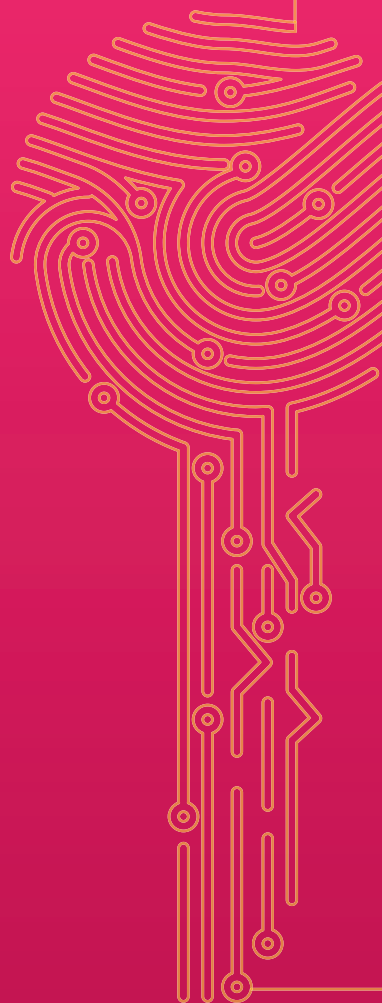


HUMANIDADES DIGITAIS, *BIG DATA* E PESQUISA CIENTÍFICA

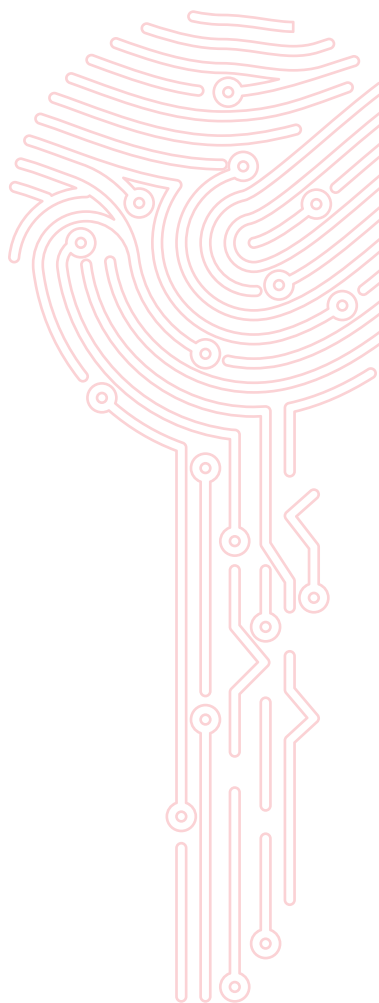
Johanna Wilhelmina Smit
José Augusto Chaves Guimarães
José Eduardo Santarem Segundo
Maria Fernanda Rollo
Ricardo César Gonçalves Sant'Ana
Widad Mustafa El Hadi



F U N D A Ç Ã O

F E R N A N D O
H E N R I Q U E
C A R D O S O

HUMANIDADES DIGITAIS, *BIG DATA* E PESQUISA CIENTÍFICA



F U N D A Ç Ã O

F E R N A N D O
H E N R I Q U E
C A R D O S O

Dezembro de 2021

FICHA TÉCNICA

Copyright © 2021 dos autores

Todos os direitos desta edição reservados à Fundação Fernando Henrique Cardoso – FFHC – Rua Formosa, 367 - 6º andar - Centro Histórico de São Paulo 01049-000 – São Paulo – SP

Site: <https://fundacaofhc.org.br>

Coordenação editorial

Fundação Fernando Henrique Cardoso
Grifo Projetos Históricos e Editoriais Ltda.

Projeto gráfico e diagramação

Lílemes Comunicação

Revisão técnica

Ana Maria Camargo (textos)
Jéssica de Almeida (referências bibliográficas)
Johanna W. Smit (tradução francês-português)

Tradução do texto de Widad Mustafa El Hadi

Brazil Translations

Este livro pode ser reproduzido livremente em parte ou na sua totalidade, sem modificações, para fins não comerciais sob a condição de que sejam devidamente indicados a publicação de origem e seus autores.



SECRETARIA ESPECIAL DA
CULTURA

MINISTÉRIO DO
TURISMO



Dados Internacionais de Catalogação na Publicação (CIP)

Seminário Internacional Humanidades Digitais, *Big Data* e Pesquisa Científica (1.: 2021: São Paulo)

Humanidades digitais, *big data* e pesquisa científica [recurso eletrônico] / Johanna W. Smit... [et al.]. – Dados eletrônicos (1 arquivo : PDF). – São Paulo : Fundação Fernando Henrique Cardoso (FFHC), 2021.

Disponível em: <https://fundacaoofhc.org.br>
ISBN 978-65-87503-11-0

1. Arquivologia. 2. Organização do Conhecimento. 3. Humanidades digitais. I. Smit, Johanna W. II. Fundação Fernando Henrique Cardoso. III. Título.

CDU – 930.25

SUMÁRIO

Apresentação <i>Sergio Fausto</i>	06
A sinergia entre a organização do conhecimento e as humanidades digitais: algumas reflexões <i>Widad Mustafa El Hadi</i>	09
Os dados no contexto da transformação digital: desafios e consequências <i>José Eduardo Santarem Segundo</i>	37
A transdução nos processos de mediação e a informação biodisponível <i>Ricardo César Gonçalves Sant'Ana</i>	58
Memória para todos: os desafios das humanidades digitais <i>Maria Fernanda Rollo</i>	78
A organização do conhecimento em uma perspectiva contextual e seus impactos na pesquisa acadêmica <i>José Augusto Chaves Guimarães</i>	99
A organização do conhecimento e as humanidades digitais: pontos de interseção <i>Johanna Wilhelmina Smit</i>	125
Referência curricular dos autores	151

APRESENTAÇÃO

Sergio Fausto

A Fundação Fernando Henrique Cardoso tem o prazer de disponibilizar o conteúdo do seminário internacional que promoveu, em 21 de setembro de 2021, em torno do tema Humanidades digitais, *big data* e pesquisa científica. O evento reuniu virtualmente seis especialistas e cerca de 400 participantes, de várias regiões do Brasil. Este livro contém a íntegra dos textos então apresentados pelos conferencistas, na expectativa de que sirvam de plataforma para outros debates sobre o assunto.

As questões que se colocam hoje em decorrência do fenômeno dos *big data* e do uso intensivo de ferramentas digitais no campo das ciências humanas têm um impacto significativo nas instituições de custódia de acervos documentais, como é o caso desta Fundação. O tema interessa tanto a curadores de arquivos, bibliotecas, museus e centros de memória quanto a pesquisadores que, com diferentes enfoques disciplinares, enfrentam massas de dados cuja procedência exige monitoramento rigoroso e constante. A pesquisa científica, nesse cenário, tem seus limites e horizontes alargados, na mesma proporção em que requer novos parâmetros para garantir a autenticidade do material empírico que valida suas hipóteses.

Questões teóricas, conceituais e éticas ligadas ao tema, sempre associadas a experiências desenvolvidas nos ambientes acadêmicos dos conferencistas convidados, formam o panorama multifacetado que esta publicação oferece.

As Humanidades Digitais se situam na viragem epistemológica provocada pela presença digital nas ciências humanas e sociais e seus desdo-

bramentos: novas ferramentas, métodos, temas de estudo e modos de comunicação. Trata-se também de viragens linguística e computacional com impacto no patrimônio documental, pois fazem emergir novos métodos de análise, sobretudo os empíricos e quantitativos, alterando os processos e práticas da pesquisa científica. Esse campo de estudos está em expansão e o artigo de Widad Mustafa El Hadi, da Universidade de Lille 3, na França, discorre sobre sua gênese e processo de institucionalização.

A exponencial produção de dados do presente passa a ser *big data* quando não se consegue mais analisá-los ou processá-los pelos meios tradicionais de acesso, servidores e *softwares* que nos são familiares. Tal fenômeno do início do século XXI, se apresenta justamente quando a informação é vislumbrada como meio de melhoria da vida do ser humano e inclusive como possibilidade de referendar decisões. O manejo dos *big data* no contexto da transformação digital é objeto da conferência de José Eduardo Santarem Segundo, da Universidade de São Paulo. Ele foca a pesquisa científica e mostra a possibilidade de erro no tratamento de dados incompletos, desbalanceados, inconsistentes ou com ruídos.

Os humanos não têm os recursos necessários para tratar grandes volumes de dados e necessitam de mediação que tornem as informações biodisponíveis. Daí a necessidade de máquinas e de transduções informacionais para se relacionar com os dados. Sobre essas questões discorre Ricardo César Gonçalves Sant'Ana, da Universidade Estadual Paulista e envereda também pela reflexão filosófica, apontando a necessidade de conscientizar a sociedade sobre a transformação em curso, atores e interesses envolvidos, geradora de um novo ser humano, por ora desconhecido.

O projeto Memória para Todos, coordenado por Maria Fernanda Rollo, da Universidade Nova de Lisboa, Portugal, é apresentado como relato de experiência e como ponto de partida para reflexões sobre a temática do seminário. Trata-se de programa de formação e pesquisa colaborativa desenvolvido no campo das ciências sociais, com fisionomia

multidisciplinar e uso intensivo de ferramentas digitais, dentro do qual é possível distinguir os eixos que singularizam as chamadas humanidades digitais.

Partindo do pressuposto de que a organização do conhecimento é um conjunto de processos, instrumentos e produtos que propiciam a mediação de culturas (do produtor, do organizador e do usuário de conhecimentos socializados e registrados), José Augusto Chaves Guimarães, da Universidade Estadual Paulista, discute os desafios e as perspectivas que se colocam à pesquisa científica nessa área, frente a novos elementos contextuais, em cenário permeado pelos *big data* e pelas humanidades digitais. Como garantir a transparência e a confiabilidade dos dados nas instituições de pesquisa, sobretudo aquelas custeadas por recursos públicos? Como preservar, tratar e dar acesso a dados inerentes a todas as fases das pesquisas, não se limitando aos resultados publicados?

Por fim são apontados os desafios que emergem das relações e dos compromissos estabelecidos entre as humanidades digitais e a organização do conhecimento. Johanna Wilhelmina Smit, da Universidade de São Paulo, evoca os ensaios pioneiros de Jean-Claude Gardin para demonstrar, de modo retrospectivo, as tensões entre o quantitativo e o qualitativo no processo de organização do conhecimento.

Os tópicos debatidos, o alto nível dos conferencistas, a atualidade e a pertinência de uma temática perfeitamente alinhada com a política cultural da Fundação configuram o acerto da iniciativa. E sinalizam a necessidade de programar outros eventos igualmente estimulantes.

Sergio Fausto

Diretor

Fundação Fernando Henrique Cardoso

A SINERGIA ENTRE A ORGANIZAÇÃO DO CONHECIMENTO E AS HUMANIDADES DIGITAIS: ALGUMAS REFLEXÕES

Widad Mustafa El Hadi

1. INTRODUÇÃO

As humanidades digitais (HD) são definidas como “um campo multidisciplinar importante, na intersecção das tecnologias digitais e das ciências humanas”. Esta interdisciplinaridade visa criar aplicações e modelos que abrem caminho a novos campos de pesquisa, tanto na área das ciências humanas quanto da informática bem como em todas as tecnologias a elas associadas. As HD estudam também o impacto dessas técnicas sobre o patrimônio cultural, as instituições memoriais, as bibliotecas, os arquivos e a cultura digital (Warwick et al., 2012). Mas como defini-las? O que as torna

unidas apesar da grande diversidade delas? No mais alto nível de generalidade, pode-se dizer que as humanidades digitais designam um diálogo interdisciplinar sobre a dimensão digital das pesquisas em ciências humanas e sociais, ao nível das ferramentas, dos métodos, dos temas de estudos e dos modos de comunicação. Este campo, caracterizado pela crescente importância da virada digital, teve um profundo impacto sobre as ciências humanas. De fato, ele modificou a natureza e as modalidades de construção do saber e fez emergir novos métodos de análise, notadamente empíricos e quantitativos, transformando assim os processos e as práticas da pesquisa científica.

A pesquisa científica em humanidades digitais é geralmente considerada uma pesquisa colaborativa e interdisciplinar que transcende as fronteiras acadêmicas tradicionais entre os cientistas da computação, os engenheiros, os bibliotecários e os profissionais da informação, bem como os pesquisadores em ciências humanas. A inclusão do computador na pesquisa em ciências humanas modifica não somente o escopo e as possibilidades desta pesquisa, mas também as próprias condições em que ela está sendo realizada (Nyhan, 2012).

Esta comunicação se propõe seguir a gênese das humanidades digitais e a institucionalização de seu campo. Ela procurará também evidenciar as ligações históricas e epistemológicas deste campo com as ciências da informação, notadamente a sua subdisciplina da organização do conhecimento no que diz respeito a áreas comuns tais como os locais de saber (bibliotecas, arquivos, museus), as aplicações, as ferramentas e as infraestruturas de pesquisa.

Também visa destacar o papel do profissional da informação no campo das humanidades digitais.

Um elemento que nos parece essencial para compreender o surgimento e a institucionalização das HD é a comparação entre a emergência e o desenvolvimento da área de HD com aquela das ciências da informação, uma disciplina científica que baseia os fundamentos teóricos e metodológicos da organização dos conhecimentos e, portanto, as bibliotecas digitais. Mesmo que o assunto ainda não tenha sido explorado, é possível destacar os problemas e os desafios comuns a essas duas áreas de estudos. Se voltarmos à história e à natureza das HD para lançar uma luz sobre algumas das questões que os conectam às bibliotecas digitais, é essencial relembrar o paralelo estabelecido entre as ciências da informação e sua relação com outras disciplinas. As ciências da informação importaram conhecimentos e métodos de outras disciplinas e, em seguida, exportaram ideias para diferentes campos, tais como a informática e a gestão. A natureza das humanidades digitais é neste aspecto semelhante na medida em que elas se desenvolveram, entre outros campos, ao lado da linguística de *corpus*, algo que se situa na fronteira entre as ciências humanas, as ciências sociais e as ciências aplicadas. A primeira manifestação “moderna” das HD foi a adoção precoce do processamento automatizado de dados textuais e linguísticos e do *corpus* conhecido hoje como *linguística computacional* e *linguística de corpus*. Voltaremos com mais detalhes sobre esta questão.

É possível considerar as HD de acordo com três pontos de viragem: um ponto de viragem epistemológico; um ponto de

viragem computacional e um ponto de viragem linguístico. Quanto às suas relações com as ciências da informação, duas entradas são possíveis: uma entrada pelas infraestruturas da pesquisa em ciências humanas e sociais e a outra pela contribuição dos precursores tais como Jean-Claude Gardin, que é considerado na França como sendo um dos precursores das HD.

2. HUMANIDADES DIGITAIS: DEFINIÇÕES, GÊNESE E INSTITUCIONALIZAÇÃO DO CAMPO

As humanidades digitais são uma temática científica dinâmica, florescente e plural, pois ela reúne temáticas disciplinares que vão desde a informática até o patrimônio, passando pelas ciências humanas e sociais. É um deslocamento de fronteiras segundo Mounier e Darcos (2012): fronteiras entre disciplinas, para as quais o recurso comum aos instrumentos informáticos pode facilitar a aproximação e fronteiras entre pesquisadores em ciências humanas e engenheiros, já que um diálogo fecundo se estabelece entre as questões científicas e as possibilidades técnicas. As humanidades digitais estão, assim, na interseção das ciências e tecnologias informáticas e das ciências humanas e sociais. A ampla disponibilidade dos recursos de informação digital e das instalações informáticas em rede, juntamente com sistemas de comunicação, de compartilhamento colaborativos levou ao desenvolvimento deste novo paradigma de pesquisa geralmente descrito como *e-ciência* ou *e-pesquisa* ou ainda *e-humanidades*. Trata-se de uma pesquisa colaborativa, entre os pesquisadores de diversas instituições

e que está amplamente distribuída geograficamente. As normas e os padrões de descrição de metadados, o armazenamento deles, os formatos de troca de dados são essenciais para permitir o compartilhamento dos recursos e dos dados, bem como o acesso aberto a esses dados. Os sistemas que suportam essas formas de trabalho são às vezes chamados de “colaboratórios”, ou “ambientes virtuais de pesquisa”, ou infraestruturas de pesquisa ou ainda ciberinfraestruturas, um tema sobre o qual daremos esclarecimentos no item (4.1.)

O termo “humanidades digitais”, que se acredita ser o equivalente do termo em inglês “*digital humanities*”, é atribuído a John Unsworth, Susan Schreibman e Ray Siemens, com a edição de *A Companion to Digital Humanities* em 2004, designando tanto a utilização das ciências humanas e sociais para estudar as tecnologias e, inversamente, as práticas digitais e a utilização dessas tecnologias para estudar os objetos culturais e sociais. Este ponto de viragem digital (*computational turn*) remonta aos meados do século 20, entretanto, quando as ciências humanas (*humanities computing*), em especial, os linguistas, os historiadores e os sociólogos (sociologia quantitativa) se apoderaram da informática para processar os dados nos seus trabalhos de pesquisa. São principalmente os pesquisadores em linguística e os especialistas da análise dos textos que foram entre os primeiros a utilizar as potencialidades da informática para conduzir suas pesquisas. O padre Roberto Busa (1913-2011) é assim considerado como um dos precursores das humanidades digitais (HD) ao lançar o Índice Thomisticus no final da década de 1940. Enquanto ele trabalhava

na edição das obras completas de São Tomás de Aquino, o padre Busa, com a ajuda de equipes da IBM, empreendeu o desenvolvimento de um índice reunindo todos os termos do *corpus*, ou seja, cerca de 11 milhões de palavras.

Outro momento forte na história das HD é a exploração das tecnologias informáticas para a constituição de vastos corpora de textos. O Brown Corpus, lançado na década de 1960 por Henry Kurcea e Nelson Francis na Universidade de Brown, nos Estados Unidos, propôs assim um *corpus* de referência para a língua americana. Esse movimento é conhecido nos anos 90 sob o rótulo de *linguística de corpus*.

Optamos por nos concentrar aqui em uma abordagem historiográfica e epistemológica das HD. O objetivo de nossa contribuição é apresentar uma visão geral dos trabalhos relativos às HD para destacar os elementos que nos parecem mais interessantes e fornecer pistas bibliográficas. Para tanto, lemos muitos textos na forma de artigos, de anais de congressos e de relatórios. Abordaremos os temas recorrentes na literatura examinada. Optamos por considerar as HD a partir de vários ângulos que os caracterizam, a saber, o ponto de viragem epistemológico; o ponto de viragem computacional; e o ponto de viragem linguístico. Também abordaremos suas ligações com as Ciências da Informação e sua subárea, a organização do conhecimento.

3. OS PRINCIPAIS PONTOS DE VIRAGEM DAS HD

3.1. O PONTO DE VIRAGEM EPISTEMOLÓGICO

O rótulo HD designa o ponto de viragem epistemológico provocado nas ciências humanas e sociais pela transição da cultura impressa para a cultura digital. O rótulo *Digital Humanities* (DH), utilizado pela primeira vez em 2001, permite que as instituições garantam essa transição, provavelmente ao longo de uma geração. As “humanidades” serão digitalizadas, avisa Claire Clivaz no seu *blog*, da Universidade de Lausanne (2012). Esta é uma nova transdisciplina que cruza e hibridiza as diversas áreas. Ela recondicionou todos os saberes e todas as práticas universitárias, trazendo novos modos de trabalho, novos locais de produção e novos públicos (Citton, 2015). As HD devem ser diferenciadas de todo um campo de experimentação cruzando em ciências humanas e sociais e informática (literary/linguistic/humanities computing). Este movimento, que nasceu em parte da digitalização e da disponibilização dos recursos, visa integrar a cultura e as tecnologias digitais nas práticas de pesquisa em ciências humanas e sociais. “O ponto de viragem digital tomado pela sociedade modifica e questiona as condições de produção e de difusão do saber”, observa Mounier (2010).

3.2. O PONTO DE VIRAGEM COMPUTACIONAL

De maneira mais precisa, certos autores como David Berry evocam um “ponto de viragem informático” (*computational turn*). As pesquisas em ciências humanas e sociais, como as outras discipli-

nas, são atravessadas por um paradigma comum baseado numa proposição epistemológica comum, aquela da teoria da informação. Em todas as disciplinas, tanto em biologia ou literatura, quanto em física nuclear ou antropologia, o objeto estudado é convertido, manipulado, analisado sob uma categoria comum: a informação, objeto de cálculos (Berry, 2011).

Este ponto de viragem tem sido objeto de diversos debates discutindo os desafios do digital para as Ciências Humanas e Sociais, um ponto de viragem que afeta consideravelmente nossa relação com a ciência, modifica profundamente as disciplinas e suas epistemologias (Berry, 2011) e que até questiona as condições de produção e de circulação do saber. Essas mutações influenciam consideravelmente as práticas de pesquisa, induzindo a uma nova relação entre a ferramenta e o pesquisador. O advento da Internet e, em seguida, da cultura Web, favoreceu o aprimoramento e o desempenho dos modelos de compartilhamento, de comunicação, de trocas. Foi a chegada da internet, depois da web nos anos 1990, que provocou a mudança da *Humanities Computing* para as *Digital Humanities*. A publicação em 2004 de *A Companion to Digital Humanities* e o lançamento no ano seguinte pela Editora da Universidade do Illinois da coleção *Topics in Digital Humanities* tornaram esse desenvolvimento visível.

3.3. O PONTO DE VIRAGEM LINGÜÍSTICO: O PROCESSAMENTO AUTOMATIZADO DE DADOS TEXTUAIS, PRIMEIRAS EXPERIÊNCIAS

Os pesquisadores em linguística e os especialistas em análise de texto foram os primeiros a usar o potencial da informática para

conduzir suas pesquisas. O padre italiano Roberto Busa (1913-2011) é considerado como sendo o precursor, e a historiografia fixa tradicionalmente o início da *Humanities Computing* no lançamento do *Índice Thomisticus* no final dos anos 1940. O padre Busa trabalhava na edição das obras completas de São Tomás de Aquino, ele empreendeu com a ajuda das equipes da IBM a elaboração de um índice reunindo todos os termos do *corpus*, ou seja, cerca de 11 milhões de palavras. Uma equipe cuida da digitação dos textos em cartões perfurados e um programa é concebido para gerar automaticamente o índice. Se a originalidade do projeto reside na implementação fundamentalmente nova, de um concorciador, sua finalidade continua sendo a edição impressa - 56 volumes, o primeiro dos quais saiu em 1974.

O índice do padre Busa revela uma relação fundamental entre as ciências da informação (CI) e mais especificamente a organização do conhecimento e as HD: o índice do padre Busa, que supostamente ajuda a compreender a obra de São Tomás de Aquino, ressoa com a própria definição dos conceitos de organização da informação ou a organização da informação tais como analisados por Bawden e Robinson, em sua *Introduction to Information Science* (2012). O assunto é denominado tanto “organização da informação” quanto “organização do conhecimento”. Esta definição nos lembra que o objetivo perseguido na construção dos índices (KWIC e KWOC) pode ser tanto o entendimento da estrutura do próprio conhecimento quanto o objetivo pragmático, ajudar a organizar ou classificar documentos físicos em estantes. Esses documentos, inicialmente, físicos em estantes tornaram-

se, recentemente, documentos digitais em um espaço virtual. O trabalho no Índice *Thomisticus* pode certamente ser considerado como sendo um trabalho precursor.

Seguindo os passos de Roberto Busa com o Índice *Thomisticus* informatizado em 1949, importantes projetos incluíram o “Brown University Corpus of American English” do professor Kucera, uma seleção analisada da prosa inglesa americana que define a norma para os corpora em língua natural, e o *Thesaurus Linguae Graecae*, que contém quase todos os textos gregos antigos, de Homero até 600 DC. Foi assim que o *Brown Corpus* foi lançado na década de 1960 por Henry Kucera e Nelson Francis na Universidade de Brown, nos Estados Unidos. Esses dois linguistas propuseram, assim, um corpus de referência para a língua americana. Esse movimento é conhecido nos anos 90 sob o rótulo de *linguística de corpus*. Este movimento constitui mais um momento forte da história das HD em conexão com o ponto de viragem linguístico, pois resulta na exploração das tecnologias informáticas para a constituição de vastos corpora de textos e as recentes evoluções acerca dos dados da pesquisa. A necessidade de perenizar os dados de pesquisa (essencialmente os corpora para a pesquisa em ciências humanas e sociais) e de compartilhá-los através da colocação dos mesmos em rede tem sido objeto de iniciativas políticas específicas voltadas para o desenvolvimento das HD (Caverni, et al., 2009).

As humanidades digitais apresentam vários desafios, mas também oferecem várias oportunidades. De fato, os meios tradicionais de processamento parecem insuficientes face a essas “grandes mas-

sas de dados” (*big data*) que exigem uma renovação da reflexão sobre os processos de extração, de armazenamento (grids, cloud computing ...), de processamento, de compartilhamento, de análise e de visualização dos dados.

Os métodos de análise automatizada parecem oferecer a possibilidade de processamento rápido e eficiente: os métodos digitais oferecem a possibilidade de processar grandes massas de dados digitalizados com apenas alguns cliques, e os resultados podem ser apresentados na forma gráfica.

As humanidades digitais são divididas em cinco áreas principais de discussão a fim de cobrir um conjunto de necessidades (*ibid.* 2009).

4. SISTEMAS DE ORGANIZAÇÃO DO CONHECIMENTO (KOS) E HUMANIDADES DIGITAIS (HD)

4.1. UMA ENTRADA PARA AS INFRAESTRUTURAS DA PESQUISA: UMA NOTA HISTÓRICA

A temática das HD merece que voltemos sobre suas ligações com as ciências da informação, em particular os sistemas de organização do conhecimento. Gostaríamos de começar pelo elo mais antigo, ou seja, a relação desta área emergente com as infraestruturas de pesquisa para as humanidades. Moulin faz remontar o nascimento das HD modernas à história das infraestruturas de pesquisa. A partir do século III aC, o imperativo de coletar, orga-

nizar e preservar saberes para o serviço do avanço do conhecimento deu origem ao primeiro “centro de informação” na forma do *Mouseion*, um centro cultural, universitário e uma biblioteca fundada em Alexandria (Egito) pelos sucessores de Alexandre o Grande. As consequências positivas deste ambicioso empreendimento tornaram-se rapidamente evidentes, através da produção de diversas infraestruturas de pesquisa na forma de gramáticas e de léxicos, ancestrais das atuais infraestruturas para os recursos linguísticos tais como o CLARIN, um elemento da galáxia de DARIAH EU (Digital Research Infrastructure for the Arts e Humanities). Como O. Olesen-Bagneux havia observado, muito pouco se sabe sobre a antiga biblioteca de Alexandria. Algumas fontes indicam que ela poderia ter guardado entre 40.000 e 700.000 rolos de manuscritos. A biblioteca fazia parte de uma instituição religiosa, o *Mouseion*, e os estudiosos eram na verdade escravos qualificados que lá tinham sido aprisionados. Qualquer tentativa de fuga podia ser punida com a morte, e uma parte da poesia escrita por esses pesquisadores atrás das grades foi realizada durante cerimônias religiosas.

