

## A opacidade nos bancos de dados digitais e a pesquisa histórica

*Opacity in Digital Databases and Historical Research*

Nelson de Paiva Bondioli



<https://orcid.org/0000-0001-9565-6990>

Universidade Estadual Paulista Júlio de Mesquita Filho

**Resumo:** Considerando o papel fundamental da computação e da internet em praticamente todos os campos de nossa vida atual, em que mais do que nunca as noções de acessibilidade e disponibilidade de materiais, acervos e etc., tomam um papel central nas discussões no meio acadêmico, voltamos nosso olhar em direção às ferramentas digitais disponíveis aos historiadores e historiadoras, focando na discussão, pouquíssimo trabalhada no Brasil, acerca da constituição e utilização de Bancos de Dados Digitais para a pesquisa histórica. Neste sentido, O presente artigo visa discutir a problemática específica da “opacidade”, encontrada no próprio projeto de desenvolvimento de bancos de dados digitais, observando como ela influi de maneira silenciosa, porém determinante, nas possibilidades de pesquisa disponíveis aos usuários destes bancos. Para tanto, apresentamos no decorrer deste estudo uma discussão teórica inicial a respeito da constituição, desenvolvimento e utilização de bancos de dados digitais, para então apresentarmos exemplos reais dos mesmos, evidenciando os elementos ligados à opacidade.

**Palavras-chave:** Banco de Dados Digitais. História. Humanidades Digitais. Opacidade

**Abstract:** Considering the pervasive role of computing and internet, the questions of accessibility and availability of resources, collections, etc., more than ever occupy a central space in the discussions of academic circles. Here, we focus the digital tools that are available to historians, a discussion, barely carried out in Brazil, concerning the development and usage of Digital Databases for Historical Research. The present article proposes to discuss the specific problematic related to the concept of opacity, which is found in the project phase itself of the development of databases, observing how it impacts silently, but fundamentally, in the possibilities of research that are made available for users of such databases. Therefore, we begin with a theoretical discussion concerning the constitution, development and usages of digital databases, and follow it with an analysis of real examples, demonstrating the issues related to opacity.

**Keywords:** Digital Databases. History. Digital Humanities. Opacity.

A utilização de bancos<sup>1</sup> de dados digitais, ou ainda das ferramentas computacionais em geral, não pode, de forma alguma, ser considerada nos dias atuais como uma novidade para o ofício do historiador. De fato, essas fazem parte de nosso cotidiano desde as décadas de 1960/70, quando diversos historiadores passaram a refletir sobre o poder da computação, primeiro para análises quantitativas de materiais e, então, para a manipulação de conjuntos de dados históricos (EDELSTEIN *et al.*, 2017).

Chamamos à atenção para, por exemplo, o fato de, já em 1982, a *American Library Association* ter lançado seu dossiê: *Computer-Assisted Reference Service in History* organizado por Joyce Falk – historiadora, bibliotecária e técnica em informática da Universidade da Califórnia. Nesse dossiê diversas discussões são travadas acerca da utilização dos bancos de dados por historiadores, tanto no quesito da organização e acesso aos catálogos, materiais e criação de meta-



Esta obra está licenciada sob uma [Creative Commons – Atribuição 4.0 Internacional](https://creativecommons.org/licenses/by/4.0/)

<sup>1</sup> Neste estudo entendemos os termos “banco de dados” e “base de dados” como sinônimos e sem qualquer distinção semântica para o contexto aqui analisado.

dados, bem como as problemáticas específicas acerca da pesquisa com fontes primárias, em que se enfatizava a questão de sua falta de disponibilidade aos historiadores (cf. KLUGMAN; KIRESEN, 1982).

A situação hoje, como sabemos, ao menos no que tange a este último ponto, encontra-se em seu completo oposto: da escassez passamos para a abundância de materiais que estão disponibilizados (BERRY, 2012): que acervos inteiros se encontram digitalizados – e muitos destes acessados de maneira gratuita e *online* –, constituídos não apenas de textos, mas também outros tipos de materiais: pinturas, moedas, fotografias, vídeos, ambientes virtuais interativos, índices geográficos, ou ainda de natureza híbrida.

Assim, se de um lado a disponibilidade de materiais para pesquisa histórica em bancos de dados encontra-se atualmente no seu patamar mais elevado – e com expectativa que esta continue em uma curva crescente, com mais acervos digitalizados e a criação de outros tantos nativamente digitais –, por outro lado, conforme apontou Mariana Flores (2015), os historiadores – especialmente brasileiros – mostram-se até agora, ausentes desta esfera de debates.

Cabe ressaltar que, além da própria Mariana Flores, existem algumas notáveis exceções de historiadores que tomaram nos últimos anos os bancos de dados como objeto de pesquisa. Podemos citar aqui o trabalho de Tiago Luís Gil (2015), e ainda mais recentemente, as pesquisas ao nível de graduação de Guilherme Sichelero (2020) e Ingrid Lucena (2020).

Todavia, conforme observado por Flores, e uma observação da qual compartilhamos, ainda são uma minoria os pesquisadores brasileiros que consideram/criticam ou problematizam a própria utilização destes bancos de dados, sendo, em geral, tomados como meros meios de se acessar fontes, como se fossem ferramentas transparentes desprovidas de aspectos subjetivos, fora de um campo discursivo.

O presente artigo, situado dentro do campo de estudos que vem consolidando-se nos últimos vinte anos sob o nome de Humanidades Digitais, mas que remonta de décadas anteriores quando ainda se aplicava a terminologia “Computação nas Humanidades”, busca fazer uma discussão que articula, aparentemente, duas problemáticas distintas: o desenvolvimento de um banco de dados digital e a sua subsequente utilização na pesquisa histórica.

