceristas e membros de banca tratarem tabelas ou gráficos no texto como se fossem elementos estranhos à atividade intelectual, algo que mais atrapalha a leitura do que explica os argumentos do autor. Nessa visão, o uso de dados é uma concessão à "empíria"; um desvio perigoso, contaminado de positivismo, que se afasta do processo puro de construção do raciocínio crítico a partir do "raciocínio crítico" de outros. Diante dessa postura, o que prevalece não é o concreto pensado, inclusive entre os que se dizem adeptos da teoria crítica, mas apenas o "pensado" pensado. Curioso também é que não se ouvem, na mesma medida, referências críticas a estudos "meramente qualitativos" ou "meramente documentais", "meramente bibliográficos", "meramente teóricos".

Contudo, este tipo de estranhamento com os dados (de qualquer natureza), que constringe ou desestimula a produção em Ciências Sociais, sobretudo entre os jovens pesquisadores, tem cedido espaço para uma postura muito mais estimulante para a pesquisa científica em Ciências Sociais. O prêmio "O Brasil que sai do Censo", estabelecido em 2012 pela Associação Nacional de Pesquisa e Pós-graduação em Ciências Sociais (ANPOCS), em conjunto com a Fundação Ford, é um sinalizador que aponta para a importância da apropriação pelos cientistas sociais do acervo de informações disponíveis sobre a realidade social. Outro projeto importante com este mesmo sentido é o Programa Observatório da Educação (OBEDUC), criado pelo Decreto Presidencial nº 5.803 de 2006, e instituído em vários Programas de Pós-graduação por meio de parceria entre a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), o Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) e a Secretaria de Educação Continuada, Alfabetização, Diversidade e Inclusão

(SECADI), todos ligados ao Ministério da Educação (MEC). O projeto OBEDUC tem como principal objetivo estimular pesquisas sobre temas relativos à educação, combinando a infraestrutura disponível das Instituições de Educação Superior – IES e as bases de dados do INEP.

Todas as entidades federais, estaduais e municipais, desde 2011, têm a obrigatoriedade de se adequarem a Lei Federal n. 12.527 que estabelece as normas para o acesso à informação no Brasil. A publicação da “Lei do Acesso à Informação” é um elemento importante para fomentar a participação democrática a partir de uma maior simetria na informação entre os diversos atores sociais, sobretudo em relação com o poder público. Além do acesso individual que cidadãos venham a fazer das informações que lhes sejam pertinentes, as Ciências Sociais têm um papel importante para apropriação qualificada e sistematizada dos dados disponibilizados para o entendimento de questões sociais de relevo de forma a tornar estas informações em instrumentos efetivos de controle social.

O conjunto de dados disponíveis por órgãos públicos têm aumentado desde 2011. Uma iniciativa do Governo Federal (www.dados.gov.br), chamada “Dados Abertos”, reúne com facilidade e de maneira indexada informações de diversas pesquisas elaboradas pelo Instituto Brasileiro de Geografia e Estatística (IBGE), Ministério do Trabalho (MTE), Ministério do Planejamento, Ministério da Saúde (Datasus), Ministério da Educação (Inep), Governos Estaduais, Poder Legislativo e Poder Judiciário. Informações que dez anos atrás se encontravam dispersas, não catalogadas e acessíveis a poucos funcionários públicos, agora estão cada vez mais abertas ao público. O próprio IBGE, além de dispor gratuitamente os microdados do Censo 2010, ainda apresenta ao público ferramentas como o Sistema IBGE de Recuperação Automática – SIDRA⁴, entre outros bancos de dados de acesso público. Assim como o Instituto de Pesquisa Econômica Aplicada – Ipea⁵ e vários outros órgãos governamentais. Somam-se a este conjunto, os acervos documentais que estão sendo digitalizados pela Biblioteca Nacional⁶, o acervo audiovisual e documental do Centro de Pesquisa e Documentação Histórica (CPDOC) da Fundação Getúlio Vargas⁷, entre muitos outros. No âmbito internacional, há também um conjunto cada vez maior de fontes de dados sistematizadas e de acesso público, como o UNDATA das Nações Unidas⁸, o World Bank Data⁹ e o PovcalNet, uma ferramenta do Banco Mundial para pesquisa *online* sobre pobreza comparada entre países¹⁰; o grande acervo da Organização Internacional do

⁴ Através do endereço <<http://www.sidra.ibge.gov.br/>>

⁵ <<http://www.ipea.gov.br/portal/>>

⁶ <<http://bndigital.bn.br/>>

⁷ <<http://cpdoc.fgv.br/>>

⁸ <<http://data.un.org/>>

⁹ <<http://data.un.org/>>

¹⁰ <<http://iresearch.worldbank.org/PovcalNet/index.htm>>

Trabalho (OIT), com o NATLEX, LABORSTA, ILOSTAT entre outros¹¹. São inúmeros exemplos de acervos de dados sistematizados, organizados e disponíveis para pesquisa social.