O sucesso precoce das infraestruturas para a pesquisa em ciências humanas (se ousarmos falar desta maneira) não foi perdido no período que seguiu o fim do Império Romano; ao contrário, ele inspirou uma série de atividades que deviam fomentar um sistema emergente de conhecimento e apontar a importância da coleta, da organização e da conservação dele a serviço do desenvolvimento científico subsequente, por exemplo, durante a criação das bibliotecas medievais e, posteriormente, quando da formação

das coleções de arte. Esses primeiros “bancos de dados” forneceram o material de base para as fases posteriores de Ris (Research Infrastructures) nas ciências humanas.

4.2. AS INFRAESTRUTURAS MODERNAS PARA A PESQUISA CIENTÍFICA EM CIÊNCIAS HUMANAS E SOCIAIS: ALGUNS ELEMENTOS DE DEFINIÇÕES

Uma infraestrutura de pesquisa científica é um conjunto de projetos, de equipes, de centros, de laboratórios e de departamentos, de associações e de sociedades científicas, etc. Trata-se, portanto, de Infraestruturas ou ciberinfraestruturas para as humanidades digitais. Para o ACLS (American Council For Learned Societies), uma ciberinfraestrutura deve ser pública, sustentável, interoperável, encorajando a colaboração entre pesquisadores e apoiando a experimentação. Os termos *ciberestrutura* ou *infraestrutura* (da informação, da pesquisa, do conhecimento), que o relatório da ESF (2011) considera como equivalentes, passam a ser então os marcadores desta evolução qualificada por Borgman (2009) de ‘*e-Research*’ (Favier, 2015).

A noção de infraestrutura é utilizada para caracterizar a evolução das ciências num ambiente digital em rede, sendo uma das consequências a possibilidade e a necessidade de compartilhar os dados da ciência bem como aquelas que garantem a perenidade deles. O exemplo dado com mais frequência é aquele de assuntos tais como o clima ou a biodiversidade, que exigem um compartilhamento dos dados em escala global. Os bancos de dados bioinformáticos, em ciências da vida (*GenBank*,

Worldwide Protein Data Bank) por exemplo, têm sido também, há muito tempo, um dos modelos desta ciência para a qual o digital e a capitalização dos dados produzidos é essencial. Desse ponto de vista, as humanidades digitais representam um fenômeno mais ligado à história das ciências do que àquela das bibliotecas digitais, às quais, no entanto, se assemelham em razão dos objetos que manipulam. O termo *infraestrutura*, antes de tudo, designando o conjunto dos meios tecnológicos, passou de um único discurso político (aquele das instâncias financiando a pesquisa) a uma noção científica qualificando esta articulação entre digital e prática científica (Favier, 2015).

Vale a pena relembrar aqui um elo conhecido entre as HD e as ciências da informação, uma disciplina que está na base da organização do conhecimento e, portanto, das bibliotecas digitais. Estas últimas têm um grande papel a desempenhar para as HD, elas são muito mais do que simples substitutos digitais para as bibliotecas clássicas já existentes. Elas são consideradas por Svensson (2010) como uma parte importante da infraestrutura das HD. Elas têm o potencial de ser ambientes virtuais complexos para a pesquisa científica ou laboratórios virtuais (*colaborteries*), ou seja, um sistema *online* que ajuda os pesquisadores a trabalharem em modo colaborativo. As características deste tipo de ambiente costumam incluir o suporte técnico, as ferramentas digitais permitindo o trabalho colaborativo (fóruns, wikis etc.), os documentos de hospedagem e ferramentas específicas à disciplina, tais como a análise dos dados, a visualização ou a gestão da simulação, a gestão das publicações e ferramentas pedagógicas. Os ambientes virtuais de

pesquisa tornaram-se importantes em áreas onde a pesquisa é principalmente realizada por equipes que abrangem instituições e até mesmo países inteiros. A capacidade de compartilhar facilmente dados e os resultados da pesquisa é valiosa.

5. LIGAÇÕES ENTRE AS HD E AS CIÊNCIAS DA INFORMAÇÃO: OS PRECURSORES COMUNS

5.1. JEAN-CLAUDE GARDIN (1925-2013)

Para os especialistas das HD, o padre Roberto Busa é considerado como sendo o precursor das HD modernas. Para os especialistas de nossa disciplina, é Jean-Claude Gardin que é considerado como sendo um dos precursores das HD modernas. De fato, alguns anos após o índice do padre Busa, em 1956, Jean Claude Gardin publicou o primeiro arquivo elaborado pelos membros da missão em uma brochura intitulada “O arquivo mecanográfico do ferramental, para descrever as ferramentas metálicas da Idade do Bronze descobertas em uma vasta área que se estende dos Balcãs até o Indo” (Gardin, 1956). Os cinco primeiros “Códigos” para a análise dos objetos arqueológicos, que diziam respeito à cerâmica, às armas e ferramentas, às moedas gregas, aos selos orientais e aos temas iconográficos, foram também criados em 1956. (Moscati, 2016; Mustafa El Hadi, 2018). Esses códigos de análise, que são conservados nos Arquivos com a documentação correspondente, só foram publicados alguns anos depois

(Christophe, Deshayes, 1964; Digard et al. 1975; Le Rider, 1975; Gardin, 1976; Gardin, 1978, citado por Moscati, 2016).

Moscati (*ibid*) dá mais detalhes sobre as atividades do Centro e sobre o enorme trabalho de Gardin que vai da arqueologia à análise do discurso a partir de textos religiosos. No início da década de 1960, a atividade do Centro¹ assentava essencialmente em três pilares principais: os objetos, a iconografia e os textos. Podemos ver aqui a relação direta com o Padre Roberto Busa, o pai das *Digital Humanities* de hoje.

Em suas áreas de orientações de pesquisa, Jean-Claude Gardin colabora ao desenvolvimento das aplicações da automação documental, das estatísticas, dos sistemas especializados em arqueologia e aos debates em epistemologia das ciências humanas. Seus anos documentais começaram quando ele tinha 27 anos. Foi nessa época que ele começou sua carreira científica. Henri Seyrig, também arqueólogo, possui uma máquina mecanográfica para a realização da indexação dos fundos documentais. Jean-Claude Gardin se encarrega deste projeto de indexação. Foi durante o domínio da mecanografia que Gardin percebeu que existiam outras formas de pesquisa documentarias (a descrição do objeto, o registro dos dados de escavação, a bibliografia, por exemplo). Nos anos 1950, o CNRS abriu-se a imensas pesquisas em informação científica e técnica, que deve levar no final à automatização das bibliotecas, e para o futuro Ministério da Cultura a documentação do patrimônio cultural, como para os inventários e os museus,

1. Referência ao Centro Mecanográfico de Documentação Arqueológica do CNRS (Nota do tradutor).

por exemplo. Jean-Claude Gardin é conhecido por suas pesquisas. Durante sua carreira, ele empreendeu pesquisas em duas áreas: a automação de pesquisa documental de informação e das linguagens documentais, desenvolvendo a linguagem “SYNTOL” (*Systagmatic Organization Language*) e, em seguida, o desenvolvimento dos trabalhos relativos à arqueologia da Ásia Central, realizando escavações no Afeganistão em sítios relativos à antiga Bactria.

GARDIN VISTO POR ÉRIC DE GROLIER (1911-1998)

Para Eric de Grolier, o SYNTOL ocupa um lugar importante na contribuição de Gardin. Ele revela que o Syntol “representa a tentativa mais ambiciosa e moderna feita na França, durante a última década, de construir uma linguagem de recuperação de informação incorporando muitos recursos considerados de valor permanente, para uma aplicação com seleção automática, de sistemas mais antigos de classificação” (1965, p. 147).

GARDIN VISTO POR JOHANNA WILHELMINA SMIT

Outra contribuição das biografias e especialistas de Gardin é aquela de Marilda Lopes Ginez de Lara, Johanna Wilhelmina Smit e Maria de Fátima Gonçalves Moreira Tálamo (Universidade de São Paulo, Brasil), tratando de um aspecto da contribuição de Gardin. Em “A organização do conhecimento a partir das propostas de J.-C. Gardin”, as autoras sublinham a contribuição fundamental deste arqueólogo para o processo de automatização do processamento da informação. Para Gardin, o trabalho intelectual

deve ser baseado em processos de classificação explícitos e formais, para que a máquina possa manipular os dados (o cálculo) e criar condições para estabelecer correlações e associações de informação. O artigo retoma as propostas de Gardin que hoje o ligam aos desenvolvimentos da web semântica e das ontologias e ao seu importante papel para os sistemas de informação e a documentação.

GARDIN VISTO POR JOSÉ AUGUSTO GUIMARÃES

Em sua conferência convidada, ISKO-France 20217, “Jean-Claude Gardin e a busca por metodologias interdisciplinares e ferramentas confiáveis para as práticas de organização do conhecimento: *análise documental e linguagem documental*”, José Augusto Guimarães (Universidade Estadual Paulista, Brasil) volta sobre o papel de Paul Otlet e sobre sua importante contribuição à elaboração de técnicas e de metodologias destinadas aos meios documentais, participando assim a promover o mais amplo acesso possível à informação. Guimarães lembra o papel fundamental do OC como mediador entre os contextos de produção e de utilização dos conhecimentos. Ele destaca também o papel fundamental de Jean-Claude Gardin na construção da dimensão epistemológica da organização do conhecimento como mediador entre os contextos de produção e de utilização do conhecimento. Ele enfatiza ainda o papel fundamental de Jean-Claude Gardin na construção da dimensão epistemológica da organização do conhecimento e do desenvolvimento de uma metodologia interdisciplinar, baseada em modelos lógicos e linguísticos, a “análise documentária”,

bem como sobre a pesquisa de ferramentas confiáveis destinadas a construir pontes sólidas entre o documento, o sistema documental e o usuário, as “linguagens documentárias”.

5.2. SUSANNE BRIET ET LE MOUVEMENT GLAM (GALERIES, LIBRARIES, ARCHIVES & MUSEUMS)

A integração dos serviços de informação na área do patrimônio cultural, que assumiu a forma de um movimento dinâmico no mundo, denominado GLAM, pode ser considerada como sendo a encarnação de uma das principais ideias de Susanne Briet relativa à integração das atividades das diversas instituições e comunidades profissionais cuidando da documentação e da difusão do saber humano. Para Robert Pagès, seu aluno, “as profissões de bibliotecário, de arquivista e de curador de museu são profissões pré-documentalistas”: um documentalista é tanto um arquivista, um bibliotecário quanto um curador de museu.

5.3 LUGAR DA ORGANIZAÇÃO DO CONHECIMENTO NAS INFRAESTRUTURAS DA PESQUISA PARA AS HUMANIDADES DIGITAIS

A organização do conhecimento, uma área de grande importância em ciências da informação, continua a ser uma área essencial, pois qualquer prática digital em rede implica, pelo menos, na produção de informação, na indexação, na classificação e na pesquisa de informação. Quanto mais a Internet evolui e se banaliza, mais a organização do conhecimento é central. No cruzamento da evolução das tecnologias Web, do movimento da abertura dos

dados e do conhecimento do uso e do usuário, a organização do conhecimento é essencial para as infraestruturas da pesquisa em ciências humanas e sociais.

6. DISCUSSÕES

OS DESAFIOS EM TORNO DOS DADOS DE PESQUISA

As humanidades digitais apresentam vários desafios, mas também oferecem várias oportunidades. De fato, os meios tradicionais de processamento parecem insuficientes frente a estas “grandes massas de dados” (Big Data) que requerem uma renovação da reflexão sobre os processos de extração, de armazenamento, de processamento, de partilha, de análise e de visualização dos dados. Os métodos de análise automatizada parecem oferecer a possibilidade de processamentos rápidos e eficientes. Os métodos digitais em constante evolução oferecem a possibilidade de processar grandes massas de dados digitalizados em apenas alguns cliques, e os resultados podem ser apresentados na forma gráfica usando ferramentas poderosas de visualização.

As humanidades digitais, portanto, visam enfrentar esse desafio e realizar o potencial em torno dos dados da pesquisa, e isso é tanto um desafio metodológico e social quanto um desafio técnico (Borgman, 2007). Como lidamos com o crescimento exponencial dos dados, que agora chamamos de *big data*? Aqui, o que está em jogo não é apenas uma questão de volume de dados, mas

também de qualidade desses dados; - quais indexações e quais enriquecimentos devem ser adotados para processar os corpora para torná-los não apenas acessíveis, mas relevantes para o pesquisador.

Segundo Caverni e Darcos (2009), as humanidades digitais são consideradas de acordo com cinco eixos principais de discussão para cobrir um conjunto de necessidades:

- **As coleções e os dados digitais:** este eixo abordará questões em torno da digitalização, a conservação, a organização e a estruturação dos dados e dos corpora multimodais (textos, imagens, sons, medidas quantitativas, etc.);
- **A organização e a disponibilização dos dados:** a navegação nos dados através de um acesso inteligente e garantindo uma grande interconexão dos dados;
- **A exploração e a análise dos dados:** o que se pretende é o processamento e a análise dos dados por meio de técnicas de mineração de dados e de descoberta do conhecimento;
- **A disseminação do saber científico:** a edição e a difusão dos conteúdos na forma de sites Internet, de livros eletrônicos, de arquivos abertos;
- **A capitalização, a difusão, a animação científica** (por meio de debates, de intercâmbios, de fóruns, das redes, das comunidades virtuais, etc.)

Uma das áreas de interesse das humanidades digitais em relação à organização da informação e do conhecimento é estudar as estratégias dos pesquisadores em ciências humanas. Estudos

realizados pela equipe da UCL de Londres mostraram que as necessidades de informação dos pesquisadores em ciências humanas diferem daquelas dos cientistas (Terras, 2014). Ao contrário dos cientistas, os pesquisadores em ciências humanas continuam a utilizar documentos impressos e objetos físicos junto com recursos eletrônicos (Warwick et al., 2008). Até recentemente, os recursos foram concebidos para a maioria dos usuários que não são oriundos das ciências humanas. O desenvolvimento de bibliotecas digitais para esta população é vital para a pesquisa futura em ciências humanas.

LUGAR DAS BIBLIOTECAS DIGITAIS NUMA REDE DE INFRAESTRUTURAS DE PESQUISA

As bibliotecas digitais são muito mais do que simples substitutos digitais para bibliotecas convencionais existentes. Elas têm o potencial de serem ambientes de pesquisa virtuais complexos (ambientes virtuais de pesquisa, um Laboratório virtual, um sistema online que ajuda os pesquisadores a colaborar).

As funcionalidades geralmente incluem fóruns de discussão e *wikis*, *blogs*, etc. a hospedagem de documentos e certas ferramentas específicas a uma disciplina, tais como a análise de dados, a visualização ou a gestão de simulação, a gestão de publicação, das ferramentas pedagógicas, tais como apresentações de *datashow*. Os ambientes virtuais de pesquisa ganharam destaque em áreas onde a pesquisa é principalmente realizada em equipes, em instituições e até mesmo em países (a possibilidade de compartilhar facilmente a informação e os dados de pesquisa).

COMO AS BIBLIOTECAS PODEM REIVINDICAR UM PAPEL NESTA TENDÊNCIA CIENTÍFICA EMERGENTE DAS HD?

O desenvolvimento das humanidades digitais levanta questões sobre o lugar das bibliotecas neste novo campo da pesquisa. As bibliotecas foram convidadas desde muito cedo a desempenhar seu papel de expertise em informação, digitalização, arquivamento e publicação online, mas este papel controverso é questionado pela concorrência de novos tipos de prestadores de informações online, tais como bancos de dados e arquivos abertos, ou os portais especializados, em razão dos quais as bibliotecas tiveram que redefinir sua identidade, oferecendo serviços adaptados à pesquisa.

- a expertise dos bibliotecários se estende a muitas questões das Humanidades digitais, a gestão de bancos de dados, sistemas de informática documental;
- a digitalização e a constituição de corpus documentais digitalizados (que mobilizam tanto as competências jurídicas, técnicas quanto organizacionais);
- a descrição dos recursos e interoperabilidade (que envolvem competências em matéria de estruturação dos dados, de normas e de protocolos de intercâmbio de dados, bem como a produção e a gestão dos metadados descritivos, administrativos, técnicos, jurídicos, e notadamente os trabalhos de intervenientes institucionais tais como a AFNOR e a ABES (Agência bibliográfica do ensino superior) ;
- o direito autoral e os princípios jurídicos da propriedade intelectual regulando a reutilização dos dados, os serviços à pesquisa e à publicação digitais;

- a publicação de conteúdos (hospedagem de uma revista online ou de atos de colóquios, constituição de um arquivo institucional, conselhos em matéria de produção de vídeos ou de gravações sonoras, disponibilização de blogs, de wikis ou de plataformas de gestão de conteúdos...);
- trazer um valor agregado à pesquisa científica.

As bibliotecas poderão assumir a missão de acompanhar as humanidades digitais, oferecendo aos professores e aos pesquisadores, iniciantes ou experientes, serviços de apoio para seus trabalhos e práticas.

RENOVAÇÃO DOS MÉTODOS DE COMUNICAÇÃO CIENTÍFICA

O movimento das humanidades digitais está levando toda uma nova geração de pesquisadores a revolucionar suas práticas de comunicação científica. A partir de agora, eles contarão cada vez mais com as mídias sociais para trocar e divulgar suas ideias da forma mais ampla possível. Uma única palavra capta perfeitamente o que constitui a unidade deste movimento: “open”.

E PARA OS PROFISSIONAIS DA INFORMAÇÃO?

As problemáticas abordadas pelos HD são semelhantes às preocupações dos principais atores da digitalização (arquivos e bibliotecas): seleção, digitalização, visualização, navegação, preservação, valorização, interoperabilidade, acessibilidade. Os profissionais da informação terão um papel importante e determinante no desenvolvimento das HD dentro das ciências humanas e sociais.

7. REFERÊNCIAS

BARRET, E. **Quel rôle pour les bibliothèques dans les humanités numériques?** Mémoire d'étude. Sous la direction d'Isabelle Westeel. 2014.

BERRY, D. M. The computational turn: thinking about the digital humanities. **Culture Machine**, v. 12, 2011.

BAWDEN, D. *et al.* **Introduction aux sciences de l'information.** London: Facet Publishing, 2012.

BORGMAN, C. The invisible library: paradox of the global information infrastructure. **Library Trends**, v. 57, n. 4, p. 652-675, 2000.

BUSA, R. Les annales de l'informatique des sciences humaines: l'Index thomisticus. **Computers and the Humanities**, v. 14, n. 2, p. 83-90, 1980.

CALHOUN, K. **Exploring digital libraries:** foundation, practice, prospects. London: Facet Publisher, 2014.

CAVERNI, J.-P.; Dacos, M. **Construire les digital humanities en France.** Des cyber-infrastructures pour les sciences humaines et sociales. Conférence des présidents d'universités (CPU). 2009.

CITTON, Y. Humanités numériques: une médiapolitique des savoirs encore à inventer. **Multitudes**, v. 2, n. 59, p. 169-180, 2015.

CRANE, G.; FUCHS, B.; IORIZZO, D. **The humanities in a global e-Infrastructure**: a web-services shopping list. UK e-science all hands meeting, 2007.

DUNN, S.; BLANKE, T. Next steps for e-science: the textual humanities and VREs. **D-Lib Magazine**, v. 14, n. 1/2, 2008.

FAVIER, L. Les humanités numériques et l'évolution des infrastructures de recherche: quels enjeux pour l'organisation des connaissances?. In: CREVRY, E. (Dir.). **Actes du 10ème Colloque ISKO-France, 5-6 novembre 2015. Systèmes d'organisation des connaissances et humanités numériques**, 2015. p. 3-16.

GUIMARÃES, J. A. Jean-Claude Gardin and the search for interdisciplinary methodologies and reliable tools for the knowledge organization practices: analyse documentaire and language documentaire. In: MUSTAFA EL HADI, W. **Fondements épistémologiques et théoriques de la science de l'information-documentation**: actes du 11e Colloque ISKO France 2017, 11-12 juillet 2017, Siège de l'Unesco, Paris [S. l.: s. n.], p. 41-59, 2018.

KOLTAY, T. Information science in the mirror of the digital humanities: some epistemological observations. **Zagadnienia Informacji Naukowej – Studia Informacyjne**, v. 52, n. 2, p. 23-31, 2014.

LARA, M. L. G. de; SMIT, J. W.; TALAMO, M. F. G. M. The organization of knowledge based on the proposals of J.-C. Gardin. In: MUSTAFA EL HADI, W. **Fondements épistémologiques et**

théoriques de la science de l'information-documentation: actes du 11e Coloque ISKO France 2017, 11-12 juillet 2017, Siège de l'Unesco, Paris [S. l.: s. n.], p. 84-92, 2018.

MOULIN, C. *et al.* **Research infrastructures in the Humanities.** Science Policy Briefing. Strasbourg: European Science Foundation, 2011.

MUSTAFA EL HADI, W. Documentation and information science: on some forgotten origins of the french contribution. **Zagadnienia Informacji Naukowej – Studia Informacyjne**, v. 56, n. 1, p. 7-20, 2018.

MUSTAFA EL HADI, W. **Fondements épistémologiques et théoriques de la science de l'information-documentation:** actes du 11e Coloque ISKO France 2017, 11-12 juillet 2017, Siège de l'Unesco, Paris [S. l.: s. n.], 2018.

MUSTAFA EL HADI, W.; FAVIER, L. Bridging the gaps between knowledge organization and digital humanities. **Advances in Knowledge Organization**, v. 14, p. 477-485, 2014.

NYHAN, J. Hidden histories: computing and the humanities: 1965-1985. **Digital Humanities Quarterly**, v. 6, n. 3, 2012.

NYHAN, J. What can modern-day digital humanities learn from studying its past? some recommandations and suggestion. *In*: 16ème Colloque International sur le Document Électronique – CiDE.16. Dispositifs numériques: contenus, interactivité et visualisation, Université de Lille 3, 21-22 novembre, p. 5-14. nov. 2013.

PALERMITI, R.; POLITY, Y. Dynamiques de l'institutionnalisation sociale et cognitive des sciences de l'Information en France. In: BOURE, R. (Ed.). **Les origines des sciences de l'information et de la communication en France: regards croisés**. Villeneuve-d'Ascq: Presses universitaires du Septentrion, 2002.

SCHREIBMAN, S.; SIEMENS, R.; UNSWORTH, J. (Eds.). **A companion to digital humanities**. Oxford: Blackwell Publishing, 2004. p. 132-144.

SVENSSON, P. The landscape of digital humanities. **Digital Humanities Quarterly**, v. 4, n. 1, 2010.

WARWICK, C.; TERRAS, M.; NYHAN, J. (Eds.). **Digital humanities**. London: Practice Facet Publisher, 2012.

WARWICK, C. *et al.* Library and information resources and users of digital resources in the humanities. **Program: electronic library and information systems**, v. 42, n. 1, p. 5-27, 2008.

WUSTEMAN, J. Virtual research environments: what is the librarian role? **Journal of Librarianship and Information Science**, v. 40, n. 2, jun. 2008.

OS DADOS NO CONTEXTO DA TRANSFORMAÇÃO DIGITAL: DESAFIOS E CONSEQUÊNCIAS

José Eduardo Santarem Segundo

INTRODUÇÃO

Inicia-se esse texto com o olhar para um cenário que vem se desenhando nas últimas duas ou três décadas. Não são necessários documentos científicos para embasar a realidade dos fatos; basta observar e comparar a quantidade de fotografias que eram tiradas em um aniversário de criança antes dos anos 2000 e observar a quantidade de disparos que uma infinidade de aparelhos celulares realiza num momento exatamente igual nos dias de hoje.

A tecnologia que permite um repique que dispara registro de imagens nos mais modernos aparelhos celulares, atualmente, permite que seja capturada uma quantidade infinitamente superior de imagens em relação ao que era reproduzido na época da tradicional máquina fotográfica.

Seria possível trazer aqui uma multiplicidade de exemplos que nos levassem a crer que o mundo está tomado por uma produção absurda de dados, mas basta olhar para o lado, ou então para a própria tela do aparelho de TV ou do dispositivo móvel que acompanha uma parcela significativa da população, para ter a certeza do transbordamento de dados.

Essa quantidade de dados que está sendo produzida na nossa frente, a olho nu, e que continua crescendo de modo desenfreado, não é necessariamente de *big data*, apesar de contribuir diretamente com esse fenômeno. É importante observar que a produção de dados gerados de forma explícita pelo ser humano, principalmente com seu aparelho celular, é apenas uma gota no oceano de dados que tem sido produzido diuturnamente.

A referência à produção explícita de um dado pode ser entendida como aquela que se efetiva com uma ação em que o usuário tem o conhecimento de que um novo dado é gerado e será armazenado. Ou seja, registra-se um fato.

Essa facilidade na criação de dados amplia-se à medida que as pessoas têm mais acesso a dispositivos como os celulares e também a própria internet. É possível ver, por meio do Quadro 1, que a quantidade de usuários que acessa a internet, e consequentemente produz conteúdo, tem aumentado no Brasil de forma relevante, chegando a quase 90% da população; e isso impacta diretamente o aumento da produção de dados.

Quadro 1 – Usuários de internet no Brasil

C2A - USUÁRIOS DE INTERNET - INDICADOR AMPLIADO

Total da população

Percentual%		Sim	Não	Não sabe	Não respondeu
Total		87	13	-	-
Área	Urbana	88	12	-	-
	Rural	81	19	-	-
Região	Sudeste	87	13	-	-
	Nordeste	85	15	-	-
	Sul	86	14	-	-
	Norte	94	6	-	-
	Centro-oeste	93	7	-	-

Fonte: <https://cetic.br/pt/tics/domicilios/2020/individuos/C2A/>

Ao contrário do que se pode imaginar, entretanto, não é exatamente a produção explícita de dados que fortalece o paradigma dos *big data*, mas justamente o dado que é criado de forma imperceptível. Uma inocente busca por livro em um *site* de *e-commerce* é capaz de gerar um conjunto de *logs* indicativos de todo o caminho percorrido pelo usuário até a finalização de sua navegação, seja ela com a execução de uma compra, um encaminhamento de compra e desistência, ou ainda um simples abandono do ambiente por não ter encontrado o que queria ou condições que gostaria para obter o livro. Esses *logs* gerados constituem uma massa de dados que é construída de forma contínua, e à revelia da vontade dos usuários.

Um passeio ou o simples ato de ir ao trabalho com a companhia do aparelho celular é capaz de gerar uma infinidade de dados.

Registra-se o deslocamento e uma outra infinidade de ações que o ser humano nem imagina por meio de todos os sensores que compõem os *smartphones*. Já parou para pensar sobre a quantidade de sensores que estão embutidos em seu aparelho celular e a quantidade de ações que ele é capaz de compreender e transformar em dados? Já tentou ler aquelas letrinhas que dizem que algumas das suas ações podem ser informadas para a fabricante do celular, de forma que ela possa “melhorar sua experiência” com o aparelho?

Além disso, há ainda todo o conjunto dos novos recursos tecnológicos que geram dados explicitamente, mas num contexto profissional, como por exemplo: captações de dados de satélites; dados gerados na área da saúde; dados gerados por sensores em esportistas, animais, automóveis e aviões, entre outros; e transações comerciais de vários tipos. O mundo passou a armazenar o fato em forma de dado, e tornou esse processo um elemento de relevância para os novos passos da humanidade.