Provocamos com o termo “aparentemente” porque, conforme buscaremos argumentar no decorrer deste texto, não é possível simplesmente separarmos a questão em um problema do tipo “técnico” – o desenvolvimento e implementação de um banco de dados – e outro de caráter “teórico-metodológico” – a pesquisa histórica e suas ferramentas. Argumentamos que é necessário, prioritariamente, compreender a especificidade e o processo de transformação que um determinado material passa para a sua disponibilidade e acesso no meio digital. Aqui reside um elemento importante pois, conforme apontaram Evans & Rees, “(...) os [diferentes] meios não simplesmente transmitem mensagens, eles afetam nossa própria relação com o mundo” (2012, p. 22).

Digitalizar um documento – seja textual ou não-textual – é uma tarefa transformadora que implica, em primeiro lugar, na desestabilização desse material. Como exemplo, notemos conforme apontou Abby Smith, há mais de duas décadas, que um grande diferencial do texto digital – mas que podemos extrapolar para imagens, e outros elementos de maneira geral – é que, ao contrário do material concreto/analógico, ele não é finito e nem final (SMITH, 1999, p. 3), isto é, abarca nele próprio várias versões e modificações. A modificação, entendida em diversos aspectos tais como edição, alteração, atualização, e etc., passa a ser uma tarefa extremamente simples e, em geral, com custo baixo em termos de recursos financeiros ou esforços, com consequências praticamente imediatas.

A flexibilidade do material é uma das principais características do meio digital, de forma que se tornou possível trabalhar com o documento de maneiras outrora inviáveis, ou mesmo não imaginadas, aplicando-se técnicas heurísticas diferenciadas – como por exemplo, a mineração de textos, ou ainda, abrindo-os como um “hipertexto”, para novas formas de organização e navegação

de seu conteúdo (LÉVY, 1996).

Assim, é preciso reconhecer como a experiência do meio digital altera de maneira substancial as possibilidades de pesquisa e entender a especificidade dessa ampliação do acesso e uso, tanto em termos de técnicas, quanto de suas ferramentas. Neste sentido, e retornando ao tema de nosso estudo, propomos observar como os bancos de dados digitais influem no trabalho de pesquisa histórica, de uma maneira que lhes é peculiar, pois como aponta Hayles:

Bancos de Dados não são necessariamente mais objetivos que argumentos, mas eles são tipos diferentes de formas culturais, encarnando diferentes modalidades cognitivas, técnicas, psicológicas e artísticas, e oferecendo diferentes maneiras de se instanciar conceitos, estruturar experiências e incorporar valores (2012, p. 56).

A partir destas considerações e buscando trabalhar com estas especificidades, nosso estudo divide-se em duas partes: a primeira em que discutiremos o que são bancos de dados digitais e as questões envolvidas em seu desenvolvimento; a segunda parte na qual evidenciaremos como estas questões aparecem na utilização destes bancos de dados pelos pesquisadores de História.

### **Bancos de Dados Digitais: Desenvolvimento, Implementação e Opacidade**

Começamos pela tarefa de definir o que são Bancos de Dados. Contudo, antes de simplesmente oferecermos a definição acadêmica, pensemos um momento a respeito de sua função: Qual o papel e para que serve um banco de dados?

Podemos responder: eles aparecem, historicamente, junto com a necessidade de determinados grupos de pessoas armazenarem itens/artefatos/documentos, de maneira tal que eles pudessem ser recuperados posteriormente. Para uma recuperação eficiente, todavia, era necessário que houvesse alguma espécie de organização racional na forma do armazenamento e de uma determinada lógica nos tipos de dados que são armazenados conjuntamente.

Percebamos assim, que os primeiros bancos de dados, obviamente, são analógicos, isto é, independiam de sistemas informatizados/computadores, para sua existência. Desse modo, e de maneira abrangente, podemos conceituar banco de dados enquanto: “um conjunto de dados relacionados, representando um pedaço ou interpretação do mundo real, que possui uma estrutura lógica com significado inerente e comumente possui uma finalidade e usuários específicos” (CAIUT, 2015, p. 1).

Podemos destrinchar alguns desses elementos da definição anterior, por meio das observações de Elmasri e Navathe sobre o assunto:

Um banco de dados representa algum aspecto do mundo real, às vezes chamado de mini-mundo (sic) ou de universo de discurso (UoD — Universe of Discourse). As mudanças no mini-mundo (sic) são refletidas no banco de dados.

Um banco de dados é uma coleção logicamente coerente de dados com algum significado inerente. Uma variedade aleatória de dados não pode ser corretamente chamada de banco de dados.

Um banco de dados é projetado, construído e populado com dados para uma finalidade específica. Ele possui um grupo definido de usuários e algumas aplicações previamente concebidas nas quais esses usuários estão interessados.” (2011, p. 3) (Grifo nosso).

Os Bancos de Dados Digitais – referidos no texto a partir de agora como BDDs – compartilham destas mesmas características gerais a que nos referimos, adicionando-se, ou ainda, privilegiando-se, de outras três características essenciais: disponibilidade, acessibilidade e escala.

Neste estudo, a partir do paradigma digital, conceituamos de maneira ampla e abrangente a disponibilidade como a questão acerca da existência ou não de um determinado recurso para o seu acesso; e a acessibilidade, por sua vez como a facilidade com que um material disponível pode, de fato, ser acessado e utilizado por uma diversidade de usuários.

O meio digital reúne ambas as características, ao menos em teoria, garantindo alta disponibilidade e acessibilidade. Essa observação é especialmente relevante se considerarmos o momento a partir da década de 1990, em que com o início do uso generalizado da internet, os dados digitais passam a ser disponibilizados globalmente e seu acesso, atualmente, é possível mesmo da palma de nossas mãos.

A questão da escala também é particularmente importante nesta discussão: O meio digital permite o armazenamento de dados – textos, imagens, vídeos, áudios etc. – com praticidade e escalas de grandeza jamais alcançadas anteriormente em que, no limite, poderíamos, por exemplo, substituir prédios e espaços físicos de bibliotecas por apenas um único computador, ou dependendo do espaço de armazenamento do dispositivo, um único *smartphone*.

