

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE FILOSOFIA E CIÊNCIAS HUMANAS
DEPARTAMENTO DE HISTÓRIA

Guilherme Giotti Sichelero

**A CONSTRUÇÃO DO BANCO DE DADOS NA PESQUISA HISTÓRICA:
análise de uma experiência**

Porto Alegre, 2020.

Guilherme Giotti Sichelero

**A CONSTRUÇÃO DO BANCO DE DADOS NA PESQUISA HISTÓRICA:
análise de uma experiência**

Trabalho de conclusão de curso apresentada como requisito parcial para obtenção do título de Licenciado em História pelo Departamento de História do Instituto de Filosofia e Ciências Humanas da Universidade Federal do Rio Grande do Sul.

Orientadora: Profa. Dra. Clarice Gontarski Speranza

PORTO ALEGRE

2020

**A CONSTRUÇÃO DO BANCO DE DADOS NA PESQUISA HISTÓRICA:
análise de uma experiência**

Trabalho de conclusão de curso apresentada como requisito parcial para obtenção do título de Licenciado em História pelo Departamento de História do Instituto de Filosofia e Ciências Humanas da Universidade Federal do Rio Grande do Sul.

Aprovado em 01 de dezembro de 2020

Conceito: A

Profa. Dra. Clarice Gontarski Speranza – Orientadora – Universidade Federal do Rio Grande do Sul (UFRGS)

Profa. Dra. Mariana Flores da Cunha Thompson Flores – Universidade Federal de Santa Maria (UFSM)

Prof. Dr. Alisson Droppa – Universidade Estadual de Campinas (Unicamp)

AGRADECIMENTOS

Em primeiro lugar, gostaria de agradecer aos meus familiares e amigos pelo suporte financeiro e emocional durante os anos de graduação. Agradeço também por ter tido a oportunidade de estudar numa universidade pública e de qualidade.

Gostaria de agradecer a minha orientadora, Profa. Dra. Clarice Gontarski Speranza, por ter me orientado na realização desse trabalho de conclusão de curso e no período de iniciação científica.

RESUMO

O presente trabalho tem como objetivo principal discutir uma experiência, durante o período de iniciação científica, de construção do banco de dados voltado a sistematização de operários imigrantes vindos do Leste Europeu, após a Segunda Guerra Mundial, para a região carbonífera do Rio Grande do Sul. O processo de construção do banco de dados passou por quatro etapas principais: a) análise preliminar das fontes; b) processo de fichamento da documentação; c) desenvolvimento inicial do banco de dados; d) construção do banco de dados. Como o banco de dados, nos dias atuais, é desenvolvido a partir da informática, esse trabalho também teve como um dos seus objetivos discutir a relação entre história e informática, defendendo a ideia que a informática é uma ferramenta importante para o trabalho do historiador. Nesse sentido, também discute-se como o historiador pode se apropriar da informática através das múltiplas possibilidades de leitura, da utilização dos documentos digitais e dos usos da internet na sua atual versão Web 2.0

Palavras-chave: Banco de Dados. Internet. Informática. Web. História quantitativa. História Serial. História serial-quantitativa. História social. História do Trabalho.

LISTA DE FIGURAS

Figura 1 - Exemplo de um arquivo no Google Docs.....	21
Figura 2 - Teorização dos documentos digitais proposta por Almeida.....	26
Figura 3 - Imagem do tweet de Matias Spektor	33
Figura 4 - Exemplo de banco de dados	36
Figura 5 - Banco de dados Slave Voyages	41
Figura 6 - Documento do da lista de comunistas	50
Figura 7 - Documento dividindo os operários imigrantes	51
Figura 8 - Relatório da vinda dos operários imigrantes	52
Figura 9 - Exemplo de fichamento de uma correspondência	53
Figura 10 - Exemplo de fichamento de um relatório produzido pelo Cadem	55
Figura 11 - Salvo-conduto de José Varga	58
Figura 12 - Exemplo de relacionamento entre as tabelas	60
Figura 13 - Exemplo do campo “dados do operário”	63
Figura 14 - Exemplo de formulário do Microsoft Access	64
Figura 15 - Tabela total ou query	65
Figura 16 - Exemplo de relatório de um operário	66
Figura 17 - Tabela de acompanhamento	68
Figura 18 - Tabela de nacionalidade e estado civil	69
Figura 19 - Tabela de ocupação profissional.....	70
Figura 20 - Tabela de processo migratório.....	71
Figura 21 - Exemplo de tabelas, campos e relacionamento numa suposta segunda versão	74
Figura 22 - Parte da carta dos operários imigrantes do Leste Europeu ao Cadem.....	78
Figura 23 - Questionário respondido pelo Cadem.....	79

LISTA DE TABELAS

Tabela 1 - Formatos de leitura digital.....	22
Tabela 2 - Sugestão sobre os documentos não-primários digitais.....	26
Tabela 3 - Web 1.0 e Web 2.0	30
Tabela 4 - Exemplo de banco de dados de uma biblioteca pessoal	37
Tabela 5 - Exemplo de cardinalidade	38
Tabela 6 - Período do processo de construção do banco de dados.....	47
Tabela 7 - Três modelos conceituais propostos por Tiago Luís Gil.....	61
Tabela 8 - Tabelas e campos do banco de dados produzido pela pesquisa	71

LISTA DE GRÁFICOS

Gráfico 1 - Nacionalidades no banco de dados	76
Gráfico 2 - Relação de ucranianos casados e solteiros	77

SUMÁRIO

1 INTRODUÇÃO.....	10
2 HISTÓRIA E INFORMÁTICA	13
2.1 Os historiadores e a informática	14
2.2 Leitura, escrita e os documentos digitais.....	17
2.3 História e internet	27
2.4 Banco de dados: uma forma de produção histórica.....	36
2.5 História serial-quantitativa e o banco de dados	42
3 O PROCESSO DE CONSTRUÇÃO DO BANCO DE DADOS.....	46
3.1 Primeira etapa: análise preliminar das fontes	48
3.2 Segunda etapa: fichamento dos documentos	53
3.3 Terceira etapa: desenvolvimento inicial do banco de dados	57
3.4 Quarta etapa: construção do banco de dados.....	61
3.5 O banco de dados concluído: versão inicial e versão final.....	66
3.7 Utilidade do banco de dados: dados quantitativos e hipótese	75
4 CONSIDERAÇÕES FINAIS	80
REFERÊNCIAS	83

1 INTRODUÇÃO

O presente trabalho tem como objetivo principal discutir um processo de construção de banco de dados na pesquisa histórica através da experiência, durante o período de iniciação científica, do desenvolvimento de um banco de dados voltado a sistematização de operários imigrantes do Leste Europeu vindos, após a Segunda Guerra Mundial, para a região carbonífera do Rio Grande do Sul.

Como a utilização do banco de dados na pesquisa histórica, nos dias atuais, se dá preferencialmente através da informática, esse trabalho também tem como um dos seus objetivos discutir a relação entre a ciência histórica e a informática, principalmente no que diz respeito aos usos do banco de dados, das possibilidades de leitura, dos documentos digitais e da utilização, por parte dos historiadores, das ferramentas da internet, principalmente na sua versão Web 2.0¹.

Buscou-se, nesse trabalho, dialogar principalmente com historiadores que desenvolveram trabalhos sobre a informática e a internet no Brasil. Nesse sentido, em relação ao banco de dados, as referências principais são os trabalhos de Tiago Luís Gil (2015) e Mariana Flores da Cunha Thompson Flores (2015). O trabalho de Gil (2015), em especial, foi de suma importância para a elaboração desse trabalho e do banco de dados, na medida que apresenta, do ponto de vista teórico e prático, o bê-á-bá da construção do banco de dados na pesquisa histórica. É um trabalho imprescindível para quem pretende utilizar essa ferramenta.

No que diz respeito à relação entre história e informática, buscou-se dialogar com alguns textos clássicos como o de Luciano Figueiredo (1997). Em relação a abordagens mais atuais, o trabalho de Célia Tavares (2012), na contextualização mais recente da relação entre história e informática, e de Fábio Chang de Almeida (2011), no desenvolvimento teórico dos documentos e fontes digitais, ajudaram muito na

¹ A Web 2.0 ou segunda internet é a atual fase do *World Wide Web*. Surgiu entre o começo e meados da década de 2000. O seu aspecto principal, em relação a formas anteriores da internet, consiste em que o usuário possui maiores possibilidades de interação com a internet, nesse contexto é que surgem as redes sociais, os sites de pesquisa e de enciclopédia online, como o Google e a Wikipédia. O conceito será aprofundado adiante, no capítulo 2, subcapítulo 3.

compreensão teórica de alguns pontos sobre a informática do ponto de vista do historiador.

Sobre a relação entre história e internet, os trabalhos de George Zeidan Araújo (2014) e Nucia Alexandra Silva de Oliveira (2014) foram de suma importância na medida que apresentaram um panorama geral e introdutório dos usos da internet por parte dos historiadores. O trabalho de Bruno Carvalho (2014), foi importante tanto no estudo da relação entre história e informática, quanto na história e internet e, em especial, a relação com as redes sociais.

Fora da área da história, o estudo de Priscila Viviane de Sousa Carvalho (2017), analisando as diferenças e semelhanças entre o PDF e EPUB e outros formatos de arquivos, contribuiu em muito em relação ao desenvolvimento das possibilidades de leitura da informática.

Durante o período da iniciação científica, a ideia de se construir um banco de dados surgiu como uma alternativa de sistematizar determinados documentos que dialogavam muito entre si e analisar aqueles dados principalmente do ponto de vista quantitativo. Contudo, durante o processo de estudo teórico e desenvolvimento empírico do banco de dados, a concepção do autor sobre a função do banco de dados ampliou-se, nesse sentido passou-se a ver esse instrumento não somente como uma ferramenta para se trabalhar dados quantitativas, mas sim uma forma ampla de produção histórica que é muito importante no diálogo com a história serial-quantitativa, mas também pode ser utilizada de diversas maneiras diferentes (no debate historiográfico, na sistematização de livros e documentos, na prática da docência). Nesse sentido, esse trabalho se justifica principalmente porque busca discutir e contribuir nos usos do banco de dados e da informática, de modo geral, na pesquisa histórica e no trabalho do historiador.

Esse trabalho está dividido em três partes principais: “História e informática”, “O processo de construção do banco de dados” e “considerações finais”.

No capítulo “História e informática”, busca-se defender a ideia de que a informática é uma ferramenta importante para o trabalho do historiador. Nesse sentido, discute-se algumas questões práticas como os usos das possibilidades de leitura e escrita e da internet pelos historiadores. Também analisa-se algumas questões teóricas, como os motivos de certa rejeição por parte dos historiadores dessa ferramenta, assim como a elaboração das fontes e documentos digitais e o conceito de banco de dados e sua relação

com a história serial-quantitativa. Nesse capítulo, buscou-se discutir alguns conceitos que foram importantes no processo de desenvolvimento do banco de dados.

No capítulo “O processo de construção do banco de dados”, são apresentadas as etapas do processo de construção do banco de dados realizado para a pesquisa, assim como aspectos da sua versão final e as suas utilidades. Foram quatro as etapas principais: a) análise preliminar das fontes; b) o processo de fichamento da documentação; c) desenvolvimento inicial do banco de dados; d) a construção do banco de dados.

Por último, nas “considerações finais”, busca-se discutir quais foram as principais contribuições e limites do trabalho e do banco de dados.

2 HISTÓRIA E INFORMÁTICA

O objetivo desse capítulo é discutir a relação entre a ciência histórica² e a informática³, atribuindo a devida ênfase a alguns conceitos e discussões que foram utilizadas no processo de construção do nosso banco de dados específico e serão aprofundados no capítulo posterior. Busca-se defender a ideia de que a informática é uma ferramenta que pode ajudar em muito no trabalho do historiador, contudo nos contrapomos a qualquer tipo de “fetichismo do computador”⁴, assim como ao “fetichismo dos livros”⁵. Entende-se que tanto as antigas ferramentas e experiências dos historiadores (a utilização de livros e documentos físicos, a vivência dentro do arquivo, etc.), quanto as novas ferramentas proporcionadas pela informática (utilização de arquivos digitais, editores de textos, páginas da internet, redes sociais, etc.) são importantes para a produção

² Para fins desse trabalho, o conceito de história adotado segue a definição do historiador francês Marc Bloch que caracteriza a história como a ciência dos homens no tempo (BLOCH, 2002). Segundo Bloch, “Diz-se algumas vezes: ‘A história é a ciência do passado.’ É [no meu modo de ver] falar errado [...] o objeto da história é, por natureza, o homem. Digamos melhor: os homens. Mais que o singular, favorável à abstração, o plural, que é o modo gramatical da relatividade, convém a uma ciência da diversidade. [...] ‘Ciência dos homens’, dissemos. É ainda vago demais. É preciso acrescentar: ‘dos homens, no tempo’” (BLOCH, 2002, p. 52-55).

³ Segundo Fernando Velloso, “Informática é informação automática, isto é, o tratamento da informação de modo automático. Portanto, informática pressupõe o uso de computadores eletrônicos no trato da informação. [...] O aspecto que se lhe pretende conferir de ciência acarreta, em busca por sua estruturação formal, o que se deve dar no terreno da semiótica: tem a informática comprometimentos tanto com a área das ciências exatas quanto com a área das ciências sociais. Pode-se considerá-la como situada na interseção de quatro áreas do conhecimento: Ciência da Computação, Ciência da Informação, Teoria dos Sistemas e Cibernética.” (VELLOSO, 2014, p. 1) A informática surge da relação entre o tratamento das informações (daí o fato de que antes da criação do conceito de informática, era denominada de *information science*, em português, ciência da informação) e o uso de computadores, ou seja, o uso de uma tecnologia formada pelo *hardware* (equipamento físico), *software* (parte “lógica”, isto é, o sistema operacional) e *peopleware* (usuário).

⁴ Na década de 1960 de o historiador francês, Emmanuel Le Roy Ladurie, escreveu “O historiador de amanhã será programador ou não será” (LADURIE, 1968 apud CARVALHO, 2014, p.168). Alguns historiadores, como François Dosse, acreditavam que por trás da história quantitativa havia um “discurso neopositivista que fetichizava o computador” (CARVALHO, 2014, p. 168). De fato, a introdução do computador na pesquisa histórica, principalmente na história serial e quantitativa, causou certa euforia entre os seus utilizadores, como exemplifica bem a frase de Ladurie (TAVARES, 2012; CARVALHO, 2014).

⁵ Ciro Flamarion Cardoso (1996), em um trabalho publicado em 1996, se preocupava com a continuidade da palavra escrita e do registro erudito da linguagem. Para Cardoso, a informática apresentava um empecilho a essa continuidade na medida que se garantia o acesso a informação sem a necessidade de ler ou realizando uma leitura limitada, além disso o historiador caracterizava a linguagem da informática e internet como “tipo de escrita empobrecida e gramaticalmente bárbara” (CARDOSO, 1996, p. 2). Segundo Cardoso, “Não compartilho, portanto, o otimismo daqueles que afirmam que a difusão dos microcomputadores fez-se a serviço da programação de textos, da escrita. Talvez seja verdade. Mas a maioria, e uma maioria que tende a continuar aumentando, dos textos digitados em computador é de nível muito baixo quanto às exigências de linguagem. É verdade que digitar na Internet constitui uma forma de escrita. E daí? A pichação de paredes e muros também constitui, mas não avança a preservação do registro erudito e sofisticado da língua [...]” (CARDOSO, 1996, p. 3).

histórica, cabendo a cada historiador utilizar o instrumento mais adequado de acordo com os seus interesses, objetivos e posições particulares.

Nesse sentido, esse capítulo está dividido em cinco subcapítulos: 1) “Os historiadores e a informática”, discutirá os motivos pelo qual os historiadores não utilizam tanto as ferramentas da informática; 2) “Leitura, escrita e os documentos digitais”, vai discutir como os historiadores podem utilizar a leitura, a escrita e as fontes digitais no fazer historiográfico; 3) “História e internet”, abordará o conceito de Internet, procurando aprofundar como o historiador pode se apropriar das ferramentas da Web 2.0; 4) “Banco de dados: uma forma de produção histórica”, irá discutir o conceito de banco de dados e os tipos existentes de BD que o historiador pode utilizar; 5) “História serial-quantitativa e o banco de dados”, discutirá a relação entre a história serial-quantitativa e o banco de dados.

2.1 Os historiadores e a informática

A grande maioria dos historiadores que trabalham com a relação entre história e informática ou história e internet levantam uma questão muito importante: por que os historiadores não utilizam ou, pelo menos, não o suficiente, as novas tecnologias⁶? Fábio Chang de Almeida (2011), tratando sobre a relação entre fontes digitais e a história, apresenta importantes argumentos sobre a rejeição dos historiadores em relação a utilização das fontes digitais. Alguns desses argumentos podem ser utilizados para o uso da informática na produção histórica de modo geral. Segundo o autor, uma das explicações para essa rejeição dos historiadores tem a ver com uma certa “tradição metódica-positivista” da historiografia que valorizava o uso de papéis em detrimento de outras formas de registro histórico. Como veremos adiante, a revolução documental

⁶ Vale ressaltar que entre os historiadores brasileiros a utilização da informática é relativamente recente. Célia Tavares cita um artigo de Ernesto Ruiz, de 1991, em que o historiador se preocupava em ensinar o básico do uso de computadores e sua relação com a pesquisa histórica, embora o objetivo do trabalho fosse mais complexo. Nesse artigo, se ensinava a diferença entre hardware e software, a função de um scanner, um editor de texto, etc. Segundo Tavares, “[...] trata-se de um texto fundador que demonstra, pelo cuidado em definir programas e equipamentos, o quanto a maioria dos historiadores desconhecia o recurso do uso de computadores, seus aplicativos e equipamentos associados” (TAVARES, 2012, p. 303). A autora ainda salienta como, durante a década de 1990, a utilização de computadores entre os historiadores ainda era precária. Ciro Flamarion Cardoso e Héctor Pérez Brignoli, pioneiramente no Brasil, dedicaram um capítulo intitulado “O uso da computação em História” em seu importante trabalho “*Os Métodos da História*”, nos anos 1970.

produzida pela Escola dos Annales rompe, em parte, com essa tradição. Segundo Almeida,

Por que os historiadores ainda relutam em aproveitar a Rede como fonte de pesquisa, especialmente enquanto fonte primária para pesquisas históricas?

Uma primeira explicação para este comportamento é de caráter histórico. Durante séculos, a historiografia baseou suas regras de validação de fontes e metodologia de análise em um suporte documental específico: o papel. Para a Escola Metódica, dita Positivista, do final do século XIX, o historiador deveria trabalhar, sobretudo, com documentos oficiais. Estes documentos eram, em última análise, textos registrados em papel: atos governamentais, tratados internacionais, códigos de leis, etc.

Mesmo atualmente, a grande maioria das fontes documentais consagradas no ofício do historiador ainda encontra sua materialidade no papel: correspondências, ofícios, requerimentos, atas, inventários, testamentos, processos, registros paroquiais, periódicos... Existe toda uma tradição historiográfica baseada nesse suporte específico. Até mesmo o estereótipo do historiador como “rato de arquivo” não dispensa a alegoria de um cenário de penumbra, onde um personagem com a postura arqueada e os óculos na ponta do nariz, analisa papéis amarelados em meio à poeira e ao mofo.

[...] Certamente tal resistência está relacionada, em parte, com a herança metodológica positivista que privilegiava os “papéis” oficiais. (ALMEIDA, 2011, p. 9-10)

O segundo motivo, segundo Almeida (2011), é de que a ausência de uma base teórico-metodológica não torna atrativo a utilização das fontes digitais pelos historiadores. Realmente, no que diz respeito a informática de modo geral, há uma certa carência em relação a trabalhos que sistematizem conceitos de informática e internet destinado aos historiadores, além disso há poucos *softwares* destinados especificamente a historiadores (THOMPSON FLORES, 2014). Isso se dá, talvez, pela novidade que representa a informática para a historiografia e por uma certa omissão dos historiadores nessa área⁷.

Aos argumentos muito bem desenvolvidos de Almeida (2011), sugiro ainda outro motivo: a mudança geracional. Os estudiosos de informática, sociólogos e pedagogos

⁷ “A dificuldade em absorver os novos recursos tecnológicos na prática historiográfica também passa, portanto, pela pouca oferta e difícil acesso no mercado de softwares específicos para a pesquisa histórica, o que poderia ser explicado pela omissão dos historiadores, em geral, nesse campo.” (THOMPSON FLORES, 2014, p. 244)

costumam dividir os usuários de internet em dois grupos principais: os nativos digitais ou *born digital* e os imigrantes digitais⁸. Os nativos digitais seriam pessoas nascidas a partir da década de 1980 que cresceram ou ainda eram adolescentes ou jovens quando ocorreu a difusão da informática e dos meios digitais, enquanto que os imigrantes digitais seriam pessoas nascidas antes de 1980 e que já eram adultas quando começaram a utilizar a informática. Não devemos nos prender tanto às datas e recortes temporais, mas sim a experiência geracional propriamente dita. Enquanto o primeiro grupo possui uma facilidade para a utilização de equipamentos digitais e está inserido em toda uma cultura digital (denominado por alguns de ciberespaço⁹), o segundo grupo não possui muitas habilidades com a informática e prefere meios analógicos, isto é, preferem o físico ao digital. Enquanto os nativos digitais são fluentes na linguagem da informática, os imigrantes digitais possuem um “sotaque” (PRENSKY, 2001)¹⁰.

Dessa diferença geracional podemos buscar compreender uma das razões do porquê os historiadores têm tanta dificuldade na utilização da informática. Segundo Thompson Flores (2015, p. 245)¹¹ mesmo os alunos nativos digitais (principalmente os nascidos nas décadas de 1980 e 1990) não tem um conhecimento tão profundo sobre a informática como se imagina, visto que sabem utilizar as redes sociais, a internet e algumas funcionalidades básicas (como, por exemplo, o Microsoft Word e o Microsoft

⁸ Esses conceitos foram sistematizados pela primeira vez por Marc Prensky no seu artigo “Digital natives, digital immigrants”, em outubro de 2001. Alguns estudiosos o consideram como o criador de ambos os conceitos.