É necessário um esforço coordenado e colaborativo para que se organize e analise esse volume de dados, aos quais se podem acrescentar uma grande quantidade de registros de estudos de casos, observações, *surveys* e coletas de documentos oriundas de pesquisadores individuais ou de grupos de pesquisa temáticos.

A União Europeia montou um projeto que pretende ter a articulação necessária para que se possa dar conta da diversidade e da potencialidade existente para as Ciências Sociais. O projeto *FuturICT Flagship*¹², iniciado em 2010, pretende combinar tecnologia da informação (ICT), Ciências Sociais e estudos sobre complexidade em diversas áreas para buscar elementos que permitam o entendimento e enfrentamento de problemas sociais vividos pelos países participantes.

É um projeto ambicioso, não apenas pelos objetivos a serem alcançados, mas, sobretudo pelo esforço de articulação entre pesquisadores de Universidades, investimento em infraestrutura computacional e predisposição para catalogação, integração e sistematização de dados para estudos em tempo real sobre as mudanças sociais e sua aplicabilidade para a elaboração de políticas públicas.

O projeto originou-se no evento organizado pela OECD, chamado de “*Global Science Forum on Applications of Complexity Science for Public Policy*”, realizado em 2008. Cientistas sociais e pesquisadores de diversas áreas propuseram a criação de um projeto amplo para a União Europeia, similar ao esforço do projeto Apollo dos Estados Unidos nos anos 1960, que levou o homem à lua, ou o laboratório CERN (Organização Europeia para a Pesquisa Nuclear) com sede na Suíça. O projeto FuturICT foi apresentado por um conjunto de pesquisadores, coordenados por Dirk Helbing, professor de sociologia do Instituto Federal de Tecnologia (ETH) de Zurique, e foi aprovado no âmbito do FET *Flagship Initiatives da Community Research and Development Information Service* (CORDIS), Comissão Europeia. O projeto FuturICT está estruturado em quatro grandes plataformas: *Living Earth Simulator Platform*, *Global Participatory Platform*, *Planetary Nervous System Platform* e *Innovation Accelerator Platform*. Os estudos desenvolvidos nessas plataformas cobrem um conjunto de áreas prioritárias, definidas como eixo para os grandes problemas enfrentados pela União Europeia: energia, redes de comunicação, economia, crime e corrupção, migração, saúde, gestão de crises. Esse projeto, que tem duração prevista de dez anos e orçamento de 1 bilhão de Euros, envolve quase 30 países, mais de uma centena de instituições, empresas, universidades e agências governamentais.

¹¹ <<http://www.ilo.org/>>

¹² <<http://www.futurict.eu/>>

A proposta do FuturICT pode ser um argumento sobre o equívoco de iniciativas do governo brasileiro, como o atual programa Ciência Sem Fronteiras, do Conselho Nacional de Pesquisa (CNPq), bem como iniciativas como o Projeto de Núcleos de Excelência - PRONEX, também do CNPq em parceria com fundações de amparo à pesquisa dos estados. O primeiro programa se firma na concepção de que determinadas áreas de conhecimento detêm o privilégio do conhecimento inovador. Fato que a história da ciência no Brasil demonstra não ser verdadeiro. Existe pseudociência tanto nas áreas de exatas, quanto nas Ciências Sociais. Assim como há produção de destaque e qualidade inovadora em cada uma dessas áreas. Não é a área de origem que define a qualidade do conhecimento, nem tampouco o que garante o retorno do investimento em capacitação. O segundo tipo de financiamento, o PRONEX, pressupõe que iniciativas surgem apenas nos locais que já receberam investimentos anteriormente, dado que a condição de elegibilidade para ter acesso aos recursos de “excelência” é ser capaz de provar a “excelência” da instituição. Ambos os programas repetem o erro de não focalizar a inovação das propostas, a qualidade dos projetos e a relevância social dos impactos a serem buscados com os investimentos públicos.