O paradigma dos *big data* é caracterizado por tudo o que já foi apresentado, mas existe uma forma bem simples de entender quando os dados deixam de ser algo do seu cotidiano e passam a ser efetivamente *big data*. É exatamente quando os dados disponíveis já não podem mais ser analisados ou processados pelos meios tradicionais de acesso a dados com que nos acostumamos nos últimos anos. É nesse momento que se experimenta os *big data*, quando já não é mais possível tratar os dados de forma amadora, ou com aquele *software* de planilha com o qual se está tão habituado.

O conceito de *big data* relaciona-se à heterogeneidade, à rápida geração e processamento e à grande quantidade de dados disponíveis digitalmente. Esse fenômeno despontou de forma mais significativa a partir do início do século XXI, quando a quantidade de dispositivos e de usuários conectados apresentou um crescimento exponencial (Mayer-Schömbberger; Cukier, 2013).

Big data é o conceito utilizado para descrever um paradigma com características não suportadas por estruturas tradicionais e por servidores e *softwares* utilizados na maioria das organizações atualmente. Refere-se aos grandes volumes de dados, que podem ser medidos em *terabytes*, *petabytes*, *exabytes* ou *zettabytes*. Estes dados são ativos organizacionais que podem ser transformados em informações úteis (Akerkar, 2014).

A revolução dos *big data* promete transformar a forma como vivemos, trabalhamos e pensamos, permitindo a otimização do processo, capacitando a descoberta de *insights* e melhorando a tomada de decisões. A realização desse grande potencial depende da capacidade de extrair valor desses dados massivos por meio da análise de dados (L'Hereux; Grolinger; Yamany; Capretz, 2017).

O principal objetivo desta revolução é fazer uso de grandes quantidades de dados para permitir a construção de conhecimentos e a melhor tomada de decisões (Mayer-Schönberger; Cukier, 2013).

Nota-se, categoricamente, que os *big data* são tidos como um paradigma em evolução e também como o insumo da vida moderna. Não por menos, Clive Humby, um matemático britânico, afirmou que “os dados são o novo petróleo”. A afirmação de Humby é facilmente comprovada ao se observar a ascensão de

valor das gigantes empresas de tecnologia americanas, superando as petrolíferas, os bancos e as empresas que foram tradicionais, por décadas passadas, em valor de mercado.

Pode-se também perceber os dados como insumos importantes quando se observa o famoso caso da Cambridge Analytica no escândalo do Facebook nas eleições dos Estados Unidos da América em 2016, e todas as consequências geradas por tal fato, como a criação da GPDR (Regulamento Geral sobre a Proteção de Dados) no cenário europeu, em 2018, e, de forma indireta, a LGPD (Lei Geral de Proteção de Dados Pessoais) no Brasil, também em 2018.

Os *big data* são uma realidade, e existe uma compreensão clara de que gerar dados e posteriormente utilizá-los para determinar ações não é mais novidade, mas uma necessária condução para o desenvolvimento humano. Assim, este texto debruça-se sobre duas questões: **Tomar decisões baseadas em dados é algo tão trivial?** e **Quais as consequências que o uso de dados tem gerado e pode gerar para a humanidade?** O objetivo é discutir os desafios e as consequências para um mundo que passou a valorizar a evidência dos dados.

O uso de dados com apelo para melhoria da vida do ser humano tem sido pauta em estudos científicos e em vários cenários de discussão nos campos empresarial, cultural e esportivo, entre outros. Ficam cada vez mais fortes o sentimento e a clareza do quarto paradigma da ciência, conforme defendia Jim Gray em seu livro *The fourth paradigm: data-intensive scientific discovery*, a *e-science*, que é o nome dado a estudos científicos baseados em dados, mas que

em vários momentos se funde com a ideia de *data-driven*, ou a cultura da utilização de dados para se tomar decisão em qualquer campo, não apenas no científico.

Em alguns contextos é possível compreender que os conceitos de *e-science*, *data science* e *data-driven* podem se confundir entre si, e realmente em determinadas situações eles são semelhantes. A *e-science* é a aplicação da cultura de *data-driven* por meio de um método científico, ou seja, é fazer ciência utilizando dados.

Ao se pensar sobre ciência, considera-se partir do contexto do senso comum, do conhecimento primário e do dia-a-dia, aquele que acompanha as tradições e os sentidos, baseando-se no contato diário do ser humano com o ambiente em que vive. De certa forma, o sol nascer sempre do mesmo lado, ainda que esse fato não seja registrado como dado, faz com que o homem aprenda qual é a trajetória da grande estrela.

O senso comum responde a várias questões do cotidiano e, em alguns casos, relaciona-se com o conhecimento advindo da religião. O homem compreende, no entanto, que é necessário determinar sua vida com algo além das sensações. Constitui, para tanto, o conhecimento aprofundado ou secundário, aquele que se dá por meio de observações sistemáticas, mais detalhadas e com mais dimensões, em busca de características incomuns.

O conhecimento é uma eterna busca, fundamentada pela geração de hipóteses e argumentos; a partir dessa busca, gera incertezas sobre o que é aceito e comprova o que não é aceito, consolidando a capacidade do ser humano de evoluir com o tempo. A ciência nasce da investigação pelo conhecimento avançado, entretanto

a busca por esse conhecimento deve ser realizada por meio de práticas sistemáticas, ou seja, o método científico.

A ciência baseada em dados, *e-science*, é então possível? Segundo Jim Gray, sim, concebendo um paradigma evolutivo a partir dos paradigmas anteriores, caracterizados pela ciência empírica, ciência teórica e ciência computacional.

Thomas Kuhn já ressaltava, em 1962, ao abordar a estrutura das revoluções científicas: “Quando um número suficiente de anomalias significativas se acumula contra um paradigma atual, a disciplina científica é lançada em um estado de crise. Durante a crise, novas ideias, talvez anteriormente descartadas, são experimentadas”. É possível, por meio dos resultados, compreender que estudos sobre dados têm tido alto impacto, com resultados surpreendentes sobre os paradigmas anteriores.

A *e-science* avança com o desenvolvimento de métodos e algoritmos para analisar grandes massas de dados, gerando conhecimento, encontrando padrões e realizando inferências que até então os paradigmas de pesquisa anteriores não tinham sido capazes de alcançar. De certa forma, seguindo conjunto de regras que caracterizam o método científico para produzir conhecimento, resultados de pesquisa baseados em dados fortalecem a criação de novos conhecimentos, coibindo a subjetividade.

Discorrer sobre senso comum, conhecimento aprofundado e ciência é sempre necessário a partir do momento em que se pretende compreender estudos e uso de dados, tanto como ciência quanto em qualquer outro contexto.

DESAFIOS E CONSEQUÊNCIAS DO USO DE DADOS

Se os *big data* são um forte aliado e insumo impulsionador, o método científico é uma regra importante; e se os resultados com estudos baseados em dados são surpreendentes, é possível compreender que o cenário está totalmente perfeito para o uso de dados?

Entende-se que utilizar dados para realizar análises e, a partir delas, tomar decisões importantes, seja no campo científico ou em outro qualquer, pode gerar consequências ainda não conhecidas e, portanto, compreende-se que há principalmente dois grandes desafios para o uso de dados, com suas devidas consequências, também importantes, que podem ser delimitados por duas grandes fases:

- a preparação e compreensão dos dados e
- a visualização dos dados analisados.

Entende-se toda a fase nomeada aqui de preparação e compreensão dos dados como um conjunto de processos que se inicia antes da coleta dos dados até o momento em que eles passarão a ser processados com algoritmos e modelos matemáticos/probabilísticos.

Em que medida é necessário pensar nos dados antes serem eles coletados? Porque o processo de coleta, que tem como princípio constituir um “lago” de dados oriundos das mais diversas fontes, com os mais variados formatos e em grandes volumes, esbarra em problemas que ocorreram na persistência dos dados.

Quando um dado é gravado, seja em uma planilha, seja em tabela de um banco de dados, seja um arquivo de texto ou uma ima-

gem, em muitos desses processos não houve preocupação em constituir metadados para esses dados, ou seja, mesmo com dados estruturados torna-se difícil de compreender o que os dados significam. Além de metadados, grande parte desses dados não tem nenhum tipo de estrutura semântica, não são acompanhados por licenças de uso e em várias situações não há nenhum tipo de registro oficial de proveniência.

O dado, antes mesmo de ser coletado, já apresenta um conjunto de problemas que podem impactar diretamente no resultado das análises. Não haver semântica nos dados pode levar a resultados que tenham um significado diferente do que os dados originais registravam.

Não haver uma licença de uso e garantia de proveniência imputa insegurança jurídica e insegurança quanto a resultados que podem não ter sido gerados a partir de processos lícitos ou que seguiram critérios adequados de qualidade.

Elementos como a não identificação do período em que os dados foram coletados, a indicação de quando voltarão a ser, com informação sobre a frequência em que isso irá ocorrer, podem ferir a ideia de se empreender esforços na análise de dados que nunca mais poderão ser expandidos.

Os desafios seguem na medida em que os dados passam a ser coletados; é muito comum encontrar inúmeros outros problemas com os dados, como: dados desbalanceados, dados incompletos, dados inconsistentes, dados redundantes, dados com ruídos.

Os dados desbalanceados ocorrem quando o número de registros varia para as diferentes classes. Isso é comum em aplicações em que dados de um subconjunto das classes aparecem com uma frequência maior que os dados das demais classes (Carvalho, 2021).

Os dados incompletos são identificados quando há registros que não contêm dados em determinados atributos. Essa ausência de valor muitas vezes pode ter sido gerada por uma falha em um equipamento de coleta, por digitação realizada por uma pessoa, por falta de obrigatoriedade na geração do dado ou ainda por desconhecimento do valor pelo gerador dos dados (Carvalho, 2021).

Dados inconsistentes são aqueles que possuem valores conflitantes em seus atributos. Essa inconsistência pode se dar entre valores de atributos de entrada (por exemplo, valor 120 para o atributo Peso e valor 3 para o atributo Idade) ou entre todos os valores dos atributos de entrada e o valor do atributo de saída (por exemplo, dois pacientes com os mesmos valores para os atributos de entrada e diagnósticos diferentes, um saudável e o outro doente). Dados inconsistentes são muitas vezes produzidos no processo de integração de dados. Por exemplo, diferentes conjuntos de dados podem usar escalas diferentes para uma mesma medida (metros e centímetros) ou codificar de forma diferente um atributo relacionado com o tamanho (Carvalho, 2021).

Um conjunto de dados pode possuir tanto objetos como atributos redundantes. Um objeto é redundante quando ele é muito semelhante a um outro objeto do mesmo conjunto de dados, ou seja, seus atributos possuem valores muito semelhantes aos atributos de pelo menos um outro objeto. No caso extremo, possui o mes-

mo valor para cada um dos atributos. Um atributo é redundante quando seu valor para todos os registros pode ser deduzido a partir do valor de um ou mais atributos (Carvalho, 2021).

Dados com ruídos são dados que contêm registros que, aparentemente, não pertencem à distribuição que gerou os dados analisados. Ruído pode ser definido como uma variância ou erro aleatório no valor gerado ou medido para um atributo (Han e Kamber, 2000). Um indicador da possível presença de ruído é a existência de **outliers**, que são valores que estão além dos limites aceitáveis ou são muito diferentes dos demais valores observados para o mesmo atributo, representando, por exemplo, exceções raramente vistas (Carvalho, 2021).

Todos esses problemas relatados atualmente representam um grande desafio para os estudos sobre dados, não apenas por gerar insegurança sobre os resultados, mas muitas vezes por não permitir que resultados relevantes sejam alcançados. É fato que a insatisfação com resultados e a dificuldade em alcançá-los tornam-se importantes, mas é alto o custo associado a todos os procedimentos que procuram mitigar os problemas nos dados.

Constituir um ambiente computacional que seja capaz de processar grandes volumes de dados demanda recursos financeiros de alto valor. O alto custo de processamento e armazenamento dos dados muitas vezes é impeditivo para que um estudo possa ser realizado. Ademais, os casos em que os problemas com dados citados anteriormente se acumulam tendem a gerar necessidade de profissionais mais capacitados para lidar com os problemas,

aumentando ainda mais os custos de um projeto de análise de dados. Esses são desafios que colocam em cheque vários estudos de dados.

Outro fator muito importante na preparação dos dados é a necessidade de se observar a relevância dos dados em sua origem. Somente profissionais muito capacitados na área de domínio dos dados são capazes de compreender se as fontes de dados não estão gerando processos discriminatórios ou enviesados já na seleção das informações. Esse é um dos problemas relevantes que podem representar impactos em tomadas de decisões.

Coneglian, Santarem Segundo e Sant'ana (2017) abordam essa questão, dando ênfase a dados de *big data*:

Questões mais graves como a indução de resultados de pesquisas científicas e modelos discriminatórios de buscadores de buscas, provam que os big data não só mantêm comportamentos discriminatórios, como também os ampliam. Torna-se necessário colocar a ciência para debater meios de utilizar tais dados, na busca de não se enviesar os resultados com a utilização de perguntas que induzem a respostas equivocadas.

É fato que os estudos sobre dados precisam de muitos cuidados para que possam realmente gerar valor e conhecimento para a sociedade.

Outro fator muito importante e desafiador para os estudos sobre dados é a necessidade de apresentá-los da melhor forma possível a quem vai consumir esse conteúdo.

A visualização é um dos pontos mais importantes de todo o pro-

cesso; portanto, é fundamental entregar a análise de forma clara para quem vai consumir e tomar decisões. Se uma pesquisa científica atingiu resultados importantes sobre um determinado medicamento, a visualização precisa deixar muito claro como esses resultados foram obtidos, de maneira que não haja dúvidas para quem está consumindo esse conhecimento.

Como os gráficos são muito mais fáceis de compreender do que textos e tabelas, é importante que haja esse tipo de possibilidade de visualização dos dados.

Segundo Aguilar (2017), “visualização de dados, informação e conhecimento são a representação por meio de imagens, gráficos ou cartografias, em que o foco central é a simplificação de conteúdos para o entendimento da ideia geral, facilitando dessa forma sua percepção”.

A visualização está diretamente relacionada à capacidade humana de enxergar conteúdos e categorizar cenas naturais. O ser humano pode processar 250 megapixels, tomar uma decisão a partir de imagem que nunca foi vista antes e até categorizar, em 150 milésimos de segundos. Somos capazes de assimilar quase que instantaneamente e sem esforço uma grande quantidade de dados, informações e conhecimento graficamente (Aguilar, 2017).

A visualização tem, portanto, um papel fundamental, porque ela finaliza o processo de análise de dados, e dessa forma, caso não seja adequada, pode gerar resultados ruins para bons processos de análise.

Não é somente a parte técnica da visualização dos dados que pode interferir nos resultados desejados; a visualização pode vir

enviesada, tendenciosa, manipulada, como resultado (que pode ter sido iniciado nas fases de preparação e análise) de fraude, falta de conhecimento técnico dos analistas e até da própria falta de conhecimento sobre o domínio em que está sendo realizada a análise.

Identificar se uma visualização mal apresentada ou induzida é um problema de dados, de análise, da fonte ou da intenção de quem quer demonstrar os resultados é um dos desafios mais latentes de tudo que envolve o uso de dados.

Por meio das Figura 1 e 2, é possível verificar dois gráficos que têm exatamente os mesmos dados; são gráficos que partiram da mesma análise, tecnicamente não há nada de errado com os gráficos, mas eles têm impacto diferente para quem está visualizando. Qual a real intenção de cada um dos gráficos? É possível identificar se houve um erro ou se houve uma intenção, ou ainda um problema que pode caracterizar-se até como ético? A simples mudança de escala no gráfico gera uma sensação completamente diferente para quem está consumindo a informação.

Figura 1 – Comissão de Venda – Opção 1

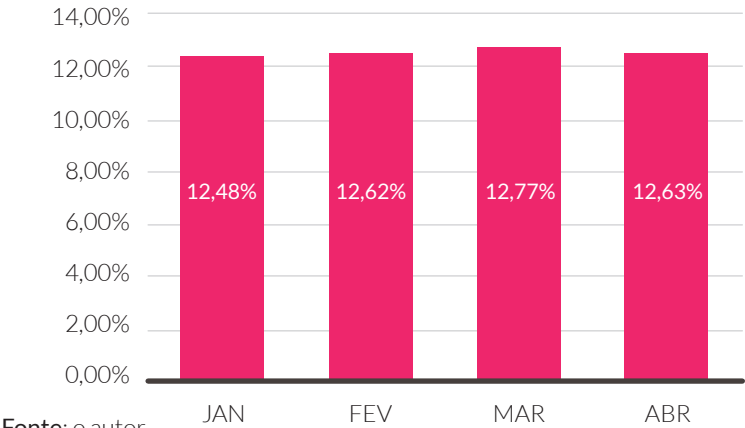
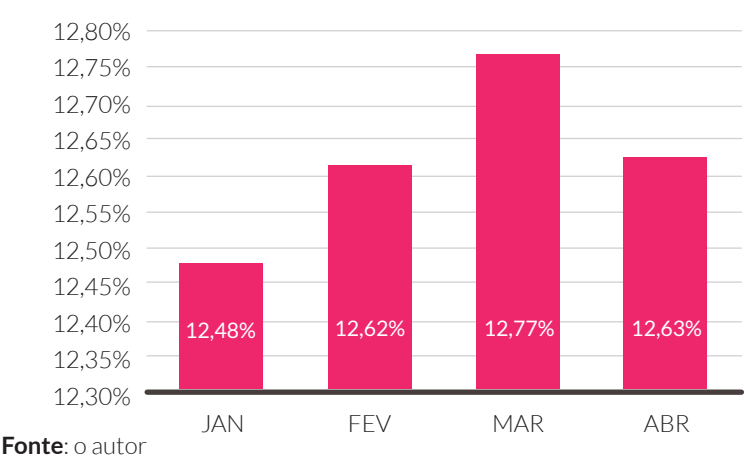


Figura 2 – Comissão de Venda – Opção 2



A Figura 3 nos apresenta um problema diferente no processo de visualização de um gráfico, visto que há claramente erros de esca-

la e até de plotagem dos pontos relevantes. Entre os erros, observa-se que os dois últimos pontos caracterizam um problema grave na demonstração dos dados. Como identificar se há um problema técnico (um erro) ou uma necessidade de indução a quem está visualizando a informação? O site WTF Visualizations (<https://viz.wtf/>) acumula imagens com visualizações inadequadas para os mais diversos tipos de gráficos e infografias.

Figura 3 – Contágios Cali



Fonte: <https://viz.wtf/image/625814555651653632>

A visualização tem se caracterizado como um grande desafio em estudos baseados em dados, com consequências que podem determinar, em muitos casos, a insegurança sobre a confiabilidade dos estudos. Daí a necessidade de desenvolver bons métodos de visualização, um processo fundamental para melhorar a boa percepção dos estudos sobre dados.

CONSIDERAÇÕES FINAIS

A ideia apresentada aqui nasce dos desafios encontrados quando os estudos sobre dados passam a fazer parte do cotidiano das pessoas, mesmo que elas não percebam isso acontecendo. O fato de os *big data* serem uma realidade não permite ter a certeza de que tudo daqui para a frente será baseado em estudos sobre dados, principalmente pelos inúmeros desafios que se têm pela frente para que a cultura de *data-driven* realmente seja incorporada nas organizações e nas pesquisas científicas, com a minimização dos problemas que foram aqui citados. Responde-se aqui a uma das perguntas apresentadas neste texto, onde fica claro que tomar decisões baseadas em dados não é algo trivial, mas carregado de desafios que ainda precisam ser resolvidos.

Ressalta-se que as empresas norte-americanas de tecnologia, como Google, Facebook, Microsoft e Amazon, que pautam e têm como insumo os dados de seus clientes, já superaram muitos dos problemas aqui apresentados, visto que em muitos casos são as geradoras dos próprios dados que processam para suas análises.

Os desafios estão apresentados. Eles têm início antes da coleta de dados e vão se multiplicando até a parte final dos estudos: a visualização dos dados.

Como em todos os paradigmas de pesquisa e todos os processos além da ciência, o estudo de dados não foge à regra; há erros básicos que precisam ser mitigados, mas há incoerência, há tendenciosidade, há discriminação, há falta de recursos, há erros técnicos e há também um alto custo envolvido em qualquer processo de análise de dados.

É fato que todos nós já somos conduzidos por resultados de algoritmos que analisam dados, e é importante reconhecer a necessidade de evolução da pesquisa e direcionar esforços para que os desafios possam ser superados. O uso de dados é um caminho sem volta. Resta saber agir da melhor forma para que a humanidade possa desfrutar disso e evoluir.

É tempo de responder a nossa segunda pergunta, que envolve as consequências que os dados têm gerado e podem gerar para a humanidade. Processar dados e constituir conhecimento baseado em evidências será, de fato, muito representativo para a evolução da humanidade; mas enquanto essa possibilidade for restrita a pequenos grupos de grandes empresas, haverá dificuldade para se compreender como evoluiremos com esses processos. É necessário pautar os desafios e desfrutar de forma límpida do oceano de dados de que dispomos.

REFERÊNCIAS

AGUILAR, A. G. *et al.* **Visualização de dados, informação e conhecimento.** Florianópolis: EdUFSC, 2017.

AKERKAR, R. (Ed.). **Big Data Computing.** Florida: CRC Press, 2019.

BELL, G. Foreword. In: HEY, T.; STWART, T.; TOLLE, K. (Eds.). **Fourth Paradigm: data-intensive scientific discovery.** Washington: Microsoft Research, 2009. Disponível em: <<http://research.microsoft.com/en-us/collaboration/fourthparadigm/>>. Acesso em: 15 abr. 2020.

FACELI, K. *et al.* **Inteligência artificial: uma abordagem de aprendizado de máquina.** 2. ed. São Paulo: LTC, 2021.

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

HAN, J.; PEI, J.; KAMBER, M. **Data mining: concepts and techniques.** Massachusetts: Morgan Kaufmann Publishers, 2000.

KUHN, T. **A estrutura das revoluções científicas.** Tradução de Beatriz Vianna Boeira e Nelson Boeira. 10 ed. São Paulo: Perspectiva, 2011.

L'HEUREUX, A. *et al.* Machine learning with Big Data: challenges and approaches. **IEEE Access**, v. 5, p. 7776-7797. 2017.

MAYER-SCHÖNBERGER, V.; CUKIER, K. **Big data**: a revolution that will transform how we live, work, and think. Boston: Houghton Mifflin Harcourt, 2013.

NÚCLEO DE INFORMAÇÃO E COORDENAÇÃO DO PONTO BR. **Pesquisa sobre o uso das tecnologias de informação e comunicação nos domicílios brasileiros**: TIC Domicílios. São Paulo: Comitê Gestor da Internet no Brasil, 2020. Disponível em: <<https://cetic.br/pt/tics/domicilios/2020/individuos/C2A/>>. Acesso em: 10 set. 2021.

A TRANSDUÇÃO NOS PROCESSOS DE MEDIAÇÃO E A INFORMAÇÃO BIODISPONÍVEL

Ricardo César Gonçalves Sant'Ana

Quando pensamos em nossa capacidade de realização, incorporamos, de forma quase intuitiva, elementos externos ao nosso corpo nos resultados possíveis, sem nos dar conta de que, na verdade, estamos ampliando nossas possibilidades por meios artificiais. Se não sou capaz de mover algo pesado, posso usar uma alavanca, uma roldana e outras extensões de meus membros motores, ampliando minha potência de agir artificialmente. Nossas competências sempre foram passíveis de ampliações como essa. Fizemos marcas em ossos e aumentamos nossa capacidade de lembrar informações, ampliando nossa memória. Transformamos materiais transparentes em lentes, ampliamos nossa visão e descobrimos o cosmos. As tecnologias nos acompanham enquanto humanos desde sempre.

Começo essa nossa reflexão propondo o porquê de algumas pessoas entenderem tão facilmente as tecnologias e outras demonstrarem tanta dificuldade, ou até mesmo verdadeira resistência em aceitá-las. Imagine a assimetria quando escolhidos eram capazes de identificar o momento certo para plantar e até mesmo como o clima se comportaria nos próximos dias ou semanas simplesmente por meio de observações feitas a partir de monólitos ou outros artefatos cuja interpretação se configurava como um ativo muito importante. O mesmo podemos pensar sobre artefatos como ábacos, que nas mãos de escolhidos permitiam a realização de cálculos com rapidez impressionante. Estas alavancas mentais compunham assim um novo ser, amálgama de funções primárias com estes recursos artificiais: um processo de hominização que pode ser contado em termos de milhares de anos.

Autores como Michel Serres (2003) nos fazem refletir sobre como incorporamos funcionalidades executadas externamente, substituindo inclusive nossas faculdades, na composição de um novo ser. No caso do acesso a dados a questão se acentua. Não temos os recursos necessários para tratar grandes volumes de dados, o que exige a participação de um tratamento maquínico para que tenhamos informações efetivamente disponíveis ao nosso ser primário, informações biodisponíveis. Este novo ser não suscita alteridade e nos deixa pouca margem de escolha no caminho entre a incoatividade da hominescência e uma condição de coatividade.

Muitas questões ontológicas emergem desse processo de composição de novas identidades, tais como as relacionadas ao paradoxo do barco de Teseu que, no entanto, escapam ao nosso es-

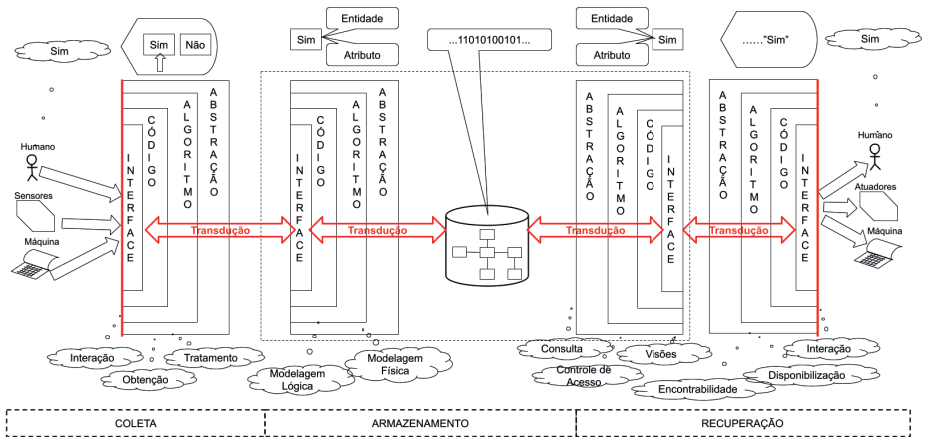
copo no momento. Nessa reflexão vamos buscar elementos sobre a mediação das tecnologias no acesso a dados e sobre nossa insciência do pouco controle que temos dessa parte importante do novo ser que agora incorpora competências digitais como suas.

O ACESSO A DADOS E A TRANSDUÇÃO INFORMACIONAL

Assim como no caso da ampliação de nossa memória por meio de registro de informações nos mais diferentes suportes, e da ampliação de nossa capacidade de cálculo por meio de artefatos como o ábaco, quando tratamos com dados nossa dependência de recursos tecnológicos cresce exponencialmente em função do volume. Nossas tabelas cresceram em tamanho e detalhes, ultrapassando, em muito, as limitações dos suportes. Já não basta registrar a informação: precisamos de alavancas que nos ajudem no processamento desses conteúdos; precisamos de máquinas para localizar, ordenar, filtrar, agrupar, calcular, transformando as grandes quantidades de dados em sínteses com que nossa cognição possa lidar, da mesma forma que as roldanas reduzem o peso das cargas, ampliando a força de nossos braços.