⁹ “O ciberespaço (...) é o novo meio de comunicação que surge da interconexão mundial dos computadores. O termo especifica não apenas a infra-estrutura material da comunicação digital, mas também o universo oceânico de informações que ela abriga, assim como os seres humanos que navegam e alimentam esse universo.” (LÉVY, 1999, p.17)

¹⁰ “É importante fazer esta distinção: como os Imigrantes Digitais aprendem – como todos imigrantes, alguns mais do que os outros – a adaptar-se ao ambiente, eles sempre mantêm, em certo grau, seu “sotaque”, que é, seu pé no passado. O “sotaque do imigrante digital” pode ser percebido de diversos modos, como o acesso à internet para a obtenção de informações, ou a leitura de um manual para um programa ao invés de assumir que o programa nos ensinará como utilizá-lo. [...] Há centenas de exemplos de sotaque de imigrante digital. Entre eles estão a impressão de seu e-mail (ou pedir a secretária que o imprima para você – um sotaque ainda “mais marcante”); a necessidade de se imprimir um documento escrito do computador para editá-lo (ao invés de editá-lo na tela; e trazer as pessoas pessoalmente ao seu escritório para ver um web site interessante (ao invés de enviar a eles a URL). Tenho certeza de que você consegue pensar em um ou dois exemplos sem muito esforço. Meu exemplo favorito é “Você recebeu meu e-mail” pelo telefone. Aqueles de nós que são Imigrantes Digitais podem, e devem, rir de nós mesmos e de nosso ‘sotaque’.” (PRENSKY, 2001, p. 2)

¹¹ “Engana-se, contudo, quem pensa que para os alunos dos cursos de história de hoje, aqueles pertencentes à geração *born digital*, possa ser redundante trabalhar esse tipo de temática na graduação. Frequentemente, ocorre que ao serem colocados diante de computadores para manejarem softwares dedicados à elaboração de banco de dados e tabelas relacionais, como Excel e Access, esses alunos nascidos na era digital tendem a se assustar, revelando que sua intimidade com as tecnologias de informação limita-se aos recursos que derivam do uso da internet ou a um manejo simples de softwares dedicados à produção de textos e apresentações de slides.” (THOMPSON FLORES, 2014, p. 245)

Power Point), mas na hora de se utilizar programas mais complexos (como o próprio Microsoft Access e o Microsoft Excel) e ferramentas mais avançadas (como a edição de arquivos de PDF, vídeos e imagens) demonstram profundas dificuldades.

De fato, grande parte dos alunos nativos digitais, embora possuam um conhecimento básico em relação a informática e a internet, não conseguem utilizar programas e ferramentas mais complexas. Contudo nota-se que uma pequena minoria possui um conhecimento intermediário ou avançado e é exatamente esse setor que está produzindo, na última década, trabalhos que dialogam com a informática e com a internet. Além disso, os estudantes com conhecimento básico podem muito bem, a partir de alguns cursos (e até algumas horas no *Youtube*), desenvolverem seus conhecimentos sobre determinados *softwares* e ferramentas, pois já estão geralmente muito mais familiarizados com a informática do que as gerações anteriores.

Embora seja correta a ideia que os jovens atuais possuem limitações, acredito que as próximas gerações de nativos digitais (principalmente os nascidos nos anos 2000 e 2010) utilizarão a informática e a internet de modo mais avançado e sistemático que os historiadores atualmente. Isso porque já desde infância utilizam aparelhos eletrônicos e estão inseridos na cultura digital ou ciberespaço. Vale salientar que os alunos nativos digitais que estão na universidade atualmente (isto é, nascidos nos anos de 1980 e 1990) só tiveram acesso amplo a informática e a internet quando já eram pré-adolescentes, adolescentes ou jovens adultos. O próprio autor só ganhou o primeiro computador pessoal, para uso coletivo de toda a família, em 2006, quando tinha dez anos de idade. Enfim, para que a informática se difunda mais na ciência histórica, é necessário que os historiadores e docentes atuais contribuam com a inclusão de disciplinas sobre a relação entre história, informática e a internet nos currículos das principais universidades do Brasil.

2.2 Leitura, escrita e os documentos digitais

A informática facilitou em muito duas das principais funções do historiador: a leitura e a escrita. Enquanto que a “leitura digital” ou a “leitura informática” acarretou

em problemas e discussões mais profundas entre os historiadores, a escrita foi facilmente aceita.

A leitura de textos foi realizada, durante muito tempo, a partir da cultura escrita e alfabética registrada no papel. Os livros, arquivos e documentos eram todos produzidos a partir do papel. Há intelectuais que acreditam que, inclusive, estamos passando por um processo de transição para uma nova cultura, a cultura digital. Segundo a historiadora Anita Lucchesi, “[...] a humanidade encontra-se na transição da cultura alfabética para a cultura digital, de modo que a forma como apresentamos as informações está se modificando” (LUCCHESI, 2014, p. 47).¹²

Nesse contexto, se faz necessário refletir sobre as novas formas de leitura que a informática oferece. A “leitura informática”, atualmente, pode ser feita sobretudo através de computadores pessoais (PC, *Personal Computer*), *notebooks*, *tablets*, *smartphones* e *e-readers*¹³ (leitores digitais). Existem alguns formatos principais: PDF (*Portable Document Format*), EPUB (*Electronic Publication*), DOC (abreviação de *document*).¹⁴

O PDF¹⁵ é provavelmente o formato mais utilizado entre os historiadores. A função desse formato é reproduzir textos, imagens e gráficos independente da resolução e dispositivo. Costuma ser o arquivo de texto mais utilizado em computadores pessoais e *notebooks*. O PDF é um arquivo em que se pode alterar a diagramação¹⁶, mas não o que está no arquivo. Portanto, não se pode mudar o tipo e tamanho da letra, nem muito menos

¹² “Na transição do analógico para o digital, como muito antes a passagem da tradição oral para uma tradição alfabética, criam-se novas formas de expressar e dispersar informações. Emergem desse caldo novas formas de escrever e ler. Debate-se o futuro do livro. Acompanha-se o surgimento de novas linguagens, tanto as de programação (HTML, por exemplo), mais invisíveis ao usuário comum de internet, quanto a largamente compartilhada linguagem informática, que não é nem verbal, nem oral, mas icônica (no Brasil ou no Japão, a imagem de cesta de lixo, de um envelope ou de uma lupa em um computador querem dizer a mesma coisa). Vemos o surgimento de novas possibilidades narrativas, menos lineares e mais hipertextuais, que intensificam potencial do virtual e podem abusar do audiovisual. Aparecem novos arranjos museográficos e se discute o que a modelagem e a impressão 3D podem fazer pelos museus” (LUCCHESI, 2014, p. 47).

¹³ O *e-reader* ou leitor digital é um aparelho ou *software* que tem como função servir de suporte físico para livros digitais (CARVALHO, 2017). Esse aparelho procura simular a página de um papel. O *e-reader* mais famoso é o Kindle, da Amazon.

¹⁴ Há vários formatos para além desses três, porém são menos utilizados. Alguns exemplos são: MOBI, AZW, IBA, DJVU, etc.

¹⁵ O formato PDF foi criado em 1991 pela empresa *Adobe Systems Incorporated* com o objetivo de servir para as empresas trocarem versões digitais de documentos.

¹⁶ “Distribuição dos elementos gráficos no espaço limitado das margens da página que vai ser criada” (CARVALHO, 2017, p. 23).

modificar a posição do texto no arquivo. Como o PDF suporta a digitalização¹⁷ de imagens, costuma ser mais utilizado entre os historiadores. É uma ótima alternativa para quem pretende digitalizar livros e documentos físicos.

O EPUB é mais utilizado na leitura de obras literárias e livros *blockbusters* e *best sellers*¹⁸. Ao contrário do PDF, o EPUB não suporta imagens¹⁹, somente textos. O arquivo é voltado para o texto, por isso é predominante entre usuários de *e-readers* e leitores que utilizam o *smartphone* e o *tablet*. No EPUB, se pode alterar o tamanho e formato no texto, além de escolher a locação do texto (se com espaçamento nas laterais ou entre as frases, por exemplo). Segundo Priscila Viviana de Souza Carvalho, “[...] EPUB foi projetado para conteúdo fluido, o que significa que a tela de texto pode ser otimizada de acordo com o dispositivo usado para leitura. O padrão é destinado a funcionar como um único formato oficial para distribuição e venda de livros digitais” (CARVALHO, 2017, p. 32). Provavelmente seja o melhor formato para leitura, mas o seu grande problema é que pressupõe que o arquivo esteja digitalizado em formato de texto, ao contrário do PDF que aceita imagens.

Para esclarecer, vejamos um exemplo. Um determinado usuário pretende digitalizar um livro físico dos anos 1970 que não existe em formato digital. Após digitalizar esse livro, o usuário terá várias imagens em *jpg* ou *jpeg* (formatos de imagens). Esse usuário poderá produzir um PDF de imagens, mas não poderá produzir um EPUB. Se, por acaso, utilizar em seu PDF de imagens uma ferramenta denominada de OCR (*Optical Character Recognition*), que tem como objetivo reconhecer caracteres de imagens, o usuário poderá digitalizar seu texto e, a partir disso, produzir um EPUB. Porém, o OCR dificilmente realiza uma digitalização perfeita, o que provavelmente deixará falhas na leitura do arquivo.

¹⁷ “Digitalizar significa transformar documentos arquivísticos tradicionais em dados binários processados por computadores usando diferentes recursos (software e hardware).” (THOMPSON FLORES, 2014, p. 246)

¹⁸ As lojas de livros digitais costumam disponibilizar os arquivos em formato EPUB e PDF (CARVALHO, 2017).

¹⁹ Aqui me refiro a imagens no sentido de que, ao contrário do PDF, não se pode fazer a leitura de uma imagem através do EPUB. Por exemplo, vamos supor que um historiador fotografou um documento de vinte páginas e agora possui 20 arquivos de imagem em formato jpeg ou jpg. Esse historiador não pode transformar essas imagens em um EPUB e realizar a leitura, mas pode transformar em PDF. O EPUB suporta imagens em seu corpo, mas não pode ser feito totalmente de imagens, como o PDF.

Enfim, a grande diferença do PDF e do EPUB, para os historiadores, é de que o primeiro é mais amplo e aceita imagens, enquanto o segundo é mais restringido e aceita somente textos. Há, um terceiro formato de transição, que é o DOC.

Embora o DOC²⁰ seja, na verdade, um formato de edição de texto e não de leitura de arquivo, sugiro a sua utilização como uma forma de transição. A função do DOC é editar os arquivos textualmente. Os historiadores conhecem bem o arquivo DOC, já que é o principal arquivo utilizado pelo Microsoft Word. Contudo, o seu caráter de transição que pretendo enfatizar aqui, é pouco utilizado.

O DOC pode servir, para além da edição de texto, para leitura propriamente dita. A sua função enquanto editor de texto pode servir para que o usuário modifique um PDF ou transforme um PDF em EPUB ou vice-versa. Por exemplo, vamos supor que o usuário possui um PDF no computador com o formato e tamanho da letra Arial 12, mas possui dificuldades na leitura do arquivo em seu computador, celular, tablet e *e-reader*. O formato DOC, a partir do Microsoft Word, pode ser utilizado para editar esse PDF, mudando sua letra, formato e até a organização da página. Um outro arquivo PDF será gerado dessa edição, mais ajustado aos interesses do usuário.

Um site que realize muito bem essa função de transição é o *Google Docs*²¹. A partir desse site, se pode fazer o *download* de um arquivo de texto para vários formatos diferentes, como: PDF, DOC, EPUB, HTML²², TXT²³, etc. A grande importância dessa ferramenta é poder salvar os arquivos diretamente em EPUB de modo simples. Vejamos um exemplo. Vamos supor que o usuário precisa ler um texto que está em formato DOC, porém, por questões pessoais, não simpatiza muito com a leitura pelo computador e nem através do *Microsoft Word*. Esse usuário em particular tem simpatia pelo suporte do *e-reader*. A maioria dos aparelhos de *e-readers* foram desenvolvidos principalmente para a leitura do formato EPUB, porém o Microsoft Word não salva arquivos nesse formato. O usuário tem duas opções: pode fazer a conversão desse arquivo DOC para EPUB

²⁰ “É o sufixo do formato pertencente ao Microsoft Word, o processador de texto produzido pela Microsoft. Utiliza atualmente como extensão padrão dos arquivos de texto: .docx” (CARVALHO, 2017, p. 31).

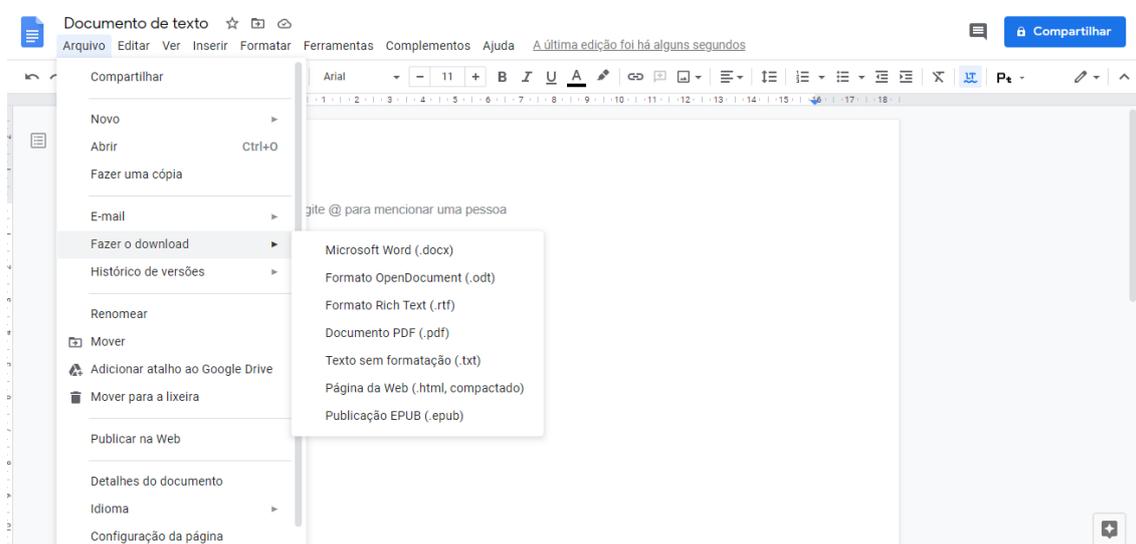
²¹ A utilização do *Google Docs* pressupõe que o usuário possua uma conta no *Google*.

²² “HTML é a abreviação para a expressão inglesa *Hyper Text Markup Language*, que significa Linguagem de Marcação de Hipertexto. É utilizada na construção de páginas e podem ser interpretadas por navegadores” (CARVALHO, 2017, p. 31).

²³ O TXT é um arquivo de texto normal, sem necessidade de formatação. No Microsoft Windows, costuma ser o formato de texto do bloco de notas ou *notepad*.

através de sites na internet ou pode enviar um e-mail para que a empresa responsável pelo aparelho faça a conversão, caso a empresa disponibilize esse serviço. Esse arquivo, por ser feito a partir de uma conversão automática, provavelmente terá erros na sua diagramação, como, por exemplo, alguns erros no tamanho das letras ou de imagens. A outra opção é utilizar o Google Docs e criar um arquivo do zero. Trazendo manualmente o conteúdo do DOC que está no Microsoft Word para o Google Docs, o usuário poderá salvar o arquivo não só em EPUB, mas em diversos outros formatos (inclusive o PDF), além de editar o arquivo como bem quiser (ver Figura 1).

Figura 1 - Exemplo de um arquivo no Google Docs



Fonte: elaboração pelo autor

Com esse exemplo o que procuro defender é a utilização de todos os formatos de arquivo pelo historiador. Não há necessidade de ficar preso somente ao PDF. Cada dispositivo possui um formato mais adequada a sua utilização e cada gênero de texto requer uma leitura. Por exemplo, é muito mais adequado a utilização do PDF no computador ou notebook na leitura de um texto científico, já que é necessário fazer anotações e marcações permanentes, em contrapartida na leitura de um texto literário ou de algum trabalho mais simples, em que não se necessita fazer tantas marcações e reflexões durante o processo de leitura, o EPUB e a utilização de *e-readers*, *tablets* e *smartphones* me parece mais favorável. Além disso, por exemplo, em arquivos PDF de imagens, dependendo do tipo de digitalização e da qualidade da imagem, talvez seja mais interessante imprimir e realizar a leitura no formato físico do papel.

Enfim, não há uma regra específica de que formato e aparelho utilizar no processo de leitura. O que busquei discutir aqui são as múltiplas possibilidades de leitura que a informática apresenta aos usuários e sugerir, de acordo com a minha experiência, algumas formas de se utilizar essas possibilidades.

Tabela 1 - Formatos de leitura digital

PDF	EPUB	DOC
Aceita imagens e textos	Aceita somente textos	Pode editar textos e imagens
O tamanho e formato da letra e a posição dos textos e das imagens não podem ser alterados no processo de leitura	O tamanho e formato da letra e espaço podem ser alterados no processo de leitura de acordo com o usuário	O tamanho e formato da letra e espaço podem ser alterados no processo de leitura de acordo com o usuário Tem a função de transição, podendo realizar alterações nos arquivos PDF
Mais favorável a leitura de trabalhos científicos	Mais favorável a leitura de trabalhos literários e leitura mais leves	Não há gênero textual mais favorável
Mais adequado ao computador pessoal e <i>notebook</i>	Mais adequado ao <i>tablet</i> , <i>e-reader</i> e <i>smartphones</i>	Mais adequado ao computador e <i>notebook</i>

Fonte: elaboração do autor a partir de CARVALHO (2017)

Se a “leitura informática” se difundiu de modo mais lento entre os historiadores e, inclusive, acarretou em debates historiográficos²⁴, a escrita informática, em

²⁴ A questão sobre a leitura no Google, na internet e na informática acarretou em alguns debates entre os historiadores. Carlos Ginzburg apresentou certa simpatia pela leitura através da internet e computadores,

contrapartida, provavelmente, de todas as possibilidades da informática, foi a que mais atraiu os historiadores e acadêmicos de modo geral.

Luciano Figueiredo (1997), em um trabalho publicado em 1997, quando a informática e a internet ainda não eram tão popularizadas no Brasil, abordou esse problema. Vejamos, nas palavras do autor:

Os editores de texto constituem para a maioria a porta de entrada do historiador no universo da informática. Através desses programas, poucos resistem ao fascínio e a facilidade que oferecem para a preparação de textos, tarefa tão freqüente no ofício.

Não é preciso lembrar os possíveis usos que têm os micros pessoais para a produção de texto, desde a transcrição de documentos nos arquivos à redação de trabalhos científicos. Os processadores estruturam-se a fim de facilitar uma série de procedimentos antes trabalhosos e muitas vezes demorados.

À medida que vai sendo digitado, o texto é armazenado na memória, onde o programa torna quase tudo possível. Se, semelhante à tradicional máquina de escrever, o princípio da digitação não muda, por trás do teclado do micro há cada vez mais poderosos programas de armazenamento e organização dos textos que são definidos. Tais programas tornam *extremamente* simples tarefas que antes exigiam certo desgaste tais quais correção de texto, revisão ortográfica, normalização editorial, confecção de índices, estabelecimento de listas e bibliografia em ordem alfabética, definição de notas e, ainda, uma adequada apresentação do texto final e sua reprodução através de impressão. A possibilidade de rearranjar a estrutura interna do texto, modificando a ordem de parágrafos, acrescentando frases no meio de sentenças já digitadas, afasta de vez qualquer resistência a seu uso.

O editor ou processador de texto possui assim, como uma de suas maiores qualidades, a possibilidade de armazenamento associada à ordenação, ou seja, os dados podem entrar da maneira mais desordenada possível que o programa dispõe de meios rápidos e eficientes para organizá-los de acordo com a estrutura final do texto que se queira adotar. (FIGUEIREDO, 1997, p. 595)

embora salientasse que a “leitura informática” não substitui os livros! Visão mais pessimista é a de Roger Chartier que argumenta que a “[...] leitura feita na tela no computador é fragmentária, algo que o Google fomentaria pela quantidade e pela velocidade da informação que se acumula” (CORREA; SCHULZ, 2011, p 270). Ginzburg contra-argumenta afirmando que toda a leitura é fragmentária, na medida que “[...] isolar frases de seu conteúdo amplo original, descontextualizar trechos e capítulos, ignorar a data de escrita, são atitudes que o leitor faz a toda hora independentemente do suporte do texto—e mesmo do tipo de texto” (CORREA; SCHULZ, 2011, p. 270).

Nota-se nas palavras de Figueiredo um certo fascínio pelos editores de textos e suas funcionalidades. Atualmente, os editores de textos (principalmente o *Microsoft Word*) já foram completamente difundidos não só entre os historiadores, mas entre os acadêmicos de modo geral. Uma possibilidade de escrita que se desenvolveu na década de 2010 é a escrita compartilhada através de ferramentas específicas na Web, como o Google Docs, ou de sites de armazenamento em nuvem (*cloud computing*), como o Google Drive. A superioridade do computador ou *notebook* sobre as velhas ferramentas de produção de escrita (como a máquina de escrever, por exemplo) fez com que muitos dos antigos produtos, inclusive, desaparecessem totalmente do mercado.

Os arquivos também se beneficiaram muito do desenvolvimento da informática. Há, atualmente, diversos acervos digitalizados na internet²⁵, sem contar os acervos pessoais²⁶ dos historiadores e pesquisadores. Pode-se fotografar a documentação a partir de fotos realizadas por *smartphones* e *tablets* pessoais. Além disso, as próprias páginas da internet podem ser utilizadas como fontes primárias e secundárias, desde que sempre se tenha muito cuidado aos utilizá-las²⁷.

Todo esse processo de informatização das fontes, necessitou que fosse teoricamente desenvolvida e sistematizada. Com isso, surgiu o conceito de fontes digitais. Segundo Fábio Chang de Almeida, a característica principal do documento digital é de que há uma “[...] dissociação entre o suporte físico e o seu conteúdo informacional. Sendo assim, é possível o descarte do suporte físico e a manutenção de seu conteúdo em um novo suporte” (ALMEIDA, 2011, p. 16). Ao contrário do documento físico, que possui uma particularidade material própria, o documento digital é independente do suporte, de modo que pode ser aberto em vários locais diferentes.

²⁵ Um exemplo é o acervo da revista do Instituto Histórico e Geográfico Brasileiro (IHGB). Há revistas desde do ano de 1839 digitalizadas em formato PDF. Ver: Revista do Instituto Histórico e Geográfico Brasileiro. Disponível em: <<https://web.archive.org/web/20150924033333/http://www.ihgb.org.br/rihgb.php?s=19>>. Acessado em: 20/10/2020.

²⁶ Há a possibilidade não somente de acervos guardados em computadores pessoais, mas também de salvar esse acervo na internet através da computação em nuvem (*cloud computing*).

²⁷ Nos últimos anos se desenvolveu o problema da “*fake news*”. As “*fakes news*” são notícias falsas produzidas por usuários ou por grupos de usuários que visam produzir notícias inventadas ou manipuladas com o intuito de atacar a imagem de personagens políticos, empresas, instituições, partidos políticos e pessoas privadas. Recentemente, em 2020, o filme do cineasta polonês Jan Komasa, *Hejter* (em inglês, *The Hater*; em português, *Rede de Ódio*), mostra como se articulam e se organizam as empresas que produzem essas “*fakes news*”. Existe, na realidade, uma verdadeira “indústria de *fake news*”.

O conceito de documento digital parte do pressuposto de que o documento é uma forma de registro da humanidade, nesse sentido o documento digital seria

aquele documento – de conteúdo tão variável quanto os registros da atividade humana possam permitir – codificado em sistema de dígitos binários, implicando na necessidade de uma máquina para intermediar o acesso às informações. Tal máquina é, na maioria das vezes, um computador (ALMEIDA, 2011, p. 17).

