

Nascido Digital: O Centro Simão Mathias de Estudos em História da Ciência – CESIMA

Resumo

Em 2018 foi colocada online a biblioteca CESIMA Digital, com documentos fundamentais para pesquisas em história da ciência em diferentes épocas e lugares. Até a data da presente publicação, a biblioteca tem sido acessada por pesquisadores de mais de 120 países. Tal lançamento, no entanto, representa a culminação de quase 30 anos de esforços pioneiros na área conhecida como Humanidades Digitais. Nesse artigo, fazemos uma breve memória do contexto histórico, científico e institucional que levou à criação da biblioteca. Na sequência, discorremos acerca de problemas particulares dos acervos especificamente desenhados para pesquisas em história da ciência e as tentativas de solução que estamos desenvolvendo.

Palavras-chave: Bases de dados digitais, história da ciência, Brasil, CESIMA.

Abstract

In 2018 we launched the online library CESIMA Digital, which comprises sources necessary to research in the history of science in different places and times. To this day, investigators from more than 120 countries have accessed our library. However, this milestone is just the culmination of almost 30 years of pioneering work in the so-called Digital Humanities. Here we briefly evocate the historical, scientific and institutional context within which we developed CESIMA Digital. Then we discuss specific problems exhibited by databases explicitly designed for studies in the history of science, as well as the solutions we have been attempting.

Keywords: Digital databases, history of Science, Brazil, CESIMA.

Junto à geração conhecida como 'digital', que nasceu em finais do século XX, nascia também digital a biblioteca do CESIMA (Centro Simão Mathias para estudos em História da Ciência),¹ em 1994.

¹ Graças à sugestão da Profa M. Helena Roxo Beltran temos o nome deliciosamente sonoro do CESIMA, ao invés das impronunciáveis sopas de letras das siglas; também devemos a ela a sugestão da belíssima imagem do 'cosmógrafo' como nosso símbolo, cf. ao final do artigo.



Figura 1: Inauguração CESIMA (1994)

Porém a gestação desta biblioteca durou muito, mas muito mesmo, pois, na verdade, teve início nos anos 80!!!! Originalmente nossa proposta era superar um enorme entrave, sempre lembrado por nosso patrono, o querido mestre Simão Mathias: o difícil acesso a originais, antigos ou raros, tão necessários a quem faz história da ciência de outras regiões e épocas. Lembremos que, ao iniciar esse processo, não havia nada parecido com *Google Books*, *Internet Archive* ou algo do gênero.²

A solução, aliás bem conhecida por mestre Mathias, era a compra, em geral pessoalmente no exterior,³ de obras caríssimas, por serem raras ou antigas; ou de seus microfilmes, sempre difíceis de encontrar ou feitos sob encomenda,⁴ e, por isso, também, caríssimos. Havia também a chamada solução 'braçal'. Algo que significava dias inteiros perdidos em grandes bibliotecas internacionais, pedindo microfilmagens – um processo cheio de entraves e favores pessoais – ou, no caso de obras mais

² *Internet Archive*, plataforma conhecida por conter um enorme volume de obras mais antigas e, portanto, sem as restrições de direitos autorais, só foi lançada em 1996, tornando-se melhor estruturada a partir de 1999; enquanto *Google Books* só surgiu em 2006 e com as conhecidas restrições de acesso às obras, especialmente, para certos lugares, como o Brasil, além de mais recentes exclusões, devidas a direitos autorais e de imagem.

³ Lembremos, mais uma vez, que nosso processo teve início na década de 1980 e plataformas, como a *AbeBooks* ou *Amazon* surgem como pequenos e quase desconhecidos sites somente em meados da década de 90, tornando-se internacionalmente viáveis somente a partir dos anos 2000.

⁴ Na época, um microfilme, entre 100 e 150 fotogramas, mas feito sob encomenda, a partir de originais, podia custar, em média, US\$500.

recentes, tirando milhares de páginas xerox, enviadas por correio, como carga, com demora de 2 até 6 meses para chegar ao Brasil.⁵ Mas, a recompensa tão esperada por esse enorme esforço (até mesmo físico), quase nunca vinha de imediato. Uma vez que, além da demora tradicional dos correios, com frequência a sua chegada gerava pilhas intermináveis de material duro de separar e catalogar, quando não com páginas faltantes etc., etc...