Na história do Brasil, iniciativas ousadas com perfil do atual projeto da União Europeia já ocorreram no passado. Investimentos maciços que articularam e mudaram o perfil da produção científica no país e a qualidade da gestão de políticas públicas, não de uma área específica do conhecimento. A criação do IBGE em 1936 é um exemplo. A formação do Centro Nacional de Processamento de Alto Desempenho (CENAPAD) em 1994, articulados no Sistema Nacional de Processamento de Alto Desempenho (SINAPD) é outro exemplo. O projeto do INEP-MEC, Observatórios da Educação (OBEDUC) em menor proporção também tem o mesmo perfil. Investimentos que aliam capacitação científica, formação e inovação de conhecimento com articulação de uma estrutura de pesquisa permanente. Nessa mesma linha se enquadra a proposta da ANPOCS de criação de um Instituto de Ciências Sociais no Ministério da Ciência e Tecnologia, nos moldes de outros já existentes, como o Instituto Nacional de Pesquisas Espaciais - INPE. Um instituto com este perfil poderá ter o papel de articular a infraestrutura necessária, congregando pesquisadores e instituições para impulsionar a pesquisa social no Brasil em moldes colaborativos similares ao FuturICT europeu. Talvez não com o mesmo propósito, mas com a mesma ambição de padronizar informações, criar meios de reprodutibilidade das pesquisas, integrar grandes volumes de dados para a pesquisa social, incentivar a formação de pesquisadores, coordenar a participação de várias IES em grandes pesquisas temáticas a partir de eixos de relevância para o país.

O artigo traduzido por Mediações, intitulado “Manifesto de Ciência Social Computacional”, de autoria de Rosaria Conte, Nigel Gilbert, Dirk Helbing (coordenador do projeto FuturICT) e outros autores tem o propósito de iniciar o debate sobre os desafios para as Ciências Sociais na era do Zettabyte. Os quatorze autores do artigo, publicado originalmente em inglês na revista *The European Physical Journal Special*

Topics (vol. 214, n. 1, de novembro de 2012), são pesquisadores de Universidades e Centros de Investigação ligados ao projeto.

A leitura do texto é um exercício de reflexão sobre interdisciplinaridade. São autores de várias áreas: sociologia, psicologia, biologia, física, computação, matemática entre outras. O artigo reflete o diálogo entre essas áreas. Três conceitos fazem a costura entre as diferentes áreas do conhecimento envolvidas nesse projeto: complexidade, emergência e evolução.

Uma visão das Ciências Sociais sobre esses três conceitos, em um primeiro momento, pode levantar a suspeição de se estar revivendo debates antigos, que remontam ao início do século XX, sobre o evolucionismo social, o determinismo biológico, o funcionalismo, todos já longamente criticados por vários autores. Mas não se trata disso. Termos que não são mais usados nas Ciências Sociais, como “leis”, “causalidade”, encontram uso em outros campos do conhecimento. Trata-se aqui de entender a linguagem em seu contexto. E o esforço de entendimento deve ser recíproco, tanto para cientistas sociais, quanto para as demais áreas.

O conceito de evolução, talvez seja o que mais crie inquietação em um primeiro momento. A premissa básica do Manifesto é a seguinte: as sociedades, a organização social, a cultura evoluem. Essa ideia não tem filiação ideológica direta a Comte ou Darwin. Não quer dizer que haja um sentido positivo para essa evolução, uma melhoria. Pode-se reproduzir a mesma ideia, mas com termos das Ciências Sociais: as sociedades, a organização social, a cultura se transformam com o tempo. Evoluir nesse contexto significa mudar com o tempo. A variável “tempo” aqui é um elemento chave para se entender a mudança. Em termos de Ciências Sociais essa mesma ideia de mudança no tempo também se constitui em um conceito chave: processo de transformação social. A ideia de processo refere-se à história, ou seja, algo que se sucede no tempo em um contexto de mudança.

A noção de emergência decorre dessa transformação social no tempo. Quer dizer que os fenômenos sociais emergem das relações sociais e não da vontade individual, ainda que os indivíduos sejam agentes racionais dessas mudanças.

Por que os indivíduos que agem racionalmente não são diretamente os promotores da mudança social? Porque os fenômenos sociais (as relações sociais que se sucedem no tempo) são complexos. O conceito de complexidade significa “inter-relacionado” ou “conectado”. Um elemento complexo é algo composto por várias partes conectadas entre si de tal modo que a alteração em uma das partes afeta as outras partes. A forma como uma mudança afetará o todo depende do nível de interconexão entre as partes.

O objetivo da Ciência Social Computacional seria o de aliar o volume de dados (que chegará a quantia de Zettabytes sob a forma de informação quantitativa, qualitativa ou de qualquer natureza) com métodos analíticos (análise quantitativa,

análise qualitativa, modelagem e simulação) para desvendar as interconexões, ou seja, a complexidade das relações sociais. Dessa maneira se poderá compreender com mais clareza a emergência dos fenômenos sociais (migração, desigualdades, etc.) e o efeito sobre as mudanças sociais ao longo do tempo (que em algumas áreas do conhecimento se chama de evolução).