Almeida (2011) ainda procura sistematizar as diversas formas de fontes digitais (ver Figura 2). Nesse sentido, o autor afirma que há fontes primárias digitais e fontes não-primárias digitais. As fontes primárias digitais²⁸ podem ser divididas em duas: documentos primários digitalizados, isto é, arquivos que existem em formato físico, como, por exemplo, documentos oficiais do estado, registros de batismo, processos crime, etc., que foram digitalizados; documentos primários digitais exclusivos, ou seja, documentos produzidos pela própria informática e internet, como os sites, blogs, redes sociais, etc²⁹.

As fontes não-primárias digitais que são arquivos produzidos pela informática e voltados somente para ela como, por exemplo, trabalhos acadêmicos, livros, artigos de jornais online, etc.

A teorização das fontes digitais produzida por Almeida, embora muito bem desenvolvida, não contempla os livros em formato físico digitalizado³⁰. Parece-me equivocado ignorar esse tipo de documento digital, visto que é um dos principais documentos digitais difundidos na internet atualmente. Talvez o autor não tenha incluído essa categoria por conta de ser considerado pirataria digitalizar livros com direitos autorais vinculados a editoras (a questão da legitimidade dos direitos autorais será discutida posteriormente). Nesse sentido, faria uma sugestão acrescentando uma divisão nas fontes não-primárias digitais: os documentos não-primários digitalizados, ou seja,

²⁸ A digitalização de acervos e fontes pode ser feito, atualmente, tanto por pesquisadores, quanto pelas instituições. Nesse sentido, mesmo quando realizado por instituições, a digitalização procura ter problemas graves, nesse sentido que Thompson Flores (2014) defende a utilização e o desenvolvimento de um método para esse procedimento.

²⁹ Também denominados de “nato digitais”.

³⁰ Nesse trabalho, optou-se por abordar uma divisão de fontes e documentos digitais que leva em consideração o seu caráter primário e não-primário. Contudo, há historiadores que analisam a questão do documento digital com base na sua origem, isto é, se é nato digital ou se tem uma origem em meios analógicos.

livros, artigos e revistas produzidas principalmente como um documento físico (um livro produzido antes da época da informática); e os documentos não-primários exclusivos que seriam trabalhos produzidos principalmente para o meio digital (ver Tabela 2).

Figura 2 - Teorização dos documentos digitais proposta por Almeida

Fontes Digitais		
Fontes Primárias Digitais		Fontes Não-Primárias Digitais
		
Documentos primários digitais		Documentos não-primários digitais
		
Documentos primários digitalizados	Documentos primários digitais exclusivos	Exemplos: Livros, dissertações, teses, <i>papers</i> e artigos em formato digital.
Documentos que existem em outro suporte, anterior à digitalização. Exemplo: pôster da II Guerra após sofrer processo de digitalização.	Documentos que não existem em outro suporte, além do digital. Exemplo: alguns <i>sites</i> da Internet.	

Fonte: Almeida, 2011.

Tabela 2 - Sugestão sobre os documentos não-primários digitais

Documentos não-primários digitais	
Documentos não-primários digitalizados	Documentos não-primários exclusivos
Documentos que existem em outro suporte e foram digitalizados. Exemplo: livros, artigos, documentos, etc.	Documentos produzidos somente ou principalmente para meios digitais. Exemplos: livros digitais, teses, dissertações e monografias, etc.

Fonte: elaboração pelo autor com base em Almeida, 2011.

2.3 História e internet

Quando se estuda as origens e o desenvolvimento da internet³¹ é necessário que se faça uma divisão entre dois períodos principais: o período das origens da internet no pós-guerra e da sua utilização para fins militares e acadêmicos durante o da Guerra Fria; e o segunda período, dos anos 1990 até atualmente, quando ocorre o desenvolvimento do *World Wide Web* (WWW ou Web) e a rede passa a ser utilizada cada vez mais para fins domésticos e se populariza no mundo todo, tornando-se parte da cultura mundial.

O surgimento da internet está relacionado com as inovações tecnológicas que surgem após a Segunda Guerra Mundial, durante o contexto da Guerra Fria. O primeiro esboço da internet atual, surgiu em 1969 como um projeto norte-americano da ARPANET³², rede da Agência de Investigação de Projetos Avançados dos Estados Unidos fundado em 1957 como resposta ao Sputnik da URSS, que visava produzir uma rede de comunicação que se manteria intacta mesmo após ataques militares (BURKE; BRIGGS; 2006; ABREU, 2009; ZEIDAN ARAÚJO, 2014).

Além dos objetivos militares, a ARPANET também serviria para fins acadêmicos. O Departamento da Defesa dos Estados Unidos criaria uma rede específica para usos militares (denominada de Milnet) e deixara a ARPANET (agora renomeada para ARPA-Internet) somente para utilização em pesquisas. Na década de 1980, seria criada a

³¹ “A Internet é uma rede de computadores de alcance global que interconecta milhões de equipamentos através do mundo. Inicialmente estes equipamentos eram essencialmente computadores de mesa, estações de trabalho baseadas em UNIX e servidores que armazenavam e transmitiam informações como sites e mensagens de e-mail. Todavia, cada vez mais equipamentos vêm sendo conectados à rede, como PDAs (Personal digital assistant, ou palmtop), televisores, notebooks, telefones celulares, automóveis, câmeras de vídeo, etc. Em função disso, o termo ‘rede de computadores’ parece estar defasado em sua aplicação à Internet” (ALMEIDA, 2011, p. 27).

³² “No início, tratava-se de uma rede limitada (Arpanet), compartilhando informação entre universidades “hi-tec” (outra palavra nova) e outros institutos de pesquisa — em 1975 havia dois mil usuários. Graças ao tipo de informação que estava sendo compartilhada, um elemento essencial de sua razão de ser era que a rede pudesse sobreviver à retirada ou destruição de qualquer computador ligado a ela, e, na realidade, até à destruição nuclear de toda a “infra-estrutura” de comunicações (“infra-estrutura” era outra palavra nova). Essa era a visão do Pentágono. A visão das universidades era que a Net oferecia “acesso livre” aos usuários professores e pesquisadores, e que eram eles comunicadores. Qualquer que fosse a posição vantajosa, de cima ou de baixo, era de importância crucial, tanto de imediato quanto a longo prazo, que a “arquitetura do sistema” (termo empregado com frequência) diferisse daquela construída para a rede telefônica. Os responsáveis pelo sistema se orgulhavam disso. Qualquer computador podia se ligar à Net de qualquer lugar, e a informação era trocada imediatamente, em “fatias” dentro de “pacotes”. O sistema de envio quebrava a informação em peças codificadas, e o sistema receptor juntava-a novamente, depois de ter viajado até seu destino. Esse foi o primeiro sistema de dados empacotados da história.” (BURKE; BRIGGS; 2006, p. 301)

NFSNET (desenvolvida pela NFS, National Science Foundation), o que faria com que a ARPANET fosse desativada em fevereiro de 1990 (ZEIDAN ARAÚJO, 2014).

A NFSNET representou quase que uma forma de transição da ARPANET para a WEB moderna. Segundo George Zeidan Araújo,

Como a tecnologia de redes de computadores era de domínio público, a NFSNET foi deixando de ser controlada pela NFS, ao mesmo tempo em que o Departamento de Defesa dos EUA financiava a fabricação de computadores incluindo o protocolo TCP/IP, que se havia tornado padrão para a conexão entre computadores. Assim, na década de 1990, a maior parte dos computadores nos EUA era capaz de se conectar a alguma rede. Após a extinção da NFSNET, em 1995, o crescimento da internet sob provedores privados de serviços de rede foi exponencial e alcançou muito rapidamente a dimensão de uma rede global de redes de computadores. O próprio projeto original da Arpanet contribuiu, sem dúvida, para que isso acontecesse, pois, além de contar com protocolos de comunicação abertos, a rede era descentralizada e reconfigurável, permitindo a adição contínua de novos nós. (ZEIDAN ARAÚJO, 2014, p. 154)

Aqui não nos interessa tanto as singularidades e particularidades do funcionamento das versões iniciais da internet, mas sim salientar o amplo processo pelo qual passou a internet até o período atual, da Web 2.0. De um sistema criado para fins militares, foi se difundido para pesquisas e o ambiente acadêmico até os anos 1990, quando passou a ser cada vez mais utilizada para fins domésticos e se difundir no mundo inteiro.

A partir da década de 1990 surge o *World Wide Web* (WWW ou Web). É um tanto quanto difícil tratar desse assunto, visto que vivemos em uma época em que a Web ainda está se desenvolvendo e estabelecer divisões temporais nessas condições são sempre muito complexas. Costuma-se dividir a Web em dois períodos principais: a primeira internet ou Web 1.0 que seria de meados dos anos 1990 até 2004; e a segunda internet ou Web 2.0, de meados dos anos 2000 até os dias atuais (ver Tabela 3). Há alguns estudiosos que afirmam existir uma Web 3.0 ou Web semântica, outros acreditam que essa seja uma nova etapa que vai surgir no futuro da Web 2.0 (ZEIDAN ARAÚJO, 2014; LUCCHESI, 2014; OLIVEIRA, 2014; ALMEIDA, 2011; TAVARES, 2012).

A Web 1.0 foi criada pelo inglês Tim Berners-Lee, em 1990³³ (ALMEIDA, 2011; TAVARES, 2012)³⁴. Essa primeira internet criou as bases para a internet atual. A Web 1.0 difundiu a utilização de *websites* de maneira relativamente simples (a partir do navegador ou *browser*), a utilização de hipertextos³⁵, *hiperlinks*³⁶ e o sistema de envio e recebimento de *e-mails*. Segundo Zeidan Araújo,

essa “primeira internet” (ou “web 1.0”) possuía uma série de limitações que seriam progressivamente superadas, até o estabelecimento, por volta de 2004, da chamada “web 2.0”, estágio no qual a internet se encontraria atualmente (ZEIDAN ARAÚJO, 2014, p. 154)

A Web 2.0 é um termo criado por Tim O'Reilly. A Web 2.0 significou um avanço em relação a Web 1.0, nesse sentido várias características que já existiam na primeira internet se aprofundaram na segunda. A grande diferença entre ambas seria de que a Web 2.0 traria uma nova forma de interação do usuário com a Web. Em sua primeira versão, o usuário comum praticamente só conseguiu visualizar dados, agora poderia produzir, ele próprio, as suas próprias páginas de internet. É nesse contexto que os blogs e as páginas pessoais de internet se difundem. Surgem enciclopédias online (como o caso da *Wikipedia*) e redes sociais de vários tipos diferentes (*Youtube*, *Orkut*, *MySpace*, *Facebook*, etc.) e sites destinados a pesquisa online (*Google*, *Yahoo*, *Uol*, etc.).

Não pretendo aqui discutir o conceito de Web 3.0 ou Web semântica, visto que, ao contrário das duas primeiras versões da Web, é um conceito ainda um tanto quanto pouco desenvolvido e não consolidado, contudo é difícil colocar a ampla difusão dos *smartphones* e dos aplicativos ocorridos durante toda a década de 2010 no mesmo conceito de Web 2.0 junto com a origem dos blogs e redes sociais durante os anos 2000³⁷.

³³ Há divergência sobre quem criou o conceito e sobre a utilização ou não da divisão Web 1.0 e Web 2.0. Segundo o próprio Tim Berners-Lee, não há razão para essa divisão, visto que a internet atual continua se baseando nos mesmos preceitos criados anteriormente (ZEIDAN ARAÚJO, 2014, p. 155).

³⁴ “Em 1991 foi publicado o primeiro website (por vezes agora chamado apenas de site ou sítio), que foi também o primeiro diretório” (TAVARES, 2012, p. 307).

³⁵ O conceito de hipertexto foi criado na década de 1970, pelo norte-americano Theodore Nelson, para caracterizar textos e leituras não-lineares realizados através da informática (CARVALHO, 2017).

³⁶ *Hiperlinks* são referências a outro arquivo de hipertexto ou outra parte dentro de um arquivo de hipertexto. Atualmente se utiliza apenas o termo “*link*”.

³⁷ Quando falamos nas várias versões da Web, deve ser levado em consideração os interesses econômicos por trás das criações dos conceitos. Pesquisando na internet, se pode achar até a definição da Web 4.0. Os cientistas da computação que desenvolvem esses conceitos muitas vezes são empresários do ramo da informática ou professores que vendem cursos sobre o assunto, daí o interesse em promover novos conceitos.

Afinal, há mais rupturas ou continuidades entre o desenvolvimento da Web nos anos 2000 e 2010? Essa é uma questão que os cientistas da computação, estudiosos da informática e historiadores do futuro terão que responder.

Tabela 3 - Web 1.0 e Web 2.0

Web 1.0	Web 2.0
1990 até 2004	2004 até os dias atuais
Predominava a visualização de informações	Surge a possibilidade da interação do usuário com os sites e a criação de informações
Difusão do website, e-mail e hipertextos e hiperlinks	Desenvolvimento dos blogs, páginas pessoais, redes sociais e de sites destinados a pesquisa online (Google)

Fonte: elaboração pelo autor com base em Zeidan Araújo, 2014; Lucchesi, 2014; Oliveira, 2014; Almeida, 2011; Tavares, 2012.

Quando procuramos discutir a relação entre história e internet, devemos levar em consideração como o historiador atualmente vai se apropriar das ferramentas disponíveis pela Web 2.0 na produção histórica. Há três ferramentas principais que o historiador pode utilizar: a pesquisa histórica na internet, as redes sociais e os aplicativos e páginas na internet.

No que diz respeito à pesquisa, a internet abriu um mar de possibilidades aos historiadores³⁸. Além de poder encontrar diversos livros e artigos em diversas línguas diferentes, o historiador pode utilizar as próprias páginas de internet e redes sociais como fontes primárias (como discutindo no subcapítulo anterior). Esse trabalho mesmo, devido a pandemia do coronavírus, foi feito quase que exclusivamente a partir de documentos digitais não-primários - de acordo com o conceito de Almeida (2011) -, principalmente livros digitalizados. Antes da era da informática e da internet, um historiador demoraria anos para formar uma biblioteca física com obras de referência não só no campo da história, mas também na literatura, sociologia, economia, ciência política, etc.

³⁸ Em relação a divulgação da produção acadêmica, duas redes sociais importantes são a Academia.edu e ResearchGate.

Atualmente, com apenas alguns cliques e *downloads*, qualquer usuário pode ter, em meses ou anos, uma biblioteca digital gigantesca.

É muito pouco discutido como os sites de armazenamento de documentos digitais revolucionaram a forma como é difundido os livros. Uma possível razão para essa ausência seria o fato de que a maioria dos livros digitais disponibilizados nesses sites são pirateados. Há, no Brasil, toda uma legislação que protege os direitos autorais de editoras e empresas, porém, na prática, essa legislação é constantemente burlada tanto nas universidades e escolas, quanto na internet. A legitimidade dessa legislação costuma ser questionada por professores e alunos, visto que representam um entrave para a educação universitária no Brasil, na medida que as bibliotecas costumam apresentar alguns problemas nos seus acervos, devido à falta de verbas.

Em relação as redes sociais, o historiador pode se apropriar dessas ferramentas de várias formas diferentes. Em primeiro lugar, como fonte primária. Segundo Bruno Leal Pasto de Carvalho, as redes sociais

são fontes de pesquisa histórica totalmente inovadoras. Uma espécie de ‘documento virtual’. Não é difícil entender isso. Grande parte das manifestações sociais e políticas na atualidade, por exemplo, antes de chegarem às ruas, passam por redes como o Twitter, o Facebook, o Youtube ou o Orkut. É nesses espaços, em grande medida, que acontece a política e o engajamento social nos dias de hoje, além de serem essas redes lugares privilegiados para a formação da opinião pública. Como será possível, por exemplo, aos historiadores do futuro (e, por que não, já do presente) compreender as manifestações de 2013 na Praça Tahrir, no Cairo, ou as manifestações populares que se espalharam por todo o Brasil em julho de 2013, sem levar em conta tudo o que foi publicado nessas redes? Ou como abordar o problema do crescimento do neonazismo no mundo sem passar pelos sites e comunidades neonazistas e negacionistas? As redes sociais não explicam todos esses fenômenos, mas certamente são parte fundamental para se entender sua forma e sua natureza dinâmica. (CARVALHO, 2014, p. 173)

É quase impossível, atualmente, se produzir uma história do tempo presente sem levar em consideração as redes sociais. Inclusive, esse é um tema que tem sido estudado pelos historiadores da geração de nativos digitais nascidos nos anos 1990. Recentemente, na UFRGS, dois trabalhos de conclusão de curso foram produzidos tendo as redes sociais como fonte primária: “*Postando o passado: A difusão da memória da ditadura civil-militar brasileira na internet através do canal do YouTube da Comissão Nacional da*

Verdade” de Bruno Grigoletti Laitano, em 2018, e “*O golpe tuitado: uma análise dos discursos produzidos no Twitter pelas principais lideranças do Golpe de 2016 (junho 2013-dezembro 2015)*” de Bruno Erbe Constante, em 2019.

Outra forma dos historiadores utilizarem as redes sociais é como uma forma de debater e divulgar a história. Vejamos dois exemplos. Em 2014, o historiador Ronaldo Vainfas publicou na sua página do Facebook um texto sobre o golpe de 1964, segundo o historiador: “Pensei em fazer um artigo pro O Globo, na página dos editoriais, mas o espaço seria mínimo e o número de leitores mais reduzido do que os do Facebook” (CARVALHO, 2014, p. 175).

Outro exemplo não é de um historiador, mas de um pesquisador da área de relações internacionais. Em 2018, Matias Spektor descobriu documentos que provavam a relação do ditador Ernesto Geisel com a tortura e a política de assassinatos do regime militar e publicou em suas redes sociais (Facebook e Twitter)³⁹. A partir das suas redes, antes de qualquer jornal produzir uma notícia, a informação já havia se espalhado entre os historiadores e pesquisadores (ver Figura 3).

³⁹ SPEKTOR, Matias. “Este é o documento mais perturbador que já li em 20 anos de pesquisa: Recém-empossado, Geisel autoriza a continuação da política de assassinatos do regime, mas exige ao Centro de Informações do Exército a autorização prévia do próprio Palácio do Planalto”. 10, maio. 2018. Twitter: @MatiasSpektor. Disponível em: < <https://twitter.com/matiaspektor/status/994587769220591617>>. Acesso em: 23/09/2020.

Figura 3 - Imagem do tweet de Matias Spektor

Fonte: elaboração pelo autor

É interessante como a questão geracional aparece mesmo na utilização das redes sociais. Enquanto os historiadores imigrantes digitais utilizam principalmente o Facebook e outras mídias digitais mais parecidas com artigos tradicionais para divulgar a história, os historiadores nativos digitais aparentemente costumam utilizar, além do Facebook, o Twitter e Youtube, e o Instagram em menor proporção.

O *Twitter* é uma rede social criada em 2006 que tem como objetivo publicar pequenas postagens de apenas 280 caracteres (o famoso *tweet*)⁴⁰. Devido a essa limitação da rede social, os autores que costumam publicar seus pensamentos e opiniões na rede postam vários *tweets* dividindo em cada *tweet* uma parte do texto. É como se fosse um artigo dividido em várias partes. Esse conjunto de *tweets* que tratam do mesmo assunto é chamado de *thread* (ou fio, em português).

Os historiadores e intelectuais nativos digitais, além do *Twitter*, costumam produzir também vídeos no *Youtube* e podcasts. Há, atualmente, vários debates que se desenvolvem entre *youtubers* pós-estruturalistas, *youtubers* marxistas, etc⁴¹. A grande maioria desses debates é feito no próprio *Youtube*, mas também no *Twitter*. Uma boa parte dos que participam desses debates são de nativos digitais, embora haja também muitos imigrantes digitais.

Quais são os problemas em utilizar essas ferramentas? O famoso escritor português, José Saramago (1922–2010), em uma entrevista para O Globo, em 2009, quando perguntado sobre o *Twitter*, na época uma novidade, respondeu o seguinte:

Entrevistador: O senhor acompanha o fenômeno do *Twitter*, já pensou em abrir uma conta no site?

J.S.—Nem sequer é para mim uma tentação de neófito. Os tais 140 caracteres refletem algo que já conhecíamos: a tendência para o monossílabo como forma de comunicação. De degrau em degrau, vamos descendo até o grunhido. (MIRANDA, 2009)

⁴⁰ Até 2017, eram apenas 140 caracteres.

⁴¹ Em novembro de 2019, quando do lançamento da edição brasileira da revista *Jacobin*, através do *Twitter* dois acadêmicos (Sean Purdy e Alvaro Bianchi) criticaram a publicação pela revista de um artigo do historiador e *youtuber* Jones Manoel da Silva. Purdy e Bianchi haviam criticado a presença de Jones Manoel da Silva na revista devido ao fato de ser, nas palavras de Purdy, um “stalinista orgulhoso”. Essa polêmica gerou todo um debate nas redes sociais e inclusive uma matéria na Folha de São Paulo intitulado “Legado de Stálin volta a inflamar debates na esquerda” sobre a censura e o papel de Stalin e do stalinismo (ver: SAYURI, Juliana. Legado de Stálin volta a inflamar debates na esquerda. Folha de São Paulo, São Paulo, 19.dez.2019. Disponível em: <https://www1.folha.uol.com.br/amp/ilustrissima/2019/12/legado-de-stalin-volta-a-inflamar-debates-na-esquerda.shtml?utm_source=twitter&utm_medium=social&utm_campaign=twfolha&utm_n=true>. Acessado em: 30/12/2019).

Em setembro de 2020, o cantor Caetano Veloso deu uma entrevista ao jornalista Pedro Bial na rede Globo. Nessa entrevista, o cantor fez um comentário elogiando o filósofo marxista Domenico Losurdo e Jones Manoel. Como Losurdo é considerado por alguns movimentos marxistas como um defensor do stalinismo, isso acarretou novamente em debates nas redes sociais sobre Stalin e o stalinismo. (ver: ZANINI, Fábio. Stálin não foi a reencarnação de Lúcifer, diz historiador que influenciou Caetano Veloso. Folha de São Paulo, São Paulo, 17.set.2020. Disponível em: <<https://www1.folha.uol.com.br/poder/2020/09/stalin-nao-foi-a-reencarnacao-de-lucifer-diz-influenciador-de-caetano-veloso.shtml>>. Acessado em: 24/09/2020)

É claro que a visão de Saramago leva em consideração também todo um debate do campo da linguística que não nos interessa aqui, mas essa mudança drástica na forma de se produzir uma redação textual pode acarretar em prejuízos futuros para a divulgação da história? Os historiadores devem se adaptar a essas novas formas (não só ao *Twitter*, mas ao *Youtube*, *Facebook* e *Instagram*) para divulgar a história ou ignora-las? Há, recentemente, um esforço por parte de alguns historiadores de utilizar as redes sociais com o objetivo de divulgação histórica, principalmente no *Youtube*⁴².

Uma rede social mais convencional no que diz respeito a produção de textos é o *Medium*. O *Medium* junta ferramentas do blog tradicional, das origens da Web 2.0, com as redes sociais mais contemporâneas. Assim, um usuário ao mesmo tempo que possui seguidores e pode seguir outros usuários, também pode publicar artigos (denominado de *story*) e produzir um blog (denominado de *publication*). É uma rede social muito utilizado por jornalistas e escritores, principalmente nos Estados Unidos.