A partir das definições que foram expostas, e tendo em mente as características dos BDDs, tomamos como pilares para a argumentação que se segue, três pontos principais: primeiro, que um banco de dados é sempre projetado/construído com uma determinada finalidade; segundo, que existe uma preconcepção de como os usuários irão utilizá-los e no que eles podem estar interessados em acessar; terceiro, que sua disponibilização representa, ao menos em teoria, acesso ao material por pessoas de qualquer lugar do mundo.

O projeto de um BDD necessita, em primeiro lugar, determinar qual será a sua finalidade (TEOREY *et al.*, 2014, p. 57). Será preciso responder de antemão, qual seu objetivo de uso. A partir deste objetivo, o projetista/desenvolvedor precisará tomar uma série de decisões sobre qual o tipo de dado que será armazenado (ex. imagens e/ou textos), bem como qual será o formato desse armazenamento e a modelagem desses dados.

A modelagem dos dados é de essencial importância e tem sérias implicações para o projeto. Vamos observar, tomando como exemplo, a forma mais comum de modelagem dos BDDs atualmente que são os “Bancos de Dados Relacionais” em que – resumidamente – dados são decompostos em diferentes colunas, formando uma linha. Cada linha forma um registro diferente armazenado no banco de dados.

Em nosso primeiro exemplo, encontrado na **Tabela 1**, o projetista/desenvolvedor quer criar um banco de dados para catalogar os livros existentes em sua biblioteca pessoal. Assim, uma coluna será responsável pela identificação única de cada registro, outra para o título, outra coluna podemos colocar o primeiro nome do autor, em outra o sobrenome, e assim por diante até que tenhamos também editora e ano de publicação.

Outras colunas poderiam ser adicionadas, como número de páginas ou número de cópias existentes, mas o projetista acreditou que essas seriam informações desnecessárias. Observemos a tabela abaixo:

**Tabela 1 – Exemplo Banco de Dados Relacional – Catálogo de Livros**

Identificador	Título	Autor, Nome	Autor, Sobrenome	Editora	Ano
1	A Hora da Estrela	Clarice	Lispector	Rocco	1977
2	Grande Sertão Veredas	José Guimarães	Rosa	José Olympio	1956

Fonte: Tabela criada pelo autor.

O modelo relacional, como indicado pelo nome, relaciona diferentes dados das colunas, formando um registro completo que poderia ser pesquisado por qualquer uma de suas colunas, facilitando assim a recuperação e a manipulação dos dados. Cabe apontar que, por se tratar de uma publicação não-voltada para a área de TI, o exemplo desta e das próximas tabelas não seguirá o rigor técnico da criação do modelo relacional, não nos preocupando com a “normalização” dos dados, ou com a construção eficiente dos relacionamentos (através da separação do conteúdo em tabelas menores e realizando relação entre as colunas por códigos de identificação). Os exemplos aqui expostos servem exclusivamente como um apoio para a compreensão das problemáticas abordadas.

Para uma massa menos homogênea de dados, por sua vez, o modelo relacional apresenta problemas específicos. Observemos os dois próximos exemplos, agora modelados para a área de história.

Um projetista/desenvolvedor deve criar um banco de dados para receber textos de autores da Antiguidade Clássica. Esse objetivo, tal qual exposto, gerará diversos problemas para sua modelagem, uma vez que, conforme sabemos, os formatos destes textos podem ser bastante diferentes. Vejamos a tentativa de criarmos um BDD, com textos de Júlio César, Plínio, o Jovem e Cícero.

Como é possível observar na **Tabela 2** existem diversas discrepâncias em cada coluna que dificultariam a alimentação do BDD. Primeiro, seria necessário decidir o que fazer com a questão da localização temporal do texto. Em um banco de dados relacional seria bastante complexo a operação de trabalhar com obras produzidas ao longo de determinado período, como no caso das Cartas de Plínio, ou ainda trabalhar com algum consenso a respeito de datas exatas de publicação, caso dos livros de Guerra da Gália de Júlio César, havendo também a opção de simplesmente abolir esta informação para consultas.

**Tabela 2 – BDD Relacional Antiguidade Clássica**

Identificador	Título	Autor	Ano	Numeração	Passagem:
1	Guerra da Gália	Júlio César	c. 58-52 a.E.C	Livro I. I	A Gália é toda dividida em três partes, uma em que é habitada pelos Belgas (belgae), outra pelos Aquitânio (aquitani) e uma terceira por um povo chamado em sua própria língua de Celtas (celtae), em latim Galli. Todos esses se diferem um dos outros na língua, nas instituições e leis (Continua...) (trad. livre do Autor).
2	Catilinárias	Marco Túlio Cícero	c. 63 a.E.C	I. I. 1	Afinal Catilina, até quando abusarás de nossa paciência? Por quanto tempo este teu furor ainda zombará de nós? Até que ponto sua audácia sem freios se voltará contra nós? Nem a guarda noturna do Palatino, nem o medo do povo, nem a afluência de todos os homens de bem, nem este local tão protegido onde o Senado se reúne, nem os rostos e expressões destas pessoas te perturbam? (Continua...) (trad. de BARBOSA, 2019)
3	Panegírico de Trajano	Plínio, o Jovem	c. 100 E.C.	1.1	Bem e sabiamente, Pais Conscritos, instituíram [nossos] ancestrais que o início das coisas a serem feitas, assim como do que deve ser dito, fosse encabeçado por uma prece: pois nada será devidamente e cuidadosamente interpretado pelos homens sem o auxílio, conselho e estima dos deuses imortais. (trad. de BONDIOLI, 2018)
4	Epístolas	Plínio, o Jovem	c. 98 – 112 E.C.	Livro X. III B	Agiste como bom cidadão e digno senador ao obedecer à justa solicitação daquela augusta assembleia: e tenho plena confiança e tenho plena confiança que desempenhará fielmente o negócio que assumires (trad. de BINATO et



disponibilizadas para a pesquisa, em que outros tipos de dados poderiam ser incluídos ou suprimidos – como por exemplo na **Tabela 3**, incluir o nome do editor do impressor do material como campo, ou na **Tabela 2**, incluir o papel do escritor no *cursus honorum* –, mas também na própria abordagem ao BDD.