Novas esperanças surgiram e nosso pesadelo pareceu começar a se desanuviar quando, em finais dos 1980 e inícios dos 1990, foi lançado, pelo *Établissement Public de la Bibliothèque de France* (EPBF), um projeto para digitalizar enormes acervos de obras, visando disponibilizá-las online ao público das salas de leitura da Biblioteca Nacional da França (BnF). E mais ainda, esse que já era um sonho de consumo, de qualquer acadêmico ou rato de biblioteca, tornou-se real em todo lugar onde houvesse internet, quando em 1997 surgiu a hoje bem conhecida *Gallica*.⁶ Embora, não seja demais lembrar que, no início da *Gallica* e de outros sites surgidos aos poucos, baixar uma obra podia tomar um dia inteiro e falhar ao final, devido ao tipo de internet discada que existia na época...

Muito embora, antes de 1994, longe ainda de sonhar com a futura *Gallica*, nosso objeto de desejo era o projeto inicial EPBF. Que mais alguém poderia desejar do que salas especialmente preparadas para leitura online, mesmo que presencial, de montanhas de obras e documentos digitalizados?⁷ Então, guardadas as devidas proporções, como poderíamos criar nosso pequeno projeto EPBF? Não seria maravilhoso digitalizar nosso sempre crescente material de pesquisa – microfiches, microfichas ou livros, obtidos a duras penas, ao longo de anos, por cada um de nós – formando uma grande base de dados para compartilhar em trabalho conjunto? Não seria ainda mais maravilhoso disponibilizar essas coleções *de e para* pesquisa, em rede local, para estudantes e outros colegas? Ou seja, toda uma série de perguntas que, fora do ambiente de nossa equipe de pesquisa, gerou muitas risadas, quando não perplexidade! Afinal, conforme constava no próprio projeto EPBF, o ponto de partida era a, então recém-turbinada, máquina digitalizadora (scanner). Um equipamento que, se dizia, talvez existisse por aqui apenas na ilha de filmagem da Globo e poderia custar até US\$ 500.000.⁸

⁵ No caso dos EUA a solução mais em conta era o conhecido Mbag (My bag), com até 100 kg, carga difícil de levar e trazer dos correios; no caso europeu, dependendo do país, a carga dividida em múltiplos pacotes também multiplicava o preço do envio e nem sempre chegava completa ao destino.

⁶ <https://gallica.bnf.fr/edit/und/a-propos>

⁷ Na época, grandes bibliotecas e centros internacionais colocavam, no máximo, seus catálogos online, muitas vezes apenas para consulta em terminais localizados em seus próprios espaços.

⁸ Embora a história de tal equipamento seja anterior, os desenvolvimentos que permitiram seu uso em grandes coleções foram em finais da década de 1980 e início dos 1990, cf. M. Lesk, "A personal history of digital libraries", *Library Hi Tech* 30, nº4 (2012): 592-603; o valor referido incluía a importação para o Brasil, sempre potencializando o custo de máquina que, na época, já pressupunha um alto investimento.

Desistir, porém, nunca foi palavra em nosso dicionário e com apoio de um projeto piloto FAPESP,⁹ e muito esforço, conseguimos dar forma a um primeiro núcleo de uma biblioteca digital, pedra basilar na fundação do CESIMA. Foram meses adaptando uma maquineta comumente utilizada em medicina para verificação de microfilmes de raio-X, como nossa primeira digitalizadora. Um processo que incluiu até o uso de lixa de unhas para tornar o suporte da máquina compatível aos nossos microfilmes.¹⁰ Ainda assim, como o programa para funcionamento da máquina também precisou de muitas adaptações e criatividade da equipe, cada frame podia demorar horas para ser digitalizado!



Figura 2: Digitalizadora criada pela equipe do CESIMA a partir de escâner de microfilmes

Mas, o processo e o produto fizeram enorme sucesso e foram divulgados por redes rádios e televisões importantes país afora, despertando até o interesse de grandes empresas em comprar o projeto – obviamente, para evitar futura concorrência acadêmica... Não vendemos, justamente pensando nisso, para assim garantir que novos equipamentos se tornassem acessíveis a acadêmicos como nós mesmos.

⁹ FAPESP, Processo no. 94/3162-9.

¹⁰ Como seria de esperar, microfilmes e microfichas eram os suportes físicos passíveis de digitalização em nossa maquineta adaptada.



Figura 3: Inauguração da Biblioteca Digital CESIMA na mídia

De toda forma, através desse empenho inicial, as risadas pararam de acontecer, pois havíamos criado a primeira biblioteca digital universitária na América Latina. Mais ainda, por ser dedicada à história da ciência e a partir da própria documentação pesquisada por cada um dos membros de nossa equipe, algo que, como veremos adiante, tornou-se uma marca de distinção até hoje. Porém, após a boa acolhida inicial, várias etapas tiveram que ser cumpridas, ao longo dos anos. A primeira delas foi adquirir, desta vez através de projeto temático FAPESP,¹¹ nossa primeira digitalizadora de fato. Também usada para material impresso, essa máquina – embora ainda cara – passaria a custar, na virada do século XXI, 1/10 daquelas usadas no começo dos 1990.