No caso do tratamento de dados, no entanto, a interação não se limita a mudanças de direção ou sentido. Estamos falando agora de trocas informacionais que requerem transformações de forma, estrutura, codificação e energia, transduções informacionais em sucessivas camadas (figura 1), exigindo funcionalidades de interface cada vez mais eficientes.

Figura 1: Transdução Informacional



Fonte: Sant'Ana, 2019.

Não se trata mais de dominar técnicas de registro em um suporte com marcações específicas que, quando observadas por outros, podem ser diretamente identificadas e interpretadas. Não basta mais a alfabetização: os conteúdos precisam ser registrados e disponibilizados por máquinas que funcionem com outras formas de energia e codificação.

Essas novas alavancas precisam coletar os dados, e assim nos orientam sobre como vamos fornecê-los, apresentando caminhos específicos em suas interfaces conosco. Podemos, por exemplo, pensar em nosso processo de interação com a máquina quando desejamos realizar a entrada de um determinado dado (o *input*). Quando temos em mente o que desejamos informar, transformamos esta informação em energia mecânica em nossos músculos,

traduzindo a informação na sequência de teclas que a máquina oferece. E este acionamento mecânico, por sua vez, é transformado em energia elétrica, que é enviada a dispositivos que vão realizar outras transformações para que a aplicação (algoritmos) possa novamente codificar esta informação de acordo com o que foi previamente definido. É bom lembrar das crescentes alternativas de coleta, dos novos dispositivos e sensores que ampliam a capacidade de coleta maquínica. Mas o mais importante para nossa reflexão é que, uma vez coletados, esses dados vão passar por sucessivas transformações sobre as quais somos totalmente inscientes enquanto usuários (Affonso & Sant'Ana, 2018).

Devidamente coletados, tais dados sofrerão novas (e muitas) transformações para que possam ser registrados em suportes digitais, cada qual com suas características - ópticas, magnéticas ou elétricas -, tudo devidamente organizado, etiquetado e endereçado. Retornando mais uma vez a nossa reflexão, ter um CD ou um *pen drive* nas mãos não nos permite ir além de alguma identificação visual de seu exterior. Tudo o que está ali registrado depende de intermediação maquínica. Mesmo assim temos a sensação de controle sobre este conteúdo, na medida em que assumimos a alavanca (o dispositivo digital) como parte do nosso novo ser.

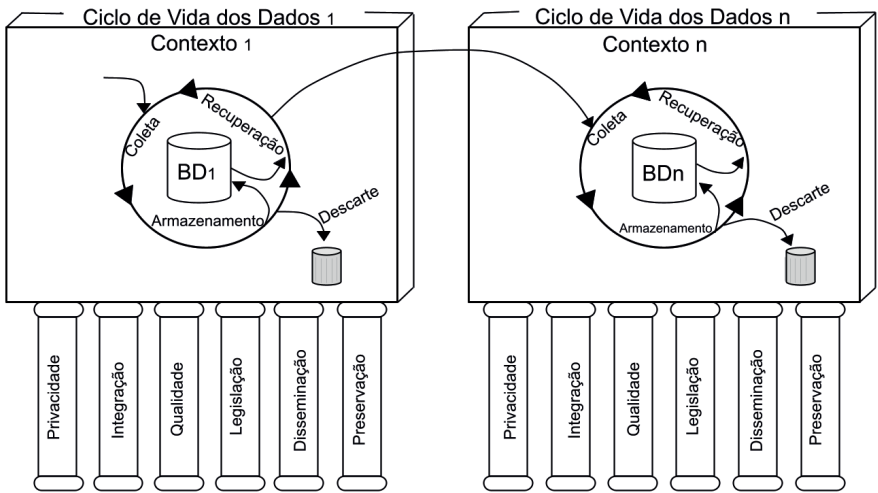
Nossos recursos nativos não são suficientes para coletar dados diretamente de suportes digitais. Depois de coletados e armazenados pela máquina, esses conteúdos precisam ser disponibilizados e, mais uma vez, uma série de transformações (transduções informacionais) é necessária para que todo aquele conjunto de informações seja apresentado, adequado a determinadas neces-

sidades e, antes de tudo, sintetizado de forma que possa, agora, ser interpretado por capacidades humanas, na condição de informações biodisponíveis.

CICLO DE VIDA DOS DADOS

Quando pensamos nos diferentes momentos envolvidos nesse processo, podemos identificar pelo menos três fases distintas: coleta, armazenamento e recuperação, podendo ainda ser estabelecido um momento em que os dados serão descartados do suporte. Com estas quatro fases é possível então desenhar um ciclo de vida dos dados (figura 2), que sinaliza ainda ações que estarão presentes, em todas as fases, de forma transversal: privacidade, integração, qualidade, legislação, disseminação e preservação (Sant'Ana, 2016).

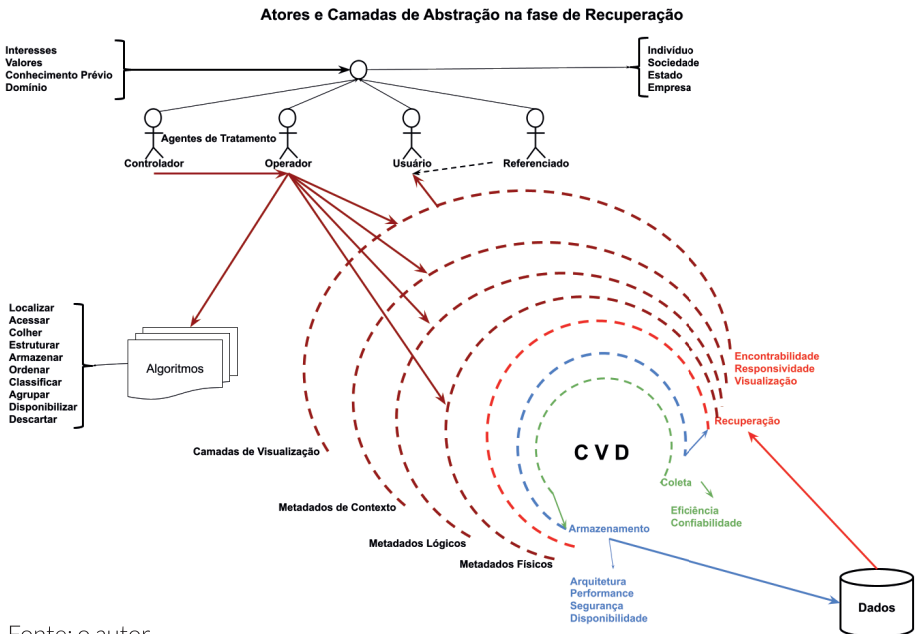
Figura 2: Ciclo de vida dos dados



Fonte: Sant'Ana, 2016.

Os ciclos se conectam por meio das fases de coleta e recuperação, propiciando a troca de informações entre eles. Tais informações são dependentes de transduções, o que não significa que não sejam definidas, modeladas e desenvolvidas por ação humana direta (pelo menos por enquanto). Assim, essas trocas envolvem atores com distintos potenciais de interação nas diferentes camadas de abstração que emergem das múltiplas transduções informacionais (figura 3), criando assimetrias no controle ou na percepção do processo como um todo.

Figura 3: Atores e camadas de abstração na fase de recuperação



Fonte: o autor.

Quando colocamos em perspectiva as fases do ciclo de vida dos dados e os atores diretamente envolvidos, percebemos que aqueles que detêm a posse dos recursos envolvidos no processo de acesso aos dados, e que em contextos como o do direito (BRASIL, 2018) recebem o nome de “controladores”, são os responsáveis pelas decisões referentes ao tratamento que será realizado nos dados.

Já as ações mais diretas, como o projeto, o desenvolvimento, a implantação e a manutenção das aplicações (e, por consequência, dos algoritmos), requerem competências técnicas específicas e configuram um outro ator, um intermediário que responde pelas

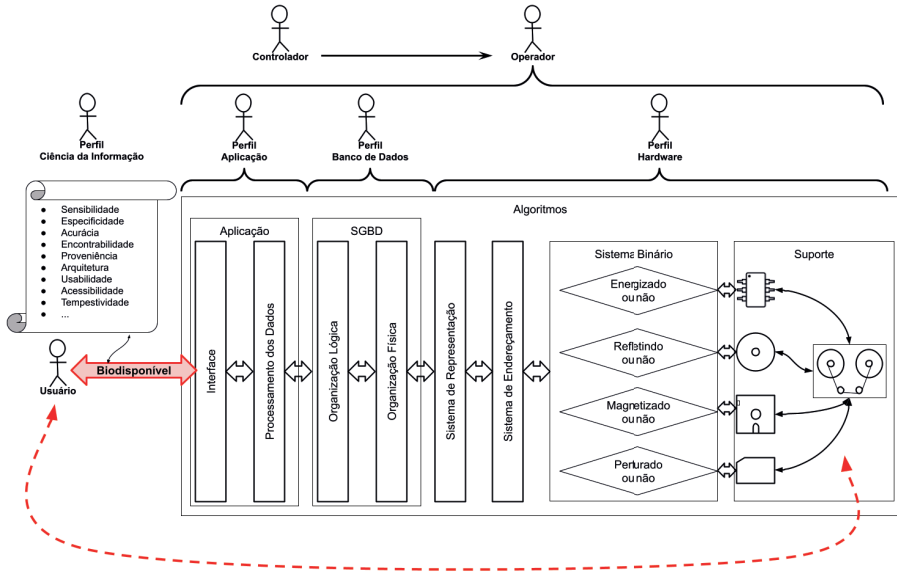
definições estabelecidas pelo controlador e que é denominado pela legislação como “operador” (Brasil, 2018). Percebe-se, ao olhar a figura 3, que é ele que tem acesso direto às camadas de abstração que são resultado do desdobramento da fase de recuperação. Trata-se de um perfil mais voltado para as necessidades de adequação da tecnologia às necessidades do controlador.

Para esta reflexão, consideramos como ator de maior relevância aquele que utiliza a aplicação em função de determinada necessidade informacional e, por isso, o identificamos, de forma mais genérica, como usuário. Quando levamos em conta o conteúdo dos dados que estão sendo tratados, no entanto, temos que considerar também aqueles a quem os dados dizem respeito, que não representam, necessariamente, o usuário. A este ator, com o qual os dados de alguma forma se relacionam, a legislação denomina “titular” (Brasil, 2018), por ser aquele a quem os dados em análise se referem. Na figura vemos, assim, ao invés do titular, a presença do usuário e do referenciado; mas cabe lembrar que os dois podem ser, eventualmente, a mesma pessoa.

Todos esses atores podem ser entendidos como pessoa natural ou jurídica, de direito público ou privado, mantendo assim a questão da identidade do ator mais aberta que aquela de cenários destinados apenas a dados pessoais.

Quando resgatamos, portanto, os elementos envolvidos no conjunto de transduções informacionais necessários para o acesso a dados, podemos perceber que o ator denominado operador tenderá a ter perfis variados, dependendo da posição que ocupará no caminho entre o usuário e o dado (figura 4).

Figura 4: Atores e informação biodisponível



Fonte: o autor.

Quanto mais próximo do suporte físico, mais próximo estará de competências relacionadas à gestão e operacionalização da parte física das máquinas, o que nos permite identificar um operador com perfil de *hardware*. Muitas vezes ele desconhece até mesmo qual será o uso final do dispositivo, considerando apenas que operações básicas deverão ser realizadas de forma eficiente e segura. Essa questão pode parecer trivial, mas aponta para um elemento novo nas tecnologias, ao romper com o modelo precedente em que “o pragmatismo do Homo faber, que constrói instrumentos para sua utilidade e máquinas para alguma finalidade” é substituído por máquinas “diferenciadas e universais [...] que transferem o

projeto de utilidade do construtor para o utilizador que as emprega para o lazer ou para o que melhor lhe pareça. Quem as constrói e produz não pode prever para quem ou para quê elas se tornarão úteis. Elas não possuem finalidade direta. Sem intencionalidade inicial, sua função é descoberta posteriormente.” (Serres, 2003, p. 75)

Em seu nível mais primário, o suporte digital será organizado sob a forma de pequenas chaves que, em sua essência, determinam duas situações possíveis: ligado ou desligado. Ou seja, um sistema binário que irá representar, por meio de cada uma destas chaves, um *bit*, o átomo da informação, e que, dependendo do suporte, irá requerer transformações específicas de energia - transduções informacionais.

Mesmo quando dependíamos de suportes mais básicos, como os cartões perfurados, dificilmente alguém, sem conhecimento mais aprofundado sobre como as posições dos furos se estruturavam, seria capaz de interpretar seu conteúdo. No entanto, tínhamos ao menos a possibilidade de visualizar os furos nos cartões, furos estes que seriam transformados em chaves liga/desliga quando lidos pelo dispositivo. Essa ainda que aparente proximidade com nossa capacidade de interpretá-los cobrava um preço alto, em termos de capacidade de armazenamento e também de registro e leitura dos seus conteúdos.

O salto tecnológico, ao adotarmos os suportes com capacidades magnéticas (discos e fitas), permitiu que as mesmas chaves fossem agora definidas em áreas muito menores dos suportes, o que ampliou, e muito, sua capacidade de armazenamento. Já não éramos capazes de perceber, sem o uso de recursos maquínicos,

onde e como os suportes estavam magnetizados, mas ganhamos também em velocidade de acesso a esses conteúdos.

Explorando a possibilidade de alguns materiais refletirem ou não a luz, passamos a adotar também suportes que, por meio de mais uma transdução informacional, longe da nossa percepção puramente humana, armazenavam e permitiam a recuperação de volumes ainda maiores e com alta velocidade de acesso.

O próximo passo seria ainda mais surpreendente: quando passamos a utilizar circuitos digitais com capacidade de manter a memória de um determinado estado, energizado ou não, mesmo quando não alimentados por energia. Os limites de volume e velocidade foram todos superados, com grandes possibilidades ainda por vir.

Retomando nossas reflexões, mesmo que tenhamos acesso direto a tais suportes, não temos recursos nativos suficientes para realizar a leitura dos conteúdos registrados. Mesmo com a percepção de que os controlamos, dependemos de milhões de linhas de código, conjuntos de instruções que podemos denominar de algoritmos, que determinam como as máquinas vão realizar a tarefa de registro e, depois, de leitura.

Seria completamente inviável tornar necessária a compreensão de como essas instruções são definidas ou mesmo como elas são executadas, e isso ocorre em nossa relação com todas as tecnologias: “quanto mais a mecânica e a termodinâmica nos introduziam a um conhecimento mais preciso das técnicas duras, de suas leis, constantes de energia e rendimento dos motores, mais amplamente ignorávamos as leis da mecânica” (Serres, 2003, p. 193).

Esse grande conjunto de chaves binárias registradas nos suportes ainda apresentam outros pontos de distanciamento com nossos recursos nativos de leitura do mundo, a exemplo da forma com que estes conteúdos binários são organizados para que possamos localizá-los. Estes dados sobre os dados, contendo os respectivos endereçamentos, também são feitos sob a forma de conteúdos registrados nos próprios suportes; são complexos e precisam de formas próprias para que possam ser interpretados.

Devidamente registradas, rotuladas e com seus endereçamentos estabelecidos, precisamos lembrar que estas chaves (ligado/desligado) deverão, mesmo com sua mínima carga semântica, representar todos os tipos de informação, o que leva à necessidade de adoção de padrões, de um sistema de representação. Serres nos propõe uma série de questionamentos relacionados com esse problema: “Como pode o real construir-se sob a forma de signos? Em que condições alguns signos que se tornam matemáticos acessam melhor o real e o representam? Que estatuto conceder àqueles que não lhe têm acesso a não ser por esse conjunto de características?” (Serres, 2003, p. 70)

Desde os primórdios da computação, ou pelo menos desde que houve a possibilidade de termos dispositivos programáveis, dividimos o universo de construção das soluções em duas grandes dimensões: dados e funcionalidades. A primeira é voltada ao conjunto dos conteúdos que serão obtidos, registrados e disponibilizados, enquanto a segunda inclui todas as ações relacionadas ao que será feito com os dados e como ocorrerá a interação tanto para obtenção quanto para disponibilização das informações.

Vamos começar pela dimensão dos dados, e temos então competências técnicas específicas, requeridas para os operadores que vão atuar nessas atividades, o que nos permite identificá-los como sendo de “perfil de banco de dados”. Eles são responsáveis pela estruturação e definição de como os dados serão registrados, tanto em termos físicos quanto lógicos.

Com relação à organização física, vale lembrar que nos níveis mais baixos de abstração os atores com perfil de *hardware* já deixam o caminho mais suave, mas mesmo assim são muitas as ações ainda necessárias para que se obtenham resultados mais eficientes e eficazes no armazenamento e recuperação dos dados.

Consideramos ainda, nesse perfil, a responsabilidade pela organização lógica dos dados por meio de identificação de estruturas semânticas, como as chamadas entidades, conjuntos de dados cujas características são estruturalmente similares e que, portanto, podem ser tratadas por meio de funcionalidades semelhantes. Tais entidades, por sua vez, são compostas por atributos que vão possibilitar sua caracterização, descrição e identificação, viabilizando, inclusive, a percepção das relações entre tais conjuntos.

Mais uma vez a distância entre os usuários e a forma como essas definições são estabelecidas, por um lado, ocultam sua complexidade; por outro, impedem uma percepção clara sobre o quê e como os dados estão sendo administrados.

Uma forma de ilustrar essa questão é pensar em um caso concreto. Se o usuário responsável pelo arquivo de um empresa não tiver acesso ao modo como a organização lógica dos dados foi implementada, ele tenderá a perceber um documento - uma nota

fiscal, por exemplo - como sendo monolítico e resultado de um conjunto único de dados, quando, na verdade, ele está distribuído por diversos conjuntos (entidades): cliente, produto, transportador, além, é claro, do próprio conjunto que responde pela nota, e de outros que respondem pela relação entre conjuntos, como aquele que permite que uma nota esteja relacionada a vários produtos e que os produtos estejam relacionados a várias notas. Esse tipo de insciência do usuário não permite que ele perceba, por exemplo, que a exclusão de um documento não implique necessariamente a exclusão dos dados do cliente ou a vinculação entre a nota e os produtos que a compunham.

Já na dimensão que inclui todas as ações que deverão ser executadas pela máquina, temos os analistas e desenvolvedores de algoritmos que vão responder pelos tratamentos necessários sobre os dados utilizados, bem como pelas interfaces de obtenção e disponibilização das informações, tanto com os usuários quanto com outros sistemas. Estamos identificando esses operadores como sendo de “perfil aplicação”. Eles estarão bem mais próximos dos usuários, mas não o suficiente para que estes últimos tenham maiores detalhes sobre como os dados serão obtidos, tratados e transformados em resposta a suas necessidades informacionais.

Aqui, a distância em relação ao modo como está feito cada um dos processos exige do usuário uma suspensão da dúvida, em função da crença generalizada de que tudo que é feito com a participação da máquina está correto. A máquina estaria acima da dúvida quando se trata de operações, principalmente se baseadas

na lógica matemática; afinal, “é comumente suposto que o trabalho em campos técnicos prossegue por meio de raciocínio lógico e racional, claramente definido.” (Agre, 1997). Quantas vezes alguém confere o resultado de cálculo de uma nota emitida pelo computador? É possível colocar em dúvida o conteúdo informado pelo funcionário (participação humana), mas o cálculo tende a merecer o pressuposto de infalibilidade.

Cabe aqui um alerta: a insciência tende a não ficar restrita somente ao usuário; os próprios operadores com “perfil aplicação” terão cada vez mais dificuldade em manter domínio sobre o grande volume de instruções necessárias para a complexidade das aplicações, com o agravante de que tais instruções podem passar a resultar de outros processos. Tornam-se inscientes, assim, todos os atores envolvidos, cabendo somente aferição dos resultados alcançados, o que é pouco para ser considerado como controle sobre o que realmente é feito com os dados.

Na definição dos dois últimos perfis, estamos admitindo uma simplificação para viabilizar a interpretação mais ampla dos papéis desempenhados pelos diferentes atores, já que existem variações crescentes nas competências necessárias para atuação nessa dimensão. Tais variações decorrem do crescente leque de tecnologias disponíveis, tanto para a gestão dos dados quanto para a construção das aplicações.

A partir da percepção dos diferentes perfis do operador, podemos incluir o controlador, que é aquele que detém a posse dos recursos necessários utilizados no contexto de um determinado ciclo de vida de dados e, por consequência, é quem será o responsável

pelas definições de como todo o processo deverá funcionar. Em conjunto com o operador, ele compõe o que denominamos de “agentes de tratamento”.

HOMINESCÊNCIA E ESCOLHA

Apesar do grande e crescente volume de dados que está à disposição para coleta, da eficácia tecnológica para registrá-los e da variedade de meios de recuperação que a potente comunicação e as novas interfaces oferecem, continuamos fisicamente os mesmos, necessitando que tais dados sejam tratados e apresentados como informação biodisponível, acessível, portanto, aos nossos recursos nativos de *input*.

Distanciados de todo o processo de acesso a dados, já que nos cabem somente as pequenas janelas de informações biodisponíveis, e considerando a pressuposta infalibilidade maquínica, corremos o risco de nos tornar vulneráveis às agendas dos agentes de tratamento, em especial dos “controladores” (detentores).

Paul Virilio (1996, p. 127) já anunciava que

definitivamente afastados da observação direta ou indireta das imagens de síntese produzidas pela máquina, para a máquina, essas imagens virtuais instrumentais serão, para nós, equivalentes àquilo que já significam hoje as representações mentais de um interlocutor estrangeiro: um enigma. [...] a prótese de percepção automática funcionará como uma espécie de imaginário maquínico do qual seremos totalmente excluídos.

Serres (2003) apontou para o processo de transformação que afeta nossos corpos, nosso mundo e nossa sociedade, uma espécie de hominização, iniciada há milhares de anos, que vem se acelerando vertiginosamente nas últimas décadas, equilibrando-se entre o sucesso e a catástrofe aos quais nossa espécie conseguiu sobreviver. Mudamos as nossas relações sociais, o mundo e a nós mesmos, mas agora nossa capacidade de ação mudou rapidamente de escala: “riscos e chances crescem tão rápido quanto nossa onipotência” (p. 20).

O mesmo autor identifica o estágio atual dessa hominização com sua volatilidade e, por que não dizer, sua característica líquida (à imagem de Bauman), chamando-a de hominescência para diferenciá-la da hominização. Acrescenta ainda:

como a luminescência ou a incandescência aumentam ou diminuem, [...] ainda que esteja sempre pronta a se extinguir [...]; assim como a eflorescência ou a efervescência designam igualmente processos marcados por essa desinência denominada ‘incoativa’, adjetivo que designa um início de floração, de efervescência ou de emoção; assim como uma planta arborescente assume pouco a pouco a forma ramificada, o porte ou a aparência de uma árvore (Serres, 2003, p. 21)

Esta hominescência flutuante, instável e imprevisível, que é marcada por sua característica **incoativa** por ser ainda incipiente ou (correndo o risco da redundância) nascente, já traz em seu rastro o protagonismo do agente maquínico em nossa relação com a informação, em especial com os dados. Tal protagonismo nos deixa pouca margem de escolha, tornando o acesso ao novo modelo

um caminho sem alternativas. Ou incorporamos a “alavanca” do tratamento maquínico dos dados, ou estamos fora do mundo. A hominescência assume então uma característica **coativa**, compelindo-nos a continuar nessa simbiose, se é que a podemos definir assim.

O novo ser que emergirá dessa transformação ainda não sabemos quem será. Cabe-nos trabalhar no sentido de conscientizar a sociedade sobre a transformação, seus atores e interesses envolvidos, e buscar caminhos que possam minimizar as assimetrias, sempre tão suscetíveis aos processos de mudança.

Para concluir, fica a provocação deixada por Michel Serres (2003, p. 21): “o processo de hominescência acaba de assumir seu lugar, embora não saibamos ainda que homem ele irá produzir, enaltecer ou assassinar. Será que alguma vez o soubemos?”

REFERÊNCIAS

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

AGRE, P. E. Toward a critical technical practice: lessons learned in trying to reform al. *In*: Geoffrey, C. B. *et al.* (Eds.). **Social science, technical systems, and cooperative work**: beyond the great divide. New York: Lawrence Erlbaum Associates, 1997.

BRASIL. **Lei n. 13.709, de 14 de agosto de 2018**. Lei Geral de Proteção de Dados Pessoais (LGDP). Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709.htm>. Acesso em: agosto de 2021.

SANT'ANA, R. C. G. Transdução informacional: impactos do controle sobre os dados. *In*: MARTÍNEZ-ÁVILA, D.; SOUZA, E. A.; GONZALEZ, M. E. Q. (Eds.). **Informação, conhecimento, ação autônoma e big data**: continuidade ou revolução? São Paulo: Cultura Acadêmica; Filoczar, 2019.

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

VIRILIO, P. A imagem virtual mental e instrumental. *In*: PARENTE, A. (Org.). **Imagem máquina**: a era das tecnologias do virtual. 2. ed. São Paulo: Editora 34, 1996.

MEMÓRIA PARA TODOS: OS DESAFIOS DAS HUMANIDADES DIGITAIS

Maria Fernanda Rollo

Vivemos num mundo digital marcado por avanços tecnológicos com ritmos e perspectivas que por vezes parecem inalcançáveis, mas que integram cada vez mais os nossos quotidianos, alterando rapidamente realidades que nos parecem ainda imutáveis, como o emprego, a organização social e as relações interpessoais, a indústria, os serviços e, de forma muito próxima e expressiva, a formação e a investigação.

Na realidade, a aproximação do conhecimento à sociedade, às empresas, às instituições sociais e culturais, num contexto de transformação tecnológica e digital vertiginosa, tem transportado novas responsabilidades, novos desafios e múltiplas oportunidades particularmente relevantes em três domínios:

1. No plano das novas fronteiras do conhecimento científico e da sua aplicação (ex. computação científica, *big data*, digita-

lização da economia, sistemas ciberfísicos, processos industriais, serviços públicos digitais, *smart cities*).

2. No domínio da formação e qualificação das pessoas, em contextos de aprendizagem inclusivos.
3. No reforço da cidadania e da sua expressão nos territórios.

A era digital amplia, sem dúvida, as possibilidades de acesso e partilha no plano da formação, da ciência e do conhecimento em geral, e constitui uma combinação de desafios e oportunidades, suscitando um esforço de formação e capacitação, e um expressivo investimento no plano científico e tecnológico, desde que estejam reunidas para o efeito as condições indispensáveis.

Se dúvidas e hesitações existissem, o contexto pandêmico que vivemos nos últimos meses exacerbou e acelerou de forma estrondosa as oportunidades e os constrangimentos que a transição digital coloca ao mundo da educação, do conhecimento, da investigação e à sociedade em geral. Entre tantos aspectos, deve salientar-se a dimensão universal e global de todos estes fenómenos, sem esquecer ou matizar como a era digital constitui um enorme desafio que pode contribuir dramaticamente, ou, pelo contrário, apoiar a mitigação do fosso de acesso e fruição à ciência/conhecimento entre países.

