



ESFORÇO PARA PERPETUAR DOCUMENTOS AUTÊNTICOS

Arquivo Público do Estado de São Paulo investe em infraestrutura computacional para preservar seu extenso acervo digital

Fabrcio Marques

O Arquivo Público do Estado de São Paulo (Apesp) lançou em novembro uma plataforma computacional que permitirá a pesquisadores e cidadãos um acesso mais refinado e abrangente ao acervo digital da instituição. O patrimônio é composto atualmente por mais de 700 mil documentos produzidos por órgãos do governo estadual ou de interesse histórico, como coleções de ex-governadores ou de personalidades que participaram da vida política e cultural paulista, de instituições como o Instituto Histórico e Geográfico de São Paulo e a Maternidade de São Paulo, de jornais e partidos políticos.

Um investimento de R\$ 6,9 milhões na ampliação da capacidade de armazenamento e processamento de dados, na compra de equipamentos e na melhoria da rede elétrica deu forma ao Arquivo Público Digital, que entrou no ar em versão beta e pode ser acessado em atom.arquivoestado.sp.gov.br. Uma plataforma única vai concentrar os acervos digitais já disponíveis e receber novas coleções. Um próximo passo será ampliar o volume de documentos disponíveis e armazená-los de forma que possam ser consultados facilmente.

Uma das coleções mais procuradas é a do Departamento Estadual de Ordem Política e Social (Deops), incorporado ao Arquivo na década de 1990, com mais de

Scanner de digitalização do Apesp (à esq.) e documentos do conselho estadual de patrimônio histórico, o Condephaat, incorporados ao núcleo cartográfico da instituição (abaixo)

300 mil fichas disponíveis para consulta (ver Pesquisa FAPESP nº 207). Parte do trabalho de digitalização foi feita em projetos de pesquisa em temas como imigração, repressão política e a presença de judeus no Brasil, alguns deles apoiados pela FAPESP, como aqueles capitaneados pelas historiadoras Maria Luiza Tucci Carneiro e Maria Aparecida Aquino, da Universidade de São Paulo (USP), entre o final dos anos 1990 e os primeiros anos deste século XXI. Agora, mais 2,5 milhões de fichas do Deops ganharão cópias eletrônicas.

Um novo repositório criado pelo Arquivo foi preparado para receber, além de reproduções de registros em papel, também os documentos chamados nato-digitais, que só existem virtualmente e vêm sendo produzidos por órgãos do Executivo paulista, a exemplo de processos eletrônicos de repartições públicas e informações fiscais. Em 2019, o governo de São Paulo criou o programa SP Sem Papel, que regula a gestão e o controle de documentos digitais – eles passaram a receber uma classificação no momento de sua produção. “Dos registros produzidos pelo Poder Executivo, 20% são históricos e têm de ser identificados e preservados, 50% podem ser eliminados em tempo curto, de até 2 anos, e 30% devem ficar disponíveis por prazos mais longos, que podem ser de 5 a 80 anos. É o Arquivo que faz a gestão de tudo isso”, explica o historiador Thiago Nicodemo, coordenador do Apesp e professor da Universidade Estadual de Campinas (Unicamp). Os prazos para disponibilização são definidos pelas chamadas tabelas de temporalidade, elaboradas pelo Arquivo com representantes dos órgãos públicos.

Parte do investimento foi aplicada na criação ou aperfeiçoamento de mais de 70 softwares de uso livre que estão ajudando a reorganizar os acervos. Os programas foram desenvolvidos na forma de *script* e permitem, com um único comando, a execução de um conjunto

extenso de tarefas repetitivas. Um dos softwares permite mudar o nome, de forma automática, de grandes volumes de arquivos, acelerando sua catalogação para torná-los mais acessíveis.

De modo geral, as coleções do Arquivo foram digitalizadas em momentos distintos e cada uma está disponível em um site específico. A nova plataforma permitirá rastrear tudo o que existe sobre um determinado assunto, uma vez que a catalogação seguirá um padrão comum. “Uma pesquisa sobre as revoluções de 1924 ou de 1932, por exemplo, dará acesso a exemplares dos jornais dos batalhões, a cartas e telegramas que estão no acervo do Instituto Histórico e Geográfico de São Paulo, a um grande processo do Tribunal de Justiça de São Paulo, que contém fotos, inclusive. Integração semelhante deverá ocorrer com acervos relativos às ferrovias, entre outros temas”, afirma o sociólogo e arquivista Carlos Menegozzo, diretor do Centro de Difusão e Apoio à Pesquisa do Apesp. Nem todas as coleções estarão disponíveis na internet. Recentemente incorporados ao Arquivo, os prontuários do Hospital Psiquiátrico do Juquery contêm informações sensíveis e não irão para a plataforma, embora possam ser consultados presencialmente.