¹¹ Desde o ano 2000 até o início de 2019, o CESIMA foi apoiado por grandes projetos Temáticos, bem como de outras agências como CNPq e a CAPES.



Figura 4: Primeira digitalizadora profissional comprada pelo CESIMA (2000)

Além disso, a proposta de nossa biblioteca digital também passou a incluir documentos pertencentes a centros e bibliotecas no exterior que, apesar de ciosos de suas coleções, colaboraram conosco, pois sabiam da importância para nossa pesquisa. Da mesma forma, conseguimos negociar outras obras de nosso interesse, tão logo companhias começaram a digitalizar profissionalmente coleções de documentos. Foi assim como, na primeira década dos anos 2000, nossa biblioteca já contava com cerca de 30.000 obras relativas as ou derivadas das pesquisas de nossa equipe, variando desde mapas e outras imagens até manuscritos, passando por livros, dicionários, diários, cadernos de laboratório e correspondências de variadas épocas e lugares.



Figura 5: Exemplo dos materiais para digitalização obtidos pelo CESIMA

Apesar disso, dois problemas permaneciam em nosso horizonte. Em primeiro lugar, a questão a ser solucionada era como transformar nossa biblioteca digital em biblioteca online. Uma vez que, dessa maneira poderia ser acessada não apenas de forma remota, por nossa equipe e nossos estudantes ou colaboradores associados, mas por pesquisadores de outros lugares. Lembremos que propostas de *open libraries* bastante completas ou imensas – como almejadas pela ‘culturonomics’,¹² de Jean-Baptiste Michel e do *Google Books* – nunca conseguiram ir adiante, particularmente devido a direitos autorais e outras questões. Nesse caso, a solução adotada foi seguir a norma de outras bibliotecas de origem física que apresentam coleções digitais. Ou seja, oferecer uma espécie de ‘chave’ para garantir o acesso apenas com permissão e responsabilizando os usuários pelo manejo dos documentos. Assim, colocada on-line a partir de 2018,¹³ a Biblioteca CESIMA *Digital*, conta com quase 12.000 acessos, desde mais de 120 países.

Enquanto isso, a documentação digitalizada, como sempre, visando à pesquisa da equipe, continuou obviamente aumentando e nos levou a segunda questão: como classificá-la, mantendo as características organizacionais que lhe seriam próprias? A resposta vem sendo formulada em duas etapas, coligadas de várias maneiras. Primeiramente, desde o início da primeira década de 2000, a equipe

¹² Jean-Baptiste Michel et al, “Quantitative Analysis of Culture Using Millions of Digitized Books,” *Science* 331, nº 6014 (2011): 176-82; esse artigo foi uma espécie de manifesto assinado por autores dos mais variados lugares e formações, incluindo membros do Google.

¹³ Graças ao trabalho dedicado e contínuo, especialmente, do Prof. J. L. Goldfarb, do Dr. O. Souza, de membros do SOBEK (UFRGS) e da DTI (PUC-SP), nossa Biblioteca foi colocada online e continua em constante desenvolvimento; cf. <http://cesimadigital.pucsp.br/>

do CESIMA debruçou-se sobre pesquisas relativas à organização e classificação do conhecimento, contando com especialistas de diversas instituições e órgãos nacionais e internacionais, incluindo a Comissão de Documentação e Bibliografia da *International Union for History of Science and Technology*. Houve, e ainda há, uma produção intensa nesse sentido, promovida através de workshops, conferências e congressos nesse campo. Assim, foi possível verificar – após estudos aprofundados sobre as diferentes formas de classificação e organização, utilizadas ao longo do tempo – que nenhuma destas se adequava devidamente à documentação em ambiente virtual. Os motivos para tanto são variados: desde as divisões obsoletas da maioria das classificações, passando pela interface crescente entre as várias áreas de conhecimento, até a impertinência da espacialidade em bases digitais. De todo modo, não seria plausível discutir aqui esses quase 20 anos de estudos de nossa equipe.¹⁴