Por último, os historiadores podem utilizar páginas na internet e aplicativos para desenvolver novas produções históricas dentro do campo da história digital ou da história virtual⁴³ (LUCCHESI, 2014). Um exemplo interessante, focado no ensino de história, é o trabalho da historiadora Anita Natividade Carneiro (2018). Utilizando o aplicativo *MyMaps* do *Google*, Carneiro desmarcou locais, na cidade de Porto Alegre, que tem relação com a memória e a história da ditadura militar. Assim, sempre que o aluno clicasse em algum ponto demarcado, apareceria uma foto do local e um breve texto explicando qual a relação com o regime militar⁴⁴.

⁴² Provavelmente o caso mais famoso seja o do canal no *Youtube*, *Leitura Obriga HISTÓRIA*. Há também outros casos famosos, como o *Café História*, uma rede social ou blog produzida para historiadores. Ver: Carvalho (2014).

⁴³ “Como a história digital ainda é algo relativamente recente, os historiadores sequer decidiram como classificá-la. Com efeito, ela tem sido considerada tanto uma tendência historiográfica, quanto uma metodologia específica, um subcampo na história, ou mesmo um novo paradigma historiográfico [...] Além disso, existe também certo desacordo entre os historiadores quanto à nomenclatura a ser adotada para essa “novidade historiográfica”, podendo-se encontrar designações como história digital, história virtual ou ainda história hipertextual. No Brasil, embora muitos historiadores empreguem o termo história digital, outros preferem história virtual” (ZEIDAN ARAÚJO, 2014, p. 155-156).

⁴⁴ Ver: Caminhos da Ditadura em Porto Alegre. Disponível em: < <https://caminhosdaditaduraemportoalegre.wordpress.com/> >. Acessado em: 04 de maio de 2020.

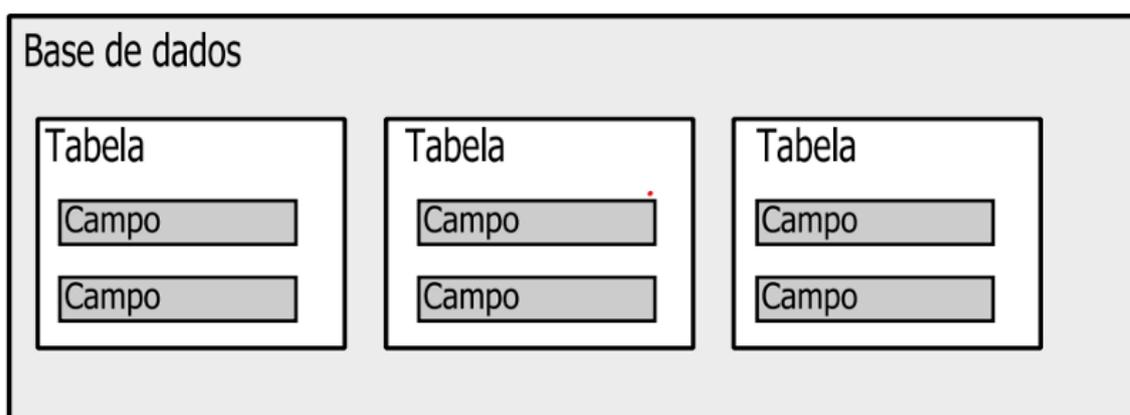
2.4 Banco de dados: uma forma de produção histórica

Quando falamos de banco de dados informatizados na pesquisa histórica⁴⁵, devemos levar em consideração o fato de que há dois sentidos: um sentido do ponto de vista da informática e outro do ponto de vista da ciência histórica.

No primeiro sentido, isto é, da informática, o desenvolvimento do banco de dados pressupõe um *hardware* (quase sempre, o computador ou *notebook*) e um *software* específico e destinado a essa função (o Microsoft Access do Microsoft Office, por exemplo).

Um banco de dados ou base de dados (*to database something*, em inglês) é um conjunto de tabelas articuladas entre si. Cada tabela possui determinados campos e registros, também denominados, respectivamente, de "colunas" e "linhas" (GIL, 2015). O conceito é bem amplo e há várias formas de produzir um banco de dados (ver Figura 4). Por exemplo, um banco de dados pode muito bem ter apenas uma tabela, assim como pode ter mais de cinquenta. Uma tabela pode ter apenas dois campos, enquanto outra tabela tem vários campos. As tabelas de determinado banco de dados podem ser ligadas por um determinado campo, o que se costuma denominar de campo de ligação, mas também as tabelas que compõe um banco de dados podem ser independentes, assim como podem ter ligações diversa.

Figura 4 - Exemplo de banco de dados



⁴⁵ O uso do banco de dados entre os historiadores possui uma história muito anterior a informática, em que se predomina o uso de cartões perfurados, como no caso do Padre Busa na década de 1940. Para aprofundar acerca desse assunto, ver Gil (2014). No caso da historiografia brasileira, muitos historiadores, como Paulo Afonso Zarth, já utilizavam o banco de dados antes da generalização da informática, sobretudo nas pesquisas sobre a escravidão.

Fonte: Gil, 2015.

Cada tabela costuma ter um objetivo próprio. Por exemplo, caso o banco de dados seja voltado a organização de uma biblioteca pessoal, cada tabela poderia representar um determinado assunto ou gênero (ver Tabela 4). Vamos supor que o autor do banco de dados escolha criar apenas duas tabelas, uma denominada de “livros de ficção” e a outra “livros de não-ficção”. Em ambas as tabelas, o autor cria dois campos iguais para cada: “título do livro” e “nome do autor”, e o campo de ligação seria o campo “nome do autor”. Os registros seriam os livros registrados em ambas as tabelas pelo autor. Vamos supor que o usuário que criou esse banco de dados quisesse descobrir todos os livros de um autor “x” em sua biblioteca a partir dos registros do banco de dados. Bastaria pesquisar pelo nome do autor “x” e logo acharia, independente da tabela. Nesse caso, teria produzido um banco de dados, com duas tabelas, quatro campos e um campo de ligação, ligando ambas as tabelas.

Tabela 4 - Exemplo de banco de dados de uma biblioteca pessoal

Banco de Dados: Biblioteca Pessoal			
Tabela 1: Livros de ficção		Tabela 2: Livros de não-ficção	
Nome do Autor <i>(campo de ligação)</i>	Título do Livro	Nome do Autor <i>(campo de ligação)</i>	Título do Livro

Fonte: elaboração pelo autor

Quando falamos de banco de dados relacionais há dois conceitos de suma importância: o de relacionamento e da cardinalidade. Um relacionamento é a ligação entre duas tabelas do mesmo banco de dados e a cardinalidade diz respeito aos graus de relações estabelecidas entre as tabelas. Há três tipos diferentes de cardinalidade (ver Tabela 5): N:N (vários-para-vários), 1:N (um-para-vários) e 1:1 (um-para-um). Vejamos um exemplo. Vamos supor que uma escola produziu um banco de dados com três tabelas: professores, alunos e turmas. A relação N:N pode ser exemplificada pela relação entre as tabelas “professores” e “alunos”, pois um professor pode dar aulas para vários alunos, e um aluno pode ter aula com vários professores. A relação 1:N pode ser entre a tabela “alunos” e “turmas”, pois um aluno pode estar somente em uma turma, mas uma turma tem vários alunos. Por último, 1:1 entre as tabelas “professores” e “turmas”, pois um

professor pode somente ter uma turma, e uma turma pode ter somente um professor (GIL, 2015).

Tabela 5 - Exemplo de cardinalidade

N=N	Várias tabelas podem se relacionar
1=N	Uma tabela pode se relacionar com várias
1=1	Uma tabela só pode se relacionar com somente uma

Fonte: elaboração pelo autor com base em Gil, 2015.

O banco de dados informatizado pode ser utilizado de milhares de formas diferentes, desde produzir uma biblioteca pessoal até registrar contatos de clientes e fornecedores de uma empresa. A partir dessa constatação podemos perceber a função essencial do banco de dados: sistematizar dados, isto é, organizar informações em um sistema que vão ao encontro dos interesses do usuário.

Quando utilizamos o banco de dados informatizado na pesquisa histórica, a sua essência enquanto uma ferramenta de organização não muda, mas ganha novas características. O banco de dados foi utilizado pelos historiadores, mesmo antes da informática, principalmente com o objetivo de sistematizar as fontes, muitas vezes de caráter quantitativo, daí sua profunda relação com a história serial-quantitativa. Segundo Luciano Figueiredo,

As bases de dados constituem uma das mais poderosas ferramentas para a história. Elas por princípio nada mais fazem do que organizar informações que são constantes e estão presentes em séries documentais (registros paroquiais, balanços de comércio, listas eleitorais, registros de impostos, legislação administrativa são algumas delas) (FIGUEIREDO, 1997, p. 597)

Mas, quando tratamos da relação entre o banco de dados e a ciência histórica não podemos somente pensar o banco de dados como só mais uma ferramenta qualquer, como se fosse um marcador de página. O banco de dados no fazer historiográfico é uma forma de produção histórica. Um autor que trabalha o banco de dados dessa perspectiva é Tiago Luís Gil. Segundo Gil e Barleta,

Desde há muito, os historiadores usam textos para contar histórias. Com o texto pode-se fazer muitas coisas: podemos construir narrativas vivas, acelerando ou diminuindo ritmos. Podemos descrever minuciosamente ou fazer afirmações curtas e retumbantes. Mas os textos não são a única forma de linguagem. Os geógrafos expressam muitas ideias na forma de mapas. Os sociólogos, na forma de redes. Os economistas, na forma de gráficos. E os artistas usam as mais diversas formas. Esta liberdade não costuma servir aos historiadores, quase sempre presos ao texto escrito. (GIL; BARLETA; 2015, p. 429)

Em outro trabalho, Tiago Luís Gil complementa:

Um banco de dados é quase uma forma de narrativa histórica. Ele obedece, perfeita ou imperfeitamente, aos preceitos e às concepções de mundo (e, dentro desses, das opiniões sobre o problema de pesquisa) do pesquisador. A primeira posição teórica é acreditar que seja possível reduzir a complexidade do social a ponto de fazê-la caber na forma de registros de uma tabela, tal como os historiadores acreditam ser possível fazer nas linhas de um texto. [...] No caso da pesquisa em História, isso inclui a escolha de objetos, de conjuntos de fontes, de metodologias e de interlocutores. (GIL, 2015, p. 11)

Existe na ciência histórica um verdadeiro “monopólio do texto”, que não é, sem sombra de dúvidas, sem razão e inútil. Entretanto, há outras formas de produzir história para além da produção textual. É nesse sentido que o banco de dados surge como uma forma de produção histórica, e não simplesmente como uma ferramenta qualquer que o historiador se utiliza para produzir a “verdadeira” forma de história que seria feita através da redação textual. Afinal, assim como na produção historiográfica de textos, o historiador, no processo de construção do banco de dados, retira as fontes de seu contexto original e procura questioná-las de acordo com o seu problema de pesquisa e com os seus objetivos.

Há, de modo geral, dois tipos diferentes de banco de dados em história: as bases de dados orientadas pela fonte ou *source-oriented* e as bases de dados orientadas pelo método ou *method-centred*⁴⁶. Enquanto a primeira procura produzir o banco de dados a partir das fontes, a segunda parte de um problema de pesquisa ou de uma metodologia

⁴⁶ É necessário enfatizar que embora divisões sejam importantes do ponto de vista teórico-metodológico, não existe um tipo de banco de dados “puro”, isto é, os bancos de dados *source-oriented* também tem algumas características do *method-centred*, e vice-versa. Nesse sentido, a grande divisão entre os dois tipos de bancos de dados reside na sua essência: enquanto um tipo é baseado na fonte, o outro é centrado na metodologia ou problema de pesquisa.

específica. A primeira procura uma transcrição estruturada da documentação, enquanto a segunda procura através do questionamento da documentação produzir o banco de dados (GIL, 2015).

Nos bancos de dados do tipo *source-oriented*, o historiador terá campos e tabelas específicos, já que sistematizará uma documentação de mesma origem. Exemplos de banco de dados orientadas pelas fontes são registros de batismos e correspondências. Por exemplo, quando vamos sistematizar um conjunto de correspondências, há determinados campos que necessariamente teremos que criar, como: remetente, destinatário, data de quando a carta foi enviada, etc.

Já no banco de dados do tipo *method-centred*, as tabelas e o campos criados pelo historiador irão ser desenvolvidos a partir do seu problema de pesquisa, da metodologia e dos seus objetivos e as fontes a serem questionadas podem ser diversas. Daí que as operações historiográficas principais desse tipo de banco de dados sejam a prosopografia, análise de redes sociais e a análise demográfica.

Por exemplo, vamos supor que o historiador pretende produzir uma base de dados voltada a sistematizar os jogadores de futebol que passaram por um clube de futebol existente em Porto Alegre nos anos 1930. As fontes a serem questionadas provavelmente serão diversas: documentos oficiais do clube e da entidade local responsável, matérias jornalísticas, depoimentos de dirigentes, torcedores e dos próprios jogadores (história oral), etc. Não há como as tabelas e os campos criados pelo historiador serem baseados em uma única fonte (por exemplo, os documentos oficiais do clube), já que ao questionar as demais fontes, muito provavelmente outras questões importantes não existentes em somente uma fonte aparecerão. Por tudo isso, é necessário que, após conhecer com profundidade a documentação, o historiador desenvolva as tabelas e os campos de acordo com seu problema de pesquisa e com as fontes que pretende questionar.

Um exemplo muito importante de banco de dados baseado no método é o *Slave Voyages*⁴⁷ (*The Trans-Atlantic Slave Trade Database*) produzido por pesquisadores da

⁴⁷ Ver: Slave Voyages. Disponível em: <<https://www.slavevoyages.org/>>. Acessado em: 04 de maio de 2020.

Emory University. O objetivo principal desse banco de dados era sistematizar as viagens do comércio transatlântico de escravos para as Américas. Segundo Tiago Luís Gil,

Trata-se de um banco de dados com mais de 35.000 viagens de navios negreiros na época moderna. Um projeto que não pretende a precisão exata do número de viagens e de seres humanos transportados. Ele pretende estimar o volume total ou, para empregar o conceito correto, a grandeza daquela prática, que não tinha nenhuma grandeza, mas contava com dezenas de milhares de viagens com mais de uma dezena de milhão de almas conduzidas para o trabalho escravo. A exatidão aqui não é o objetivo nem poderia ser. A crítica documental foi pouco considerada e há motivos para isso. Se fosse bem feita, como quem faz uma tese usando uma única fonte, o projeto jamais seria concluído. E uma crítica bem feita pode mudar o número exato de navios e escravos que cruzaram o Oceano Atlântico, mas não a grandeza. (GIL, 2015, 98-99)

Figura 5 - Banco de dados Slave Voyages

The screenshot shows a web interface for a database titled 'Comércio Transatlântico de Escravos - Base de Dados'. It features a navigation bar with various filters and a main table displaying voyage records. The table has columns for voyage identity, ship name, departure port, acquisition location, disembarkation location, arrival year, number of slaves, and captain's name.

Identidade da viagem	Nome da embarcação	Porto onde a viagem começou*	Principal lugar de aquisição de escravos*	Principal local de desembarque de escravos	Ano chegou com escravos	Escravos chegaram ao 1º porto	Nome do capitão
1	Pastora de Lima	Rio de Janeiro	Mozambique	Bahia, port unspecified	1817	290	Dias, Manoel José
2	Tibério	Bahia, port unspecified	Mozambique	Bahia, port unspecified	1817	223	Mata, José Maria da
3	Paquete Real	Bahia, port unspecified	Cabinda	Bahia, port unspecified	1817	350	Ferreira, José dos Santos
4	Bom Caminho	Bahia, port unspecified	Quilimane	Bahia, port unspecified	1817	342	Dias, Domingos Francisco
5	Benigretta	Bahia, port unspecified	Cabinda	Bahia, port unspecified	1817	516	
6	Voador	Bahia, port unspecified	Mozambique	Bahia, port unspecified	1817	515	
7	Formiga	Bahia, port unspecified	Malembo	Bahia, port unspecified	1817	204	Viana, Isidoro António
8	Vigilante Africano	Pernambuco, port unspecified	Luanda	Bahia, port unspecified	1817	374	Amorim, José Gomes de

Fonte: *Slave Voyages*

Um grande problema fruto da produção do banco de dados é aquilo que Figueiredo (1997) denominou de “síndrome da incompatibilidade”, ou seja, depois da produção do banco de dados pelo historiador e da sua correspondente utilização na pesquisa, o banco de dados perde, digamos assim, “razão de ser”. Em outras palavras, ele serve apenas para aquela pesquisa, sendo incompatível (por motivos metodológicos, historiográficos ou técnicos) com outras pesquisas na área. Em relação a reutilização de banco de dados, Tiago Luís Gil (2015) enxerga dois problemas fundamentais: um de caráter técnico e outro de caráter teórico-metodológico.

O primeiro diz respeito as constantes mudanças de sistemas, formato de arquivos e programas. Em relação a esse problema, o historiador deve ter uma preocupação em preservar os arquivos e estar sempre atualizado em relação as novidades. De qualquer modo, atualmente há várias formas na internet de converter um arquivo de um formato para o outro, embora essas conversões, por serem realizadas de modo automático, podem acarretar em alguns erros nos arquivos⁴⁸. O segundo problema é mais complexo. Sabemos que a historiografia constantemente amplia ou simplesmente muda as tendências historiográficas, os métodos, temas e paradigmas. Nesse caso, é necessário ter em mente o limite do banco de dados, assim como de um trabalho textual. É natural que todos os trabalhos, em uma hora ou outra, sejam “datados”, e o banco de dados talvez também não possa escapar disso.

2.5 História serial-quantitativa e o banco de dados

Criada pelos historiadores franceses Marc Bloch e Lucien Febvre, a Escola dos Annales representou uma nova forma de se produzir história. Foi, nas palavras de Peter Burke (1991), uma “Revolução Francesa da historiografia”. Os historiadores não buscavam mais uma história dos grandes homens e dos grandes fatos, mas sim uma história mais ampla, que buscasse explicar as estruturas econômicas, sociais, políticas, psicológicas, mentais, etc. Se a história metódica, dita positivista, predominante no século XIX, era uma história narrativa e política, a história dos Annales do século XX é uma história explicativa ou uma história-problema, tratando a investigação histórica através de problemas de pesquisa e explicações.

Há um consenso na historiografia de que haveriam três gerações principais dos Annales. A primeira geração (1929–1949) seria dominada pelas obras e inovações de seus fundadores, Marc Bloch e Lucien Febvre. A segunda geração (1946–1968) seria marcada pela predominância do historiador Fernand Braudel, que traria inovações interessantes ao Annales, como uma nova temporalidade (teoria das três durações), uma maior relação entre história e geografia (o fazer historiográfico a partir do Mediterrâneo) e os seus estudos sobre o capitalismo já no final da vida. É na segunda geração que novas

⁴⁸ A conversão de arquivos do Microsoft Access para o Microsoft Excel e vice-versa se realiza de modo mais simples, na medida que são produtos da mesma empresa. Há, inclusive, a opção importar e exportar os arquivos em ambos os programas.

tendências surgem, como a história serial-quantitativa, história demográfica, uma relação mais profunda entre o movimento dos Annales e os marxistas, etc.

A terceira geração (1968–1989) não teria uma figura principal, mas sim várias figuras como Jacques Le Goff, Georges Duby e Michèle Perrot. A história das mentalidades e a história cultural seriam a força motriz dessa geração, mas outras tendências também aparecem ou se desenvolvem como a história serial-quantitativa, a história das mulheres, história da família, etc. A história serial-quantitativa que se desenvolveu na segunda geração dos Annales mais para discutir com a história econômica, história social e a história demográfica, na terceira geração iria dialogar com a emergente história das mentalidades e a história cultural (BURKE, 1997; BARROS, 2012).

Como consequência dessa mudança radical no fazer historiográfico, a forma como se analisava as fontes e os documentos também mudou. Segundo Febvre,

A história faz-se com documentos escritos, sem dúvida. Quando estes existem. Mas pode fazer-se, deve fazer-se sem documentos escritos, quando não existem. Com tudo o que a habilidade do historiador lhe permite utilizar para fabricar o seu mel, na falta das flores habituais. Logo, com palavras. Signos. Paisagens e telhas. Com as formas do campo e das ervas daninhas. Com os eclipses da lua e a atrelagem dos cavalos de tiro. Com os exames de pedras feitos pelos geólogos e com as análises de metais feitas pelos químicos. Numa palavra, com tudo o que, pertencendo ao homem, depende do homem, serve o homem, exprime o homem, demonstra a presença, a atividade, os gostos e as maneiras de ser do homem. (FEBVRE apud LE GOFF, 1992, p. 540)

A partir dessa ampliação da concepção de documento, como expressão da atividade do homem, produzida pela Escola dos Annales, surgiu o conceito de série. Segundo José D'Assunção Barros,

Trata-se de considerar os documentos ou as fontes históricas não mais em sua perspectiva singular, como documento único, a ser analisado nos seus próprios limites, mas sim como partes constituintes de uma grande cadeia de fontes de mesmo tipo. A série é necessariamente formada por fontes homogêneas, comparáveis, capazes de serem apreendidas no interior de uma continuidade (seja uma continuidade temporalizada ou especializada), e que permitam uma 'abordagem de conjunto' através de aportes metodológicos como a quantificação, a análise tópica, a identificação de recorrências ou mudanças de padrão no decorrer da série. O objetivo da análise de séries, neste sentido, é

perceber tanto as permanências como as oscilações e variações. (BARROS, 2012, p. 205)

O conceito de série, introduzido pela história serial, seria muito utilizado pela segunda e terceira geração dos Annales, primeiro se relacionando com a história econômica, a história social e a história demográfica e, na terceira geração, com a história das mentalidades. Principalmente na segunda geração, a história serial seria utilizada junto com a história quantitativa⁴⁹, formando, assim, a história serial-quantitativa⁵⁰ (BARROS, 2012; BURKE, 1997).

A diferença entre a história serial e a história quantitativa é tênue⁵¹. Enquanto a história serial pressupõe a série, isto é, uma documentação que possui certa homogeneidade e que pode ser comparada, a história quantitativa tem necessariamente relação com a noção de “número” e “quantidade” (BARROS, 2011; BARROS, 2012). Nesse sentido, se pode fazer uma história serial que não seja quantitativa, mas não se pode fazer uma história quantitativa que não seja serial, visto que a “[...] quantificação pressupõe a serialização - se não de fontes, pelo menos de dados. O inverso é que não ocorre. Posso trabalhar com séries de fontes sem estar necessariamente interessado no número. Estarei interessado em verificar recorrências, mas não necessariamente quantidades”⁵² (BARROS, 2011, p. 166).

⁴⁹ A história quantitativa já era utilizada pela história dos preços desde o século XIX (BARROS, 2012).

⁵⁰ Há correntes historiográficas que discordam do uso do termo quantitativo, visto que a pesquisa quantitativa seria uma abordagem das ciências sociais e não da ciência histórica. Nesse sentido, o termo quantitativo seria redundante. Nesse trabalho, optou-se por utilizar o conceito de história serial-quantitativo de Barros (2011; 2012), mas sempre levando em consideração as críticas feitas a esse conceito. Ver: Farinatti (2008).