Percebamos que enquanto nos primeiros exemplos, **Tabela 1 e 2**, as colunas foram concebidas por um modelo autoral, isto é, como uma forma de organizar, armazenar e recuperar dados a partir de autores e suas obras, o terceiro exemplo utiliza uma construção cronológica, sequencial e com palavras-chave, mas ambos poderiam ter sido modelados a partir de outras prioridades e contextos.

O desenvolvedor/projetista poderia ter decidido durante a criação desta ferramenta que a finalidade do BDD fosse possibilitar pesquisas temáticas, em que o processo de alimentação dos dados fosse organizado e direcionado a partir de temas como religião e assuntos religiosos, militares e conflitos bélicos, realeza e classes altas, e infinitas outras combinações, que se adequassem ao uso preconcebido de seus usuários.

Nesse caso, o BDD poderia ser menos útil para pesquisadores especializados, que necessitassem de uma clara demarcação autoral de frases e passagens, e mais útil talvez a um público leigo interessado em textos e passagens sobre algum tema como a “amizade”, “felicidade” ou a “importância da religião”.

Outras formas de organização também poderiam ser imaginadas, bem como outras colunas a serem exploradas pelos usuários, levando-nos então ao ponto central desta nossa discussão, a saber, a problemática da opacidade dos Bancos de Dados.

Definimos como Opacidade, justamente, todas as questões que aparecem no projeto de banco de dados e das quais seus usuários não fazem parte ou não tem conhecimento. Opacidade é esse sem-número de decisões que são tomadas previamente e que inevitavelmente direcionam as formas e possibilidades de pesquisa/utilização de qualquer BDD.

Cabe destacar que embora existam autores que toquem nesta questão da importância de nos atentarmos às decisões de construção e desenvolvimento de ferramentas digitais, inclusive sob outros nomes tais como *black-boxing* (RIEDER; RÖHLE, 2012, p. 75-77) ou ainda fazendo discussões sobre opacidade e transparência (cf. VITALE, 2016), não encontramos nenhuma análise específica para o conceito de opacidade em bancos de dados, e portanto, construímos aqui nossa própria compreensão e utilização para o termo.

A opacidade pode assim, tomar diferentes formas. Na **Tabela 1**, comentamos expressamente a decisão do projetista/desenvolvedor de não adicionar determinados campos por não os considerar pertinentes ou necessários, exemplificando questões de caráter subjetivo do desenvolvedor para a disponibilização de dados. Na **Tabela 2**, apontamos como decisões de viabilidade técnica e custo em tempo/esforço para a alimentação de um BDD podem afetar a inclusão ou exclusão de conteúdos/colunas que poderiam ser disponibilizadas para pesquisa.

A **Tabela 3**, permite observar diferentes formas de organização de um determinado BDD, sempre a partir da finalidade que foi atribuída previamente pelo projetista/desenvolvedor, tendo em vista a preconcepção de utilidade aos usuários.

As questões de opacidade ligam-se, portanto, diretamente, às observações que Manovich realizou a respeito dos bancos de dados: acreditar que os dados são “passivos” a serem acessados é um grande erro, pois estes não existem por si próprios, mas precisam ser gerados (2007, p. 43). O projetista/desenvolvedor precisa coletar e organizar os dados, ou ainda, na impossibilidade de uma digitalização, criá-los no ambiente digital a partir do zero. De qualquer forma, os dados sofrerão diferentes processos de organização, limpeza e edição a fim de serem alimentados nos bancos de dados e então disponibilizados aos usuários.

## Bancos de Dados Digitais Históricos: Interesses, Usos e Atualização

Encontramos a partir das décadas de 1990 uma grande concentração de projetos de digitalização de acervos e criação de banco de dados, visando sua disponibilização gratuita e *online*. Dentre os vários exemplos disponíveis, escolhemos quatro casos para analisarmos, devido à suas especificidades: o projeto *Perseus* (*online* a partir de 1995), *LacusCurtius* (*online* a partir de 1997), bem como na década seguinte, os projetos *Catálogo y Biblioteca Digital de Relaciones de Sucesos* (*online* a partir de 2001) e *Mapping the Republic of Letters* (*online* a partir de 2008).

O Projeto *Perseus*<sup>2</sup> – *Perseus Digital Library* –, sediado na Universidade Tufts, é um dos maiores, senão propriamente o maior projeto de digitalização e disponibilização de fontes clássicas já realizado. Atualmente, conta também com um extenso catálogo que ultrapassa aquela primeira temporalidade, em que encontramos fontes árabes da Antiguidade Tardia, documentos da Renascença, do século XIX, bem como imagens de sítios arqueológicos e outros artefatos. Coordenado pelo Dr. Gregory Crane, professor de Estudos Clássicos da Universidade de Harvard, o projeto teve início em 1987, gerando um primeiro resultado em CD-ROM em 1992. A partir de 1995 começaram os trabalhos de disponibilização *online* desse banco de dados, que continuam nos dias atuais (DARLACK, 2016).

Acerca das fontes clássicas, as encontramos no seu idioma original – Latim ou Grego – e, por vezes, traduções para o inglês. A maior parte das traduções disponibilizadas, no entanto, foram retiradas da *Loeb Classical Collection*, da Harvard University, apresentando, assim, muitas traduções do início a meados do século XX ou ainda anteriores. Os textos que possuem a tradução para o inglês apresentam anotações que, devemos observar, também foram retiradas das traduções da *Loeb Classical*.