No entanto, a necessidade de resposta levou a uma segunda etapa. Desta vez, um mergulho nas pesquisas sobre as chamadas Humanidades Digitais, voltadas, justamente, ao trabalho com grandes corpora digitais. Lembremos que, especialmente para a história da ciência e campos coligados, a identificação das mudanças nos ramos do conhecimento quase sempre apresenta vínculos com a transformação de seus conceitos, ao longo do tempo. Ao ser uma pesquisa desafiadora, sobretudo quando individual, a equipe do CESIMA vem trabalhando de forma conjunta, há muitos anos, nessa proposta. Por exemplo, nossos estudos históricos sobre diferentes tipos de materiais, tais como balsamos, corantes, metais, etc., demonstram que por detrás de uma mesma palavra pode haver transformações conceituais e saltos entre variados campos do conhecimento, conforme a época e local.¹⁵ Trata-se, porém, de uma análise feita de modo quase artesanal e por sobre uma documentação imensa, pois junta diversas pesquisas individuais da equipe. Mas, trata-se também de uma pesquisa considerada, em princípio, possível de realizar através da chamada mineração computacional em grandes corpora. Algo que atraiu especialistas em Humanidades Digitais, tanto pela natureza do trabalho, quanto pela singular formação da própria biblioteca CESIMA *Digital* – feita *por* pesquisas e *para* a pesquisa. Tomada como modelo para desenvolver e aplicar instrumentos de busca e outras ferramentas computacionais, nossa biblioteca vem sendo objeto de estudo, especialmente, para duas equipes voltadas as Humanidades Digitais.

Uma delas é a equipe liderada pelo Prof. Eliseu Reategui, da Universidade Federal do Rio Grande do Sul, que desenvolveu um instrumento de mineração chamado *SOBEK*.¹⁶ Baseado no sistema conhecido como *text analytics*, tal instrumento de busca permite a identificação de caracteres em textos

¹⁴ Esse longo processo gerou e foi sustentado por uma série enorme de produções e eventos, cuja enumeração seria demasiada aqui, mas sugere-se aos interessados consulta aos currículos Lattes dos pesquisadores do CESIMA.

¹⁵ Por representar um período de mais de três décadas, o volume das produções é ainda maior do que no caso da nota acima e, novamente, sugere-se consulta aos currículos de nossa equipe.

¹⁶ <http://sobek.ufrgs.br/index-en.html>

digitais,¹⁷ e a partir disto compõe as famosas nuvens de palavras, muito usadas no ensino. O *SOBEK* foi o primeiro instrumento desse tipo criado no Brasil e tem sido usado, principalmente, para facilitar a leitura e compreensão de textos, com grande sucesso. No caso da biblioteca *CESIMA Digital*, em que a equipe do *SOBEK* vem trabalhando já há algum tempo, a adaptação desse minerador produziu várias nuvens de palavras que possibilitam identificar a prevalência de certas palavras/conceitos numa obra e a sua comparação com outras. Um resultado bastante útil, pois, embora ainda demande muito desenvolvimento, ajuda a verificar sintaticamente palavras que envolvem conceitos, enquanto a identificação destes últimos não seja possível de forma direta, através de equipamentos que, no futuro, permitam análises semânticas dos textos.

Uma outra equipe, em cujo projeto internacional o *CESIMA* tem o privilégio de participar, pertence ao Consorcio *Digital History and Philosophy of Science*. Tal consorcio internacional desenvolve, desde 2017, um projeto chamado *Research Coordination Network* que, financiado pela *Nacional Science Foundation*, se dedica à integração de bibliotecas digitais em e para a história da ciência. Como também se dedica ao desenvolvimento de instrumentos computacionais que possibilitem tratar a temporalidade e a contingência inerentes à documentação de nosso campo de estudos.¹⁸ Como esperado, novamente um dos pontos centrais de nossa participação nesse grande consorcio é a biblioteca *CESIMA Digital*, devido às possibilidades modelares que oferece quanto à natureza de sua formação e desenvolvimento *em e para* a pesquisa da história da ciência.

Existe, portanto, em ambos os casos assinalados, uma articulação harmônica para o desenvolvimento de estudo em Humanidades Digitais. Por um lado, contamos com aqueles que elaboram novos instrumentos de buscas e similares (cientistas da computação e da informação), enquanto por outro com aqueles que podem nutrir e oferecer ideias para novos usos de tais instrumentos computacionais (historiadores da ciência). Continuamos, assim, desenvolvendo nossas pesquisas há mais de um quarto de século, através da *CESIMA Digital*, que agora conta também com uma equipe de alunos e egressos, de nosso Programa de Pós-Graduação em História da Ciência, treinados para alimentar e levantar seus dados.



SOBRE AS AUTORAS:

¹⁷ Ou seja, imagens digitalizadas transformadas em textos digitais, através de softwares, como o OCR, assim permitindo o reconhecimento de seus caracteres.

¹⁸<https://digitalhps.org>

https://nsf.gov/awardsearch/showAward?AWD_ID=1656284&HistoricalAwards=false

Ana Maria Alfonso-Goldfarb

aagold@dialdata.com.br

Silvia Waise

silviawaise15@gmail.com

Márcia Ferraz

mhferraz@me.com

Centro Simão Mathias de História da Ciência

Programa de Estudos Pós-Graduados em História da Ciência