⁵¹ “A História Serial propriamente dita, conforme já foi discutido, refere-se ao uso de determinado tipo de fontes (homogêneas, do mesmo tipo, referentes a um período coerente com o problema a ser examinado), que permitam uma forma específica de tratamento historiográfico (a serialização de dados, a identificação de elementos ou ocorrências comuns que permitam a identificação de um padrão e, em contrapartida, uma atenção às diferenças, às vezes graduais, para se medir variações). Já com relação à História Quantitativa, esta deve ser definida através de outro critério, que é o seu campo de observação. O que a História Quantitativa pretende observar da realidade está atravessado pela noção de “número”, “quantidade”, valores a serem medidos. As técnicas utilizadas pela abordagem quantitativa serão estatísticas, ou baseadas na síntese de dados através de gráficos diversos e curvas de variação a serem observadas de acordo com eixos de abscissas e coordenadas.” (BARROS, 2012, p. 207)

⁵² “A “série” é o que canaliza a atenção do historiador na modalidade da História Serial; o “número” ou a medida é o que canaliza a atenção do historiador no caso da História Quantitativa” (BARROS, 2011, p. 166).

O banco de dados na pesquisa histórica foi utilizado principalmente se relacionando com a história serial-quantitativa. Entretanto o banco de dados é uma ferramenta que apresenta múltiplas possibilidades e não necessariamente tem relação com a pesquisa quantitativa. Embora cumpra muito bem funções quantitativas, a característica principal do banco de dados é sistematizar dados, isto é, organizar informações em um sistema. Nesse sentido, para além da análise documental, pode ser utilizado no debate historiográfico ou simplesmente como uma ferramenta de organização. Segundo Gil,

as bases de dados podem ser usadas em qualquer pesquisa, e não, apenas, em estudos quantitativos ou seriais. É certo que estes últimos dependem de bases para ser realizados, mas a recíproca não é verdadeira. Podemos usar bases para qualquer tipo de estudo. É claro que nem todos os estudos ou abordagens precisam desse tipo de ferramenta, mas o simples fato de organizar as informações, mesmo notas de pesquisa, torna as bases de dados úteis para qualquer situação. (GIL, 2015, p. 100)

No capítulo seguinte, discutiremos a nossa experiência na produção de um banco de dados voltado a sistematizar operários imigrantes na região do então município de São Jerônimo, na zona carbonífera do Rio Grande do Sul, durante a década de 1940.

3 O PROCESSO DE CONSTRUÇÃO DO BANCO DE DADOS

O objetivo desse capítulo é discutir uma experiência de construção e utilização do banco de dados na pesquisa histórica. Esse trabalho tem relação com o projeto de pesquisa “*Solidariedade e conflito: experiências e identidades entre mineiros de carvão no Rio Grande do Sul (1850-1950)*” da Dra. Clarice Gontarski Speranza, do qual fui bolsista de iniciação científica. O objetivo principal do projeto de pesquisa era estudar os grupos sociais e étnicos que participaram do processo de formação da classe trabalhadora mineira no Rio Grande do Sul de 1850 até 1950. A região carbonífera do Rio Grande do Sul localizava-se no Baixo Jacuí, onde atualmente estão localizadas as atuais cidades de Arroio dos Ratos, São Jerônimo, Butiá, Charqueadas e Minas do Leão⁵³.

Nessa região, uma imigração significativa de europeus veio formar, junto com a população brasileira já residente na região, a classe trabalhadora mineira no século XX. Nas primeiras décadas do século passado diversas nacionalidades vieram compor o quadro social da região, com uma predominância de espanhóis vindos de localidades próximas das Astúrias (no norte da Espanha), que possuía uma tradição mineradora desde o início do século XIX. Após a Segunda Guerra Mundial, a imigração veio principalmente de países do Leste Europeu, que haviam sido mais atingidos durante a guerra (SPERANZA, 2018).

O banco de dados foi construído durante o período de iniciação científica (agosto de 2018 até agosto de 2019) a partir de documentos oficiais principalmente do Cadem⁵⁴, consórcio que dominava o ramo da mineração no Rio Grande do Sul. O principal objetivo do banco de dados foi sistematizar a documentação buscando reconstituir a trajetória de

⁵³ “No Rio Grande do Sul, a Região Carbonífera do Baixo Jacuí compreende os municípios de Arroio dos Ratos, Barão do Triunfo, Butiá, Charqueadas, Eldorado do Sul, General Câmara, Minas do Leão, São Jerônimo e Triunfo. Esta área está compreendida na região centro-sul, com cinco municípios incluídos na região metropolitana da capital Porto Alegre: Arroio dos Ratos, Charqueadas, Eldorado do Sul, São Jerônimo e Triunfo. Neste território, encontra-se ainda uma subdivisão: a chamada região carbonífera tradicional, formada pelos municípios onde o carvão foi descoberto e que deram origem à exploração do minério em escala industrial: Arroios dos Ratos, Butiá, Minas do Leão e São Jerônimo. Todos tiveram sua economia e sociedade forjadas a partir da extração do carvão mineral” (WITKOWSKI; MORIGI, 2019, p. 192)

⁵⁴ “Consórcio Administrador de Empresas de Mineração, união das duas principais mineradoras (Companhia Estrada de Ferro e Minas São Jerônimo, de Arroio dos Ratos, e Companhia Carbonífera Riograndense, mais tarde Companhia Carbonífera Minas do Butiá, de Butiá). Surgiu em 1936 e virtualmente monopolizou a produção até o final dos anos 1950. Em 1964, foi transformado na empresa Copelmi, ainda hoje atuante.” (SPERANZA, 2018, p. 13)

operários imigrantes do Leste Europeu (poloneses, russos, ucranianos, etc.) que vieram para a zona carbonífera riograndense após a Segunda Guerra Mundial.

O processo de construção do banco de dados passou por quatro etapas principais⁵⁵: a) análise preliminar das fontes, em agosto de 2018; b) fichamento da documentação, entre agosto e setembro de 2018; c) desenvolvimento inicial do banco de dados, entre outubro de 2018 e janeiro de 2019; d) construção do banco de dados, de janeiro até abril de 2019.

Tabela 6 - Período do processo de construção do banco de dados

Atividade	Época
Análise preliminar das fontes	Agosto de 2018
Versão final do fichamento	Setembro de 2018
Construção Inicial do Banco de Dados (Modelo lógico e Modelo físico)	Outubro de 2018 até Janeiro de 2019
Construção do Banco de Dados (processo de sistematização das fontes)	Janeiro de 2019 até Abril de 2019
Versão Inicial do Banco de Dados	Abril de 2019

Fonte: elaboração do autor

A construção do banco de dados ocorreu entre agosto de 2018 e abril de 2019. Foi a primeira experiência profissional do autor com a pesquisa documental e também com a utilização do banco de dados. Na etapa final da construção do banco de dados, em março de 2019, se iniciou uma pesquisa mais teórica sobre o banco de dados que resultou na apresentação, em outubro de 2019, no Salão de Iniciação Científica (SIC) da UFRGS, de um trabalho intitulado “*A utilização do banco de dados na pesquisa histórica: uma experiência*”⁵⁶. Desde então venho pesquisando mais a fundo aspectos teóricos da construção e desenvolvimento do banco de dados na pesquisa histórica.

⁵⁵ Essas etapas foram estabelecidas para fins didáticos e organizativos para a apresentação da experiência de construção do banco de dados. O processo de construção do banco de dados não necessariamente precisa passar por etapas específicas, mas vai se constituindo a partir da constante relação entre a documentação e o banco de dados.

⁵⁶ SICHELERO, Guilherme Giotti. **A utilização do banco de dados na pesquisa histórica: uma experiência**. Salão de Iniciação Científica (31.: 2019 out. 21-25: UFRGS, Porto Alegre, RS).

Esse capítulo está dividido da seguinte maneira: 1) “Primeira etapa: análise preliminar das fontes”, que busca apresentar um breve panorama da documentação que seria sistematizada; 2) “Segunda Etapa: fichamento da documentação”, pretende apresentar o processo de fichamento; 3) “Terceira etapa: desenvolvimento inicial do banco de dados”, busca discutir com o autor chegou a opção pelo banco de dados e o seu desenvolvimento inicial; 4) “Quarta etapa: construção do banco de dados”, busca discutir alguns problemas e facilidades descobertas pelo autor no processo final de construção do banco de dados; 5) “Banco de dados concluído: versão inicial e final”, pretende apresentar o cada tabela do banco de dados especificadamente, discutindo sobre a sua escolha e elaboração; 6) “Equívocos e correções”, discutirá alguns equívocos do banco de dados e as supostas correções que uma segunda versão teria; 7) “Utilidade do banco de dados: dados quantitativos e hipótese”, buscará discutir as duas principais utilidades do banco de dados produzido na pesquisa.

3.1 Primeira etapa: análise preliminar das fontes

Em agosto de 2018 iniciou o período de iniciação científica e uma das primeiras atividades solicitadas era a de produzir um fichamento sobre uma determinada documentação. Tratava-se de uma seleção de documentos sobre a imigração nas minas de carvão que havia sido digitalizada e estava em cinco arquivos de PDF, tendo sido repassada à pesquisa pelo mestrando estadunidense Cody Williams⁵⁷. Portanto, de acordo com Almeida (2011), podemos classificá-los como documentos primários digitalizados. A documentação física está localizada no Museu Estadual do Carvão⁵⁸, na cidade de Arroio dos Ratos (Rio Grande do Sul).

A grande maioria da documentação era formada por documentos oficiais, gerada pelo Cadem ou por órgãos governamentais, principalmente do Departamento Nacional de Imigração. Há também correspondências, principalmente oficiais e trocadas entre o

⁵⁷ Williams defendeu dissertação sobre os trabalhadores acidentados na mineração na Universidade de Oklahoma em 2018. O trabalho, intitulado “*“To sacrifice the vital rhythms of their being”: impairment, resistance, and industrial crisis in São Jerônimo, Rio Grande do Sul, 1944-1964*”, foi desenvolvido com período sanduíche no Rio Grande do Sul.

⁵⁸ O Museu Estadual do Carvão, em Arroio dos Ratos, está localizado no antigo complexo industrial do Poço 1. Uma equipe organizou o acervo documental da mineração entre novembro 2014 e agosto de 2018. O acervo é composto por documentos das antigas empresas mineradoras da região do final do século XIX até a década de 1990 (WITKOWSKI; MORIGI, 2019).

Cadem e a polícia e o governo. Os próprios operários produziram algumas correspondências, como é caso da carta das esposas dos operários a Getúlio Vargas, uma carta dos operários imigrantes do Leste Europeu agradecendo ao Cadem e uma carta de um operário alemão residente na Alemanha procurando emprego no Brasil. A maior parte dos documentos eram relatórios sobre os operários imigrantes, tanto relatórios de imigração, quanto relatórios sobre a atividade dos trabalhadores no local de trabalho.

Vejamos, resumidamente, os aspectos principais de cada arquivo.

O arquivo “Caixa 3, Cartas do Sindicato, 08.01.1”, possuía sete páginas e haviam somente dois documentos. O primeiro documento era uma carta escrita pelas esposas dos operários para Getúlio Vargas requerendo melhorias na região e nas condições de vida após a greve dos trabalhadores em 1946. O segundo documento era um relatório do interventor nas minas e capitão do exército⁵⁹, Ebenezer Cabral de Mello, que havia ficado responsável pela região após as greves de 1946, ao Ministro do Trabalho, Indústria e Comércio elaborado após a visita do militar a região de São Jerônimo, visando analisar as condições dos operários. Era uma resposta à carta das esposas dos operários.

O arquivo “Sílvia Greve de 1946 e CADEM 1946-1948” tinha seis páginas e dois documentos. O primeiro era a solicitação da prisão de alguns grevistas que participaram greve de 1946 por parte da direção do Cadem ao então Presidente da República, Eurico Gaspar Dutra. O segundo era uma lista de supostos comunistas que trabalhavam nas minas de Butiá desenvolvida pelo próprio Cadem (ver Figura 6).

⁵⁹ Ver: SPERANZA, Clarice Gontarski. Cavando direitos: as leis trabalhistas e os conflitos entre trabalhadores e patrões nas minas do Rio Grande do Sul nos anos 40 e 50. Porto Alegre: UFRGS, 2012. 272 f. Tese (Doutorado em História) - Programa de Pós-graduação em História, Instituto de Filosofia e Ciências Humanas, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2012. p. 179.

Figura 6 - Documento do da lista de comunistas

RELAÇÃO DOS COMUNISTAS DAS MINAS DE BUIÁ

Nome	Local de trabalho
João Gomes de Oliveira	Oficina
Gregório Ripoll	"
Antônio Ortiz Ruiz	"
Adão Teixeira de Oliveira	Usina
Edmundo Schornes	"
Alcides Gomes da	Fogo 2
Stanislau Zavalick	"
Tomaz Pakula	"
Afonso Piazza	"
Armando Flores da Silva	"
Adolfo Manoel Rufino	"
Laurentino Luis da Cunha	"
Antônio R. da Silva	"
Silvano Nascimento Ambrosio	"
Alexandre Javgraki PA	"
Horival Bolina	"
Lotario P. da Silva	"
Henrique João José da Silva	"
João Rodrigues	"
Adão Lessa	"
Francisco Souto	"
Oscar Abreu	"
Pedro Pereira	Fogo 3
Alcino Batista Oliveira	"
Antônio Manoel da Silva	"
Angélico Peixoto da Silveira	"
Vicente Rodrigues Stauberatier	"
Gabriel Quadros	"
Laudelino Pereira Nunes	"
Amaro Cunha	"
Margarito Antônio Duarte	"
Honorio Figueira	"
Claudio Perez Cifuentes	"
Umbelino Caetano	"
Ubirajal Domingos da Silva	"
Capitão Alves	"
Artur Lúcia	Fogo R-1
Antônio Batista Dorneles	Sem ocupação
Waldo José Figueira da Silva	"

Fonte: arquivo Museu do Carvão

O arquivo "Silvia Box (1943-1945) GC AHM-MEC" com seis páginas e seis documentos. São correspondências entre a direção do Cadem e a polícia pedindo o retorno ao trabalho nas minas de operários alemães afastados durante a Segunda Guerra Mundial e relatando sobre alguns tumultos da greve de 1946.

O arquivo "Silvia Box CADEM 1946-1948 (GC) AHM-MEC" com quatorze páginas e treze documentos. Esse arquivo trata, praticamente, sobre a imigração de operários vindos do Leste Europeu. Há um relatório, produzido pelo consórcio, que busca dividir os operários imigrantes em três categorias principais: "bons operários", "maus

operários” e “operários regulares” (ver Figura 7). Além disso, há alguns documentos muito interessantes sobre um operário chamado José Varga, de origem húngara. Varga fora preso em 1943 acusado de “ser súdito do eixo”. Sua prisão provavelmente foi motivada pelo fato de participar ativamente de assembleias dos trabalhadores e ter certa influência entre os operários. Há algumas correspondências da diretoria do Cadem o caracterizando como “anarquista” e perigoso.

Figura 7 - Documento dividindo os operários imigrantes

**RELAÇÃO DOS OPERÁRIOS IMIGRANTES QUE ESTÃO TRABALHANDO NA
COMPANHIA CARBONÍFERA MINAS DE BUTIÁ**

N o m e	Nacionalidade
BONS OPERÁRIOS:	
ANTONOV Fedor	Russa
CURAK Wasył	Ucraniana
MYCHAILIO Pablo	"
MIAS Edward	Polonesa
SAWITSCH Nikola	Russa
PAWLUK Kiril	Ucraniana
SLOBODANIUK Iwan	"
MILLYAN Jan	"
SZCZEPANSKI Josef	Polonesa
CHARCZENKO Jakob	Ucraniana
SAWOSZ Kyzma	"
GLIŹ Simão	Polonesa
OCZYCZENKOM Paulo	"
GOBRATCHEVSKI Constantino	"
TYLAK Floriano	"
OPERÁRIOS REGULARES:	
SUCHECKY Steben	Polonesa
CHULZ Bronislawe	"
KOMSTAS Ch Timoteo	Ucraniana
ROKICKI Alexander	Polonesa
LYSYK Woldsimiers	"
SCHMANTZ Nicolai	Ucraniana
SNIEDZE Erika	Lituausa
ZELTANSVIS Janis	"
KUSNECOFF Leonid	Russa
SKLAR Mikola	Ucraniana
MÁUS OPERÁRIOS:	
HAZAROW Michail	Russa
DUDKO Nicolas	Ucraniana
MIRONENKO Ivan	Russa
SOKOLOWSKIY Nikolai	"
MASANKO Ivan	"

Fonte: arquivo Museu do Carvão

O "Dossiê dos Imigrantes (Europeus), CADEM Caixa 296, AHM" reunia com sessenta e duas páginas e mais de 30 documentos digitalizados. Nesse arquivo,

predominavam os documentos contendo os relatórios de imigrantes (principalmente vindos do Leste Europeu). Esses documentos traziam alguns dados como nome, idade, nacionalidade, familiares, estado civil, profissão inicial, bagagem, etc (ver Figura 8). Havia, além disso, alguns relatórios do Cadem sobre as funções profissionais dos trabalhadores na empresa, frequência no trabalho, os imigrantes que saíram da empresa, salários, etc.

Figura 8 - Relatório da vinda dos operários imigrantes

1 - ANTONOV Fedor	40 anos	C	russo	tecnico de minas
" Antonina	29 "	C	"	domestica
" Vasili	18 "	S	"	estudante
" Zenadi	3 "	S	"	"
2 - MEZERYA Stefan	39 "	C	Ucraniano	operario
" Ludmyla	9 "	S	"	menor
" Katarina	29 "	C	"	domestica
3 - LYSENKO João	40 "	C	"	operario
" Maria	34 "	C	"	domestica
" Nina	12 "	S	"	menor
" Walter	6 "	S	"	"
" Lydia	3 "	S	"	"
4 - KOWAL Pedro	33 "	C	"	operario
" Anna	23 "	C	"	domestica
" Wira	2 "	S	"	menor
5 - LENIO Fedor	22 "	S	"	operario
6 - HLYNCZUK Nikola	25 "	C	"	operario
" Maria	28 "	C	"	domestica
" Cristina	3 "	S	"	menor
" Jan	1 "	S	"	menor
7 - WOLKOVICH Pedro	42 "	C	"	Engenheiro de minas
" Anna	42 "	C	"	domestica
" Wladimir	23 "	S	"	Eletricista
8 - SUCHECKY Steban	31 "	S	Polonês	operario
9 - CERELY Pedro	39 "	C	"	"
" Stanislaw	26 "	C	"	domestica
10 - CHULZ Bronislawa	29 "	C	"	operario
" Ana	32 "	C	"	domestica
" Jadwiga	9 "	S	"	menor
11 - DZUGAN Michel	37 "	S	Ucraniano	operario
12 - CURAK Wasyl	25 "	C	"	operario
" Natalia	26 "	C	"	domestica
" Bimon	10 meses	S	"	menor
13 - MYCHAILIO Pablo	25 anos	C	Ucraniano	operario
" Janina	25 "	C	"	domestica

Fonte: arquivo Museu do Carvão

É a partir sobretudo dos dois últimos arquivos que o banco de dados foi construído. Os documentos mais utilizados no BD foram sobretudo os relatórios de imigração e os relatórios sobre os operários imigrantes produzidos pelo Cadem.

3.2 Segunda etapa: fichamento dos documentos

Após realizar a análise preliminar da documentação, iniciou-se o processo de fichamento dos documentos através do Microsoft Word⁶⁰. Os campos⁶¹ escolhidos para interrogar e sistematizar as fontes foram os seguintes: “documento” (nome do documento já existente no próprio ou escolhido pelo autor), “localização” (localização do documento em específico no arquivo PDF, com numeração de página e nome do arquivo), “data” (a data de produção do documento), “autoria” (autor do documento), “destinatário” (destinatário do documento), “tema” (escolha de um tema para fins organizativos), “resumo” (breve resumo do documento) e “observações” (observações do documento em específico que se relacionavam com o resto das fontes).

Figura 9 - Exemplo de fichamento de uma correspondência

DOCUMENTO: Documento da Delegacia de Polícia ao chefe de escritório da CADEM;

LOCALIZAÇÃO: Silvia Box CADEM 1946-1948 (GC) AHM-MEC – P. 5;

DATA: 11 de fevereiro de 1948;

AUTORIA: Delegacia de Polícia;

DESTINATÁRIO: Humberto Lupinacci, Chefe de Escritório do CADEM;

TEMA: Imigrantes;

RESUMO: A delegacia de polícia envia um documento a Lupinacci lhe informando que o operário estrangeiro Floriano Tilak havia sido levado à delegacia e entrado em “entendimento direto” com o Chefe de Polícia. Além disso, caso se fizesse necessário a presença de Lupinacci, o mesmo seria chamado.

OBSERVAÇÕES: Floriano Tilak aparece no documento anterior (página 4, Relação dos operários imigrantes que estão trabalhando na Companhia Carbonífera Minas de Butiá). É considerado um “bom operário” e é de origem polonesa.

⁶⁰ Há programas específicos para a realização do processo de fichamento, como é o caso do Mendeley.

⁶¹ Não confundir com os campos do banco de dados, como discutido no capítulo anterior.

Fonte: imagem elaborada pelo autor

Alguns desses campos variavam de acordo com a documentação sistematizada. Contudo, o padrão dos campos desenvolvido (autoria, remetente, data, etc.) levava em consideração as correspondências trocadas entre o Cadem e órgãos oficiais, visto que era a tipo de documento dominante (ver Figura 9). Na medida que alguns desses campos perderam sua razão de ser, de acordo com a documentação sistematizada, passaram a não ser utilizados. Por exemplo, em um documento denominado “Relação dos imigrantes que ainda então trabalhando em Butiá” (ver Figura 10), um documento oficial produzido pelo Cadem, não há porque os campos “destinatário” e “remetente” serem utilizados, visto que são campos criados especificadamente para sistematizar correspondências.

O Microsoft Word, por ser um editor de texto, apresenta ao usuário esse tipo de mobilidade no momento do fichamento e sistematização das fontes. Ao contrário do banco de dados, os campos não precisam ser fixos, de modo que o usuário pode utilizar os campos que mais lhe agradam, de acordo com o documento específico, sem, inclusive, seguir um determinado padrão.

Figura 10 - Exemplo de fichamento de um relatório produzido pelo Cadem

DOCUMENTO: Relação dos operários imigrantes que estão trabalhando na Companhia Carbonífera Minas de Butiá;

LOCALIZAÇÃO: Silvia Box CADEM 1946-1948 (GC) AHM-MEC – P. 4;

DATA: 21 de dezembro de 1948;

AUTORIA: CADEM;

TEMA: Imigrantes;

RESUMO: O documento trata-se de uma lista de imigrantes que trabalhavam nas minas de Butiá. A lista é dividida entre bons operários, operários regulares e maus operários.

OBSERVAÇÕES: Todos os operários imigrantes que aparecem na lista são do Leste Europeu. São trinta operários ao total, sendo dez de origem polonesa, onze ucranianos, sete russos e dois lituanos. Dos cinco imigrantes considerados “maus operários”, quatro são russos. Dos quinze operários considerados “bons operários”: setes são ucranianos, seis são poloneses e apenas dois são russos. Entre os “operários regulares”: quatro são poloneses, três são ucranianos, dois lituanos e apenas um russo.