A organização desse acervo na página inicial dos textos clássicos<sup>3</sup> se dá de maneira alfabética por nome do Autor, embora seja possível realizar pesquisa por nomes, títulos e termos/frases tanto dentro de todo o catálogo, quanto em passagens específicas, nos idiomas Inglês, Grego, Latim, Inglês Antigo e Nórdico Antigo. Assim, um dos diferenciais do projeto *Perseus* para outros projetos de bancos de dados, de acordo com seus próprios idealizadores, seria que, com a maturidade atingida ao começo da década de 2000, houve uma preocupação a respeito da utilização de seus recursos e soluções. Isto é, enquanto outros acervos digitais propunham-se somente a disponibilizar seu conteúdo deixando aos usuários toda a iniciativa de pesquisa, o projeto *Perseus* passava a focar na criação de ferramentas para que “os usuários pudessem entender o que a biblioteca digital dava a eles”<sup>4</sup>: automatização de links e serviços de extração e visualização de informações.

As questões ligadas à opacidade deste BDD podem ser observadas assim a partir dos seguintes elementos: primeiro, na escolha de utilização, majoritariamente, de textos da *Loeb Classical* na alimentação do banco de dados. Essa escolha tem grande implicações aos pesquisadores, quando observamos que o projeto, iniciado ao final da década de 1980, disponibiliza principalmente traduções bastante antigas dos textos clássicos, traduções estas que nem sempre se preocupam com o tratamento filológico apropriado dos termos originais, focando, amiúde, em torná-lo mais narrativo/agradável a leitura.

Segundo ponto importante a observamos diz respeito à ênfase colocada nas ferramentas de auxílio à manipulação dos dados do BDD. Durante seu desenvolvimento, diferentes materiais foram incorporados de modo a tornar possível pesquisas que até mesmo entrecruzassem, por exemplo, textos e imagens. Entretanto, devemos observar que existe uma clara preconcepção a respeito daquilo que seria mais útil ou importante aos usuários: pesquisas a partir de termos específicos, em

---

<sup>2</sup> *Perseus Digital Library* - <http://www.perseus.tufts.edu/hopper/>. Acessado em 25/02/2021

<sup>3</sup> <https://www.perseus.tufts.edu/hopper/collection?collection=Perseus:collection:Greco-Roman>. Acessado em 25/02/2021

<sup>4</sup> <https://www.perseus.tufts.edu/hopper/opensource>. Acessado em 25/02/2021.

latim ou outras línguas. Voltaremos, adiante, a este tópico, entretanto, cabe agora mostrarmos como que outras previsões de pesquisa não foram contempladas, pois no projeto de desenvolvimento não se considerou a possibilidade de pesquisar fontes a partir de temas, datas de escrita ou publicação, tipos de manuscritos etc.

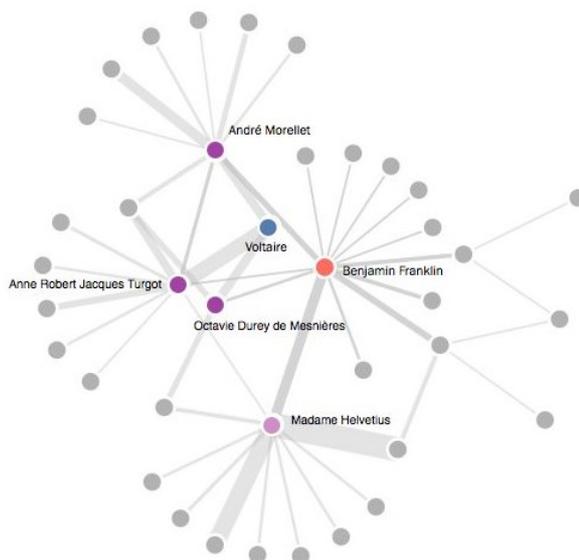
A partir da percepção da importância das ferramentas de visualização, encontramos outro projeto, que fez desse elemento uma das suas principais preocupações: *Mapping the Republic of Letters*<sup>5</sup>, sediado na Universidade de Stanford.

Criado em 2008 e coordenado por Paula Findlen, Dan Edelstein, contado com Nicole Coleman como co-investigadora, *Mapping the Republic of Letters* é concebido como uma série de estudos de casos, acerca das redes de correspondência de intelectuais entre os séculos XV e XIX (cf. MEDICI, 2019). O objetivo deste projeto é mapear as redes de correspondência entre pensadores, cientistas e estudiosos, observando como essas redes se formavam, sua extensão e como se modificavam ao longo do tempo. Tal abordagem visa permitir uma exploração espacial do contexto das redes entre intelectuais no passado, levando em consideração, entre outros elementos, rotas e o tempo gasto para a troca de correspondências. O projeto tem como ponto central a criação de ferramentas de visualização de dados, isto é, a criação de representações gráficas permitindo acesso visual rápido e dinâmico ao seu conteúdo.

A página inicial dos casos de estudo<sup>6</sup> apresenta vários personagens cujos estudos de sua correspondência estão finalizados ou em andamento, como Galileu, Voltaire, Benjamin Franklin ou ainda um mapa intelectual do Império Espanhol entre 1600-1810.

A página do estudo de caso de Voltaire<sup>7</sup>, por exemplo, mostra diversos dados quantitativos e qualitativos acerca de suas correspondências: frequência, idioma, alcance e o mapeamento da rede. Observem a **Figura 1**.

**Figura 1:** Redes de Correspondência de Voltaire e Benjamin Franklin



Fonte: *Mapping the Republic of Letters*. Disponível em:  
<http://republicofletters.stanford.edu/casestudies/voltaire.html>. Acesso em 25/02/2021.

A **Figura 1** é um exemplo de como as ferramentas de visualização podem nos ajudar a fazer inferências a partir dos dados de um determinado BDD. Neste mapeamento da rede de Voltaire, observa-se que, apesar de este pensador e Benjamin Franklin não possuírem nenhuma troca de cartas entre eles, ambos estavam conectados à uma mesma rede, inclusive trocando cartas com

<sup>5</sup> *Mapping the Republic of Letters* - <http://republicofletters.stanford.edu/>. Acessado em 25/02/2021

<sup>6</sup> <http://republicofletters.stanford.edu/casestudies/index.html>. Acessado em 25/02/2021

<sup>7</sup> <http://republicofletters.stanford.edu/casestudies/voltaire.html>. Acessado em 25/02/2021

diversas figuras em comum, tais como André Morelett, Anne Turgot, Octavie des Mesnières.