Falamos de novas formas de aprender, de trabalhar, de comunicar e até de pensar, tendo consciência de que, na história do mundo, poucas áreas de desenvolvimento (se houver) tiveram um impacto tão grande na sociedade nestas poucas décadas e com tal dimensão e ritmo.

São impressionantes as experiências e imensos os exemplos que, no âmbito desta avassaladora, mas nem sempre visível, transformação digital da sociedade ocorreram de forma tão inovadora quanto surpreendente neste tempo de pandemia. De repente, até os processos de investigação e colaboração científica e acadêmica se alteraram, revelando possibilidades incríveis, mas também enormes desigualdades e deixando muitos de fora. Claro que tudo isto nos tem obrigado a ajustar, a alinhar opções e estratégias e a definir prioridades – nos planos globais/institucionais, mas também ao nível individual. A pandemia, sem dúvida, agudizou e tornou mais imperativa e inexorável a questão de como dominar, acompanhar a velocidade, profundidade e dimensão da transformação digital.

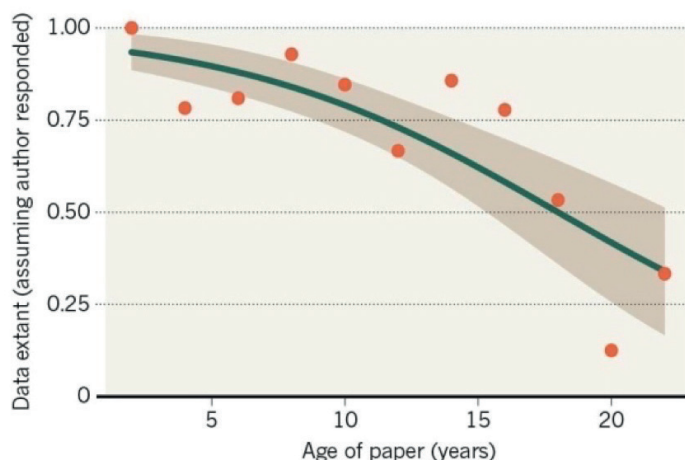
Convoco apenas dois tópicos que afetam diretamente o nosso campo da ciência e do conhecimento. O primeiro tem a ver com a quantidade da informação disponível. Veja-se o gráfico em baixo. Espelha bem o gigantismo do que está em causa. É claro que esperamos que a tecnologia dê resposta à enormidade da produção de dados e que até seja capaz de a organizar. A inteligência artificial supera já a nossa imaginação mais criativa, estando longe de ser ficção.



Fonte: <https://www.slideshare.net/sfamilian/working-with-big-data-jan-2016-part-1>

O segundo aspecto reporta-se à questão da preservação e do patrimônio em geral, compreendendo a perda de informação/de dados e a vulnerabilidade e exigências colocados pelo patrimônio digital.

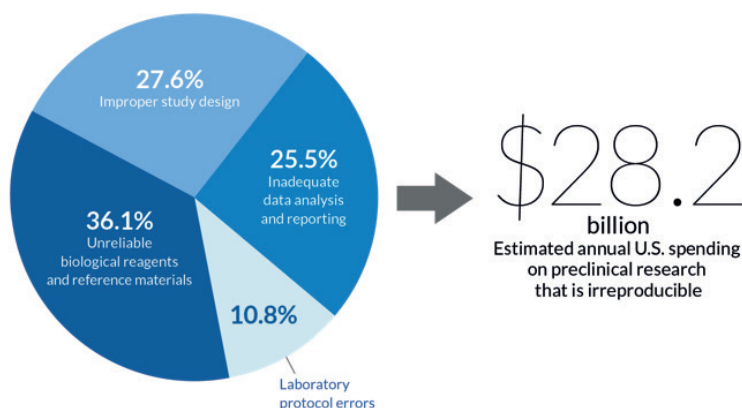
Veja-se o que está também a acontecer.



Fonte: "Scientists losing data at a rapid rate", in Nature News, 19/12/2013
<https://www.nature.com/news/scientists-losing-data-at-a-rapid-rate-1.14416>

Mantendo a expectativa e a esperança na ciência e na tecnologia, também será possível evitar a perda de informação. Mas é necessário que essa vontade se imponha e que acompanhe a consciência da perda de conhecimento e de patrimônio digital, que já está a acontecer.

Esperemos que essa preservação ocorra num sentido amplo e responsável, para além do que, evidentemente, a percepção de perda de valor econômico, que o gráfico em baixo espelha, suscitará.



Fonte: The Economics of Reproducibility in Preclinical Research. **PLOS Biology**, v. 16, n. 4. 2015.

A verdade é que nunca produzimos tanto, e nunca desperdiçamos tanto. Não só dados, como bem sabemos.

A perda é ainda mais condenável considerando o seu significado em termos de desperdício de oportunidades, prejudicando possibilidades e oportunidades de reprodutibilidade e reaproveitamento de informação/dados/conhecimento que possam gerar mais

conhecimento e beneficiar outros contextos e mais pessoas... lesando, é claro, recursos e um patrimônio comum. Deve acrescentar-se à preservação, curadoria e reutilização da informação produzida um outro tópico de reflexão e preocupação, relativa à veracidade e credibilidade dos dados e da informação em geral, prosseguindo propósitos de verdade, cumprindo boas práticas, garantindo a transparência e o respeito pela propriedade intelectual, hoje uma das angústias mais relevantes no plano da formação e da produção científica.

Enfim, o desafio é enorme, envolvendo e atingindo outros atores que não só as instituições dedicadas à ciência e à formação, colocando, de qualquer forma, estas entidades de produção de conhecimento numa posição vital no sentido de contribuir para o superar.



É elevado o risco da perda de informação, da nossa história, memória e identidade, no que integra como testemunho do passado e, cada vez mais, do patrimônio digital, hoje em permanente composição. Cumpre-nos criar condições de proteção desse legado, herança/patrimônio digital, não descuidando quanto à sua vulnerabilidade e efemeridade. Importa assumir como responsabilidade coletiva a preservação do patrimônio digital em benefício das gerações presentes e futuras. Assim o tem indicado, entre outras instituições internacionais, a própria UNESCO, que lhe dedicou precisamente, em 2003, uma *Charter on the preservation of digital heritage*.

Está dado o mote, enunciado o caminho: “the access to digital heritage materials, especially those in the public domain, should be free of unreasonable restrictions. At the same time, sensitive and personal information should be protected from any form of intrusion.” E, como inspiração: “*Member States may wish to cooperate with relevant organizations and institutions in encouraging a legal and practical environment which will maximize accessibility of the digital heritage. A fair balance between the legitimate rights of creators and other rights holders and the interests of the public to access digital heritage materials should be reaffirmed and promoted.*”

As instituições de ensino superior e investigação constituem um espaço vital na produção do conhecimento e, claro, na formação. São agentes catalisadores e repositórios das nossas formações e expressões culturais, pelo que constitui o seu patrimônio, pelo que estudam e investigam, pelo que cuidam e conservam. A maior parte do que atualmente ocorre e circula nos seus contextos

acontece em ambiente digital, que cumpre preservar e, quanto a mim, partilhar, acreditando nas instituições culturais e científicas, na formação, na partilha e no acesso mais equitativo ao conhecimento e à investigação, como um dos caminhos mais eficazes para a igualdade, o desenvolvimento e o bem-estar.

Deve, todavia, salientar-se que o desafio da preservação, e mesmo da curadoria, não só é superável se, para além da ação e das condições proporcionadas pelas instituições, for acompanhada pela atuação individual ou de grupo dos académicos, alunos, investigadores, e cidadãos em geral.

Isso significa e implica que o domínio das competências digitais não pode estar exclusivamente dependente das tecnologias de informação e comunicação, nem deve estar dissociado da capacidade de interpretar a tecnologia em todos os seus contextos. Por outras palavras, a Ciência da Computação já não é apenas um desafio de engenharia, requerendo o desenvolvimento e alargamento de competências de forma transversal às diversas áreas científicas/disciplinares; além disso, a forma como influencia e penetra os diversos contextos e todas as dimensões do quotidiano das pessoas implica a assunção de responsabilidades públicas – e das instituições culturais e de ensino superior e investigação em particular – no sentido de democratizar a sua acessibilidade. Acresce a indispensabilidade da percepção e atuação individual. Como exemplo, veja-se o impacto do enorme crescimento dos resultados do *machine learning* e dos diversos componentes da inteligência artificial, já tão presentes no dia-a-dia. Na realidade, o confronto permanente com a realidade digital exige de cada um

de nós uma capacidade renovada de conhecer, aprender, reaprender e, porventura, um papel ativo como cidadãos, utilizadores, investigadores e professores.

De resto, estima-se que em 2025 a maioria dos postos de trabalho exigirá formação superior e/ou especialização técnica, sendo que neste preciso momento cerca de 70 milhões de cidadãos europeus carecem ainda de competências adequadas de leitura e de escrita. Os desafios são muitos: cerca de metade da população não tem competências digitais básicas e praticamente 1/3 nunca utilizou a internet. Quando observamos com mais detalhe, percebemos que há micro-realidades – o designado fosso digital – às quais precisamos dar resposta urgentemente: défice de competências digitais nos grupos mais vulneráveis, nos mais velhos, nas pessoas com carências económicas e também naquelas com baixos níveis de escolaridade.

Acrescente-se a esta reflexão a compreensão das transformações em curso nas múltiplas dimensões da formação e da investigação e produção de conhecimento em geral. Assinale-se, entretanto, o impacto do que, para sintetizar, podemos colocar sob o chapéu da Ciência Aberta, e o tanto que se está a processar em termos de evolução e renovação do ensino superior e da produção científica, introduzindo novos métodos e novas aprendizagens baseadas num modelo de colaboração, cocriação e transferência multilateral de conhecimentos, contrariando o paradigma tradicional da transferência unilateral de conhecimento.

No que respeita às humanidades digitais, admitindo a sua afirmação, salientam-se seis eixos fundamentais (sem a pretensão de

esgotar a sua esfera de atuação) através dos quais a sua contribuição é mais evidente e relevante para a atividade científica e a sociedade em geral: salvaguarda e preservação do patrimônio digital; educação e formação; organização da informação; articulação com outras áreas científicas; acessibilidade, disseminação e partilha do conhecimento; e compromisso com a sociedade.

O programa Memória para Todos convoca os eixos referidos, ilustrando as humanidades digitais para além da aplicação do digital na investigação na área das ciências humanas.

No que respeita aos objetivos do presente Seminário, o caso do programa Memória para Todos proporciona uma reflexão direta, de matriz multidisciplinar, sobre a gestão do patrimônio documental, os desafios suscitados pela crescente expansão das novas tecnologias da informação, a criação de repositórios digitais, o desenvolvimento de ferramentas dedicadas à sua organização e gestão, curadoria e preservação, e seu impacto nas práticas da formação e da pesquisa científica. Refira-se ainda ao modo como convoca diversas dimensões da relação da ciência/academia com a sociedade em geral, refletindo novas formas de compreensão do processo de criação do conhecimento.

Memória para Todos (<https://memoriaparatodos.pt>) é um programa empenhado na promoção do estudo, organização e disseminação do patrimônio histórico, cultural, tecnológico e digital, desenvolvido em estreita relação com arquivos e bibliotecas, instituições da administração pública, municípios e autarquias, entidades privadas, escolas e associações locais, compreendendo diversos projetos e domínios/temáticas de pesquisa.

Sob a consigna *Faça História partilhando a sua*, Memória para Todos regista, preserva e partilha histórias e memórias, incluindo testemunhos orais, documentação e objetos pessoais e familiares, valorizando as histórias de vida e das comunidades.

Parto dos princípios e metodologias que o regem e dos resultados alcançados para refletir sobre humanidades digitais, *big data* e pesquisa científica, convocando a sua perspetiva transdisciplinar, com especial enfoque nas práticas colaborativas de investigação e construção do conhecimento. O programa compreende diversos projetos e iniciativas, criados por nós ou sugeridos pela sociedade, que envolvem o público no seu processo de desenvolvimento. É, por natureza, um projeto colaborativo, contando com uma vasta equipe de colaboradores da academia e da sociedade em geral.

Os conteúdos reunidos (como objetos, fotografias, testemunhos áudio e vídeo, sons e outros registos), com a participação e envolvimento dos cidadãos e instituições, são disponibilizados *on-line* em acesso aberto, bem como os resultados da pesquisa associada.

As primeiras ações realizadas remontam a 2009-2010 no âmbito das Comemorações para o Centenário da República¹, lançando o desafio à comunidade portuguesa para a partilha de memórias, testemunhos e objetos que retratassem a implantação da República em Portugal (5 de outubro de 1910) e a história da I

1. A Comissão para as Comemorações do Centenário da República foi nomeada pelo presidente da República, com a seguinte composição: Artur Santos Silva (presidente), Maria Fernanda Rollo (comissária executiva), Francisco Sarsfield Cabral, Raquel Henriques da Silva e João Serra, substituído por Rui Vieira Nery.

República (1910-1926) em efetivo ambiente de ciência cidadã. Foi então possível apreender a riqueza e extensão de objetos, em especial documentos e fotografias, que existiam na posse de particulares e, muito especialmente, a(s) memória(s) que perduravam entre herdeiros, familiares de pessoas que ainda viveram os acontecimentos de outubro de 1910 e nos anos seguintes. Ficou muito clara a relevância desses contributos para o aprofundamento e enriquecimento da história do período da I República. Importa assinalar que à I República sucedeu, depois de um período de ditadura militar/nacional, o Estado Novo, e como este regime autoritário impôs um contexto de rejeição e censura da herança e do legado político e cultural do ideário e prática republicanos, tornando assim mais preciosos os testemunhos físicos que então ‘reapareceram’. O legado dessa recolha, que ilustrou um conjunto muito vasto de exposições, publicações, vídeos e *performances*, foi partilhado com várias instituições e foi acolhido, virtualmente, pelo Centro República (<https://www.centrorepública.pt>).

A partir de então, e especialmente a partir do projeto Portugal 1914 (a que já faremos referência mais detalhada), o programa formalizou-se e desenvolveu-se, assumindo como missão a recolha, partilha e preservação de memórias e testemunhos, a democratização do acesso às ferramentas da investigação histórica, a promoção de dinâmicas colaborativas de investigação, a produção e disseminação de conhecimento e a criação de novas fontes para a investigação, centrando-se nas dinâmicas sociais da construção da memória, enquanto fator essencial para a definição de identidade. O programa Memória para Todos assume ainda, como

propósitos, a promoção da educação patrimonial, a literacia, o desenvolvimento de competências digitais e a democratização do acesso às ferramentas da investigação histórica.

Memória para Todos inclui atualmente um amplo conjunto de projetos, constituindo um programa interdisciplinar e colaborativo, distinguindo-se pelos métodos de envolvimento das comunidades, a aplicação de diagnósticos e soluções de base científica sobre desafios sociais e o desenvolvimento de produtos em acesso aberto, partilháveis e reutilizáveis.

Neste sentido, adota metodologias de registro, organização e divulgação de conteúdos específicos, de acordo com a natureza dos materiais, suportes e informação em causa, nomeadamente testemunhos orais, histórias de vida, espólios e coleções privadas e familiares, através da criação de bases de dados e de processos digitais de conservação e disponibilização de dados. Convoca e aplica várias metodologias, associadas às ciências documentais, à história, à museologia, ao património, às ciências sociais em geral, sendo de salientar as práticas de história oral, método biográfico, prosopografia, história local e georreferenciação, compondo um ambiente evidentemente comum e frequentado pelas humanidades digitais.

No desenvolvimento do programa, os cidadãos têm sido convidados a partilhar as suas memórias pessoais e familiares relativas a determinados momentos e processos históricos e, em muitos casos, desafiados a envolver-se e a fazer parte ativa do programa; as comunidades têm sido convidadas e suscitadas no mesmo sentido, compreendendo e assumindo contribuir para a história da

sua própria comunidade (ver, em particular, os projetos Memórias das Avenidas e Memórias das Aldeias).

Por tudo isso, e constituindo um programa de atividades, o Memória para Todos aplica também metodologias específicas de envolvimento comunitário, dinamização social e produção de conhecimento, tendo em vista o registro, a sistematização e o mapeamento de informação e dados provenientes da população em geral.

O leque de metodologias convocadas é ainda mais amplo, correspondendo às especificidades de cada projeto, incluindo domínios menos afins às áreas das humanidades e das ciências sociais, nomeadamente no campo das ciências da vida.

Depreende-se a sua natureza interdisciplinar e multidisciplinar, tanto ao nível da concepção e desenvolvimento do programa em geral, quanto no quadro específico de cada projeto que promove.

A dimensão multidisciplinar do Programa amplia-se quando se trata de explorar os materiais recolhidos e, noutro plano, se captam, acompanham e exploram os conteúdos, as relações e o enquadramento social em que se desenvolve. Projetos como Memórias das Avenidas, Memórias de São Domingos e Memórias das Aldeias, entre outros, convocam as áreas das neurociências e da psicologia, especialmente a propósito do impacto da participação, no programa Memória para Todos, de pessoas / comunidades específicas.

O projeto Portugal 1914 (www.portugal1914.org) foi lançado em 2012, compreendendo um portal destinado a disponibilizar e re-

colher informação sobre a história e o impacto do envolvimento português na I Guerra Mundial. Incluiu várias atividades editoriais, exposições e um forte componente de recolha de memórias e conteúdos, amplificado através das redes sociais, e a organização de ações específicas de recolha de memórias a que passamos a designar Dias da Memória (períodos específicos em que as pessoas são convidadas a deslocar-se a determinados locais para partilhar as suas memórias e os objetos associados, e onde são recebidos por uma equipe composta por vários especialistas, munidos dos recursos necessários para o registro de áudio e vídeo das entrevistas e a digitalização ou fotografia dos diversos materiais). A ação de recolha de memórias decorreu em escala nacional e virtualmente em França. Os primeiros Dias da Memória realizaram-se na Assembleia da República - Parlamento português, entre 17 e 19 de outubro de 2014, com a participação de dezenas de pessoas cujas contribuições foram registradas em áudio e vídeo; os materiais que partilharam foram integralmente digitalizados ou fotografados. A descrição e a caracterização de todos esses conteúdos foram integradas em bases de dados relacionais. Seguiram-se iniciativas semelhantes em Tomar (fevereiro de 2015), Amadora (outubro de 2015) e Vila Franca de Xira (setembro de 2016), seguindo o interesse e o convite de vários municípios. Os seus resultados integram o programa internacional *Europeana 1914-1918*. Terá sido a primeira iniciativa de *crowdsourcing* desta envergadura realizada em Portugal, tendo contado com a colaboração de várias entidades e a divulgação através da televisão nacional, que elaborou um vídeo promocional apresentando pela primeira vez a consigna do programa *Faça história partilhando a*

sua. O projeto foi um sucesso, pelo envolvimento que suscitou, pela dimensão, riqueza e diversidade do patrimônio reunido (todo ele reproduzido e uma parte, por solicitação dos seus proprietários, doada a museus, arquivos e bibliotecas) e ainda, sob o ponto de vista historiográfico, permitindo aprofundar o conhecimento sobre a história do envolvimento na I Guerra, nomeadamente através do acesso a fontes até então inéditas. Envolveu uma vasta equipe de investigadores e de voluntários, em Portugal e em França. Além disso, promoveu a colaboração de várias escolas e a participação de alunos que desenvolveram projetos de recolha no âmbito das suas atividades escolares. O desenvolvimento do Portugal1914 envolveu já o desenvolvimento de um conjunto de ferramentas digitais de gestão da recolha, organização e disponibilização da informação – bases de dados, edição de conteúdos, análise de informação.

A partir de então são diversos os projetos realizados por iniciativa do programa ou por desafio de entidades. Destaque-se, no contexto desta apresentação, o recurso às ferramentas digitais que têm acompanhado e até de certa forma contribuído para o desenvolvimento e a sucessiva superação de objetivos e desafios no âmbito do programa Memória para Todos.

Podemos de certa forma replicar o enunciado deste Seminário, começando por salientar o patrimônio digital existente ou constituído a partir da recolha de memórias e testemunhos, através do registro digital oral ou em vídeo, de sua organização e edição, compondo um renovado legado cuja salvaguarda e preservação cumpre garantir. A par desses testemunhos, a extensão, diversida-

de e riqueza de objetos associados, cuja organização e reprodução digital se tem realizado. Tudo quanto se tem recolhido, incluindo a caracterização dos entrevistados, é registrado, descrito, classificado em bases de dados relacionais, adicionando ferramentas de classificação e organização relacionando os registos em áudio e vídeo. Acrescente-se o processo de contextualização das informações recolhidas, cumprindo preceitos rigorosos de classificação e adotando ontologias e semânticas normalizadas. O conjunto alargado de investigadores associados aos diversos projetos partilha um contexto de formação científica e técnica adquirida previamente ou no âmbito do programa em diversas áreas – bases de dados e edição digital em particular. Em vários projetos algum grau de formação ou aquisição de competências digitais envolve intervenientes externos ao programa – incluindo estudantes de escolas secundárias ou pessoas mais velhas e em alguns casos já reformadas, como nos projetos Memórias das Avenidas ou Memórias de São Domingos. Qualquer um dos projetos pressupõe uma articulação dinâmica ao nível da comunicação via digital.

De salientar também, constituindo um desafio permanentemente renovado e ampliado, as formas de colaboração a distância, via digital. Para além da submissão de conteúdos, através de formulários específicos, cada vez mais elaborados, como ocorre nesta fase com um dos projetos mais recentes dedicado à recolha de testemunhos de emigrantes e de trabalhadores portugueses, assumiremos em breve o desafio de, na fase mais precoce possível, envolver a comunidade académica, os investigadores que estudem temas semelhantes, a participar num cenário de investigação colaborativa.

Conforme referido, uma das condições essenciais do programa é a sua divulgação em acesso aberto e a promoção da máxima acessibilidade, respeitando rigorosamente o regulamento da proteção de dados e os princípios da propriedade intelectual, o que significa o esforço permanente de atualização e adaptação aos normativos/procedimentos que proporcionem a sua integração e/ou interoperabilidade com um conjunto de plataformas, desde logo portais das entidades parceiras e infraestruturas mais ou menos especializadas. Todas as fases de desenvolvimento de qualquer um dos projetos, e o programa em geral, evocam cenários de multidisciplinaridade, nas áreas da arquivística, ciências documentais, informática, ciências da comunicação, *design* e o amplo conjunto de disciplinas diretamente associadas a projetos específicos.

Vários projetos focam conjunturas relevantes da história contemporânea de Portugal; alguns, da iniciativa do Memória para Todos, têm sido desenvolvidos, como o Portugal1914, com o apoio de entidades de âmbito nacional. Refira-se o projeto dedicado às memórias da Revolução de Abril e da Descolonização (acompanhado pela RTP e articulado com um programa editorial de quase duas centenas de episódios televisivos, organizando Dias da Memória no quartel histórico do Carmo, da Guarda Nacional Republicana) e o projeto Memórias da Resistência e da Liberdade, em articulação com o Museu do Aljube, recolhendo memórias e testemunhos de presos políticos durante o Estado Novo e, é esse o nosso propósito mais recente, dos seus familiares. Destaque-se, a propósito, uma das vertentes do programa, dedicada à promoção da utilização dos conteúdos recolhidos e organizados por parte da comu-

nidade acadêmica e a sociedade em geral. Neste caso, para além da contribuição para o conteúdo expositivo do Museu do Aljube, a disponibilização de informação para trabalhos de investigação, nomeadamente teses e dissertações académicas, compondo o escopo de ferramentas digitais à sua disposição.

Uma nota para referir o conjunto de projetos dedicados ao conhecimento e ao registro de memórias associadas ao mundo do trabalho – Memórias da Fábrica de Sacavém, da Cortadoria Nacional do Pelo, do centro fabril de São João da Madeira (envolvendo diversos sectores de atividade), ente outros, e a sua articulação com o novo projeto *Europe at Work* a apresentar brevemente pela Europeiaana.

Alguns projetos têm tido como foco comunidades específicas, devendo destacar-se Memórias das Avenidas, que conta com um grupo de colaboradores – os Maiores das Avenidas – muitos dos quais moradores desse bairro histórico da cidade de Lisboa, tal como acontece com o projeto Memórias de São Domingos. Está em curso um projeto de recolha de Memórias das Aldeias, a propósito do qual cumpre salientar a intenção do projeto no sentido de contribuir, para além da identidade das comunidades, para a promoção e valorização do desenvolvimento e coesão social e territorial. Neste caso acrescenta-se ainda a disponibilidade de reutilização de conteúdos para enriquecer a promoção turística. Qualquer um destes projetos tem tido impacto social em outros domínios, inclusive nas áreas da saúde, contribuindo para o combate às doenças da memória e ao isolamento e desenvolvendo o acompanhamento do envelhecimento ativo.

Por fim, cumprindo a missão definida e prosseguindo os objetivos apontados, os projetos que constituem o programa Memória para Todos decorrem da estratégia, do percurso de investigação e do compromisso social e científico prosseguidos. Assim, parte dos projetos mais recentes prossegue como propósito a contribuição para a Agenda 2030 e para a superação de um dos desafios sociais mais complexos que hoje se colocam à humanidade, associado às alterações climáticas. É nesse contexto que se enquadram projetos como o dedicado à Lagoa de Óbidos, ao Parque da Gorongosa em Moçambique, às reservas da biosfera, ou o projeto Living Rivers, procurando, a partir da recolha de memórias de tempos passados e das perceções que persistem, contribuir para o conhecimento e a consciencialização das alterações em curso.

O compromisso do Programa Memória para Todos é, em suma, baseando-se na prática científica, como é a sua matriz, para com as pessoas, as comunidades e o contexto que habitam, envolvendo-as em diversos planos e procurando uma dinâmica colaborativa que promova a preservação e valorização da memória, a criatividade e a criação de novo conhecimento, garantindo a sua partilha e promovendo a sua acessibilidade - constituindo, portanto, uma interface e uma intermediação entre a academia e a sociedade.

O recurso às ferramentas digitais acontece, natural e inevitável, procurado e desenvolvido, refletindo e apropriando-se do contexto digital pelas possibilidades que proporciona e pela forma como molda e condiciona a produção do conhecimento e a sociedade em geral, como oportunamente se poderá perceber, também, através das Memórias do Digital.

Memória para Todos é, assim, um espaço em que a história acontece e o patrimônio cultural é promovido, salvaguardado e compartilhado, beneficiando-se do contexto e das ferramentas digitais em sinergia e convergência entre comunidades científicas e a sociedade em geral.

A ORGANIZAÇÃO DO CONHECIMENTO EM UMA PERSPECTIVA CONTEXTUAL E SEUS IMPACTOS NA PESQUISA ACADÊMICA

José Augusto Chaves Guimarães

INTRODUÇÃO

Um aspecto que sempre permeou os saberes e os fazeres das ciências humanas e sociais reside na recuperação, preservação e disponibilização da memória da sociedade em suas mais diversas facetas. Como destaca Tedesco (2014), ao referir-se à contribuição seminal de Maurice Halbwachs, essa memória busca resgatar e preservar elementos do passado de modo a fornecer subsídios para que a sociedade tenha continuidade e para que os cidadãos que a integram possam encontrar, nessa memória, aspectos para reconhecer e afirmar sua própria identidade.