O desafio para as Ciências Sociais está em ter infraestrutura para coletar e armazenar esse volume de dados, ter recursos de tecnologia da informação para processar estes dados, dispor de métodos analíticos que possam oferecer respostas nítidas aos problemas da complexidade social e propor políticas e indicadores de monitoramento para intervir sobre as mudanças sociais. Aqui não cabe o juízo ingênuo de que melhores dados resultarão em melhores políticas públicas, dado que o conflito de interesses está presente na própria complexidade a ser estudada.

Os demais artigos que compõem o Dossiê mostram os avanços no uso de métodos quantitativos em pesquisa social no Brasil e em Portugal. Marília Patta Ramos (UFRGS, Brasil) apresenta uma reflexão sobre a lógica dos métodos quantitativos e sua aplicabilidade à pesquisa em Ciências Sociais. A autora discute o modelo, a importância e a utilidade da pesquisa quantitativa para o entendimento dos fenômenos sociais. O artigo mostra também com exemplos como as Ciências Sociais incorporam, apesar das resistências, os métodos quantitativos no Brasil. Susana da Cruz Martins, Rosário Mauriti e António Firmino da Costa (CIES-IUL, Portugal) discutem as fontes para a pesquisa quantitativa através dos usos de microdados em pesquisa nas Ciências Sociais.

O artigo de José Eustáquio Diniz Alves e Suzana Marta Cavenaghi (ENCE/IBGE, Brasil) aborda a aplicação temática da análise de dados para na construção de indicadores, referência para a formulação e análise de políticas públicas. Os autores discutem a questão da desigualdade de gênero no Brasil. Ana Sílvia Volpi Scott e Dario Scott (Unisinos, Brasil) apresentam uma reflexão sobre fontes de dados que mostra a amplitude e as possibilidades de estudos quantitativos, com a construção de indicadores, a partir de fontes paroquiais. O texto de José Roberto Vicente (IEA, Brasil) faz uma aplicação temática dos estudos quantitativos, discutindo a questão da pobreza rural e os impactos que resultam dos serviços públicos no Brasil. O artigo de Dorian Borges (DPCIS/UERJ, Brasil) mostra como um tema como o sentimento de insegurança e a vitimização pode ser analisado empiricamente e conceitualmente, ao mesmo tempo em que faz uma contextualização dos debates sobre o assunto. Os autores Julimar Santos Pinto e André Junqueira Caetano (PUC-Minas, CEDEPLAR/UFMG, Brasil) demonstram empiricamente como desenvolver um modelo baseado em lógica *fuzzy*, usando o método *grade of membership*, para a categorização de perfis de vulnerabilidade social da juventude. O artigo é um exemplo de aplicação de programação em R, linguagem que tem se tornado recurso padrão da computação científica. João Sebastião, Joana Campos, Sara Merlini e Mafalda Chambino (CIES-IUL, Portugal) apresentam um artigo que discute um tema e um método de análise de grande relevo atualmente para as

Ciências Sociais, os estudos de redes sociais, focalizando nos elementos de violência infanto-juvenil. Maísa Faleiros da Cunha (NEPO-Unicamp, Brasil) também apresenta uma interessante discussão sobre fontes para a pesquisa com ênfase em fontes documentais da cidade Franca, São Paulo, para o estudo de população e de famílias escravas. O artigo de Henrique Frey e Ednelson Mariano Dota (PUC-Campinas, Brasil) faz uma análise exploratória sobre mobilidade espacial da população na região metropolitana de Campinas a partir dos dados do Censo 2010.

O dossiê pretende trazer novas reflexões e mostrar diferentes abordagens quantitativas, análise de dados e modelos de construção de indicadores em Ciências Sociais. O que se espera é que o debate permita levar as Ciências Sociais em língua portuguesa a dar um passo adiante, vencer dicotomias antigas e cumprir os desafios teóricos e metodológicos da era Zettabyte. O tamanho da tarefa exige mais do que o artesanato intelectual, requer infraestrutura, cooperação e investimento em larga escala para as Ciências Sociais. Não se trata de evitar olhar os clássicos ou abandonar tradições intelectuais. Todo conhecimento científico é cumulativo e depende de releituras constantes. Mas, não se pode dar as costas ao presente e se postar e ficar à sombra dos cânones. É necessário se colocar de pé sobre os ombros dos gigantes e ver mais longe, encarar a pesquisa atual com todos os meios disponíveis e dar um passo com as Ciências Sociais para dentro do século XXI.