Novamente questões de opacidade tornam-se mais transparentes quando analisamos, justamente, a previsão de usos na fase de desenvolvimento para esse BDD. *Mapping the Republic of Letters* é idealizado como fonte de consulta para pesquisas quantitativas, ou até mesmo estatísticas a respeito das redes que se formavam entre intelectuais. O BDD parte do pressuposto que os interessados vão trabalhar com a distribuição, volume, e outros dados relativos à correspondência entre os indivíduos, de modo que, se, de um lado serve como um importante índice, por outro, não oferece os documentos em si para consulta e pesquisa online.

Neste ponto, podemos observar como são construídos de maneira quase que diametralmente opostas, por exemplo, os BDD *Perseus* e *Mapping the Republic of Letters*, que a partir de seus próprios conjuntos e propostas de uso, oferecem recursos diferentes e diferenciados aos seus usuários.

O BDD *LacusCurtius*<sup>8</sup>, por sua vez, traz um caso bastante singular. Criado em 1997 por William Thayer, o BDD é hospedado no domínio institucional da Universidade de Chicago, não havendo, porém, qualquer relação institucional entre essa iniciativa e a universidade.

*LacusCurtius* foi criado a partir do, e mantido exclusivamente pelo, interesse pessoal de Thayer no passado greco-romano. Ele próprio reconhece em uma página específica do site falando sobre suas credenciais<sup>9</sup>, que não possui treinamento ou instrução acadêmica em História, Estudos Clássicos, Arqueologia etc.

Apesar de sua natureza “amadora”, o *LacusCurtius* disponibiliza para os interessados – especialistas ou não – diversos textos da Antiguidade Clássica, alguns em idioma original – Latim e Grego – e outros apenas contando com traduções em inglês. Tal qual o projeto *Perseus*, o *LacusCurtius* também se utiliza principalmente dos textos e traduções da coleção *Loeb Classical*.

Das diversas comparações que poderíamos fazer a respeito do projeto *Perseus* e o *LacusCurtius*, dentro do panorama que está sendo traçado nesse estudo, voltamos à questão da importância das ferramentas de pesquisa na utilização dos bancos de dados. Enquanto o *Perseus* construiu sua própria gama de ferramentas para serem utilizadas na pesquisa e visualização de seus documentos, o *LacusCurtius* conta somente com o ferramental de pesquisa em sites produzido e distribuído gratuitamente pelo *Google*. Desta forma, podemos notar a grande diferença nos resultados e visualizações que esses BDDs oferecem aos seus usuários. Observem as **Figuras 2 e 3**, em que o termo “*Pater Patriae*” é buscado em ambos os BDDs:

O *Perseus* apresenta uma interatividade muito maior com seus usuários – e possibilidades de pesquisa dos dados – a partir de suas próprias ferramentas, enquanto a **Figura 3** mostra como, apesar de utilizar uma caixa de pesquisa dentro do site do *LacusCurtius*, o resultado é direcionado pela e para a página de pesquisa do *Google*. Cabe, certamente, frisarmos que o escopo e natureza de cada um desses BDDs são completamente distintos, contando com recursos e possibilidade que, de muitas maneiras, refletem estas características de suas implementações.

---

<sup>8</sup> *LacusCurtius* - <https://penelope.uchicago.edu/Thayer/E/Roman/home.html>. Acessado em 25/02/2021

<sup>9</sup> Credentials. - <https://penelope.uchicago.edu/Thayer/E/HELP/credentials.html>. Acessado em 25/02/2021

Figura 2: Pesquisa frase *Pater Patriae* no Projeto Perseus

The screenshot shows the Perseus Project search interface. At the top, there is a search bar with the text "Pater Patriae" and a "Search" button. Below the search bar, there are navigation links: Home, Collections/Texts, Perseus Catalog, Research, Grants, Open Source, About, and Help. The main content area displays "Showing 1 - 10 of 16 document results in Latin." and lists several results, including "Francis Glass, *Washingtonii Vita* (ed. J.N. Reynolds)", "P. Ovidius Naso, *Tristia*", "Florus, Lucius Annaeus, *Epitome Rerum Romanorum*", and "P. Ovidius Naso, *Fasti* (ed. Sir James George Frazer)". On the right side, there is a "Refine This Search" panel with options for Language (Latin), Required words, Required phrase (Pater Patriae), Allowed words, and Excluded words. At the bottom right, it says "All Matching Documents (16)".

Fonte: Captura de tela pelo autor. 25/02/2021.

Figura 3: Pesquisa frase *Pater Patriae* no *LacusCurtius*

The screenshot shows a Google search for "pater patriae site:https://penelope.uchicago.edu/Thayer/". The search results are displayed on the LacusCurtius website. The first result is "LacusCurtius • Roman Law — Patria Potestas (Smith's ...)" with a date of Jan 26, 2020. The second result is "Augustus (by John Buchan) • Book IV, Chapter 3" with a date of Feb 19, 2020. The third result is "Augustus (by John Buchan) • Book IV, Chapter 1" with a date of Feb 12, 2019. The fourth result is "Augustus (by John Buchan) • Book IV, Chapter 6" with a date of Feb 12, 2019. The fifth result is "Augustus (by John Buchan) • Book IV, Chapter 4" with a date of Feb 12, 2019.

Fonte: Captura de tela pelo autor. 25/02/2021.

O último BDD que selecionamos para análise neste estudo é o *Catálogo y Biblioteca Digital de Relaciones de Sucesos – CBDRS*<sup>10</sup>. Iniciado como um projeto interdisciplinar na Universidade de Coruña em 1994, se desenvolveu como iniciativa para o acesso livre e pela internet destes impressos de notícia – as *Relações de Sucesso* – coordenado por Dra. Nieves Pena Sueiro e Sagrario López Poza, ambas professoras da Universidade galega.