Ao longo do tempo, buscou-se construir artefatos que pudessem ir além das limitações da memória humana, de modo a abrigar um volume cada vez maior de dados, informações e conhecimentos. Assim, desde a invenção da imprensa, no século XV, a partir da composição por tipos móveis, que permitiu a multiplicação de conteúdos e seu armazenamento em distintos locais, passando pela *Encyclopédie* de Diderot e D'Alembert, no século XVIII, e pelos esforços de Paul Otlet, no final do século XIX, relativamente ao Repertório Bibliográfico Universal, entre outros, a busca por elementos extensores da memória humana foi uma preocupação recorrente, aspecto que culminou, em meados do século XX, com a criação do MEMEX, a cargo de Vannevar Bush e sua equipe, em um primeiro intento de configuração de uma memória automatizada, artificial, que pudesse abrigar um volume consideravelmente maior de informações que aquele até então possível ao ser humano; uma memória

... in which an individual stores all his books, records, and communications, and which is mechanized so that it may be consulted with exceeding speed and flexibility. It is an enlarged intimate supplement to his memory. (BUSH, 1945, p. 101)

Se, no pós-Segunda Guerra Mundial, a preocupação com o volume de dados, informações e conhecimentos já se fazia notar de forma sistemática, em nossos dias, e notadamente após o desenvolvimento da World Wide Web, esse cenário cresceu – em volume e em complexidade – em progressão efetivamente geométrica, a ponto de se chegar a um panorama em que o exponencial volume de dados – os então denominados *big data* – já não pode

mais ser tratado por sistemas tradicionais (manuais ou automatizados), exigindo complexas estratégias de inteligência artificial que consigam subsidiar a busca por informações e a construção de conhecimentos.

Como destacam Hey, Tansley e Tolle (2009), os *big data* trouxeram uma nova perspectiva para o avanço científico, rompendo fronteiras de conhecimento a partir do uso de sofisticadas e inovadoras tecnologias para coletar, tratar, organizar e disponibilizar dados de modo que a informação deles oriunda pudesse trazer um valor agregado.

Esse novo desafio foi mais rapidamente incorporado pelo fazer científico das denominadas “ciências duras”, haja vista sua premente necessidade de tratamento de grande volume de dados de pesquisa pois, como destaca Guimarães (2018, p. 1):

Em tempos de cada vez mais acelerado desenvolvimento científico e tecnológico, permeado por uma alta competitividade, há de se garantir elementos que legitimem as atividades de pesquisa, dando-lhes credibilidade e condições de replicabilidade.

No entanto, a questão da transparência e da confiabilidade dos dados é algo que atinge as instituições de pesquisa como um todo, notadamente aquelas custeadas por recursos públicos, uma vez que lhes cabe prestar contas à sociedade sobre o financiamento recebido. Com isso, a preservação, o tratamento e a disponibilização dos dados inerentes a todas as fases da pesquisa – e não apenas aos resultados publicados – torna-se premente e, indo além das ciências exatas e biológicas, atinge as ciências humanas

e sociais, onde assume um caráter mais específico e complexo¹, abrindo espaço para as denominadas humanidade digitais, que se constroem na intersecção – e no diálogo – entre as tecnologias e as ciências humanas e sociais e, mais especificamente, no campo da memória e da documentação, trazem uma importante contribuição aos arquivos, bibliotecas, museus e centros de documentação digitais.

À vista disso, e considerando a organização do conhecimento como um campo do saber em que se investiga um conjunto de processos, instrumentos e produtos que propiciam a mediação de culturas (do produtor, do organizador e do usuário de conhecimentos socializados e registrados), discutem-se os desafios e perspectivas que se colocam à pesquisa científica nessa área, tendo em vista os novos elementos contextuais, em um cenário permeado pelos *big data* e pelas humanidades digitais.

A ORGANIZAÇÃO DO CONHECIMENTO COMO UNIVERSO INVESTIGATIVO

A Organização do Conhecimento - OC, entendida por Esteban Navarro e Garcia Marco (1995) como um campo de estudos de natureza interdisciplinar² que agrega e coloca em diálogo teorias

1. Pois aqui os dados podem ser acordes musicais em rascunhos de partituras, estudos preliminares para uma obra de arte etc.

2. Veja-se, a esse respeito, Szostak, Gnoli e López-Huertas (2016)

e metodologias oriundas de áreas como a Administração, a Arquivística, a Biblioteconomia, a Computação, a Comunicação, a Diplomática, a Documentação, a Linguística, a Lógica, a Psicologia e a Sociologia, entre outras, parte do pressuposto de que o conhecimento se constrói a partir da análise e articulação de informações, em um dado contexto, informações essas que, por sua vez, decorrem de conjuntos de dados previamente coletados, preservados, organizados e disponibilizados³.

Indo além, esse conhecimento insere-se no *continuum* dos processos (produção, organização e uso) que se efetiva em uma dinâmica helicoidal (e não cíclica, como durante muito tempo se advogou), pois é um sistema aberto e que se retroalimenta (Smit; Barreto, 2002; Guimarães, 2008). Tem-se, assim, a **organização** como processo central, uma vez que estabelece a mediação entre os contextos de **produção** (e materialização) do conhecimento e seu **uso** e apropriação, para geração de um novo conhecimento.

A configuração da OC decorre, outrossim, da necessidade humana de sistematização e consolidação de um conhecimento (conjunto de saberes) verificável em uma dada sociedade em um dado momento histórico, com um objetivo de *transmissão*. Registrem-se, aqui, os trabalhos de filósofos ligados à teoria do conhecimento (Aristóteles, Platão, Porfírio, Bacon e Harris, Lineu) e ainda de educadores, como Comênio. Como consequência, em um nível mais pragmático, verifica-se a necessidade de resgate do conhecimento

3. Acerca dos aspectos teórico-conceituais da Organização do Conhecimento, veja-se: Dahlberg (1993, 1995, 2006); Jaenecke (1994); San Segundo (1996); Hjørland (2008, 2017) e Gnoli (2016), entre outros.

registrado em documentos, visando ao seu acesso com um objetivo de *recuperação*, em cujo âmbito se destacam os esforços de documentalistas como Calímaco, Otlet e La Fontaine, entre outros. Mas foi efetivamente a partir de Henry Evelyn Bliss, nos Estados Unidos, que o termo Organização do Conhecimento se efetivou a partir de seus estudos sobre a necessidade de se analisar e se representar sistematicamente o fenômeno do conhecimento humano produzido, o que foi consolidado em *The Organization of Knowledge and the system of the science*, publicada em 1929 (Bliss, 1929).

É importante sinalizar que tipo de conhecimento constitui objeto da OC para que não seja confundido, por exemplo, com os processos que são objeto das ciências cognitivas⁴, em que se tem por conhecimento:

... the subjectively and objectively fairly well-founded certainty of somebody about the existence of a fact or a matter. This knowledge is not transferable, it can only be elaborated by somebody's own personal reflection. (Dahlberg, 2006, p. 11)

No caso ora em espécie, trata-se de um conhecimento cujo principal pressuposto reside em sua socialização por meio de registros, de modo a que possa ter portabilidade no espaço e permanência no tempo e, destarte, constituir elemento para a preservação da memória (Baturité, 2001; Smit; Barreto, 2002). Com isso, a OC refere-se aos processos, produtos e instrumentos voltados à descrição, representação e disponibilização de conhecimentos produzi-

4. "If you concentrate on *personal* knowledge to which only the individual has direct access, knowledge organization indicates a *mental* and therefore internal process of structuring or transforming (e.g., visualizing) knowledge" (Reinmann, 2012).

dos pela sociedade, a partir da atuação humana ou de aparatos tecnológicos (Hjørland, 2008), pois, como destacam Anderson e Pérez-Carballo (2001) a OC comporta tanto as abordagens *human-based* quanto as *machine-based*. Enquanto campo de estudos, a OC é caracterizada, por García Marco (1995), como disciplina de natureza tanto social (dado seu caráter aplicado) quanto científica (por estar sujeita ao método científico), revelando natureza inter e transdisciplinar, e se apresenta, no dizer de Esteban Navarro (1995, p. 66), “como uma plataforma de integração daquilo que ele denomina como *ciências documentais*”⁵.

Essa natureza eminentemente registrada e socializada do objeto da OC converge com aquilo que Hjørland (2015, 2017) entende como concepção estrita do campo – o da organização de um conhecimento socializado – mas que convive, por sua vez, com uma dimensão mais abrangente, voltada aos sistemas conceituais e mesmo aos campos de atuação na sociedade.

Partindo dessa concepção social, materializada e cíclica de conhecimento⁶, observa-se que o foco investigativo da OC recai sobre a busca pela compreensão, organização e representação desse conhecimento, de tal forma que possa tornar-se disponível e aces-

5. Miranda (1999, p. 69), ao referir-se sobre o caráter científico da disciplina, argumenta que a mesma pressupõe análise, reflexão e aplicação de fundamentos científicos.

6. Tem-se, no caso, aquilo que Dahlberg (1993, p. 214) denomina como “conhecimento em ação”, ou seja, algo acerca do qual existe um certo consenso social, um conhecimento registrado e socializado, cuja organização e representação será desenvolvida de modo que, a partir dele, possa ser gerado novo conhecimento (Guimarães, 2001).

sível a um número maior de pessoas⁷. Como destacam Esteban Navarro e García Marco (1995, p. 147), tem-se

... la disciplina dedicada al estudio y desarrollo de los fundamentos y técnicas de planificación, construcción, gestión, uso y evaluación de sistemas de descripción, catalogación, ordenación, clasificación, almacenamiento, comunicación y recuperación de los documentos creados por el hombre para testimoniar, conservar y transmitir su saber y sus actos, a partir de su contenido, con el fin de garantizar su conversión en información capaz de generar nuevo conocimiento. Se trata, por lo tanto, de una ciencia tridimensional, ya que se ocupa de los principios, métodos e instrumentos puestos en acción para la gestión del conocimiento humano desde una triple perspectiva: su representación, su organización y su comunicación documental.

Desse modo, trabalha-se com um conhecimento que:

- a) só faz sentido se entendido como parte de um conjunto, e em intrínseca relação com seus congêneres;
- b) materializa-se e socializa-se por meio do registro que, por sua vez, vai além de mera forma para refletir uma lógica de estruturação de conteúdos para servir a fins específicos (como a garantia de credibilidade, de autenticidade, de procedência e mesmo de valor testemunhal ou histórico);
- c) possui natureza eminentemente contextual, aspecto que passa sua criação, organização, disseminação, apropriação e seu uso.

7. Veja-se, em Pinho (2006), uma cuidadosa e fundamentada abordagem do conceito e do objeto da Organização do Conhecimento.

Tentando oferecer subsídios ao processo de construção epistemológica da área, Barité (2001) propõe um conjunto de premissas⁸, nas quais destaca que o conhecimento, enquanto produto, necessidade e dínamo social, se realiza a partir da informação (e, ao socializar-se, nela novamente se transforma) e possui uma estrutura e um processo de comunicação abertos. Seu registro e socialização ocorre em documentos (conjunto organizado de dados disponíveis), se expressa por meio de conceitos e tem seu processo de organização operado por meio de sistemas de conceitos (com fins científicos, funcionais ou de documentação) e pautado por uma tônica artificial, provisória e determinista, com o fim último de promover seu melhor aproveitamento individual e social. A isso se alia, ainda, o fato de essa organização ser eminentemente intencional (Briet, 1951; Buckland, 1991). Em síntese, pode-se dizer que se trata de um campo de estudo e de atuação que pressupõe uma dimensão coletiva e orgânica, materializada em um registro específico que articule forma e conteúdo, e em um contexto de produção e de uso (Tognoli; Rodrigues; Guimarães, 2019).

Tendo por objetivo a preservação e a promoção do acesso, a OC estabelece pontes entre a produção e o uso do conhecimento, fazendo-o por meio de instrumentos denominados sistemas de organização de conhecimento que, como já mencionado, constituem sistemas de conceitos criados para fins científicos, funcionais ou de documentação (pesquisa). Tais instrumentos (normas de descrição, esquemas de classificação, taxonomias, tesauros,

8. Observe-se, igualmente, em Hjørland (1994), um conjunto de princípios para a OC, esta entendida como um produto historicamente desenvolvido.

ontologias, tipologias documentais, vocabulários controlados etc.), por sua vez, subsidiam o desenvolvimento de processos (identificação, análise diplomática, classificação, descrição, indexação etc.) e resultam na geração de produtos que atuam como representações que podem ser consideradas como “substitutos do conhecimento” (*surrogates of knowledge*), tais como quadros de classificação, tabelas de temporalidade, índices, notações, descritores, catálogos, inventários etc. (Olson, 2002).

Enquanto universo investigativo, a OC vem tradicionalmente se desenvolvendo em três dimensões de pesquisa: uma dimensão epistemológica, voltada para a consolidação de teorias, metodologias, paradigmas e do reconhecimento dos diálogos interdisciplinares; uma dimensão tecnológica ou aplicada, voltada para o desenvolvimento de novas ferramentas/instrumentos e produtos (a partir de novos contextos de produção e de uso); e uma dimensão cultural, voltada para contextos e atividades/processos de mediação entre esses contextos (os sujeitos, as culturas, os valores etc.). Tais dimensões encontram espaço de divulgação e interlocução notadamente no contexto da International Society for Knowledge Organization – ISKO, criada em Frankfurt, em julho de 1989, por Ingetraut Dahlberg, a partir dos trabalhos de mais de duas décadas da Society for Classification, hoje integradas por mais de seiscentos pesquisadores, de mais de cinqüenta países, oriundos de diferentes áreas como Ciência da Informação, Filosofia, Linguística, Ciência da Computação, Arquivística e Cientometria, bem como domínios específicos, a exemplo da Informática Médica.

A ISKO, em busca da viabilização da pesquisa teórica e aplicada em OC, subsidia a comunicação e o intercâmbio de experiências investigativas entre seus membros, atuando ainda em colaboração com instituições relacionadas à organização conceitual e ao processamento do conhecimento, tais como: UNESCO, Comissão Europeia, ISO (International Organization for Standardization), IFLA (International Federation of Library Associations and Institutions), ASIS&T SIG/CR (Special Interest Group on Classification Research of the American Society for Information Science and Technology), NKOS (Networked Knowledge Organization Systems/Services) e Infoterm (International Information Centre for Terminology, dentre outros (ISKO, 2021).

Na atualidade, a ISKO se encontra formalmente presente nos seguintes capítulos nacionais e/ou regionais: África Ocidental, Alemanha/Áustria/Suíça, Brasil, Canadá/Estados Unidos, China, Espanha/Portugal, França, Índia, Irã, Itália, Países Baixos, Maghreb (Tunísia, Argélia e Marrocos), Polônia e Singapura, tendo como veículos científicos a revista *Knowledge Organization* e a série *Advances in Knowledge Organization* (coletâneas dos *papers* apresentados nos congressos internacionais bienais realizados desde 1990). Dentre os grandes temas até então investigados no âmbito da ISKO, pode-se destacar a relação da OC com paradigmas cognitivos, a gestão de qualidade, o mundo em mudança, estruturas e relações, desafios na integração de fronteiras, dinamismo versus estabilidade, sociedade global da informação e do conhecimento, questões de cultura e identidade, paradigmas conceituais, padrões históricos e perspectivas futuras, busca por um mundo

sustentável, contexto de um mundo digital, interfaces e articulações com disciplinas, domínios, serviços e tecnologias.

Como se pode observar, a pesquisa em OC vem, por um lado, buscando firmar seu estatuto epistemológico para se sedimentar cada vez mais como uma disciplina que possa identificar as teorias e metodologias que lhe são inerentes e, por outro, procurando avançar no que tange ao aporte tecnológico, com especial ênfase na inteligência artificial, para fazer frente a um cenário cada vez mais complexo e a um volume de dados em crescimento exponencial. Tudo isso, no entanto, vem se construindo sem deixar de lado o elemento contextual que permeia a própria concepção de OC, com um olhar muito atento para os vieses culturais e para os problemas ligados à procedência e à confiabilidade do que se organiza.

Outro aspecto a destacar é a relação cada vez mais estreita com os distintos universos documentais, em espaços voltados para as “instituições coletoras de cultura” (Homulos, 1990) que promovem a “gestão institucional dos saberes” (Fernandes, 2006), tais como arquivos, bibliotecas, centros de documentação, museus etc.

DESAFIOS E PERSPECTIVAS PARA A PESQUISA EM ORGANIZAÇÃO DO CONHECIMENTO

Com um universo investigativo dinâmico e complexo, a OC necessita manter um olhar concomitante no ontem, no hoje e no amanhã (López-Huertas, 2008). Nesse sentido, um primeiro de-

safo reside na cada vez mais necessária intensificação de esforços para promover a consolidação teórica e metodológica da disciplina, evidenciando sua própria historicidade, suas “escolas de pensamento”⁹, seus diálogos multi e interdisciplinares, e reconhecendo, outrossim, os domínios de conhecimento específicos que a integram, como, por exemplo, a denominada Organização do Conhecimento Arquivístico, que vem sendo cada vez mais objeto de pesquisa na área, dadas as peculiaridades que a permeiam (Tognoli; Rodrigues; Guimarães, 2019).

Outro desafio se coloca no desenvolvimento de ferramentas cada vez mais potentes (e aqui a dimensão tecnológica se faz mais e mais premente) para evitar a recuperação de “lixo informacional”. Para tanto, necessárias se tornam ferramentas cada vez mais amigáveis que possam garantir a confiabilidade e a autenticidade dos registros em ambiente digital, mormente em um contexto de *big data* e, em especial, em tempos permeados pela *disinformation* (quando a falsidade de conteúdo ocorre de forma deliberada com o intuito de prejudicar uma pessoa, um grupo social, uma organização etc.), pela *misinformation* (quando a falsidade não é intencional e decorre de erro ou equívocos) e pela *malinformation* (que possui a sutileza da “aparência de verdade” a partir de fontes minimamente credíveis, mas criada com intuito doloso), caracterizando um amplo espectro das denominadas *information disorders* (Wardle, 2017; Lazer et al., 2018).

9. Pesquisadores como Birger Hjørland, Jens-Erik Mai (Universidade de Copenhagen), Joseph Tennis (Universidade de Washington), Claudio Gnoli (Universidade de Pavia) e Richard Smiraglia (Universidade de Wisconsin-Milwaukee), entre outros, têm envidado robustos e sistemáticos esforços nesse sentido.

Em direta relação com a eventual presença das citadas *information disorders*, aponta-se o desafio de promover a garantia cultural do conhecimento, de modo a que seja organizado em consonância com seu próprio contexto – e, por decorrência, com os valores que permeiam as culturas envolvidas, - no sentido de se evitarem proselitismos, dominações culturais, preconceitos e censura, muitas vezes inerentes a um “poder de nomear” tacitamente atribuído pela sociedade a um profissional ou a um sistema informatizado. (Olson, 2002; Guimarães, 2006; Guimarães et al., 2008)

Por outro lado, instigantes perspectivas se colocam, notadamente em um contexto permeado pelas humanidades digitais, tais como a mais ampla possibilidade de diálogo com outras disciplinas, o aumento da interoperabilidade dos sistemas e uma recuperação cada vez mais rápida, eficaz e útil, capaz de promover uma ética “trans-cultural de mediação” (Garcia Gutierrez, 2002), com “garantia cultural” (Beghtol, 2002, 2005) e em respeito à dinâmica dos distintos domínios de conhecimento (Thellefssen & Thellefssen, 2004).

CONSIDERAÇÕES FINAIS

À vista do exposto, pode-se observar que a OC vem, ao longo do tempo, consolidando seu universo de pesquisa, agregando preocupações de ordem epistemológica, tecnológica e cultural, e com impactos cada vez mais efetivos no universo das instituições de preservação da memória coletiva que, como destaca Halbwachs (2013), se forma a partir do processo de interação social dos

indivíduos e se constitui de elementos julgados importantes e que são guardados como a memória oficial da sociedade, nos denominados “lugares da memória”.

Nesse contexto, o processo de organização, mediador das instâncias de produção e de uso/apropriação, vai muito além das dimensões do dado e da informação para atingir o conhecimento enquanto algo estruturado, contextualizado, com nítidas referências culturais, intencionalmente produzido e socializado por meio de documentos, que, uma vez institucionalmente abrigados, assumem portabilidade no espaço e permanência no tempo.

Com isso, o conceito de documento se expande, e forma e conteúdo deixam de ser instâncias distintas para se integrarem em um espectro mais amplo e coeso¹⁰. Assim, o documento, na bibliote-

10. Infelizmente o conceito de documento, ao longo do tempo, foi sofrendo algumas limitações conceituais, sendo muitas vezes reduzido tão somente à dimensão de registro de informações, com clara ênfase na questão do suporte: “unidade de registro de informações, qualquer que seja o suporte ou suporte formato” (Arquivo Nacional, 2005, p. 73). Nesse sentido, fundamental se torna recordar a concepção de Otlet (1934) para quem o documento concentra e difunde o pensamento – aqui entendido como conhecimento – coadunando conteúdo e continente. Nessa linha, Briet (1951, p. 7) vê no documento uma “base de conhecimento fixada materialmente e suscetível de ser utilizada”, aspecto que se completa com a visão de Buckland (2017), para quem o termo documento corresponde a conhecimento registrado cujo conteúdo deve ser compreendido a partir do contexto social de seu produtor, cotejado com o contexto social do usuário. Destaca-se, outrossim, um aspecto fundamental do documento, qual seja: o desempenho de uma efetiva função social (Brown & Duguid, 2017). No âmbito arquivístico, vale recordar as palavras de Heredia Herrera (1991, p. 121, citada por Rondinelli, 2013, p. 46), para quem o documento “abarca tudo o que pode transportar o conhecimento humano (...) e que esteja submetido à intervenção de uma inteligência ordenadora”. Vejam-se ainda, sobre o documento, os trabalhos de Rodríguez Bravo (2002), de Silva (2009) e de Ortega e Lara (2010).

ca, deixa de ser representativo apenas de um ou mais assuntos para incorporar, em seu conteúdo, o autor (quem é, de onde provém, o que pensa, com quem colabora ou interage etc.), o espaço (qual a instituição produtora), o tempo (qual o momento histórico em que o documento foi produzido) e, ainda, aspectos ligados à fiabilidade do conteúdo (mormente em tempos de *fake news*, em que os instrumentos de *fact checking* acabam sendo fundamentais). Da mesma forma, o documento de arquivo, como uma articulação efetiva de *actio* e *conscriptio*, evidencia um conhecimento socialmente produzido em que se fundem sua função, seu tema, sua procedência, sua organicidade com seus congêneres, as condições de sua produção – aqui incluídos os valores culturais envolvidos –, sua fórmula diplomática e, não menos importante, sua autenticidade e fiabilidade¹¹.

Nessa concepção de conhecimento (socialmente produzido e registrado) a ser organizado, o aspecto contextual assume caráter determinante e, se estes tempos de *big data* parecem desafiadores, as humanidades digitais aí estão para promoverem o acesso *global*, ou seja, em uma perspectiva global, mas sem desconsiderar as especificidades locais (Andersen; Skouvig, 2017).

Como destacam Mustafa El-Hadi e Favier (2014), os arquivos, bibliotecas e museus digitais ainda padecem com problemas como a falta de interfaces devidamente adaptadas, a falta de interoperabilidade e a ausência de tecnologia semântica adequada, mas há de se considerar que muito do caminho já foi trilhado e as

11. Acerca da complexidade conceitual que cada vez mais atinge o documento veja-se, para tanto, Silva (2009).

perspectivas se mostram cada vez mais promissoras¹².

Tem-se, pois, que a OC, na perspectiva apresentada, vai muito além de um aparente *modismo terminológico*, que um olhar mais incauto poderia sugerir, para constituir efetivamente um campo de estudos cujas origens remontam às categorias aristotélicas e cuja formalização, enquanto tal, remonta aos estudos de Henry Evelyn Bliss (1929), resultando em uma disciplina que conta não apenas com instâncias científicas próprias e consolidadas (a ISKO e suas publicações científicas, já mencionadas), como, e principalmente, com uma concepção bastante clara e cientificamente fundamentada daquilo que tem por objeto de estudo: o conhecimento socializado e contextualizado, indo além de uma noção reducionista de documento como mero suporte de informações para abranger uma dimensão social a partir dos contextos envolvidos. Tal aspecto, por si mesmo, demonstra a própria evolução da episteme da área que, partindo de um olhar ontológico, se amplia para incorporar elementos de natureza sociocognitiva e cultural.

Nessa tônica, cabe encarar o desafio de promover processos, instrumentos e produtos de OC que, acima de tudo, considerem o

12. Vale recordar as indagações de Claudio Gnoli (2017) aos pesquisadores da área de OC (KO, em inglês): *Can KO principles be extended to a broader scope? Can ontological and epistemological approaches be reconciled? Can any ontological foundation be identified? Should disciplines continue to be the structural base of KO? How can viewpoint warrant be respected? How can KO be adapted to local collection needs? How can KO deal with changes in knowledge? How can KOSs represent all these dimensions? How can software and formats be improved to better serve these needs? Who should do KO?*

contexto¹³, sem o que a modalidade de conhecimento, que é seu objeto, torna-se isolada, anacrônica e sem permeabilidade – e utilidade - social.

13. A questão do contexto vem sendo cada vez mais investigada na organização do conhecimento, em especial a partir das abordagens de análise de domínio. Veja-se, para tanto: Hjørland e Albrechtsen, 1995; Beghtol, 1995; Hjørland, 2002, 2004; Tennis, 2003, 2012; Mai, 2005; Smiraglia, 2012, 2015; Albrechtsen, 2015; Guimarães e Tognoli, 2015; López-Huertas, 2015; Castanha, Rosas e Gracio, 2016, entre outros.

REFERÊNCIAS

- ALBRECHTSEN, H. This is not domain analysis. **Knowledge Organization**, v. 42, n. 8, p. 557-561, 2015.
- ANDERSEN, J.; SKOUVIG, L. **The organization of knowledge caught between global structures and local meaning**. Bingley: Emerald Publishing Limited, 2017.
- ANDERSON, J. D.; PÉREZ-CARBALLO, J. The nature of indexing: how humans and machines analyze messages and texts for retrieval: research, and the nature of human indexing. **Information Processing & Management**, v. 37, n. 2, p. 231-254, 2001.
- ARQUIVO NACIONAL (Brasil). **Dicionário brasileiro de terminologia arquivística**. Rio de Janeiro: Arquivo Nacional, 2005.
- BARITÉ, M. **Glosario sobre organización y representación del conocimiento, clasificación, indización, terminología**. Montevideo: Comisión Sectorial de Investigación Científica, 1997.
- BARITÉ, M. Organización del conocimiento: un nuevo marco teórico-conceptual en bibliotecología y documentación. In: CARRARA, K. (Org.). **Educação, universidade e pesquisa**. Marília, SP: Unesp Marília Publicações; São Paulo: FAPESP, 2001. p. 35-60.
- BEGHTOL, C. Domain analysis, literary warrant, and consensus: the case of fiction studies. **Journal of the American Society for Information Science**, v. 46, n. 1, p. 30-44, 1995.

BLISS, H. E. **The organization of knowledge and the system of the sciences**. New York: Henry Holt and Company, 1929.

BRIET, S. **Qu'est-ce que la documentation?** Paris: Éditions Documentaires Industrielles et Techniques, 1951.