Um imigrante de origem polonesa considerado “bom operário” aparece com o nome de Constantino Gobratचेvski; no documento anterior (página 2, Imigrantes vindos pelos vapor “Itaimbê” e chegados Pôrto Alegre no dia 9 de junho de 1948), de junho do mesmo ano, um imigrante que havia chegado a Porto Alegre e sido enviado a região de Butiá chamava-se Constant Gorbakgski. Talvez seja o mesmo imigrante apenas com o nome alterado.

Fonte: imagem elaborada pelo autor

Sem sombra de dúvidas, o fichamento a partir do Microsoft Word ou de algum programa específico para essa função, é muito útil e pode ajudar em muito o trabalho do historiador ao lidar com a documentação, principalmente no que diz respeito a fontes qualitativas. Contudo, no processo de análise das fontes e sistematização dos documentos, o fichamento e o Microsoft Word apresentaram algumas limitações que levaram o autor a buscar uma outra forma de sistematizar as fontes, o banco de dados.

Duas foram as limitações principais apresentadas pelo Microsoft Word. Em primeiro lugar, a forma um tanto quanto burocrática de se registrar os documentos, tendo que copiar e colar, a todo momento, os campos básicos padrão (documento, localização, data, etc.), atrapalha em muito o fluxo do processo de fichamento. Claro que se pode deixar no documento os campos básicos antes de preenchê-los ou criar, de acordo com a documentação, campos específicos. Mas, de qualquer modo, ter que levar em

consideração esse problema sempre que se vai fichar um documento, torna o processo profundamente cansativo.

Em segundo lugar, a impossibilidade de se “cruzar” a documentação. A partir do campo de “observações” se buscou construir uma relação entre os documentos, contudo esse campo não dava conta da complexidade das relações entre as fontes, como pode ser notado observando os campos de “observações” nas figuras 9 e 10. Embora se pudesse estabelecer uma certa conexão entre os documentos, essa conexão era um tanto quanto frágil, na medida que facilmente se poderia perder alguma informação, visto que a percepção das semelhanças entre os documentos dependia única e exclusivamente do olhar subjetivo do autor.

É necessário enfatizar que o *software* não tem como objetivo principal sistematizar nada, mas sim servir como um editor de texto. Nesse sentido, não se pode cobrar do Microsoft Word funções que não são suas. O fichamento é uma importante ferramenta para a sistematização de documentos, principalmente no que diz respeito a fontes qualitativas, contudo nessa pesquisa em específico apresentou limitações no que diz respeito aos objetivos do autor com a documentação, isto é, de sistematizá-la buscando constituir as trajetórias dos operários imigrantes.

O arquivo final do fichamento foi concluído em setembro de 2018. O arquivo em formato DOC (formato tradicional do Microsoft Word) chegou a 21 páginas e em torno de 61 documentos fichados.

Como discutido no capítulo anterior, entendo o banco de dados, seguindo as considerações de Tiago Luís Gil (2015), como uma forma de produção histórica. Nesse sentido, a produção do fichamento da documentação também pode ser considerada uma forma de produção histórica. Afinal de contas, assim como o banco de dados e a redação textual, busca sistematizar e interrogar as fontes a partir de determinados campos que vão ao encontro dos interesses e problemas de pesquisa do historiador. Retira a fonte do seu contexto original e busca, a partir de seus objetivos e questionamentos, sistematizá-la de acordo com seu problema de pesquisa ou com o próprio formato da fonte.

3.3 Terceira etapa: desenvolvimento inicial do banco de dados

Após a conclusão do processo de fichamento, em setembro de 2018, a partir de discussões com a orientadora, surgiu a ideia de sistematizar esses documentos de maneira mais organizada, visto que, no processo de análise e fichamento da documentação, percebemos que as fontes dialogavam muito e parecia interessante tentar produzir, através de algum *software*, uma forma de perceber e organizar as relações entre os operários.

Pode-se dizer que tanto no processo de análise preliminar das fontes, quanto no processo de fichamento, dois tipos de documentos estimularam o desenvolvimento e a escolha pelo banco de dados. O primeiro deles foi a documentação referente ao operário José Varga. Ao analisar a documentação sobre Varga, percebeu-se que poderia se conhecer algumas etapas pelo qual passou o operário da chegada ao Brasil até sua prisão durante a Segunda Guerra Mundial. Os documentos eram diferentes da maioria que seria utilizada na construção do banco de dados. Eram três documentos principais: salvo-conduto (ver Figura 11), atestado de tempo de trabalho do Cadem (contendo período que trabalhou no consórcio e os salários) e correspondências entre o engenheiro e o diretor do consórcio.

Não se conseguiu desenvolver tanto quanto se gostaria as etapas da trajetória de Varga, mas a análise de seu caso em específico foi muito importante para que a ideia de se utilizar um banco de dados para tentar sistematizar alguns aspectos da imigração e da vida dos operários surgisse durante a pesquisa.

Figura 11 - Salvo-conduto de José Varga

REPARTIÇÃO CENTRAL DE POLÍCIA
EXTRA REGIÃO POLICIAL
Delegacia de Polícia

Rio Grande do Sul
BRASIL

São Jerônimo 5 de Junho de 1942.

Salvo - Conduto N.º 90
Válido somente por 60 dias.

Nome: JOSÉ VARGA
Nacionalidade: HUNGARO
Idade: 42 anos Solteiro
Profissão: Mineiro
Residência: Minas do Batia-32 distrito
Destino: Porto Alegre

Assinatura do portador:
José Varga

SELOS

OBSERVAÇÕES

Pessoa conhecida, identificada com carteira profissional N.º 26972.
Como estrangeiro encontra-se registrado no Livro N.º 2, fls. 94, desta
Delegacia de Polícia.

Heitor Trautke
DELEGADO DE POLÍCIA
com. p. m. p. sub. emp. l. m. t. u. a. l.

Tip. de R. C. P. - Mod. 2

Fonte: acervo documental Museu Estadual do Carvão

O segundo tipo de documento foram os relatórios produzidos pelo Cadem e pelo Departamento Nacional de Imigração. Percebeu-se que os relatórios do Cadem (principalmente sobre a divisão de operários e a lista de comunistas) e os relatórios de imigração se relacionavam. A partir disso, o banco de dados seria produzido para buscar, principalmente, a relação entre os documentos de imigração e os documentos produzidos pelo Cadem, visando investigar as trajetórias dos imigrantes vindos do Leste Europeu. Nesse processo, haverá muitas relações entre os relatórios de imigração e os relatórios

dos operários produzidos pelo Cadem, porém a lista dos comunistas acabará por não ter nenhuma relação com os documentos sobre os imigrantes vindo do Leste Europeu.

Escolhida a opção pelo banco de dados, a primeira etapa foi procurar um *software* para sistematizar essa documentação. O programa escolhido foi o Microsoft Access, do pacote Office da Microsoft. A escolha do programa ocorreu devido ao fato de ser compatível com o Windows e ter sido produzido pela mesma empresa do sistema operacional, a Microsoft. Por tudo isso, pareceu ser a escolha mais adequada no momento.

Aprender as funções básicas do Microsoft Access foi relativamente simples. Bastou algumas horas assistindo vídeo aulas no *Youtube* e testando aquilo que foi aprendido. Há duas razões principais para isso. Em primeiro lugar, a experiência prévia do autor com a informática ajudou em muito. Em segundo lugar, a facilidade do programa em si. Ao contrário de outros programas (como o próprio Microsoft Excel, por exemplo), o Microsoft Access é mais simples, pelo menos nas suas funcionalidades básicas. Por tudo isso, o *software* conseguiu suprir a demanda.

Em contrapartida, o que foi profundamente mais complexo foi o desenvolvimento inicial do banco de dados. Quais tabelas criar? Quais campos escolher? Qual campo escolher para ser o campo de ligação? Tudo isso já costuma ser, naturalmente, trabalhoso e complexo, para alguém sem nenhuma experiência na construção de um banco de dados se torna ainda mais difícil.

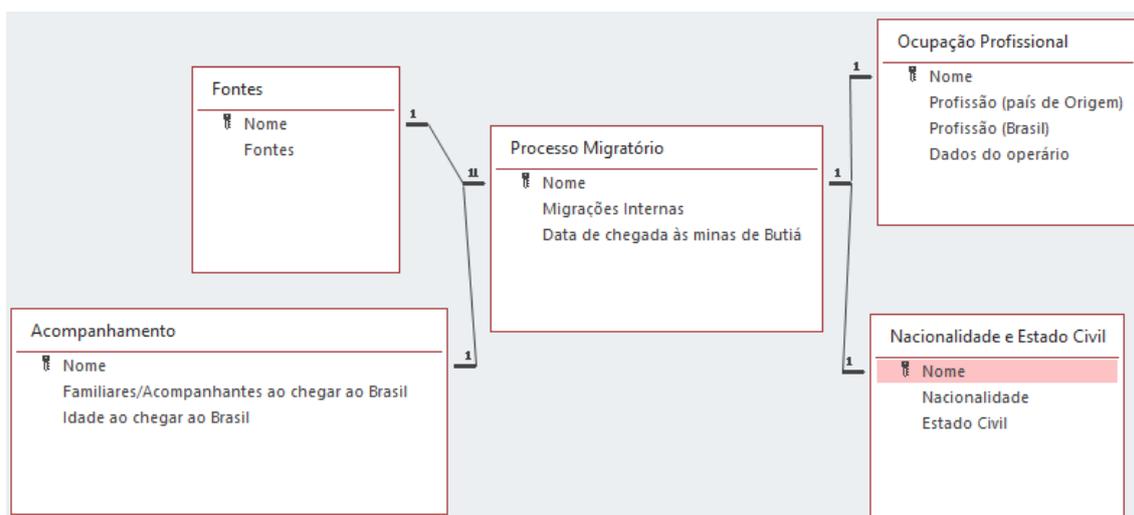
Tiago Luís Gil (2015), nesse processo inicial da construção do banco de dados, acredita que há três modelos importantes (ver Tabela 7). Os dois primeiros são o modelo conceitual e o modelo lógico. Ambos são mais abstratos. No primeiro modelo, vai se decidir o objetivo do banco de dados, qual tipo de banco de dados será feito (se *source-oriented* ou *method-centred*) e se escolherá as tabelas e os campos específicos. No segundo modelo, se estabelecerá as relações e a cardinalidade entre as tabelas e os campos e, se for o caso, se escolherá um campo de ligação. Ambos os modelos são como se fossem rascunhos do banco de dados.

Na prática, ambos os modelos foram desenvolvidos juntos em nossa experiência. Nesse sentido, se decidiu que o objetivo do banco de dados seria sistematizar, a partir da documentação, as trajetórias dos imigrantes, vindos do Leste Europeu. Se optou por,

inicialmente, usar a grande maioria da documentação, buscando tirar o máximo possível das relações entre os documentos. Não houve, nesse processo inicial, uma discussão profunda sobre o tipo de banco de dados que se procurava construir, mas como se pode notar, o banco de dados que acabou sendo construído é do tipo *method-centred*. Foram escolhidas cinco tabelas principais: “acompanhamento”, “fontes”, “nacionalidade e estado civil”, “ocupação profissional” e “processo migratório” (ver Figura 12). Em subcapítulos posteriores discutiremos com mais profundidade os campos e as funções de cada tabela.

Optou-se também por criar um campo de ligação que ligaria toda as tabelas do banco de dados. O campo “nome” (referente ao nome do operário) foi o escolhido para essa função. Essa escolha foi relativamente simples, levando em questão que o objetivo do banco de dados era sistematizar os documentos a partir de cada operário, buscando constituir sua trajetória. Contudo, nos estudos teóricos e práticos posteriores, se percebeu que não necessariamente um banco de dados necessita ter um campo de ligação que liga todas as tabelas de um banco de dados, mas sim vários tipos de ligações diferentes. De qualquer modo, essa escolha inicial não foi equivocada e nem acarretou em prejuízos futuros mais graves ao banco de dados.

Figura 12 - Exemplo de relacionamento entre as tabelas



Fonte: elaboração pelo autor

O terceiro modelo é o modelo físico. E o banco de dados informatizado propriamente dito, ou seja, o desenvolvimento do banco de dados no computador. Isso

pressupõe, logicamente, dois aspectos iniciais: a escolha de um *software* (no nosso caso, o Microsoft Access) e um conhecimento prévio de como se deve utilizar o programa (GIL, 2015).

Sobre o desenvolvimento do modelo físico, segundo Tiago Luís Gil:

Parece-me importante, tendo pronto o nosso modelo físico, gastar mais algumas horas em planejamento. Importa tomar um conjunto de fontes para testar o modelo. Não é preciso ter muitas, eu diria que uma amostra de 1% dos casos, se forem milhares e longos, já ajuda. Se forem milhares e curtos, eu recomendaria testar com 3% dos casos. Se forem centenas e longos, recomendaria 5%, se forem curtos, 10%. Mas não tome esses valores como receitas de bolo. Esses testes devem ajudar você a deixar sua base de dados flexível na medida certa e não para criar confusão. (GIL, 2015, p. 61-62)

O processo de desenvolvimento inicial do banco de dados, desde os primeiros rascunhos até últimos testes durou em torno de três meses, de outubro de 2018 mais ou menos até dezembro e janeiro de 2019. Durante esse período, o autor já havia dominado o Microsoft Access o suficiente para utilizar as principais funções e as tabelas e campos já haviam sido escolhidos e criadas no *software*.

Tabela 7- Três modelos conceituais propostos por Tiago Luís Gil

Modelo Conceitual	Desenvolvimento abstrato e inicial das tabelas e campos
Modelo Lógico	Desenvolvimento abstrato de relacionamento entre as tabela e campos
Modelo Físico	Desenvolvimento físico do banco de dados. Requer escolha de <i>software</i> e conhecimento prévio de como se utilizar o programa

Fonte: elaboração pelo autor com base em Gil, 2015.

3.4 Quarta etapa: construção do banco de dados

O processo de construção do banco de dados durou de janeiro até abril de 2019, quando foi finalizada a versão inicial do banco de dados. Durante esses meses, ocorreu a

sistematização da documentação. Nesse processo, surgiram alguns problemas e soluções que não foram previstas ou foram previstas de maneira muito abstrata.

Dois foram os problemas principais que surgiram no processo de construção do banco de dados: em primeiro lugar, o “problema do campo extra”, também conhecido como “observações”, “notas”, “comentários”, etc.; em segundo lugar, o “problema do desconhecido”.

Difícilmente se pode construir um banco de dados sem acabar apelando para um campo que busque registrar informações consideradas pelo autor como secundárias ou talvez não tão importantes. É nesse sentido que o campo “observações” ou “extra” costuma ser utilizado. Mas, afinal de contas, esse campo é realmente necessário? É fruto de um equívoco no planejamento inicial ou, de fato, faz parte da essência de um banco de dados? Caso se opte por não utilizá-lo, pode se perder algumas informações importantes?

Segundo Tiago Luís Gil,

Geralmente usado como um ‘saco de gatos’, o campo ‘observações’ e seus amigos ‘comentários’ e ‘notas’ mais atrapalham do que ajudam. Eles, comumente, são a vala comum onde todas as falhas do projeto vão parar. E muitas dessas falhas poderiam ter rendido grandes ideias e conclusões. Nosso foco precisa estar em evitar o campo ‘observações’. Antes de preenchê-lo, seria melhor discutir com a equipe, ou pensar sozinho, se for o caso, na possibilidade de criar algum campo ou solução que resolva aquele problema e outros potenciais, vindouros, da mesma natureza.

Esse campo, contudo, é indispensável. Ele pode ser importante para anotações temporárias. Nesse caso, sua utilidade é bem grande. Mas eu não o chamaria de “observações”, nome convidativo para procedimentos preguiçosos de trabalho, e criaria um campo chamado de ‘notas temporárias de pesquisa’, mais apropriado para sua real e valorosa função. (GIL, 2015, p. 81-82)

Inicialmente, para se dar conta de informações secundárias dos operários, se utilizou o campo “dados do operário” da tabela “ocupação profissional”. Esse campo foi criado sobretudo para registrar os dados do documento que dividia os operários imigrantes entre “bons operários”, “operários regulares” e “maus operários” e documentos semelhantes, contudo, na medida que apareciam outras informações consideradas secundárias (como por exemplo, alguma relação do operário com a polícia),

passou-se a utilizar esse campo com a função de campo extra. Acredito que, no desenvolvimento do banco de dados, utilizar o campo “dados do operário” acabou sendo uma boa ideia (ver Figura 13).

Figura 13 - Exemplo do campo “dados do operário”

Dados do operário
Considerado "operário regular" em dezembro de 1948
Trabalhava nas minas de Butiá em novembro de 1948
Desconhecido
Desconhecido
Considerado "operário regular" em dezembro de 1948
Desconhecido
Desconhecido
Considerado um "bom operário" em dezembro de 1948
Desconhecido
Trabalhava nas minas de Butiá em novembro de 1948
Desconhecido
Desconhecido

Fonte: elaboração pelo autor

O segundo problema que surgiu foi o “problema do desconhecido”. Na medida que se busca interrogar uma determinada documentação a partir de perguntas que se materializam, no banco de dados, através de tabelas e campos, é natural que alguns dos campos não sejam preenchidos. Sendo assim, um termo (ou não) deve ser utilizado naquele registro vazio. No processo de construção do banco de dados, optei por utilizar o termo “desconhecido”. Poderiam ter sido outro termo (como “sem informação”, “não se aplica”, etc.) ou até deixar o registro sem informação alguma⁶², isto é, vazio.

Algumas questões levantadas no problema anterior também valem para o “problema do desconhecido”: afinal, faz parte da essência do banco de dados ou é um erro do planejamento inicial ou de um campo específico? Deve-se utilizar um termo ou deixar o registro vazio?

⁶² O Microsoft Access oferece a opção de se utilizar um termo específico (no caso, o termo “desconhecido”) sempre que um registro é deixado vazio pelo autor no formulário.

Durante a construção do banco de dados, três facilidades foram encontradas e desenvolvidas: em primeiro lugar, a construção de um formulário; em segundo lugar, a elaboração de uma tabela geral; e em terceiro lugar, a utilização de relatórios.

Figura 14 - Exemplo de formulário do Microsoft Access

The image shows a screenshot of a Microsoft Access form titled "Processo Migratório". The form contains the following fields and values:

Nome	Somen Gluz	Familiares/Acoi	Imigrou com esposa e um filho
Migrações Inter	Junho de 1948 em Porto Alegre	Idade ao chegar	30
Data de chegada	1948	Fontes	Silvia Box CADEM 1946-1948 (GC) AHM-MEC
Profissão (país c	Mineiro		
Profissão (Brasi	Mineiro		
Dados do operá	Considerado um "bom operário" em dezembro de 1948		
Nacionalidade	Polonês		
Estado Civil	Casado		

At the bottom of the form, there is a status bar showing "Registro: 14 de 192", a filter icon, "Sem Filtro", and a search button labeled "Pesquisar".

Fonte: elaboração pelo autor

A primeira facilidade desenvolvida foi a utilização de um formulário para acelerar o processo (ver Figura 14). Através do formulário se poderia registrar os operários de modo muito mais rápido e simples, além disso encontrar um registro em específico se tornou mais simples também, já que o formulário do Microsoft Access contém uma pequena barra de pesquisa que é muito útil para encontrar registros específicos. Em alguns bancos de dados, principalmente voltado ao público, o formulário costuma ser utilizados pelos clientes ou público de modo geral e as tabelas e campos pelo técnico. É interessante

como se pode ressignificar⁶³ algumas ferramentas do *software* visando sempre os objetivos do autor do BD.

A segunda facilidade diz respeito a criação de uma tabela total contendo todos os campos de todas as tabelas (ver Figura 15). A partir dessa tabela total (denominada de *query* ou consulta) se pode ter uma visão ampla do banco de dados. Isso ajuda muita durante o processo, na medida que, através dessa ferramenta, se poder perceber alguns erros, principalmente erros de digitação nos registros ou até algum campo equivocado.

Figura 15 - Tabela total ou *query*

Nome	Migrações Internas	Data de che.	Profissão (p	Profissão (B	Dados do operário	Nacionalida	Estado Civil	Fam
Alexander Kof	Junho de 1948 em Porto Alegre	1948	Mineiro	Mineiro	Desconhecido	Desconhecido	Casado	Imig
Pawel Oczycze	Junho de 1948 em Porto Alegre	1948	Mineiro	Mineiro	Considerado um "bom operário" em de	Polonês	Casado	Imig
Somen Gluz	Junho de 1948 em Porto Alegre	1948	Mineiro	Mineiro	Considerado um "bom operário" em de	Polonês	Casado	Imig
Costant Gorbal	Junho de 1948 em Porto Alegre	1948	Mineiro	Mineiro	Considerado um "bom operário" em de	Polonês	Casado	Imig
Simon Schbegi	Junho de 1948 em Porto Alegre; Ao chegar e	1948	Mineiro	Desconhecido	Desconhecido	Desconhecido	Casado	Imig
Ismailov Mich	Junho de 1948 em Porto Alegre; Ao chegar e	1948	Mineiro	Desconhecido	Desconhecido	Desconhecido	Casado	Imig
Nicolau Hlyncz	Deixou as Minas de Butiá em setembro de 1	Desconhecido	Desconhecido	Mineiro	Desconhecido	Desconhecido	Casado	Desc
Miguel Dembil	Deixou as Minas de Butiá em setembro de 1	Desconhecido	Desconhecido	Mineiro	Desconhecido	Desconhecido	Casado	Desc
Wladislaw Kuh	Deixou as Minas de Butiá em setembro de 1	Desconhecido	Desconhecido	Mineiro	Desconhecido	Desconhecido	Casado	Desc
João Furmanis	Deixou as Minas de Butiá em setembro de 1	Junho de 1947	Operário	Mineiro	Desconhecido	Lituano	Casado	Imig
Erwin Wudel	Deixou as minas de Butiá em dezembro de 1	Desconhecido	Desconhecido	Mineiro	Desconhecido	Desconhecido	Casado	Desc
José Kurgan	Deixou as Minas de Butiá em setembro de 1	Julho de 1947	Mineiro	Mineiro	Desconhecido	Polonês	Solteiro	Imig
Stive Boric	Deixou as Minas de Butiá em setembro de 1	Desconhecido	Desconhecido	Mineiro	Desconhecido	Desconhecido	Casado	Desc
Rmao Dokuckii	Deixou as Minas de Butiá em setembro de 1	Desconhecido	Desconhecido	Mineiro	Desconhecido	Desconhecido	Casado	Desc
Rndel Saarepe	Deixou as Minas de Butiá em setembro de 1	Desconhecido	Desconhecido	Mineiro	Desconhecido	Desconhecido	Casado	Desc
Fedor Antonov	Desconhecido	Junho de 1947	Técnico de Mir	Mineiro	Considerado "bom operário" em de	Russo	Casado	Imig
Waayl Curak	Desconhecido	Junho de 1947	Operário	Mineiro	Considerado "bom operário" em de	Ucraniano	Desconhecido	Che
Pablo Mychaili	Desconhecido	Junho de 1947	Operário	Mineiro	Considerado "bom operário" em de	Ucraniano	Desconhecido	Che
Edward Mias	Desconhecido	Junho de 1947	Operário	Mineiro	Considerado "bom operário" em de	Polonês	Desconhecido	Che
Nikola Sawitsc	Desconhecido	Junho de 1947	Operário	Mineiro	Considerado "bom operário" em de	Russo	Solteiro	Imig
Kiril Pawluk	Desconhecido	Junho de 1947	Operário	Mineiro	Considerado "bom operário" em de	Ucraniano	Casado	Imig
Iwan Slobodar	Desconhecido	Desconhecido	Desconhecido	Mineiro	Considerado "bom operário" em de	Ucraniano	Desconhecido	Desc
Jan Millyan	Desconhecido	Julho de 1947	Mineiro	Mineiro	Considerado "bom operário" em de	Ucraniano	Casado	Imig
Josef Szczepan	Desconhecido	Julho de 1947	Mineiro	Mineiro	Considerado "bom operário" em de	Polonês	Casado	Imig
Jakob Charczel	Desconhecido	Julho de 1947	Mineiro	Mineiro	Considerado "bom operário" em de	Ucraniano	Casado	Imig
Kyzma Sawosz	Desconhecido	Setembro de 1	Desconhecido	Mineiro	Considerado "bom operário" em de	Ucraniano	Casado	Imig
Florian Tylak	Desconhecido	Desconhecido	Desconhecido	Mineiro	Em fevereiro de 1948, o operário teve	Polonês	Casado	Desc
Steban Suchec	Desconhecido	Junho de 1947	Operário	Mineiro	Considerado "operário regular" em de	Polonês	Desconhecido	Imig
Bronislawe Ch	Desconhecido	Junho de 1947	Operário	Mineiro	Considerado "operário regular" em de	Polonês	Desconhecido	Imig
Alexander Bek	Desconhecido	Julho de 1947	Mineiro	Mineiro	Considerado "operário regular" em de	Polonês	Solteiro	Imig

Fonte: elaboração pelo autor

A terceira facilidade foi a utilização dos relatórios (ver Figura 16). O relatório é uma ferramenta do Microsoft Access que apresenta individualmente cada registro específico das tabelas. No caso do banco de dados em questão, cada registro diz respeito a um operário imigrante. A partir dessa ferramenta podia-se observar melhor como havia ficado o registro da trajetória de um operário em específico.