O principal objetivo do *CBDRS* era reunir e disponibilizar para qualquer pessoa interessada na cultura do “século de ouro” espanhol:

De forma gratuita e livre a descrição bibliográfica de todas as *Relações de Sucesso* impressas entre os séculos XVI-XVIII, de que se tenha conhecimento, complementada com a identificação das edições, sua localização de exemplares e sua edição digital (fac-símile do texto transcrito) (SUEIRO; GARCÍA, 2014, p. 338).

O trabalho, que tomou mais de duas décadas enfrentou uma série de problemas na sua implementação: as tecnologias mudaram rapidamente de forma que, embora tenha começado com elementos de ponta e passado por atualizações de software, as tecnologias utilizadas na criação

<sup>10</sup> *Catálogo y Biblioteca Digital de Relaciones de Sucesos – CBDRS (siglos XVI-XVIII)*. <https://www.bidiso.es/CBDRS/>. Acessado em 25/02/2021.

do banco de dado encontraram-se ultrapassadas. Além disso, conforme o trabalho foi se desenvolvendo, percebeu-se que havia um volume muito maior de dados do que havia sido pensado inicialmente, e que agora precisavam ser geridos de maneira eficiente<sup>11</sup> (SUEIRO; PLACES, 2019, p. 81).

A partir dessas questões, o *CBDRS* encontrou-se carecendo de uma reengenharia de seu BDD, em que, conforme sua criadora comenta, precisou resolver não apenas os problemas de performance técnica, mas também se atentar a um importante elemento ligado à sua opacidade: “mudou o perfil de seus usuários que esperam [hoje em dia] mais usabilidade e maiores benefícios de catálogos digitais” (SUEIRO; PLACES, 2019, p. 81).

Observamos aqui, talvez com maior clareza do que nos outros exemplos, como que uma reconfiguração das intenções dos usuários, levou também diretamente à necessidade de uma reengenharia do BDD de modo a tornar possível os tipos de pesquisa que os usuários imaginavam.

Assim, após um planejamento técnico e uma (re)visão ampla da utilização desse BDD por pesquisadores, uma série de intervenções de reestruturação foram levadas a cabo, chegando ao quadro atual do *CBDRS* que permite a pesquisa de seu acervo a partir de uma grande variedade de parâmetros que podem ser observados na **Figura 4**.

**Figura 4:** Opções de busca pelo acervo do *CBDRS*

Búsqueda simple de ediciones de relaciones de sucesos Buscador avanzado

Ayuda para las búsquedas

General

Título de edición

Persona  Autor, editor, impresor, traductor y otros roles.

Lugar de edición   Sin lugar de edición

Año de edición  Desde  Hasta   Sin año de edición

Relaciones con más de una edición localizada

Últimas incorporaciones

Modificadas recientemente

Con algún ejemplar digitalizado

Limpiar Buscar

« 1 2 3 4 5 6 7 8 9 10 ... 122 »

Ediciones recuperadas: 6.081 Ordenación alfabética / ordenación cronológica

1622. A los 15 de enero. Capitulaciones y conuenciones entre su Magestad, y las dos Ligas Grisa, y Cade, y Señoría de Mayenfelt, y los

Fonte: Captura de tela pelo autor. 25/02/2021.

Devemos também destacar que um dos principais elementos desta reengenharia do *CBDRS* foi refletir sobre o conceito de escalabilidade, isto é, de pensar não apenas no grande número de dados que precisam ser geridos por este BDD, mas também sobre como ele pode crescer ainda mais ao passar do tempo, com a alimentação de novos dados.

A questão da escalabilidade é particularmente importante pois, conforme nos referimos ainda na introdução deste estudo, a tendência é de encontrarmos uma linha crescente no que tange a materiais digitalizados e aqueles nativamente digitais, disponibilizados para o pesquisador, de modo que essa abundância de dados precisa ser gerida de maneira eficiente não apenas pensando

<sup>11</sup> Atualmente, o *CBDRS* conta com 6430 edições e mais de 9000 exemplares, contando com a reprodução digital de mais de 2000 deles, que são conservados de maneira física em cerca de 175 bibliotecas ao redor do mundo (SUEIRO; PLACES, 2019, p. 81).

no momento atual, mas efetivamente planejando-se os BDDs para o futuro.

## Considerações finais

Ao longo deste estudo, fizemos uma série de apontamentos a respeito do desenvolvimento e a implementação de bancos de dados digitais, preocupando-nos, sobretudo, com aqueles utilizados para a pesquisa histórica.

A principal problemática que trouxemos tratou do conceito de opacidade, ou seja, as decisões que estruturam um BDD e que direcionam os usuários deste sistema frente suas possibilidades, na maior parte das vezes, sem o seu conhecimento.

Nos exemplos apresentados na segunda parte do trabalho, é possível notarmos claramente como a questão da opacidade se entrelaça com as possibilidades de pesquisa: como que a partir das decisões de finalidade e uso, de desenvolvimento e implementação de um determinado BDD, diferentes problemas podem ser tratados e pesquisados, assim como diferentes são as formas de fazê-lo.

Se, como aponta Manovich (2007), os Bancos de Dados contemplam tanto formas narrativas quanto não-narrativas, isto é, podem ser considerados como listagens de informação ao mesmo tempo em que podem produzir um significado narrativo a partir de sua interatividade com o usuário, é necessário observar que eles se encontram, em ambos os lados, dentro do campo discursivo. Os bancos de dados são formas culturais específicas que dependem de uma série de pressupostos para constituir-se. Qualquer que seja o banco de dados, antes de ser criado, precisa responder, no mínimo, às seguintes questões: O que será feito, para que será feito, como será feito e para quem será feito.

No exato momento em que estas perguntas são respondidas, nasce a problemática da opacidade, e os pesquisadores que vierem a utilizar este banco de dados devem conscientizar-se sobre as formas como essa problemática se relaciona com seus próprios estudos.