Claudia Bauzer Medeiros, pesquisadora do Instituto de Computação da Uni-

camp e membro da coordenação do Programa FAPESP de Pesquisa em eScience e Data Science, diz que a iniciativa do Arquivo chama a atenção pelo volume de informações e pela metodologia rigorosa adotada para organizá-las. “Há uma quantidade enorme de documentos de todos os tipos que precisam ser classificados usando normas internacionais que garantam a preservação de longo prazo. Ao mesmo tempo, eles precisam ser disponibilizados de forma eficiente, inclusive pela internet, sem desrespeitar eventuais direitos de privacidade e confidencialidade de pessoas mencionadas.” Segundo ela, repositórios criados por universidades e outras instituições enfrentam desafios semelhantes para armazenar e preservar os dados de pesquisa. “Não basta guardá-los. Eles precisam estar acompanhados de uma documentação que siga padrões internacionais, para que sejam facilmente identificados e possam ser cruzados com outras bases”, explica Medeiros. “A disponibilização de dados de forma aberta e bem documentada faz parte das práticas da ciência aberta, hoje preconizada como parte das boas práticas de pesquisa em todo o mundo. A Assembleia Geral da Unesco já votou em novembro de 2021 um documento sobre isso, com recomendações a todos os países-membro. A FAPESP é pioneira em iniciativas de ciência aberta na América Latina.”



LISTA GERAL DE PASSAGEIROS

Para o porto de SANTOS (Estado de S. Paulo-Brasil)

Matricula	Procedencia	Nome do Capitão						Nome do Medico		Tonelagem		
		SEXO		ESTADO				Profissionalidade	Profissão	Religião		Sede
EXTENSO	Parentesco com o chefe da família	1	2	3	4	5	6			7	8	

Listas de bordo de imigrantes que chegaram a São Paulo no século passado, em fase de digitalização

A manutenção das coleções impôs um aprendizado aos gestores do Arquivo. “No início, o acervo digital era armazenado em pastas, sem seguir padrões e critérios técnicos predefinidos”, diz a socióloga Camila Brandi, diretora técnica do Departamento de Preservação e Difusão do Acervo do Apesp. “Ficou claro que precisaríamos de soluções tecnológicas e uma política de governança capazes de garantir que os documentos fossem bem preservados e sua autenticidade aferida. Eles precisam estar em um formato de preservação e os procedimentos executados devem ser controlados e documentados. Só assim será possível auditar para sabermos se alguém entrou naquela pasta, se apagou algo sem querer ou por querer, e reconstituir suas características originais.”

Já no caso dos registros em papel digitalizados, a exemplo dos exemplares de jornais como o *Última Hora*, os prontuários relativos à história da hanseníase no estado de São Paulo ou as fichas da Delegacia de Estrangeiros, o desafio do Arquivo é tratá-los e disponibilizá-los de forma que possam se manter íntegros por muito tempo e sejam acessíveis. Algumas coleções, por exemplo, foram processadas utilizando ferramentas de Reconhecimento Óptico de Caracteres (OCR), que permitem a busca por palavras em seu

conteúdo. Outras não foram – nesses casos, é muito mais difícil para o usuário ou o pesquisador encontrar e extrair a informação que está procurando. “Após mais de duas décadas de projetos de digitalização, sabemos mais sobre riscos e benefícios da tecnologia”, diz Brandi. “Todo mundo já perdeu um pendrive com informações importantes ou uma foto de família que foi corrompida e ficou inutilizada. Com as instituições, isso não foi diferente. Elas aprenderam que não basta digitalizar acervos. É fundamental cuidar deles de modo perene para que mantenham sua autenticidade e integridade.” De acordo com Thiago Nicodemo, esse aprendizado está ajudando a discutir e a refletir melhor sobre o que deve ou não ser convertido para o formato eletrônico. “Já gastamos muito dinheiro no Brasil e no mundo digitalizando acervos e vemos agora que é necessário fazer isso com muito mais critério.”

A historiadora Ieda Bernardes, diretora do Departamento de Gestão do Sistema de Arquivos do Estado de São Paulo, ressalta que a transformação do Apesp tenta evitar a repetição de erros do passado. “A ideia de que a tecnologia ajudaria a organizar a bagunça dos acervos não é nova. Nos anos 1950, a microfilmagem surgiu para resolver os problemas de acesso, mas eles continuaram. A digitalização promovida nos últimos 20 anos seguiu no mesmo caminho. O problema é que os acervos foram reproduzidos da

mesma forma desorganizada das versões em papel, como um espelho do mundo físico”, afirma. Segundo ela, alguns gestores cometeram o equívoco de descartar arquivos físicos, substituindo-os por suas versões eletrônicas, que acabaram se corrompendo ou ficaram inacessíveis pela obsolescência dos suportes tecnológicos.

Bernardes explica que uma política para preservar documentos digitais pode trazer ganhos de eficiência, mas custa bem mais caro do que manter o arquivo em papel. “Quem pensa que vai economizar está equivocado. É necessário financiar continuamente a atualização tecnológica, já que computadores e softwares ficam ultrapassados em poucos anos, e monitorar o estado dos documentos”, diz.

Medeiros, da Unicamp, conta que parte significativa do orçamento da Agência Espacial Norte-Americana, a Nasa, é gasta com a preservação dos informações obtidas em pesquisas e viagens espaciais. “Há alguns anos conversei com um diretor da Nasa que relatou as dificuldades para preservar esse patrimônio e contou que 30% dos recursos da agência são aplicados em dados. É preciso investir em atualização das tecnologias e principalmente em recursos humanos especializados.” De acordo com ela, trata-se de um investimento essencial. “Pode não custar barato, mas, se perdemos os registros, o prejuízo é enorme. Jogamos fora um tesouro.” ■