⁶³ Em relação a esse problema de ressignificação ou adaptação dos programas de banco de dados a pesquisa histórica: “entre as bases mais comumente utilizadas (Excel, Access, FileMaker) nenhuma é dedicada especialmente a constituir um instrumento para a pesquisa histórica, ou seja, são softwares para construção de bancos de dados pensados para os mais variados fins (empresariais, governamentais) e o que nos toca fazer é adaptá-los o máximo possível a fim de que respondam às nossas demandas de pesquisas” (THOMPSON FLORES, 2015, p. 243).

Figura 16 - Exemplo de relatório de um operário

Nome	Michael Zason
Migrações Internas	Desconhecido
Data de chegada às minas	Julho de 1947
Profissão (país de Origem)	Telegrafista
Profissão (Brasil)	Mineiro
Dados do operário	Desconhecido
Nacionalidade	Polonês
Estado Civil	Solteiro
Familiares/Acompanhante	Imigrou sozinho
Idade ao chegar ao Brasil	31
Fontes	Dossiê dos Imigrantes (Europeus), CADEM Caixa 296, AHM

Fonte: elaboração pelo autor

3.5 O banco de dados concluído: versão inicial e versão final⁶⁴

A versão inicial do banco de dados ficou pronta em abril de 2019. Possuía as mesmas tabelas e campos da versão final, a única diferença é que tinha 284 nomes de operários registrados. Após a produção da versão inicial, averiguou-se que os operários imigrantes presentes no relatório sobre os comunistas não possuíam relação nenhuma com a o restante da documentação de imigração. Nesse sentido, foram retirados todos do banco de dados. Além disso, o caso de José Varga e de alguns alemães que haviam sido presos durante a Segunda Guerra Mundial também foram retirados. De modo geral, todos aqueles registros que demonstraram não ter nenhuma relação com os documentos de imigração foram retirados.

⁶⁴ A documentação sistematizada está disponível em: <<https://onedrive.live.com/?authkey=%21APXqspB6vspG3Pg&id=7F21F67F0E038081%216217&cid=7F21F67F0E038081>>. O banco de dados na sua versão final e o fichamento estão disponíveis em: <<https://drive.google.com/file/d/1J2lz8sDTxbQt5PS3p2oW7EJDO0NONYPB/view?usp=sharing>>. Qualquer problema em relação ao *download* dos arquivos, entrar em contato com o autor pelo e-mail: guilhermesichelero@gmail.com.

A versão final foi um arquivo em formato ACCDB (*Microsoft Access Database*) de um banco de dados de tipo *method-centred* (centrado no método) contendo cinco tabelas (“acompanhamento”, “fontes”, “nacionalidade/estado civil”, “ocupação profissional” e “processo migratório”) com 192 registros. Cada uma dessas tabelas possui campos específicos unidos pelo campo de ligação “nome”. Vejamos com profundidade cada uma das tabelas em específico.

A tabela “acompanhamento” (ver Figura 17) foi desenvolvida com o intuito de registrar os acompanhantes dos operários ao chegar no Brasil. Entretanto, na medida que a maioria dos documentos que continham os nomes dos operários e seus familiares também apresentavam a idade do trabalhador, resolvi por acrescentar o campo “idade”, ao invés de criar uma tabela específica. Sendo assim, os campos dessa tabela são: “nome” (campo de ligação), “familiares/acompanhantes ao chegar ao Brasil” e “idade ao chegar no Brasil”.

Figura 17- Tabela de acompanhamento

Acompanhamento			
	Nome	Familiares/Acompanhante	Idade ao chegar ao Brasil
+	José Kurgan	Imigrou sozinho	32
+	José Rohaljski	Desconhecido	Desconhecido
+	Josef Hokko	Imigrou sozinho	27
+	Josef Rogalski	Imigrou com a esposa e o filho	22
+	Josef Szczepanski	Imigrou com esposa e o filho	29
+	Junis Zeltauskis	Desconhecido	Desconhecido
+	Jurko Senczuk	Imigrou com a esposa e dois	47
+	Kiril Pawluk	Imigrou com a esposa	47
+	Kurt Braczak	Imigrou sozinho	18
+	Kuzma Sawocz	Imigrou com a esposa e filho	26
+	Kyzma Sawosz	Imigrou com a esposa e o filho	26
+	Leonid Kusnecoff	Desconhecido	Desconhecido
+	Leonidio Kusnecoff	Desconhecido	Desconhecido
+	Luciano Kurowski	Imigrou sozinho	23
+	Manoel Dzuguna	Desconhecido	Desconhecido
+	Michael Bondarczuk	Imigrou com sua esposa	34
+	Michael Dembicki	Imigrou com sua esposa	49
+	Michael Noi	Imigrou com a esposa e o filho	29
+	Michael Zason	Imigrou sozinho	31
+	Michail Nasorow	Imigrou com a esposa e filho	29
+	Michail Nazarow	Imigrou com a esposa e filho	29
+	Michat Zazan	Desconhecido	Desconhecido
+	Michel Dzugan	Imigrou sozinho	37
+	Miguel Bondartchik	Desconhecido	Desconhecido
+	Miguel Dembicki	Desconhecido	Desconhecido
+	Mikola Sklar	Imigrou com a esposa e o filho	23
+	Mykola Sklar	Imigrou com a esposa e filho	25
+	Nacolau Jaroschenko	Desconhecido	Desconhecido

Fonte: elaboração pelo autor

A tabela “fontes” foi feita apenas para se registrar os documentos em que o operário aparecia. Dois campos foram utilizados: “nome” (campo de ligação) e “fontes”. O documento de referência foi o nome do arquivo de PDF.

A tabela “nacionalidade e estado civil” (ver Figura 18) foi desenvolvida, inicialmente, para se ter uma compreensão mais profunda das características pessoais do

operário⁶⁵. Dessa maneira, os campos da tabela eram: “nome” (campo de ligação), “nacionalidade” e “estado civil”.

Figura 18 - Tabela de nacionalidade e estado civil

Nacionalidade e Estado Civil			
	Nome	Nacionalidade	Estado Civil
+	Michael Dembicki	Ucraniano	Casado
+	Michael Noi	Ucraniano	Casado
+	Michael Zason	Polonês	Solteiro
+	Michail Nasorow	Desconhecido	Casado
+	Michail Nazarow	Russo	Casado
+	Michat Zazan	Desconhecido	Desconhecido
+	Michel Dzugan	Ucraniano	Solteiro
+	Miguel Bondartchik	Desconhecido	Desconhecido
+	Miguel Dembicki	Desconhecido	Desconhecido
+	Mikola Sklar	Ucraniano	Casado
+	Mykola Sklar	Desconhecido	Casado
+	Nacolau Jaroschenko	Desconhecido	Desconhecido
+	Nicolai Schmantz	Ucraniano	Desconhecido
+	Nicolai Yaroshenko	Ucraniano	Casado
+	Nicolau Berezuckij	Desconhecido	Desconhecido
+	Nicolau Chitry	Desconhecido	Desconhecido
+	Nicolau Hlynczuk	Desconhecido	Desconhecido
+	Nicolau Schmautz	Ucraniano	Casado
+	Nicolau Sterlandnikow	Desconhecido	Desconhecido
+	Nikola Berezuckij	Ucraniano	Casado
+	Nikola Hlynczuk	Ucraniano	Casado
+	Nikola Pawluk	Ucraniano	Casado
+	Nikola Radion	Polonês	Casado
+	Nikola Sawitsch	Russo	Solteiro
+	Nikola Sterladnikow	Ucraniano	Solteiro
+	Nikolai Chytry	Polonês	Casado
+	Nikolai Sokolowskiy	Russo	Solteiro
+	Nikolaj Sokolowskiw	Desconhecido	Solteiro

Fonte: elaboração pelo autor

⁶⁵ Como um dos objetivos do projeto de pesquisa era estudar as relações étnicas e raciais dos operários da região carbonífera riograndense, bem no começo da construção do banco de dados os campos “etnia” e “raça” foram criados, contudo após uma revisão da documentação e averiguação de que esse dado não existia nas fontes, foi retirado.

A tabela “ocupação profissional” (ver Figura 19) procura dar conta do aspecto profissional do operário. Os campos escolhidos são: “nome” (campo de ligação), “profissão (país de origem)”, “profissão (Brasil)” e “dados do operário”. O campo “dados do operário”, como já discutido anteriormente, serviu como função de “extra” da tabela, registrando aspectos secundários da vida profissional e pessoal do operário que não estavam previstas inicialmente.

Figura 19 - Tabela de ocupação profissional

	Nome	Profissão (país de Origem)	Profissão (Brasil)	Dados do operário
+	Timoteo Komstas CH	Operário	Mineiro	Desconhecido
+	Alexander Karaman	Desconhecido	Desconhecido	Desconhecido
+	Alexander Kohut	Mineiro	Mineiro	Desconhecido
+	Alexander Rokicki	Marceneiro	Mineiro	Considerado "operário regular" e
+	Alexandre Marcauk	Desconhecido	Mineiro	Desconhecido
+	André Kalanyr	Serralheiro	Mineiro	Desconhecido
+	Andreas Bentkowski	Operário	Mineiro	Desconhecido
+	Antos Klos	Luminador	Mineiro	Desconhecido
+	Basilio Antonoff	Desconhecido	Mineiro	Desconhecido
+	Basilio Jurkivo	Desconhecido	Mineiro	Desconhecido
+	Basilio Marczuk	Desconhecido	Mineiro	Desconhecido
+	Basilio Qurak	Desconhecido	Mineiro	Trabalhava nas minas de Butiá em
+	Bronislaw Gunther	Mineiro	Mineiro	Desconhecido
+	Bronislaw Chulz	Operário	Mineiro	Considerado "operário regular" e
+	Bruno Szulo	Desconhecido	Mineiro	Trabalhava nas minas de Butiá em
+	Casimir Pazderski	Operário	Mineiro	Desconhecido
+	Cassemiro Pazderski	Desconhecido	Desconhecido	Mineiro
+	Ch Timoteo Komstas	Desconhecido	Mineiro	Considerado "operário regular" e
+	Ciril Malahuza	Operário	Mineiro	Desconhecido
+	Cirilo Halauza	Desconhecido	Mineiro	Desconhecido
+	Costant Gorbakgski	Mineiro	Mineiro	Considerado um "bom operário"
+	Demetrio Kasper	Desconhecido	Mineiro	Desconhecido
+	Demetrio Zikalek	Desconhecido	Mineiro	Trabalhava nas minas de Butiá em
+	Dmytre Zukalbk	Operário	Mineiro	Desconhecido
+	Dmytro Kasper	Operário	Mineiro	Desconhecido
+	Domingo Szyzsko	Desconhecido	Desconhecido	Mineiro
+	Dominiqua Czisko	Operário	Mineiro	Desconhecido
+	Edward Mias	Operário	Mineiro	Considerado "bom operário" em

Fonte: elaboração pelo autor

Por último, a tabela “processo migratório” (ver Figura 20). Produzida para tentar registrar o processo de migração dos trabalhadores. Inicialmente, se buscava registrar as datas de saída dos operários de seus países de origem, contudo a ausência dessa informação acabou por tornar esse campo obsoleto. Nesse sentido, essa tabela possuía três campos: “nome” (campo de ligação), “migrações internas” e “data de chegada às minas de Butiá”. O campo “migrações internas” buscava registrar o processo migratório

do operário até chegar à região de São Jerônimo, mas acabou se tornando mais ampla, na medida que também registrou o período de saída.

Figura 20 - Tabela de processo migratório

Processo Migratório		
Nome	Migrações Internas	Data de chegada às minas de Butiá
Timoteo Komstas CH	Desconhecido	Junho de 1947
Alexander Karaman	Desconhecido	Setembro de 1948
Alexander Kohut	Junho de 1948 em Porto Alegre	1948
Alexander Rokicki	Desconhecido	Julho de 1947
Alexandre Marcauk	Deixou as minas de Butiá em janei	Desconhecido
André Kalanyr	Deixou as minas de Butiá em janei	Julho de 1947
Andreas Bentkowski	Retornou ao Rio de Janeiro em jull	Junho de 1947
Antos Klos	Retornou ao Rio de Janeiro em jull	Julho de 1947
Basilio Antonoff	Deixou as minas de Butiá em janei	Desconhecido
Basílio Jurkivo	Deixou as minas de Butiá em janei	Desconhecido
Basílio Marczuk	Deixou as minas de Butiá em janei	Desconhecido
Basilio Qurak	Desconhecido	Desconhecido
Bronislaw Gunther	Junho de 1948 em Porto Alegre	1948
Bronislawe Chulz	Desconhecido	Junho de 1947
Bruno Szulo	Desconhecido	Desconhecido
Casimir Pazderski	Desconhecido	Junho de 1947
Casemiro Pazderski	Deixou as minas de Butiá em març	Desconhecido
Ch Timoteo Komstas	Desconhecido	Desconhecido
Ciril Malahuza	Desconhecido	Junho de 1947
Cirilo Halauza	Deixou as minas de Butiá em janei	Desconhecido
Costant Gorbakgski	Junho de 1948 em Porto Alegre	1948
Demetrio Kasper	Deixou as minas de Butiá em janei	Desconhecido
Demetrio Zikalek	Desconhecido	Desconhecido
Dmytre Zukalbk	Desconhecido	Junho de 1947
Dmytro Kasper	Desconhecido	Junho de 1947
Domingo Szyszko	Deixou as minas de Butiá em març	Desconhecido
Dominiqua Czisko	Desconhecido	Julho de 1947
Edward Mias	Desconhecido	Junho de 1947

Fonte: elaboração pelo autor

Tabela 8 - Tabelas e campos do banco de dados produzido pela pesquisa

Banco de dados produzido	
Tabelas	Campos
Processo migratório	Nome, Migrações Internas, Data de chegada às minas de Butiá

Acompanhamento	Nome, Familiares/Acompanhantes, Idade ao chegar no Brasil
Ocupação Profissional	Nome, Profissão (país de origem), Profissão (Brasil), Dados do operário
Nacionalidade e Estado Civil	Nome, Nacionalidade, Estado Civil
Fontes	Nome, Fontes

Fonte: elaboração pelo autor

3.6 Equívocos e correções

Após todo o processo de construção de banco de dados, identificamos alguns equívocos cometidos, contudo avaliamos que esses equívocos não inviabilizam a utilização do banco de dados por historiadores e pesquisadores que pretendam, futuramente, estudar essa documentação sistematizada. Apresentamos aqui estes problemas, no entanto, como forma de reflexão sobre o processo de desenvolvimento da metodologia do banco de dados.

Há dois equívocos principais que, embora não comprometam a utilização do banco de dados em futuras pesquisas, poderiam ter sido melhor desenvolvidos: em primeiro lugar, a escolha das tabelas; em segundo lugar, o “problema do desconhecido”.

Ao final do processo, avaliei que houve um exagero na escolha das tabelas. As tabelas “acompanhamento” e “nacionalidade e estado civil” poderiam estar em uma mesma tabela já que ambas possuíam poucos campos que se relacionavam (duas cada, sem contar o campo de ligação). Além disso, a tabela “fontes” (com apenas dois campos) ficou, digamos, assim meio isoladas das outras tabelas, afinal de contas era praticamente uma tabela com apenas um campo.

Outro equívoco tem a ver com o “problema do desconhecido”. Evidentemente que esse é um problema que dificilmente se pode evitar, principalmente quando o usuário tem pouca experiência na construção de banco de dados, contudo há formas, como será discutido adiante, de evitar-se esse problema. Além disso, se poderia ter utilizado termos diferentes mais adequados ao questionamento do campo em específico. Por exemplo,

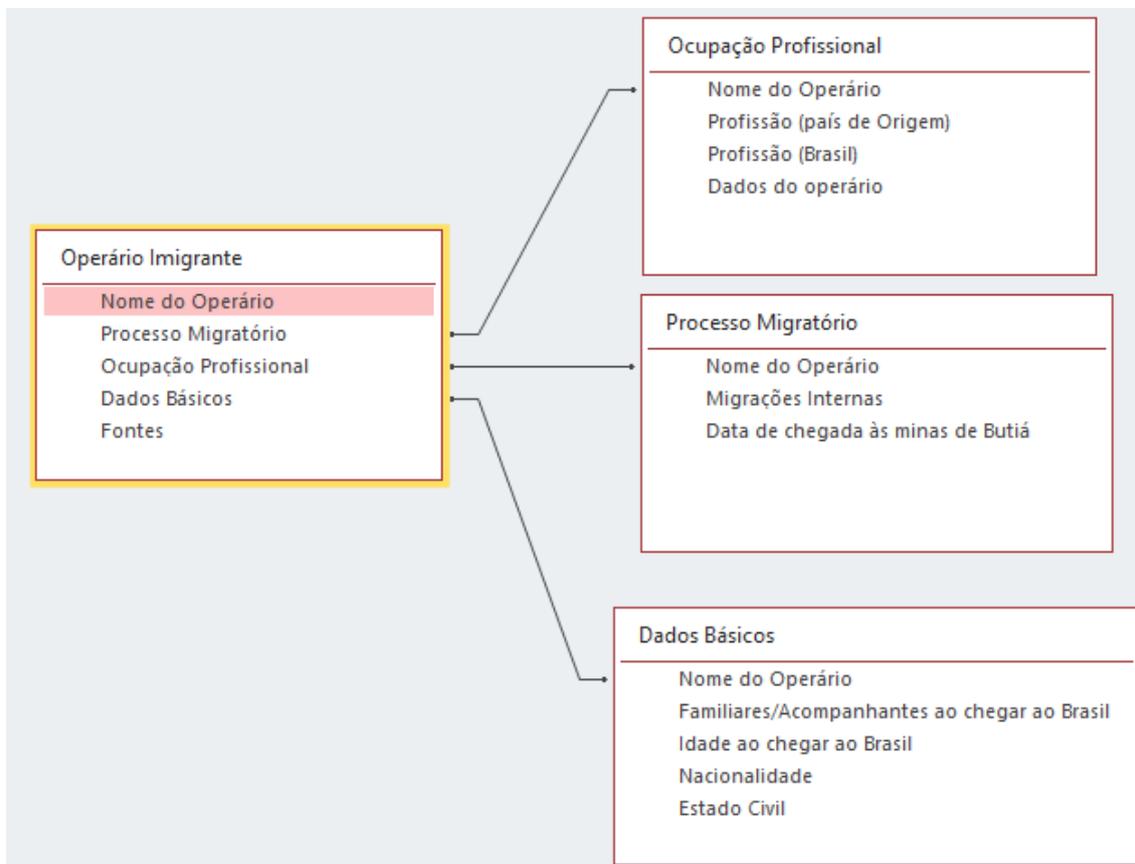
utilizado o termo “desconhecido” na maior parte dos campos (já que, de fato, eram informações desconhecidas) e o termo “não se aplica” no campo “dados do operário”. Isso ajudaria a fazer uma distinção mais afinada entre quando não há a informação ou quando o campo não se aplica aquele operário.

Após todo o processo empírico de construção do banco de dados e a reflexão teórica, tanto para a apresentação no SIC de 2019, quanto para este trabalho de conclusão de curso, em 2020, avalio que produziria o banco de dados de outro modo.

Entre as modificações, criaria uma tabela principal denominado “operário imigrante” contendo cinco campos: “nome do operário”, “processo migratório”, “ocupação profissional”, “dados pessoais” e “fontes”. Os campos “nome do operário” e “fontes” registrariam, logicamente, o nome do operário e as fontes utilizadas, mas os outros três campos seriam somente informando se há ou não dados sobre determinado assunto, portanto seria campos de somente duas escolhas: “sim” ou “não”.

A partir dessa tabela principal criaria outras três tabelas: “processo migratório” e “ocupação profissional”, semelhantes a existentes atualmente, e a tabela “dados pessoais” iria incluir os campos das atuais tabelas de “nacionalidade e estado civil” e “acompanhamento”.