Do exposto, acreditamos restar claro que os bancos de dados não são simplesmente consumíveis ou consultáveis, mas que, em sua própria natureza, implicam escolhas e delimitações que podem tanto auxiliar quanto criar empecilhos para determinados usuários e seus objetivos.

## Referências

BERRY, David. Introduction: Understanding the Digital Humanities. In: BERRY, David (ed.). *Understanding Digital Humanities*. New York: Palgrave Macmillan, 2012, p. 1-20.

CAIUT, Fábio. *Administração de Bancos de Dados*. Rio de Janeiro: Escola Superior de Redes, 2015.

DARLACK, James Marion. Web Review: Perseus Digital Library. *Theological Librarianship*, v. 9, n. 1, p. 11-15, 2016.

EDELSTEIN, Dan *et al.* Historical Research in a Digital Age: Reflections from the Mapping the Republic of Letters Project. *The American Historical Review*, v. 122, n. 2, 2017, p. 400-424.

ELMASRI, Ramez; NAVATHE, Shamkant. *Sistemas de Bancos de Dados*. São Paulo: Pearson, 2011.

EVANS, Leighton; REES, Sian. An Interpretation of Digital Humanities. In: BERRY, David (ed.). *Understanding Digital Humanities*. New York: Palgrave Macmillan, 2012, p. 21-41.

FLORES, Mariana Flores da Cunha Thompson. Os bancos de dados, os arquivos digitais e o papel do historiador. *Acervo*, v. 28, n. 2, p. 240-251, 2015.

GIL, Tiago. *Como se faz um banco de dados (em história)*. Porto Alegre: Ladeira Livros, 2015.

- HAYLES, Katherine. How We Think: Transforming Power and Digital Technologies. In: BERRY, David (ed.). *Understanding Digital Humanities*. New York: Palgrave Macmillan, 2012, p. 42-66.
- KIRESEN, Evelyn-Margaret; KLUGMAN, Simone. The use of Online Databases for Historical Research. *RQ*, v. 21, n. 4, p. 342-351, 1982.
- LÉVY, Pierre. *O que é o Virtual?* Rio de Janeiro: Editora 34, 1996.
- LUCENA, Ingrid Guimarães de. *Papel, digitalização, conjuntos de dados ou a demolição dos exames censórios do Conservatório Dramático Brasileiro (1843-1864)*. 2020. 33 f. Trabalho de Conclusão de Graduação em História - Universidade de Brasília, Brasília, 2020.
- MANOVICH, Lev. Database as Symbolic Form. In: VESNA, Victoria (ed.). *Database Aesthetics*. Minneapolis: University of Minnesota Press, 2007, p. 39-60.
- MEDICI, Catherine. Mapping the Republic of Letters. *Renaissance and Reformation / Renaissance et Réforme*, v. 42, n. 4, p. 180-184, 2019.
- RIEDER, Bernhard; RÖHLE, Theo. Digital Methods: Five Challenges. In: BERRY, David (ed.). *Understanding Digital Humanities*. New York: Palgrave Macmillan, 2012, p. 67-84.
- SMITH, Abby. *Why Digitize?* Washington: Council on Library and Information Resources, 1999.
- SICHELERO, Guilherme Giotti. *A construção do banco de dados na pesquisa histórica: análise de uma experiência*. 2020. 86 f. Trabalho de Conclusão de Graduação em História – Universidade Federal do Rio Grande do Sul, Porto Alegre, 2020.
- SUEIRO, Nieves Pena; GARCÍA, Sandra Álvarez. El Catálogo y Biblioteca digital de relaciones de sucesos: bases de datos bibliográficas, textos e imágenes. *Janus*, Anexo 1, 2014, p. 335-345.
- SUEIRO, Nieves Pena; PLACES, Ángeles Saavedra. Obsolescencia y resiliencia en Humanidades digitales El caso de la Biblioteca Digital de Relaciones de Sucesos. *Artnodes*, n. 23, p. 79-88, 2019.
- TEOREY, Toby et al. *Projeto e Modelagem de Bancos de Dados*. São Paulo: Elsevier, 2014.
- THALLER, Manfred. A Draft Proposal for a Standard for the Coding of Machine Readable Sources. *Historical Social Research / Historische Sozialforschung*, n. 40, p. 3-46, 1986.
- VITALE, Valeria. Transparent, Multivocal, Cross-disciplinary: The Use of Linked Open Data and a Community-developed RDF Ontology to Document and Enrich 3D Visualisation for Cultural Heritage. In: BODARD, Gabriel; ROMANELLO, Mateo (eds.). *Digital Classics Outside the Echo-Chamber: Teaching, Knowledge Exchange & Public Engagement*. London: Ubiquity Press, 2016.

### Notas de autoria

**Nelson de Paiva Bondioli** é Doutor em História (UNESP). Bolsista de pós-doutorado PNPd-CA-PES do Programa de Pós-Graduação em História/UNESP. Tecnólogo em Banco de Dados (UNIFRAN). Tecnólogo em Análise e Desenvolvimento de Sistemas (UNINTER). Pesquisador do #Veredas\_Digitais – Laboratório de Tecnologias e Humanidades – UNESP; Pesquisador do CPEP – Centro de Pesquisas e Estudos Plinianos. E-mail: nelson.bondioli@unesp.br.

### Como citar esse artigo de acordo com as normas da revista

BONDIOLI, Nelson de Paiva. A opacidade nos bancos de dados digitais e a pesquisa histórica. *Saeculum – Revista de História*, v. 26, n. 45, p. 125-139, 2021.

### Contribuição de autoria

Não se aplica

**Financiamento**

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

**Consentimento de uso de imagem**

Não se aplica.

**Aprovação de comitê de ética em pesquisa**

Não se aplica.

**Licença de uso**

Este artigo está licenciado sob a [Licença Creative Commons CC-BY](#). Com essa licença você pode compartilhar, adaptar, criar para qualquer fim, desde que atribua a autoria da obra.

**Histórico**

Recebido em 24/03/2021.

Modificações solicitadas em 13/08/2021.

Aprovado em 13/09/2021.