BROWN, J. S.; DUGUID, P. **The social life of information**. Boston: Harvard Business Review Press, 2017.

BUCKLAND, M. K. Document theory. In: HJØRLAND, B.; GNOLI, C. (Eds.) **Encyclopedia of knowledge organization**. [S.l.]: ISKO, 2017.

BUCKLAND, M. K. Information as thing. **Journal of the American Society for Information Science**, v. 45, n. 5, p. 351-360, 1991.

BUSH, V. *As we may think*. **Atlantic Monthly**, v. 176, p. 101-108, 1945.

CASTANHA, R., ROSAS, F.; GRACIO, M. C. C. The complementarity of Hjørland's and Tennis's proposals to domain analysis under bibliometrics. In: GUIMARÃES, J. A. C.; MILANI, S. O.; DODEBEI, V. **Knowledge organization for a sustainable world: challenges and perspectives for cultural, scientific, and technological sharing in a connected world**. Würzburg: Ergon, 2016. p. 219-226.

DAHLBERG, I. Current trends in knowledge organization. In: GARCIA MARCO, F. J. Organización del conocimiento en sistemas de información e documentación 1. **Actas del I Encuentro de ISKO-España**, 1993, Zaragoza: Librería General, 1995. p. 7-25.

DAHLBERG, I. Knowledge organization: a new science? **Knowledge Organization**, v. 33, n. 1, p. 11-19, 2006.

DAHLBERG, I. Knowledge organization: its scopes and possibilities. **Knowledge organization**, v. 20, n. 4, p. 211-222, 1993.

ESTEBAN NAVARRO, M. A.; GARCÍA MARCO, F. J. Las primeras jornadas sobre organización del conocimiento: organización del conocimiento y información científica. **Scire**, v. 1, n. 1, p. 149-157, 1995.

FERNANDES, G. C. O objeto de estudo da ciência da informação. **Informare**, v. 1, n. 1, p. 25-30, 1995.

FERNANDEZ MOLINA, J. C. *et al.* Aspectos éticos de las nuevas tecnologías de la información y la comunicación: y su reflejo en la organización y representación del conocimiento. In: GASCÓN, J.; BURGUILLOS, F.; PONS, A. (Eds.). **La dimensión humana de la organización del conocimiento**. Barcelona: Universitat de Barcelona, 2005. p. 177-186.

GARCIA MARCO, F. J. Avances en organización del conocimiento en España: los II Encuentros sobre Organización del Conocimiento en sistemas de información y documentación. In: GARCIA MARCO, F. J. Organización del conocimiento en sistemas de información y documentación 2. **Actas del II Encuentro de ISKO-España**, 1995, Zaragoza: Librería General, 1997.

GNOLI, C. Ten long-term research questions in knowledge organization. **Knowledge Organization**, v. 35, n. 2/3, p. 137-149, 2007.

GNOLI, C. What is knowledge organization about? **Knowledge Organization**, v. 43, n. 8, p. 668-9, 2016.

GUIMARÃES, J. A. C. Ciência da informação, arquivologia e biblioteconomia: em busca do necessário diálogo entre o universo teórico e os fazeres profissionais. In: FUJITA, M. S. L.; GUIMARÃES, J. A. C. (Orgs.). **Ensino e pesquisa em biblioteconomia no Brasil: a emergência de um novo olhar**. São Paulo: Cultura Acadêmica, 2008. p. 33-44.

GUIMARÃES, J. A. C. Perspectivas de ensino e pesquisa em organização do conhecimento em cursos de biblioteconomia: uma reflexão. In: CARRARA, K. (Org.). **Educação, universidade e pesquisa**. Marília: Unesp Marília Publicações; São Paulo: FAPESP, 2001. p. 61-74.

GUIMARÃES, J. A. C. **Propetip 22 - Gestão de dados de pesquisa**. São Paulo: PROPE/UNESP, 2018.

GUIMARÃES, J. A. C.; TOGNOLI, N. B. Provenance as a domain analysis approach in archival knowledge organization. **Knowledge Organization**, v. 42, n. 8, p. 562-569, 2015.

HALBWACHS, M. **A memória coletiva**. Tradução de Beatriz Sidou. 2.ed. São Paulo: Centauro, 2013.

HEY, T.; STWART, T.; TOLLE, K. (Eds.). **Fourth paradigm: data-intensive scientific discovery**. Washington: Microsoft Research, 2009. Disponível em: <<http://research.microsoft.com/en-us/collaboration/fourthparadigm/>>. Acesso em: 01 out. 2019.

HJØRLAND, B. Domain analysis: a socio-cognitive orientation for information science research. **Bulletin of the American Society for Information Science and Technology**, v. 30, n. 3, p. 17-21, 2004.

HJØRLAND, B. Domain analysis in information science: eleven approaches - traditional as well as innovative. **Journal of Documentation**, v. 58, n. 4, p. 422-462, 2002.

HJØRLAND, B. Fundamentals of knowledge organization. **Knowledge Organization**, v. 30, n. 2, p. 87-111, 2003.

HJØRLAND, B. Knowledge organization. In: HJØRLAND, B.; GNOLI, C. (Eds.) **Encyclopedia of Knowledge Organization**. ISKO, 2017.

HJØRLAND, B. Nine principles of knowledge organization. In: ALBRECHTSEN, H.; OERNAGER, S. **Knowledge organization and quality management**. Frankfurt: Indeks, 1994. p. 91-100.

HJØRLAND, B. What is knowledge organization (KO)? **Knowledge Organization**, v. 35, n. 2/3, p. 86-101, 2008.

HJØRLAND, B.; ALBRECHTSEN, H. Toward a new horizon in Information science: domain-analysis. **Journal of the American Society for Information Science**. v. 46, n. 6, p. 400-425, 1995.

HOMULOS, P. Museums to libraries: a family of collectinmg institutions. **Art Libraries Journal**, v. 15, n. 1, p. 11-13, 1990.

INTERNATIONALSOCIETYFORKNOWLEDGEORGANIZATION (ISKO). **ISKO's mission**. 2021.

JAENECKE, P. To what end knowledge organization? **Knowledge Organization**, v. 21, n. 1, p. 3-11, 1994.

LÓPEZ-HUERTAS, M. J. Domain analysis for interdisciplinary knowledge domains. **Knowledge Organization**, p. 42, n. 8, p. 570-580, 2015.

MAI, J. E. Analysis in indexing: document and domain centered approaches. **Information Processing and Management**, v. 41, p. 599-611, 2005.

MIRANDA, M. L. C. A organização do conhecimento e seus paradigmas científicos: algumas questões epistemológicas. **Informare**, v. 5, n. 2, p. 64-77, 1999.

MUSTAFA EL-HADI, W.; FAVIER, L. Birdging gaps between knowledge organization and digital humanities. In: BABIK, W. **Knowledge organization at the 21st century: between historical patterns and future prospects**. Würzburg: Ergon, 2014. p. 477-485.

ORTEGA, C. D.; LARA, M. L. G. de. A noção de documento: de Otlet aos dias de hoje. **DataGramaZero**, v. 11, n. 2, 2010.

OTLET, P. **Traité de documentation: le livre sur le livre, théorie et pratique**. Bruxelles: Muindaneum, 1934.

PINHO, F. A. **Aspectos éticos em representação do conhecimento: em busca do diálogo entre Antonio García Gutiérrez, Michèle Hudon e Clare Beghtol**. 2006. 116 f. Dissertação (Mestrado em

Ciência da Informação) - Faculdade de Filosofia e Ciências - Universidade Estadual Paulista, Marília, 2006.

REINMANN, G. Knowledge Organization. In: SEEL, N. M. *et al.* (Eds.). **Encyclopedia of the sciences of earning**. Berlin: Springe, 2009.

RODRIGUEZ BRAVO, B. **El documento**: entre la tradición y la renovación. Gijón: Trea, 2002.

RONDINELLI, R. C. **O documento arquivístico ante a realidade digital**: uma revisão conceitual necessária. Rio de Janeiro, Ed. FGV, 2013.

SAN SEGUNDO, R. **Sistemas de organización del conocimiento**: la organización del conocimiento en las bibliotecas españolas. Madrid: Universidad Carlos III de Madrid, Boletín Oficial del Estado, 1996.

SILVA, R. R. **A face oculta do documento**: tradição e inovação no limiar da Ciência da informação. 2009. 331 f. Tese (doutorado) - Universidade Estadual Paulista, Faculdade de Filosofia e Ciências de Marília, 2009.

SMIRAGLIA, R. P. **Domain analysis for knowledge organization**: tools for ontology extraction. Oxford: Chandos, 2015.

SMIRAGLIA, R. P. Epistemology of domain analysis. In: SMIRAGLIA, R. P.; LEE, H. L. (Eds.). **Cultural frames of knowledge**. Würzburg: Ergon, 2012, p.111-124.

SMIT, J. W.; BARRETO, A. A. Ciência da informação: base conceitual para a formação do profissional. In: VALENTIM, M. L. (Org.). **Formação do profissional da informação**. São Paulo: Polis, 2002. p. 9-23.

SZOSTAK, R.; GNOLI, C.; LÓPEZ-HUERTAS, M. J. **Interdisciplinary knowledge organization**. Switzerland: Springer, 2016.

TEDESCO, J. C. **Nas cercanias da memória**: temporalidade, experiência e narração. 2. ed. Passo Fundo, RS: Ed. Universidade de Passo Fundo, 2014.

TENNIS, J. T. Two axes of domains for domain analysis. **Knowledge Organization**, v. 30, n. 3/4, p. 191-195, 2003.

TENNIS, J. T. What does a domain analysis look like in form, function, and genre? **Brazilian Journal of Information Science**, v. 6, n. 1, p. 3-14, 2012.

TOGNOLI, N. B.; RODRIGUES, A. C.; GUIMARÃES, J. A. C. Definindo o conhecimento arquivístico: estruturas conceituais. **Informação & Informação**, v. 24, n. 2, p. 58-75, 2019.

WARDLE, C. **Information disorder: toward an interdisciplinary framework for research and policymaking**. Strasbourg: Council of Europe, 2017.

A ORGANIZAÇÃO DO CONHECIMENTO E AS HUMANIDADES DIGITAIS: PONTOS DE INTERSEÇÃO

Johanna W. Smit

AS HUMANIDADES DIGITAIS

As humanidades digitais (HD) seriam um novo nome para velhos desafios? Tentarei responder à pergunta, mas necessário se faz iniciar pelo detalhamento de alguns conceitos acerca destas HD.

Na literatura encontram-se sinônimos para as HD: computação em humanidades, interseção da computação com as disciplinas das ciências humanas (e eu acrescentaria sociais), e afirmações tais como “consequência da opção da sociedade pelo digital” ou, simplesmente “as humanidades do presente atual” (Sousa, 2011). As HD são sempre conceituadas como uma transdisciplina ou interdisciplina, ou ainda como um conjunto de práticas que incor-

pora todos os métodos, sistemas e perspectivas heurísticas relacionadas ao digital nas ciências humanas e sociais (Dacps, 2011). Algumas conceituações vão além: “atividade interdisciplinar que transfere para os meios digitais o trabalho tradicional com textos, objetos culturais e outros dados, com isso estendendo radicalmente seus usos potenciais e impactos de textos, objetos culturais e outros dados” (O'Donnell citado por Sousa, 2011). O manifesto das HD, datado de 2011, convoca “à integração da cultura digital na definição da cultura geral do século XXI” (Dacos, 2011). Como essa proposta é muito ambiciosa, e vai muito além de minha capacidade de reflexão, proponho voltar ao nosso cotidiano - o atual modo de produção do conhecimento na comunidade das ciências humanas e sociais (CHS).

A junção das CHS com a tecnologia pode ser entendida como um processo decorrente tanto do contínuo progresso na área tecnológica como na concepção do papel que pode ser atribuído à tecnologia pelas CHS. Passa-se de uma visão utilitária (como a tecnologia pode ajudar as CHS) para uma visão integrativa, de verdadeira simbiose entre as duas áreas.

A visão utilitária pode ser caracterizada pela **justaposição**, na qual os esforços são concentrados na utilização da tecnologia para otimização do trabalho nas CHS - uma decorrência lógica do desenvolvimento da tecnologia -, que passou a estar presente em praticamente todas as esferas de atividades da sociedade, inclusive no mundo da pesquisa, tanto das ciências duras, exatas e precisas, quanto das CHS. Interessante notar que as primeiras experiências de utilização da computação nas

CHS se deram nas áreas da literatura e linguística, arqueologia e demografia (Alves, 2016).

A bibliografia das HD enfatiza a tecnologia digital, como seu intitulado deixa claro. No entanto, as experiências que citam a interação entre as CHS e a tecnologia voltam no tempo para a era pré-digital. Refiro-me a experiências que hoje podem parecer folclóricas, mas que a partir dos anos 50 do século XX recorreram a cartões perfurados ou ao sistema *peek-a-boo* na condição de recursos mecanográficos. Concebidos como recursos para potencializar as correlações entre dados ou informações, seu alcance era, evidentemente, limitado, mas não me parece justo ignorar importantes experiências realizadas com esta tecnologia pré-digital, até porque os desafios são os mesmos, como tentarei detalhar adiante. Citado como precursor, o padre jesuíta Roberto Busa desenvolveu análises linguísticas e literárias dos escritos de São Tomás de Aquino com a ajuda de cartões perfurados e computadores de grande porte da IBM: o Index Thomisticus, datado de 1949 (Alves, 2016; Pimenta, 2019; Rockembach, 2019).

Parece lógico, e não há nenhuma avaliação crítica nesta constatação, que o ingresso da tecnologia nas CHS inicialmente foi valorizado pela facilidade que a tecnologia, e depois a *web*, representavam

- a) para a pesquisa por informações, à medida que o acesso à informação não ficava mais restrito à presença física do pesquisador;
- b) para a divulgação de conteúdos, tornada instantânea e também não mais restrita a limites geográficos (mas ainda hoje, em parte, restrita a limites idiomáticos);

c) no que diz respeito às ciências da informação, a flexibilidade conferida pela tecnologia para registrar, representar, ordenar e reordenar os dados - propiciando diferentes acessos à mesma informação - foi um marco divisor importantíssimo para a área. A organização de bases de dados e, posteriormente, o desenvolvimento de ferramentas cada vez mais amigáveis para a busca e recuperação da informação, hoje naturalizados, significaram, no entanto, um importante progresso. O advento da *web* propiciou novos desenvolvimentos no que diz respeito à representação, à recuperação e ao acesso às informações, ao introduzir recursos para o estabelecimento de relações entre os dados de forma dinâmica (com destaque para a *web* semântica e a conexão entre dados estruturados) e a organização de repositórios digitais (Coneghan; Segundo, 2017).

O progresso destas atividades mediadas pela tecnologia é tão evidente quanto importante, e não deve ser menosprezado de forma alguma; mas creio que os aspectos positivos obscureceram por certo tempo a percepção de seus limites ou limitações. A comunidade das CHS passou a ser confrontada com novos problemas, ainda hoje candentes, tais como a questão dos direitos autorais, direito à imagem e à privacidade, questões muito complexas relacionadas à preservação digital com requisitos de integridade e autenticidade de dados e documentos, a rápida obsolescência de muitos dados e *links* digitais (Rockembach, 2019), a conservação do patrimônio digital (Rollo, 2020), além do convívio cotidiano com o fenômeno das *fake news*.

Embora este início da convivência entre as CHS e a tecnologia não possa ser restrito ao aspecto quantitativo de produção, divulgação e acesso a dados, documentos e objetos, certamente esse aspecto determinou o sucesso da tecnologia em sua interface com as CHS, apesar de certas resistências que, até hoje, podem ser observadas. Não se pode considerar que o aspecto qualitativo fosse ignorado, mas a adesão à tecnologia foi impulsionada pelas facilidades que ela representava, tanto quantitativas quanto espaciais (borrando os limites espaciais e geográficos) e temporais (acesso instantâneo).

Ao lado de uma exigência maior em relação às possibilidades oferecidas pela computação, nota-se também a conscientização em relação a questões quantitativas, desveladas pelo início da interação entre HD e CHS: refiro-me às discussões sobre **big data**, ou seja, conjuntos de dados grandes demais para serem analisados por formas tradicionais de pesquisa, e a discussões, ainda hoje muito frequentes, sobre as metodologias e ferramentas para conseguir navegar neste mundo quantitativamente ilimitado. As diferentes métricas e os diferentes algoritmos continuam sendo debatidos e aperfeiçoados.

Até aqui estamos falando de desafios que apareceram com o advento das HD, advento este que tanto significou a otimização de processos relacionados à coleta, organização, divulgação e mediação de conhecimentos como também a convivência com novos problemas, demandando novas soluções (proteção de dados, *fake news* etc.). A distinção entre fontes de dados primários e critérios para sua representação e manipulação visando à análise estão até

hoje na ordem do dia e são, mais do que nunca, muito pertinentes. Desnecessário destacar que a produção de dados, o acesso a dados e a publicação de dados, incluindo suas representações e categorizações (ou classificações) não são neutros, embora extremamente importantes! O recurso à tecnologia (hoje digital) é fundamental e não se concebe mais não recorrer a ela. A organização de bases de conhecimento e motores capazes de correlacionar os dados, característicos da inteligência artificial, também estão na ordem do dia, na condição de recursos passíveis de enriquecimento dos trabalhos nas CHS. Discussões sobre dados abertos, ciência aberta e metodologias abertas, muito promissoras, são consequências naturais deste momento, renunciando (nem sempre de forma explícita ou consciente) as HD em seu potencial máximo.

Nota-se uma evolução natural nas atividades das CHS:

não se trata de encomendar um banco de dados, uma plataforma, um desenho de visualização a um 'especialista em computação'; trata-se, efetivamente, de conceber classificações, indicadores e formas de leitura em conjunto com os profissionais da área da computação (Sousa, 2011, p. 5).

Os projetos não demandam mais uma justaposição, mas uma verdadeira simbiose entre duas ou mais especialidades, sendo uma a da computação, cujos avanços avassaladores induzem uma exigência cada vez maior em relação aos procedimentos clássicos de acesso, uso, produção, publicação, mediação e divulgação de dados e informações.

O volume de dados, critérios para a agregação de dados e a correlação de dados passam a receber maior atenção e impõem a necessidade de um trabalho colaborativo, no qual os especialistas das diferentes CHS devem interagir com os especialistas da área tecnológica, numa **simbiose** que produz algo novo, algo que os diferentes pesquisadores não teriam como visualizar isoladamente. Esta união de especialidades (e não a subordinação de uma abordagem tecnológica a uma ciência humana ou social) teve como consequência a necessidade de inventar um **novo modo de trabalho**. Ao volume de dados contrapõe-se uma abordagem mais qualitativa, pois a união com a abordagem tecnológica fez emergir a necessidade de se ter clareza acerca dos dados, de como foram produzidos, como se relacionam ou como os objetivos da pesquisa sugerem relacionamentos entre os mesmos. A introdução da tecnologia no trabalho das CHS supõe uma consequência muito mais evidente do que a simples otimização do trabalho, à medida que introduz uma abordagem no modo de trabalho da produção científica que, para simplificar, chamarei **lógica**. Em resumo, além das vantagens representadas pelo início da computação nas CHS, a progressão para a fase atual representa um desafio – imenso para muitos – da passagem de um modo de produção e escrita do conhecimento caracterizado pela **retórica**, para um modo de produção subordinado à **lógica**. As HD, vistas em todo seu potencial, significam uma revolução no formalismo acadêmico (Alves, 2016), reconfigurando as CHS. As “trilhas de citação” são substituídas por argumentos claramente explicitados e correlacionados a outros argumentos que repetem ou contradizem (parcialmente, integralmente) outros argumentos. Trata-se, por-

tanto, de explicitar dados descritos de forma padronizada e tipos de correlações lógicas que podem ser estabelecidas entre dados e conclusões.

Esses projetos demandam, assim, uma profunda intercompreensão entre esses pesquisadores, todos eles, neste movimento, tornando-se sujeitos da tecnologia - no caso dos humanistas, aprendendo, efetivamente, a “pensar como seus computadores” (Sousa, 2011, p. 5).

Em artigo que relata a implantação de um Laboratório em Rede de Humanidades Digitais, residente no IBICT, Pimenta, após elencar uma série de ações que formataram o laboratório, afirma que estes pontos

parecem auxiliar-nos a compreender o quanto o impacto das novas tecnologias da informação e comunicação (NTIC), notadamente aquelas que são mediadas pelos recursos digitais e computacionais, está diretamente relacionado às novas formas metodológicas mediadas pelo digital para sua subsequente produção de conhecimento. Ademais, é nesse contexto que se sugere refletir e lançar novos questionamentos sobre o que chamamos aqui de uma nova “escrita” e sua respectiva “gramática” para a produção do conhecimento que se torna tangível a partir dessas práticas laboratoriais (Pimenta, 2019, p. 6).

Voltarei mais tarde às questões acima elencadas: “pensar como os computadores” e a “nova escrita” com sua respectiva “gramática”. O fato é que muitos, nas CHS, não trabalham assim hoje e tampouco é assim que eu estou trabalhando neste momento.

A seguir descreverei um exemplo das HD através do trabalho de um arqueólogo que recorreu à computação ainda na década de 50 do século passado, adotou as propostas das HD nos anos 80 (embora naquele momento a expressão HD ainda não fosse usual) e elaborou importante reflexão sobre os limites e as dificuldades inerentes ao conceito e ao exercício das HD nas CHS.

AS HD NA ARQUEOLOGIA: UM EXEMPLO

Jean-Claude Gardin (1925-2013) fuge da França aos 15 anos, indo para a Inglaterra para se alistar na Marinha Francesa Livre sob comando do general de Gaulle. Cuidou de torpedos e no final da guerra tinha a patente equivalente à de tenente. Recebeu 3 condecorações, sobre as quais nunca falou. Na Inglaterra, formou-se em Economia, voltou para a França onde se formou em linguística, história das religiões, etnologia, línguas orientais vivas e língua persa (Djindjian, 2016).

A reflexão sobre a arquitetura das construções científicas (ou acadêmicas) nas disciplinas históricas e literárias constitui o eixo de sua atuação acadêmica, complementado por provocações relacionadas ao modo de produção do conhecimento nas CHS quando este se apresenta distanciado do pensamento formal. Em seu conjunto, sua trajetória profissional fornece uma ilustração exemplar do potencial e das dificuldades inerentes às HD.

O período de 1952 a 1969 resume as primeiras incursões da tecnologia no trabalho arqueológico. Tendo sido enviado a Beirute

em 1952, para participar dos trabalhos arqueológicos no Instituto Francês de Beirute, em seguida é enviado ao Afeganistão para participar de prospecções em dois sítios arqueológicos. Em 1955 ele volta para Beirute e explica a seu chefe (Henry Seyrig, um conhecido arqueólogo) que pensa em largar a arqueologia porque o caráter repetitivo do registro de dados de descrição e indexação dos achados arqueológicos não o atraía. Seyrig então pede que ele explique o problema e Gardin propõe utilizar a mecanografia¹ para a indexação do acervo. Ainda naquele ano de 1955 o CNRS² cria o Centro Mecanográfico de Documentação Arqueológica³, sob direção de Gardin.

Entre 1955 e 1969 o centro reuniu diferentes especialistas e elaborou diferentes códigos descritivos (hoje diríamos ontologias) para análise de objetos arqueológicos, iconografia e textos⁴. O projeto de análise conceitual do Alcorão com fichas perfuradas (Allard; Elziere; Gardin; Hours, 1963) reuniu uma equipe pluri-disciplinar; este trânsito entre diferentes disciplinas, concebidas como parceiras e sem hierarquia entre disciplinas mais ou menos importantes, caracterizou o trabalho de Gardin desde o início, renunciando o discurso interdisciplinar atual sobre as HD.

1. Depois chamada informática, computação ou simplesmente tecnologia.

2. Centre National de la Recherche Scientifique francês, correspondendo ao CNPq brasileiro.

3. Depois denominado Centro de Análise Documentária em Arqueologia – CADA.

4. A título de exemplo, cito quatro códigos para descrição padronizada de artefatos (Gardin, 1958).

O investimento em mecanografia, recorrendo a computadores de grande porte da IBM, corresponde a uma tendência geral do pós-guerra e nos remete à publicação do *Index Thomisticus*, de 1949 (citado anteriormente), considerada a primeira utilização da tecnologia nas humanidades.

Em 1954 Gardin obtém recursos da Rockefeller Foundation visando o intercâmbio com arqueólogos dos Estados Unidos, Grã-Bretanha e Alemanha para detalhar aspectos metodológicos do trabalho arqueológico. Nas palestras Gardin chamou a atenção para o fato de que a introdução da mecanografia não constituía o aspecto mais importante de sua proposta, mas o reconhecimento das implicações teóricas da metodologia na pesquisa arqueológica e nas fundamentações das reconstruções históricas. A informatização pressupunha a explicitação completa de dados e raciocínios, e esse era o aspecto por ele valorizado.

Naquela época ele discorre sobre a importância da abertura de dados de pesquisa, conservando os dados sob forma analítica, colocando-os à disposição dos estudiosos, para que, hipoteticamente, sínteses diferentes pudessem ser geradas, e para evitar que os mesmos materiais fossem descritos repetidamente (Gardin, 1962). As discussões acerca dos dados abertos continuam presentes hoje, incluindo aspectos econômicos, para evitar o financiamento de coleta de dados já levantados por outra pesquisa.

Além da introdução da mecanografia, outra introdução importante nos assuntos que nos interessam aqui também data daquela época: a atenção às questões lógico-linguísticas e semânticas presentes na produção de dados e textos e em sua organização e

recuperação. A introdução dos aspectos linguísticos e semânticos inseminou de forma duradoura as discussões sobre a organização do conhecimento, ao reconhecer que a descrição, indexação e classificação para a recuperação de informações passam, obrigatoriamente, pela linguagem natural. Muitas referências podem ser citadas, mas enumero aqui apenas os textos iniciais da abordagem linguística: Gardin, 1955; 1966; 1967; 1970; 1973.

Os problemas inerentes à introdução da informática nas CHS ficam evidenciados na estrutura organizacional do CNRS: entre 1970 e 1975, o centro é rebatizado, tornando-se uma Unidade de Pesquisa em Análise Documentária e Cálculo Arqueológico. Empregando analistas de sistemas, matemáticos, arqueólogos e outros especialistas, a concepção pluridisciplinar leva a fortes embates entre dois grupos de profissionais: os profissionais de informática e estatísticos, de um lado, e de outro lado os demais especialistas. Os primeiros queriam desenvolver algoritmos, e o segundo grupo, de diferentes especialistas, queria fazer com que o primeiro trabalhasse em seus projetos, apesar de perseguirem objetivos de pesquisa pouco claros e excessivamente hermenêuticos, na opinião dos primeiros. A dualidade anteriormente apontada entre o modo de produção do conhecimento retórico contraposto a um modo de produção lógico gerou uma dificuldade de diálogo entre os dois grupos e levou ao fechamento da unidade.