Figura 21 - Exemplo de tabelas, campos e relacionamento numa suposta segunda versão



Fonte: elaboração pelo autor

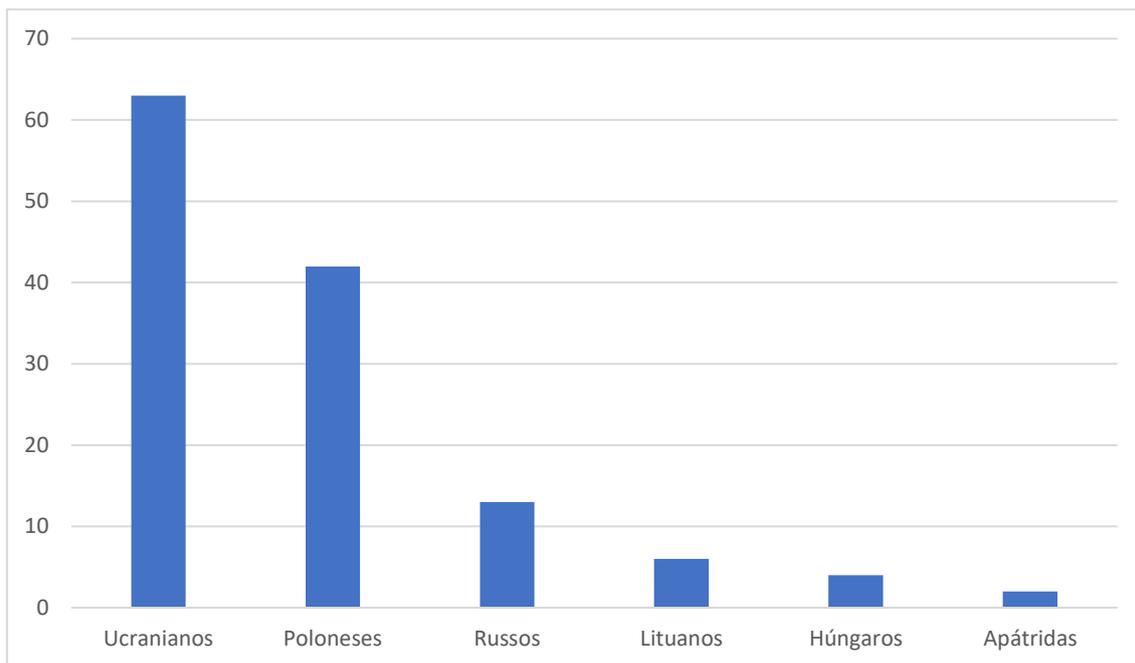
Essa segunda versão do banco de dados seria melhor e mais desenvolvida porque, em primeiro lugar, é mais organizada e acurada que a anterior e, em segundo lugar, procura diminuir a frequência do “problema do desconhecido”. Por exemplo, se buscarmos dados sobre o operário ucraniano Níolas Dudko, não encontraremos na documentação informações sobre o processo migratório do operário, mas há dados sobre estado civil, nacionalidade, idade, etc. Nesse sentido, todos os campos da tabela “processo migratório” tiveram que ser preenchidos com o termo “desconhecido”. A segunda versão evitaria esse equívoco, na medida que já na tabela “operário imigrante” o campo “processo migratório” seria registrado a opção “não”.

3.7 Utilidade do banco de dados: dados quantitativos e hipótese

O banco de dados produzido pela pesquisa possui duas utilidades principais: em primeiro lugar, para estabelecer relações entre os dados quantitativos dos operários imigrantes; em segundo lugar, para ajudar no desenvolvimento de uma hipótese sobre a relação entre o Cadem e a vinda de imigrantes do Leste Europeu.

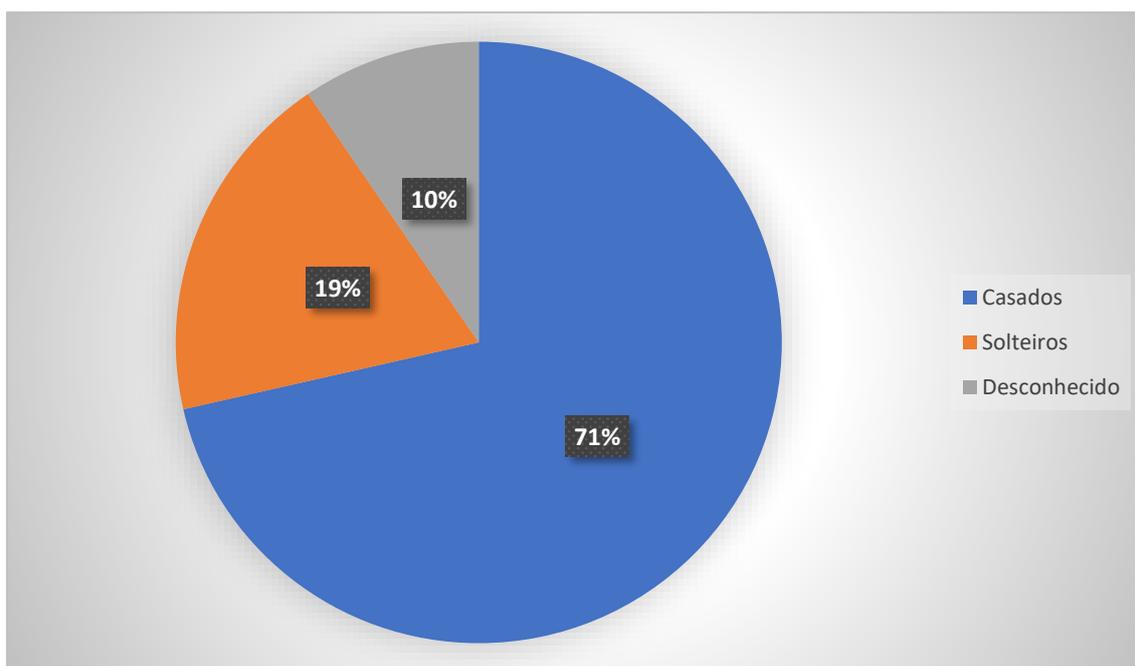
Em relação aos dados quantitativos⁶⁶ do banco de dados, vejamos um relatório resumido das principais informações: dos 109 operários que temos dados sobre a nacionalidade (ver Gráfico 1), as principais nacionalidades a imigrar foram ucranianos (63 ao total: 45 casados e 12 solteiros, ver Gráfico 2), poloneses (42 ao total: 15 casados e 19 solteiros), russos (13 ao total: 7 casados e 3 solteiros) e lituanos (6 ao total: 1 casado). Dos 119 operários que temos dados sobre o estado civil, 81 eram casados e 38 solteiros. Dos 119 operários que temos dados sobre a data de chegada a região carbonífera, 95 chegaram em 1947 e 24 em 1948. Dos 124 operários que temos dados sobre os acompanhantes, 38 imigraram sozinhos e 85 imigraram acompanhados.

⁶⁶ Os dados quantitativos do banco de dados contribuíram no trabalho da Profa. Dra. Clarice Gontarski Speranza no artigo “*O afluxo de refugiados de guerra nas minas de carvão do Rio Grande do Sul ao fim da II Guerra*”, no prelo.

Gráfico 1 - Nacionalidades no banco de dados

Fonte: elaboração pelo autor

Outros relacionamentos poderiam ser realizados entre os dados, como por exemplo a relação de operário ucranianos casados que imigraram acompanhados dos cônjuges e filhos e chegaram em 1947 e 1948, nesse exemplo três tabelas se relacionariam (a tabela “Nacionalidade e Estado Civil”, “Acompanhamento” e “Processo Migratório”). Outro exemplo, poderia-se estabelecer a relação entre os operários poloneses e sua profissão no seu país de origem, nesse caso duas tabelas se relacionariam (as tabelas “Nacionalidade e Estado Civil” e “Ocupação profissional”). São múltiplas as relações entre os campos que podem ser feitas a partir da tabela total ou *Query*.

Gráfico 2 - Relação de ucranianos casados e solteiros

Fonte: elaboração pelo autor

A partir do banco de dados e da análise documental uma hipótese pode ser levantada sobre a relação entre a vinda dos operários imigrantes do Leste Europeu e o Cadem. Ao contrário das imigrações do início do século XX, onde vieram vários operários adeptos de supostas “ideias perigosas”, ou seja, com militância anterior anarquista ou comunista, como pode ser o caso do operário José Varga, o Cadem optou por trazer, após a Segunda Guerra Mundial, operários principalmente do Leste Europeu, que fossem contrários a essas concepções.

Pode-se embasar essa hipótese com base em três evidências principais levantadas a partir da documentação. Em primeiro lugar, a ausência de relação entre a lista de comunistas elaborado pelo consórcio e os documentos de imigração, como atesta o banco de dados. Em segundo lugar, uma carta de agradecimento ao Cadem elaborada pelos operários imigrantes, agradecendo por salvá-los do “maldito regime vermelho totalitário comunista de Stalin” (ver Figura 22). Em terceiro lugar, a caracterização, por parte do consórcio, de “russos brancos” (isto é, russos não adeptos do comunismo) como uma das melhores nacionalidades para migrar, no questionário de imigração respondido pelo Cadem (ver Figura 23).

Figura 22 - Parte da carta dos operários imigrantes do Leste Europeu ao Cadem

Ilustríssimo Senhor
 Dr. Fernando Sacouck,
 Sr. D. Eng. Chefe das Minas de Butiá.

Hoje, dia 23 de Junho de 1948, cumprimos com júbilo um ano de trabalho, e comemoramos também o primeiro aniversário de nossa chegada neste tão hospitaleiro e tão generoso Brasil.

Nós, imigrantes políticos ucranianos, russos-brancos, russos, poloneses, estonianos, letonianos e jugoslavos, deixamos a nossa pátria, não porque não gostássemos dela ou porque ela não fosse preciosa para nós, mas porque o „maldito regime vermelho totalitário comunista de Stalin” pisoteia tudo o que é belo, que é liberdade, democracia, cristianismo.

Só até hoje lá se efetua o extermínio físico, não só dos melhores cidadãos da Rússia subjugada, mas também da Europa, dos Balcãs, do Extremo Oriente e da China. Lá os comunistas fuzilam ou matam lentamente suas vítimas, martirizando-as e torturando-as nos múltiplos campos de concentração de camponeses, operários, funcionários e intelectuais.

Só agora na escala internacional começa a se revelar em toda a sua plenitude o bestial sistema comunista, sem moral e sem escrúpulos, como é sua verdadeira face.

Fonte: acervo documental Museu Estadual do Carvão

Figura 23 - Questionário respondido pelo Cadem

QUESTIONÁRIO IGCR. SOBRE D.P.s

INDÚSTRIA

Nome da Indústria (empresa) :- Cia. Carbonífera Minas de Butiá

Município:- São Jerônimo

Estado:- Rio Grande do Sul

Espécie de Indústria:- Mina de Carvão

Quantas famílias de D.P.s recebeu ?:- 86

Quantas pessoas ao todo ? ...:- 186

Nacionalidades das pessoas...:- Ucrânianos, Russos, Polacos e Húngaros.

Qual a nacionalidade que lhe parece melhor ? (x):- Alemã, polonesa, Ucrâniana e russos brancos.-

Está satisfeito com o rendimento do trabalho dos D.P.s ?...:- Sim.-

Está satisfeito com o comportamento dos D.P.s ?:- Sim.-

Gostaria de receber mais famílias de D.P.s em sua indústria ?:- Sim.-

De que nacionalidade preferiria ?:- Das indicadas em (x)

Espaço para Observações..... Podemos receber mais 300 operários, entre solteiros e casados.

Assinatura do Responsável

Fonte: acervo documental Museu Estadual do Carvão

O banco de dados produzido pela pesquisa, além disso, dialoga com vários campos e modalidades da história: em primeiro lugar, com a história social e a história do trabalho, na medida que foi produzido para uma pesquisa dentro dessa área e buscava sistematizar operários imigrantes; em segundo lugar, com a história serial-quantitativa, na medida que dialogou tanto com a serialização, quanto com a quantificação da documentação; em terceiro lugar, com a história demográfica, visto que trabalha com processos migratórios e populacionais.

4 CONSIDERAÇÕES FINAIS

Buscou-se, nesse trabalho de conclusão de curso, discutir uma experiência de construção de um banco de dados na pesquisa histórica durante o período de iniciação científica. Esse banco de dados buscava sistematizar dados de operários imigrantes vindos do Leste Europeu para a região carbonífera do Rio Grande do Sul após a Segunda Guerra Mundial. Além disso, procurou-se discutir os aspectos teóricos e práticos das várias ferramentas que a informática e a internet disponibilizam aos historiadores no processo do fazer historiográfico.

Inicialmente, constatou-se a eficácia da utilização dessas ferramentas, não só do banco de dados, mas também do fichamento, dos arquivos de leitura de arquivo (PDF, EPUB, DOC, etc.), do *Google Docs*, dos sites de armazenamento de arquivo, etc. Ao longo desta análise, procuramos sugerir como o historiador pode utilizar a informática e a internet como ferramentas no seu trabalho profissional, assim como são os livros e os arquivos e documentais físicos.

Há três contribuições principais desse trabalho e do banco de dados. Em primeiro lugar, o próprio banco de dados. O historiador ou pesquisador que no futuro buscar estudar essa documentação terá a sua disposição além do banco de dados e do fichamento da documentação, uma apresentação de trabalho em um salão de iniciação científica (SIC 2019) e, especialmente, um trabalho de conclusão de curso que reflete e analisa o próprio processo de construção do banco de dados.

Em segundo lugar, esse trabalho contribui para a ampliação e a difusão dos estudos e da utilização do banco de dados entre os historiadores. Há poucos trabalhos, no Brasil, que proponham uma reflexão teórica sobre a utilização de banco de dados entre os historiadores e menos ainda sobre experiências de construção. Tanto que as referências do autor nessa discussão se basearam principalmente em trabalhos recentes de dois autores: Tiago Luís Gil (2015) e Mariana Flores da Cunha Thompson Flores (2015). Nesse contexto, discutir uma experiência concreta de construção de um banco de dados durante a iniciação científica de um graduando no curso de história pode ajudar principalmente a outros alunos, pesquisadores e historiadores que pretendam utilizar ferramentas da informática em suas pesquisas.

Em terceiro lugar, acredito ser importante que os historiadores reflitam sobre a utilização dos tipos diferentes de arquivos de leitura (PDF, EPUB, DOC, etc.) e do *Google Docs*, assim como das redes sociais e dos sites de armazenamento de arquivo. Por serem todas essas ferramentas muito novas, surgidas ou desenvolvidas durante o período da Web 2.0 (principalmente na década de 2010), a análise destas entre os historiadores ainda é incipiente. A geração atual de nativos digitais que frequentou a universidade durante a década de 2010 (a grande maioria de nascidos no final da década de 1980 e durante toda a década de 1990) produziu, sobretudo na segunda metade dessa década, trabalhos inovadores que buscavam dialogar com as redes sociais e com a internet. Contudo, não encontrei análises sobre os usos de documentos digitais na prática, tanto no que diz respeito aos tipos de arquivos (se PDF ou EPUB), quanto em relação aos aparelhos (se computador ou *e-reader*). Nesse sentido, uma das contribuições desse trabalho foi tentar dar conta dessa relação entre a “leitura informática” e o historiador.

Provavelmente a maior contribuição seja subjetiva, visto que o conhecimento e a experiência que o autor adquiriu nesse processo o qualificaram a utilizar o banco de dados como ferramenta não só na atividade como historiador, mas também como professor, na prática da docência, e na própria vida pessoal.

Dois foram os principais limites desse trabalho. Em primeiro lugar, o fato de que o conhecimento teórico e a experiência do autor na construção do banco de dados eram limitados. Como já discutido anteriormente, o autor não possuía, antes de iniciar a iniciação científica, nenhuma experiência mais profunda com a utilização do banco de dados e a pesquisa documental, de modo que foi durante esse período que pela primeira vez tive contato com as discussões teóricas sobre o banco de dados e a crítica documental. Foi a partir daí que conheci utilizar o banco de dados na prática, sistematizando uma determinada documentação.

Em segundo lugar, busquei ao máximo no trabalho debater com historiadores que de alguma forma estudaram com mais profundidade a informática e a internet, de modo que os pesquisadores da área da informática e internet ficaram em segundo plano. Talvez o estudo e discussão com esses autores de outras áreas pudessem ajudar no desenvolvimento ou maturação de alguns conceitos e na própria construção do banco de dados, principalmente no que diz respeito a utilização mais avançada de algumas *softwares* e aparelhos físicos.

Luciano Figueiredo (1997) em seu artigo “História e informática: o uso do computador”, descrevia, nos anos 1990, com precisão, como funcionava um editor de texto. Atualmente, em 2020, os historiadores e pesquisadores já dominam essa ferramenta, pelo menos no que diz respeito às suas funcionalidades básicas. De modo que, o relato detalhado de Figueiredo já não parece mais tão necessário. Com certo otimismo, guardadas as devidas proporções, espero que as descrições e discussões realizadas nesse trabalho de conclusão de curso sobre o banco de dados passe por processo semelhante.

REFERÊNCIAS

Fontes Primárias:

Dossiê Imigrantes - Arquivo (1946-1948). Museu Estadual do Carvão (Arroio dos Ratos).

Bibliografia:

ABREU, Karen Cristina Kraemer. **História e usos da Internet**. BOCC, 2009. Disponível em: <http://www.bocc.ubi.pt/pag/abreu-karen-historia-e-usos-da-internet.pdf>. Acessado em: 20/05/2020.

ALMEIDA, Fábio Chang de. **O historiador e as fontes digitais: uma visão acerca da internet como fonte primária para pesquisas**. Aedos, n. 8, v.3, jan./jun. 2011.

BARROS, José D'Assunção. **A história serial e a história quantitativa no movimento dos Annales**. História Revista (UFG. Impresso), v. 17, p. 203-222, 2012.

BARROS, José D'Assunção. **História Serial, História Quantitativa e História Demográfica: uma breve reflexão crítica**. Revista de C. Humanas, Vol. 11, Nº 1, p. 163-172, jan./jun. 2011.

BLOCH, Marc. **Apologia da História ou O Ofício de Historiador**. Rio de Janeiro: ed. Zahar, 2002.

BURKE, Peter. **A Revolução Francesa da historiografia: a Escola dos Annales 1929-1989**. São Paulo: Editora Universidade Estadual Paulista, 1991.

BURKE, Peter; BRIGGS, Asa. **Uma história social da mídia: de Gutenberg à internet**. Rio de Janeiro: Jorge Zahar, 2006.

CARDOSO, Ciro Flamarion (org.). **Novos domínios da história**. Rio de Janeiro: Elsevier, 2012.

CARDOSO, Ciro Flamarion. **No Limiar do Século XXI**. Tempo, Rio de Janeiro, vol. 1, nº. 2, 1996, p. 7-30.

CARDOSO, Ciro Flamarion; BRIGNOLI, Héctor Pérez. **Os métodos da história:** introdução aos problemas, métodos e técnicas da história demográfica, econômica e social. São Paulo: Graal, 2002.

CARDOSO, Ciro Flamarion; VAINFAS, Ronaldo (orgs.). **Domínios da história:** ensaios de teoria e metodologia. Rio de Janeiro: Campus, 1997.

CARNEIRO, Anita Natividade. **Caminhos da Ditadura em Porto Alegre:** Ensino de História através da tecnologia digital. UFRGS, 2018. 67 f. Trabalho de conclusão de curso- Instituto de Filosofia e Ciências Humanas, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2018.

CARVALHO, Bruno Leal Pasto de. **Faça aqui o seu login:** os historiadores, os computadores e as redes sociais online. Revista História Hoje, v. 3, nº 5, p. 165-188 - 2014.

CARVALHO, Priscila Viviane de. **Estudo analítico sobre a estrutura de livros digitais nos formatos PDF e EPUB.** Universidade Federal do Piauí, 2017. 62 f. Dissertação (Mestrado em Letras) - Centro de ciências humanas e letras, Universidade Federal do Piauí, Teresina, 2017.

CONSTANTE, Bruno Erbe. **O golpe tuitado:** uma análise dos discursos produzidos no Twitter pelas principais lideranças do Golpe de 2016 (junho 2013-dezembro 2015). UFRGS, 2019. 78 f. Trabalho de conclusão de curso- Instituto de Filosofia e Ciências Humanas, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2019.

CORRÊA, Dênis Renan Corrêa; SCHULZ, Marcos. **Conhecimento histórico e Internet:** uma conversa com Carlo Ginzburg. Aedos. Num.8, vol. 3, Janeiro - Junho 2011.

FARINATTI, Luís Augusto. **Construção de séries e micro-análise:** notas sobre o tratamento de fontes para a história social. Anos 90, Porto Alegre, v. 15 n. 28, p. 57-72, dez. 2008.

FIGUEIREDO, Luciano. História e Informática: o uso do computador. In: CARDOSO, Ciro Flamarion; VAINFAS, Ronaldo(orgs.). **Domínios da história: ensaios de teoria e metodologia.** Rio de Janeiro: Campus, 1997.

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

GIL, Tiago Luís; BARLETA, Leonardo. **Formas alternativas de visualização de dados na área de História**: algumas notas de pesquisa. Rev. hist. (São Paulo), n. 173, p. 427-455, jul.-dez., 2015.

LAITANO, Bruno Grigoletti. **Postando o passado**: a difusão da memória da ditadura civil-militar brasileira na internet através do canal do YouTube da Comissão Nacional da Verdade. UFRGS, 2018. 82 f. Trabalho de conclusão de curso- Instituto de Filosofia e Ciências Humanas, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2018.

LE GOFF, Jacques. **História e memória**. São Paulo: Editora da UNESP, 1992.

LÉVY, Pierre. **Cibercultura**. São Paulo: Editora 34, 1999

LUCCHESI, A. **Por um debate sobre História e Historiografia Digital**. Boletim Historiar, v. 1, p. 45-57, 2014.

MIRANDA, André. **A internet não veio para salvar o mundo**: entrevista com José Saramago. O Globo, 2009. Disponível em: <<http://www.observatoriodaimprensa.com.br/armazem-literario/a-internet-nao-veio-para-salvar-o-mundo/>>. Acessado em 29/09/2020.

OLIVEIRA, Nucia Alexandra Silva de. **História e internet**: conexões possíveis. Revista Tempo e Argumento, vol. 6, núm. 12, maio-agosto, 2014, pp. 23-53.

PRENSKY, Marc. **Nativo Digitais, Imigrantes Digitais**. NCB University Press, Vol. 9 No. 5, Outubro 2001. Disponível em <https://colegiongeracao.com.br/novageracao/2_intencoes/nativos.pdf>. Acessado em: 04/05/2020.

SPERANZA, Clarice Gontarski. **Cavando direitos**: as leis trabalhistas e os conflitos entre trabalhadores e patrões nas minas do Rio Grande do Sul nos anos 40 e 50. UFRGS, 2012. 272 f. Tese (Doutorado em História) - Programa de Pós-graduação em História, Instituto de Filosofia e Ciências Humanas, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2012.

SPERANZA, Clarice Gontarski. **Solidariedade e conflito**: experiências e identidades entre mineiros de carvão no Rio Grande do Sul (1850-1950). UFRGS, 2018. Projeto de Pesquisa - Programa de Pós-graduação em História, Instituto de Filosofia e Ciências Humanas, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2018.

TAVARES, Célia. História e Informática. In: CARDOSO, Ciro Flamarion (org.). **Novos domínios da história**. Rio de Janeiro: Elsevier, 2012.

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

VELLOSO, Fernando. **Informática: conceitos básicos**. Rio de Janeiro: Elsevier, 2014.

WITKOWSKI, Alexsandro; MORIGI, Valdir José. **Território e identidade na região carbonífera do Baixo Jacuí/RS: o acervo documental de mineração do arquivo histórico do museu estadual do carvão**. *ParaOnde!?*, Porto Alegre, v.12, n.1, p.191-210, 2019.

ZEIDAN ARAÚJO, George. **Ler, pesquisar e escrever história em tempos de internet: desafios e possibilidades**. *Revista Tempo e Argumento*, vol. 6, núm. 12, maio-agosto, 2014, pp. 151-164.