Gardin então volta ao trabalho de campo arqueológico, longe dos gabinetes, ao criar em 1974 o laboratório “Arqueologia da Ásia Central”, coordenando escavações em dois sítios no Afeganistão. A invasão militar soviética no Afeganistão, em 1978, redundou

no encerramento do trabalho local, mas os dados científicos coletados foram enviados à França, dando início a uma nova etapa que novamente nos aproxima das HD, quando ele reflete sobre a construção do conhecimento em arqueologia. A proposta de uma **análise logicista**, que se opõe às construções hermenêuticas baseadas em conceitos iniciais insuficientemente explicitados, prioriza a validação lógica do raciocínio que sustenta a construção do conhecimento em arqueologia. Ou seja, sua proposta objetiva a verificabilidade das argumentações pela adoção de uma linguagem formal desenvolvida a partir da linguagem natural na qual as CHS tradicionalmente constroem seu conhecimento. A proposta, em síntese, se desdobra em três partes (Gardin, 1974; 1979):

- uma base de fatos ou dados, formalmente descritos;
- um conjunto de regras de inferência (os motores de inferência dos sistemas especialistas), igualmente explicitados e formalizados e que são mobilizados para estabelecer as relações entre fatos ou dados. Interessante notar que já naquela época a complexidade das inferências que subjazem na construção do conhecimento arqueológico era considerada muito mais complexa do que inicialmente suposto;
- a clara distinção entre fatos ou dados descritos pela pesquisa e argumentos provenientes de outras pesquisas e respectiva bibliografia.

A formalização das três partes acima enumeradas permitiria, segundo Gardin, produzir um conhecimento verificável, que pudesse ser conferido ou auditado e, portanto, seria “científico”. Desnecessário dizer que a proposta encontrou muita resistência, até

porque ela ignora uma prática secular de construção do conhecimento nas CHS, na qual a retórica (a construção do conhecimento através da linguagem natural) prevalece, mas também por receio de que os textos elaborados pelos pesquisadores poderiam ser auditados e potenciais erros de raciocínio apontados. Em resposta aos pesquisadores receosos ou francamente contrários ao logicismo, ainda na década de 80, uma bibliografia se dedicou a demonstrar a fragilidade de muitas construções nas CHS e que Gardin resumia pela expressão “anything goes”⁵.

A ênfase na formalização representa tanto o maior aporte que Gardin trouxe para as CHS como está na base das atuais discussões sobre as HD, embora na década de 80 a expressão HD ainda não fosse utilizada. Enormes progressos foram verificados na inteligência artificial ao longo das últimas décadas, os sistemas especialistas ficaram mais espertos e as regras de inferência mais potentes, mas o desafio da formalização do conhecimento nas CHS permanece como um objetivo dificilmente alcançado. Essa constatação, no entanto, não deve ser entendida como uma confissão de impossibilidade e tampouco como um argumento para menosprezar os imensos progressos alcançados no entrelaçamento de dados, palavras ou objetos com sua descrição, organização e possibilidades de acesso.

Muito embora a formalização do conhecimento e do pensamento não fosse novidade na teoria, sua aplicação prática se demonstrou muito mais complexa do que poderia parecer inicialmente, o que

5. Análises críticas de estudos de obras literárias, poéticas e musicais, além das construções arqueológicas, dão uma ideia do potencial destrutivo da aplicação de uma lógica formal: cf. Gardin, 1982; 1987; Gardin; Lagrange; Martin; Molino; Smit, 1981.

levou Gardin a abrir nova trincheira, ao chamar a atenção para a obesidade informacional. Consequência de uma construção do conhecimento que se caracteriza, nas CHS, pela prevalência do aspecto cumulativo pelo qual x análises são possíveis, sem que se veja claramente a relação entre elas e o quanto houve de acréscimo no conhecimento, uma constatação se impõe e ela ilumina uma faceta do estado atual das construções em CHS: a saber, a impossibilidade de acompanhar toda a bibliografia de qualquer assunto, por mais que seu âmbito tenha sido muito bem delimitado. A prática das publicações científicas leva Gardin a constatar que a organização do conhecimento, através da indexação, é refém de uma indústria de publicações científicas nas CHS, repletas de estudos pobremente fundamentados, nos quais a retórica prevalece sobre a formalização do raciocínio. O desequilíbrio entre a produção de textos científicos e o potencial – cognitivo e financeiro – de seu consumo é tratado em uma palestra de abertura do encontro da ISKO-França (Gardin, 2001), retomando ideias anteriormente expostas na British Library (Gardin, 1987).

O logicismo incorporado às publicações científicas corresponde à necessidade de transparência das construções em CHS e ataca ao mesmo tempo o problema da inflação de publicações (em número e volume). Inspirado no positivismo lógico, contrapõe-se aos impasses gerados pela busca utópica da exaustividade, ao separar claramente a representação dos fatos empíricos elementares (constituição de bases de dados) - que sempre será única e dependente da interpretação buscada (hipótese de pesquisa) – e a representação das operações lógicas exercidas sobre os fatos

empíricos para chegar a conclusões. O paralelo com as HD, que pregam que o pesquisador “pense como um computador” é evidente quando a formalização é vista como condição necessária para que o processamento digital possa ser efetivado.

Finalizo pela descrição de dois exemplos práticos de HD nas CHS, inspirados nas propostas de reescrita na produção de textos científicos experimentados por Gardin e que sintetizam as diferentes fases de sua trajetória nas HD: o projeto Arkeotek e a “tradução” de um livro para um CD-ROM.

O **projeto Arkeotek**, com sua respectiva revista (*The Arkeotek Journal*)⁶, existiu entre 2007 e 2010. A revista era dedicada à publicação de *corpora* experimentais e etnográficos com valor referencial para a identificação e caracterização tecnológica de vestígios pré-históricos. Os *corpora* compreendem diferentes materiais (pedras, ossos, cerâmicas, fios, conchas, madeira etc.). O projeto representa uma síntese das propostas que, de certa forma, já vinham sendo desenvolvidas por Gardin desde 1955, ao voltar sua atenção para novas formas de publicação (que tornem a indexação e a recuperação da informação mais eficientes), bem como para a questão da cumulatividade nas CHS, com especial ênfase na falta de “capitalização do conhecimento teórico”: o contínuo recomeço na produção científica, ignorando o que já foi assentado. O projeto pode ser considerado o testamento de Gardin, que lança mão de uma interface multimídia ao publicar estudos arqueológicos que separam claramente os dados iniciais da pesquisa, as interpretações e outros dados, mobilizados para enriquecer

6. (<http://www.thearkeotekjournal.org/>)

e diversificar a argumentação, de acordo com os princípios do logicismo (Gardin; Roux, 2004).

the
arkeotekjournal

revue de l'association européenne d'archéologie des techniques

Accueil

À propos de la revue

Soumettre un article

Corpus

LES JARRES DE MARIAGE DÉCORÉES DU DELTA INTÉRIEUR DU NIGER (MALI) : ESSAI DE DÉLIMITATION ARCHÉOLOGIQUE D'UN TERRITOIRE ETHNIQUE

DONNÉES CONTEXTUELLES ET CORPUS

[+] Contexte social

[+] Contexte économique

[+] La céramique: corpus et déplacements

LES CÉRAMIQUES RICHEMENT DÉCORÉES

[+] Mode d'acquisition

[+] Lieux d'acquisition et distances

[+] Lieux d'acquisition et approvisionnement

LES CÉRAMIQUES COMMUNES

[+] Mode d'acquisition

[+] Lieux d'achat et distance

[+] Lieux d'achat et dispersion

STRUCTURE SPATIALE ET ETHNIE

[+] Structure spatiale

[+] Zones de répartition et Archéologie

Accéder aux illustrations de l'article

all.diagrammes

LES JARRES DE MARIAGE DÉCORÉES DU DELTA INTÉRIEUR DU NIGER (MALI) : ESSAI DE DÉLIMITATION ARCHÉOLOGIQUE D'UN TERRITOIRE ETHNIQUE

Alain GALLAY

Pour citer cet article

Mots-clés

Mali, céramiques, spécialisation artisanale, réseaux de distribution, territoire ethnique, cartographie.

INTRODUCTION

De 1988 à 1995, le Département d'Anthropologie et d'Ecologie de l'Université de Genève a mené, au Mali, une série d'enquêtes ethnoarchéologiques sur la céramique traditionnelle parmi les diverses populations occupant le Delta intérieur du Niger. L'enquête, dirigée par l'auteur de ce travail en collaboration avec Eric Huysecom, a porté sur une centaine de villages, plus de 300 potières et quelque 150 complexes d'habitation, totalisant près de 6500 céramiques documentées et, pour la plupart, dessinées (Gallay 1993, 2000, 2005, Huysecom, Mayor 1993, de Ceuninck 1993, 2000, Gallay et alii 1996, 1998, Burri 1996, 2003).

La présente démonstration s'inscrit dans le cadre de la recherche et de l'élaboration de modèles permettant de délimiter l'espace géographique occupé par une population particulière (ci-après « ethnies ») à partir de vestiges archéologiques. Elle repose sur l'observation qu'il existe dans la zone considérée des traditions céramiques parfaitement individualisées, dont les composantes matérielles, les céramiques, permettent de délimiter deux zones concentriques, l'une, propre aux céramiques de mariage, superposable à l'aire occupée par le groupe ethnique, l'autre, plus large, propre aux céramiques communes et caractéristique des réseaux de distribution en relation avec une économie à marchés périphériques (Bohannan, Dalton 1962).

La démonstration s'articule en quatre volets :

1. Contexte d'observation

Le premier (P01 à P08) présente le contexte techno-économique et social caractéristique du Delta intérieur du Niger et situe le contexte d'application ou d'actualisation du modèle proposé.

2. Approvisionnement des concessions de la zone de production en céramiques richement décorées

Le second volet (P09 à P017) concerne l'origine des céramiques présentes dans une habitation occupée par une famille étendue (concession dans la terminologie consacrée) et la mise en évidence des mécanismes assurant l'approvisionnement d'une concession en céramiques, notamment pour ce qui concerne les céramiques richement décorées souvent offertes à la nouvelle épouse. Les divers mécanismes considérés permettent de délimiter une zone d'extension restreinte correspondant à la zone de production de la tradition céramique considérée.

3. Diffusion des céramiques communes dans la zone de production

Le troisième volet (P018 à P024) est consacré aux modalités de diffusion de la céramique commune dans la zone de

Figura 1 – extrato de página da revista de 2007, com um artigo sobre os jarros decorados usados em casamentos no Mali. À esquerda se encontram os *links* para os dados iniciais e o *corpus*, cerâmicas ricamente decoradas e cerâmicas comuns, bem como os *links* para as conclusões sobre a distribuição espacial dos jarros. Fonte: <http://www.thearkeotekjournal.org/>

HUMANIDADES DIGITAIS, BIG DATA E PESQUISA CIENTÍFICA | 141

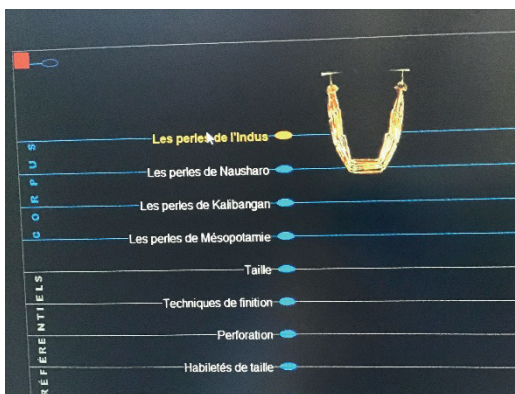
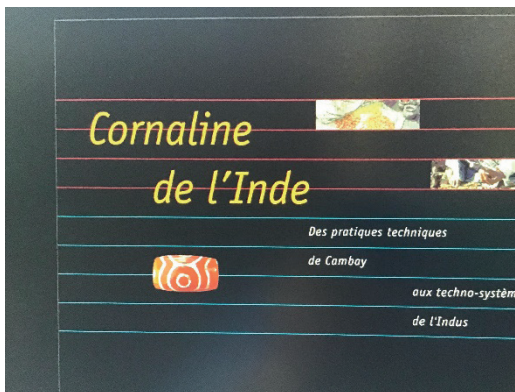
As **peças de cornalina**, usadas para elaborar colares e outros objetos de adorno, localizadas em sítios arqueológicos da idade do bronze, identificados no vale do rio Indo, no Paquistão, foram estudadas e resultaram em duas publicações:

- uma publicação clássica, sob forma de um livro de cerca de 500 páginas (Roux, 2000) no qual diferentes autores discorrem a respeito do significado das peças encontradas, a partir de diferentes pontos de vista: tecnologia para fabricação das peças, divisão do trabalho envolvido, estratificações sociais, economia, comércio internacional etc.
- um CD-ROM encartado no livro que, sem perda de informação, expõe a base de conhecimento (dados e regras de inferência), inspirada no modelo logicista, objetivando responder às necessidades de pesquisa de futuros consulentes e recorrendo à vantagem das facilidades de navegação hipertextual (Gardin; Roux, 2004).

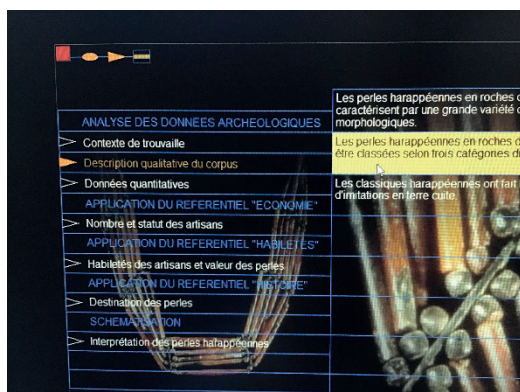
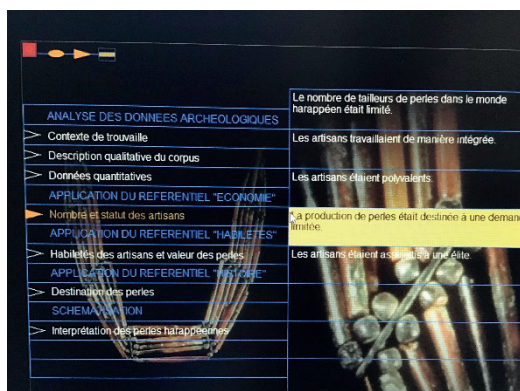
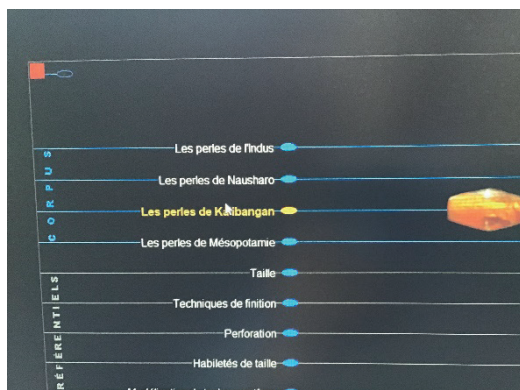
Pela experiência, os leitores do livro são explicitamente diferenciados dos consulentes do CD-ROM. Os autores do CD-ROM, que trabalharam em conjunto com os arqueólogos e especialistas em computação editorial, manifestaram dúvida se o livro teria muitos leitores, tanto pelo seu caráter retórico (o que impede uma visualização clara de dados e inferências), como também pelo crescimento exponencial de publicações, impedindo que qualquer especialista consiga ler tudo o que foi publicado sobre seu campo de interesse. Em resumo, o CD-ROM é formatado para a consulta e o livro, para a leitura. Os autores acreditam que a consulta seja

muito mais frequente do que a leitura⁷.

O CD-ROM, que apresenta o conteúdo do livro sob forma “consultável”, constitui tanto uma exemplificação da argumentação de Gardin sobre as publicações em CHS e respectivo modo de produção do conhecimento, como também suas limitações: a tecnologia empregada em 2000 para elaborar o CD-ROM já está, parcialmente, ultrapassada e, portanto, não mais acessível via computador (em especial, curtos filmes que mostram o trabalho de fragmentação das pedras e polimento das peças)...



7. Essa afirmação data de 2004, mas certamente o volume de publicações foi exponencialmente aumentado neste meio tempo.



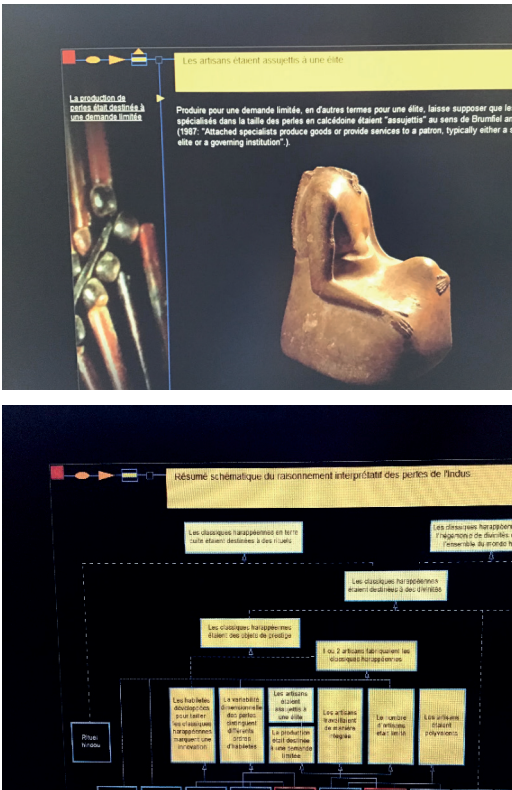


Figura 2 – telas do CD-ROM. Fonte: Roux; Blanco, 2000.

O DESAFIO DAS HD PARA AS CHS

Se Gardin foi reconhecido pela sua influência na arqueologia na segunda metade do século XX, tanto na França como em outros países, com ênfase na introdução da informática para tornar os procedimentos de construção do conhecimento verificáveis, suas propostas relacionadas à descrição formalizada e à recuperação

da informação são menos lembradas. O objetivo de Gardin consistiu, desde o início, na elaboração de representações formalizadas dos textos científicos, ultrapassando o estágio retórico ao preservar o caráter multidimensional das pesquisas, visando potencializar múltiplas combinatórias entre objetos, dados e fatos e as relações que os unem. A resistência a sua proposta de reescrita – embutido nas HD – aponta para uma questão sem resposta, ao que consta: o modo de produção de conhecimento nas CHS deve ser alterado? É importante ou necessário alterá-lo?

Em resumo, repetindo Riobaldo no *Grande sertão: veredas*,
me inventei neste gosto de especular ideia.

REFERÊNCIAS

ALLARD, M. *et al.* **Analyse conceptuelle du coran sur cartes perforées**. Paris: Mouton, École Pratique des Hautes Études, 1963.

ALVES, D. As humanidades digitais como uma comunidade de práticas dentro do formalismo acadêmico: dos exemplos internacionais ao caso português. **Ler História**, n. 69, p. 91-103, 2016. Disponível em: <<http://journals.openedition.org/lerhistoria/2496>>. Acesso em: 3 ago. 2021.

CONEGHIAN, C. S.; SANTAREM SEGUNDO, J. E. Europeana no Linked Open Data: conceitos de web semântica na dimensão aplicada das humanidades digitais. **Encontros Bibli**, v. 22, n. 48, p. 88-99, 2017.

DACOS, M. Manifesto das humanidades digitais. *In*: HUMANIDADES digitais: grupo de pesquisas Universidade de São Paulo. São Paulo, set. 2011.

DJINDJIAN, F.; GARDIN, J.-C. Un archéologue libre! **Les Nouvelles de l'Archéologie**, n. 144, p. 4-9, 2016.

GARDIN, J.-C. Document analysis and linguistic theory. **Journal of Documentation**, v. 29, n. 2, p. 137-168, 1973.

GARDIN, J.-C. Documentation sur cartes perforées et travaux sur ordinateurs dans les sciences humaines. **Revue Internationale de Documentation**, v. 29, n. 3, p. 84-92, 1962.

GARDIN, J.-C. Éléments d'un modèle pour la description des lexiques documentaires. **Bulletin des Bibliothèques de France**, n. 5, p. 171-182, 1966.

GARDIN, J.-C. **Expert systems and scholarly publications**. London: The British Library, 1987.

GARDIN, J.-C. Four codes for the description of artefacts: an essay in archeological technique and theory. **American Anthropologist**, v. 60, n. 2, p. 335-357, 1958.

GARDIN, J.-C. *et al.* **La logique du plausible**: essais d'épistémologie pratique en sciences humaines. Paris: Éditions de la Maison des Sciences de l'Homme, 1981.

GARDIN, J.-C. La logique, naturelle ou autre, dans les constructions de sciences humaines. **Revue Européenne des Sciences Sociales**, v. 25, n. 77, p. 179-195, 1987.

GARDIN, J.-C. Lectures plurielles et sciences singulières de la littérature. **Diogène**, n. 118, p. 3-14, 1982.

GARDIN, J.-C. **Les analyses de discours**. Neuchâtel: Delachaux et Niestlé, 1974.

GARDIN, J.-C. Problèmes de documentation. **Diogène**, n. 11, p. 107-124, 1955.

GARDIN, J.-C. Procédures d'analyse sémantique dans les sciences humaines. *In*: POUILLON, J.; MARANDA, P. **Échanges et commu-**

nications: mélanges offerts à **Claude Lévi-Strauss à l'occasion de son 60^{ème}** anniversaire. Mouton: La Haye, 1970. p. 628-657.

GARDIN, J.-C. Recherches sur l'indexation automatique des documents scientifiques. **Revue d'Informatique et de recherche opérationnelle**, v. 1, n. 6, p. 27-46, 1967.

GARDIN, J.-C. **Une archéologie théorique**. Paris: Hachette, 1979.

GARDIN, J.-C. Vers un remodelage des publications savantes: ses rapports avec les sciences de l'information. *In*: CHAUDIRON, S.; FLUHR, C. **Filtrage et résumé automatique de l'information sur les réseaux**. Nanterre: Université de Paris-X, 2001.

GARDIN, J.-C.; ROUX, V. The Arkeotek project: a european network for knowledge bases in the archaeology of techniques. **Archeologia e Calcolatori**, n. 15, p. 25-40, 2004.

PIMENTA, R. M. Das iniciativas em humanidades digitais e suas materialidades: relato de um laboratório em construção contínua. **Memória e Informação**, Rio de Janeiro, v. 3, n. 1, p. 1-14, 2019.

ROCKEMBACH, M. Arquivamento da web no contexto das humanidades digitais: da produção à preservação da informação digital. **Liinc em Revista**, Rio de Janeiro, v. 15, n. 1, p. 131-139, 2019.

ROLLO, M. F. Desafios e responsabilidades das humanidades digitais: preservar a memória, valorizar o patrimônio, promover e disseminar o conhecimento. O programa Memória para Todos.

Estudos Históricos, Rio de Janeiro, v. 33, n. 69, p. 19-44, 2020.

ROUX, V. (Ed.). **Cornaline de l'Inde**: des pratiques techniques de cambay aux techno-systèmes de l'Indus. Paris: Éditions de la Maison des Sciences de l'Homme, 2000.

ROUX, V.; BLASCO, P. **Cornaline de l'Inde**. Paris: Éditions de la Maison des Sciences de l'Homme, 2000.

SOUSA, M. C. P. de. Humanidades digitais? Um breve panorama. *In*: HUMANIDADES digitais: grupo de pesquisas Universidade de São Paulo. São Paulo, set. 2011.

REFERÊNCIA CURRICULAR DOS AUTORES

Sergio Fausto é diretor da Fundação Fernando Henrique Cardoso.

Widad Mustafa El Hadi é professora de Ciências da Informação e da Comunicação na Universidade de Lille 3 (França), junto ao Departamento de Ciências da Informação e da Documentação, e membro do Laboratório Gériico (Grupo de Estudos e de Pesquisa Interdisciplinar em Informação e Comunicação), no qual é responsável científica pelo eixo 4 (“Circulação da informação e organização do conhecimento”). Seus principais temas de pesquisa são as aproximações teóricas da organização do conhecimento; o impacto que a língua e a cultura exercem sobre ela; e, mais recentemente, seus aspectos éticos. Participa ativamente da ISKO (International Society for Knowledge Organization) desde 1996, tendo sido fundadora, juntamente com Jacques Maniez, e primeira presidente (1996-2001) de sua seção francesa. É a atual presidente da ISKO-França, desde novembro de 2015.

Ricardo César Gonçalves Sant’Ana é docente da Universidade Estadual Paulista, atuando na Faculdade de Ciências e Engenha-

ria, *campus* de Tupã, e no Programa de Pós-Graduação em Ciência da Informação da Faculdade de Filosofia e Ciências, *campus* de Marília. Graduado em Matemática e Pedagogia, mestre e doutor em Ciência da Informação e livre-docente em Sistemas de Informações Gerenciais. Atuou no setor privado como consultor, integrador e pesquisador de novas tecnologias informacionais (1988-2011). Pesquisa temáticas relacionadas a ciclo de vida dos dados, campo informacional, algoritmização, transdução e biodisponibilidade informacional.

José Eduardo Santarem Segundo é docente da Universidade de São Paulo, na Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, e no Programa de Pós-Graduação em Ciência da Informação da Universidade Estadual Paulista, *campus* de Marília. Mestre e doutor em Ciência da Informação pela Universidade Estadual Paulista, *campus* de Marília, é pós-doutor pela Faculdade de Engenharia da Computação da Western University of Canada e livre-docente em Informação e Tecnologia pela Universidade de São Paulo. Bolsista de Produtividade em Pesquisa PQ-2 do CNPq, é coordenador do GT8 - Informação e Tecnologia, da Associação Nacional de Pesquisa e Pós-Graduação em Ciência da Informação (ANCIB).

Maria Fernanda Rollo é professora catedrática da Faculdade de Ciências Sociais e Humanas da Universidade Nova de Lisboa, onde coordena o Doutorado em História, a Pós-Graduação em Políticas e Gestão de Ciência e Tecnologia e o projeto Histó-

ria, Territórios e Comunidade. É membro do Research Council do Instituto Europeu de Florença; investigadora da cátedra UNESCO Biodiversidade e Conservação para o Desenvolvimento Sustentável e do Laboratório Associado TERRA; e administradora da Fundação Mário Soares e Maria Barroso. Tem como principais áreas de investigação a história contemporânea de Portugal; a história da economia, da sociedade e da inovação; Europa, movimentos de cooperação e integração; história e organização da ciência; ciência aberta.

José Augusto Chaves Guimarães é professor titular da Universidade Estadual Paulista junto ao Departamento de Ciência da Informação da Faculdade de Filosofia e Ciências e ao Programa de Pós-Graduação em Ciência da Informação, *campus* de Marília. Bacharel em Biblioteconomia e em Direito, mestre e doutor em Ciência da Comunicação pela Universidade de São Paulo, fez pós-doutorado em Documentação na Universidad Carlos III de Madrid. Foi docente do Curso de Especialização em Organização de Arquivos do Instituto de Estudos Brasileiros da Universidade de São Paulo. É pesquisador 1-A do CNPq, onde atua como membro do Comitê de Artes, Ciência da Informação e Comunicação, desde julho de 2020.

Johanna Wilhelmina Smit atuou como docente de graduação e pós-graduação junto ao Departamento de Cultura e Informação da Escola de Comunicações e Artes da Universidade de São Paulo. Dirigiu o Arquivo Geral da Universidade de São Paulo até 2013.

Graduada em Biblioteconomia, tem mestrado em Documentação pela École Pratique des Hautes Études (Paris) e doutorado em Análise do Discurso pela Universidade de Paris. Representou a área da Ciência da Informação junto à CAPES.

