

ACESSO E USO DE ARQUIVOS DA WEB: ANÁLISE DOS ASPECTOS ÉTICOS E LEGAIS

Lúcia Andréia Nunes de Oliveira Nunes (Universidade Federal do
Rio Grande do Sul)

1 INTRODUÇÃO

Com os avanços das tecnologias de Internet e o desenvolvimento rápido da *web*, a quantidade de conteúdo disponibilizado nesse ambiente aumentou significativamente, se constituindo em uma importante fonte de pesquisa. Mas com a mesma velocidade que a informação é produzida, ela se perde e se apaga e, devido a esta característica efêmera, a necessidade de arquivamento da *web* torna-se reconhecida.

O arquivamento da *web* pode ser definido como um “processo que compreende coletar, armazenar e disponibilizar a informação retrospectiva da *World Wide Web* para futuros pesquisadores” (ROCKEMBACH, 2017, p. 9), devendo ser pensado de forma sistêmica, o que inclui o estabelecimento de políticas de arquivamento (ROCKEMBACH; PAVÃO, 2018).

Baseado no modelo de ciclo de vida do arquivamento da *web*, que apresenta as fases do arquivamento, destaca-se a etapa de acesso e uso da informação, que envolve aspectos sensíveis como privacidade e direitos autorais.

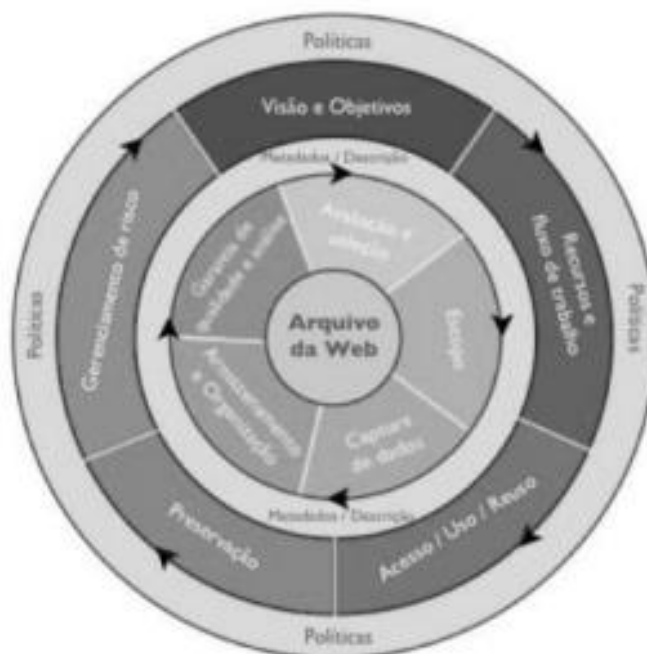
A questão que norteou a pesquisa foi: como iniciativas de arquivamento da *web* fornecem acesso às informações por elas preservadas considerando os aspectos éticos e legais? O objetivo foi analisar os aspectos éticos e legais pertinentes ao acesso e uso das informações de *websites* arquivados a partir dos documentos termos de uso e das políticas de privacidade disponibilizadas nas plataformas digitais de iniciativas de arquivamento da *web*.

Muitas iniciativas de arquivamento da *web* vêm surgindo pelo mundo com o intuito de preservar o conteúdo disponibilizado nesse ambiente e, tendo em vista que a temática ainda é recente no cenário nacional, analisar as formas de acesso e uso estabelecidas por iniciativas de arquivamento da *web*, considerando os aspectos legais e éticos, poderá contribuir para os estudos sobre o arquivamento da *web* brasileira.

2 ASPECTOS ÉTICOS E LEGAIS NO ACESSO E USO DE ARQUIVOS DA *WEB*

Desde a sua criação, a *web* evoluiu de forma muito rápida, ganhando novas ferramentas que a tornaram mais dinâmica. Essa natureza dinâmica e efêmera, traz a necessidade de refletir sobre o armazenamento dos conteúdos disponibilizados nesse ambiente, impondo desafios para a sua preservação; o ponto de partida para enfrentar esses desafios é o planejamento, a metodologia e as plataformas para coletar, armazenar e disponibilizar o conteúdo, etapas que correspondem ao ciclo de vida do arquivamento da *web* (ROCKEMBACH, 2018 e 2019).

O ciclo de vida do arquivamento da *web* é um modelo que apresenta as fases do arquivamento (FIGURA 1); dentre as etapas individuais, está o acesso-uso-reuso da informação, onde é determinado como será dado o acesso à informação arquivada e como o uso será monitorado (BRAGG; HANNA, 2013, p. 28). Essa etapa, envolve aspectos sensíveis das políticas de arquivamento da *web*, como legislação sobre privacidade, proteção de dados pessoais e direitos autorais (BRAYNER, 2019).

Figura 1 – Ciclo de vida do arquivamento da *web*

Fonte: Extraído de Bragg e Hanna (2013 *apud* ROCKEMBACH, 2018, p. 13).

A internet favoreceu o surgimento de um espaço interativo de acesso e compartilhamento de informações, alterando o processo de produção, disseminação e uso da informação. As possibilidades de interação, a superação das distâncias geográficas, a produção de informação descentralizada, o compartilhamento de conteúdo, são fatores que influenciam a nova organização social e redefinem novos padrões culturais, políticos e éticos.

2.1 Contexto da pesquisa e procedimentos metodológicos

O *International Internet Preservation Consortium*, fundado em 2003, congrega 56 instituições que trabalham de forma colaborativa na preservação de conteúdo da *web*. Dentre essas instituições, foram selecionadas 19 plataformas digitais para o desenvolvimento da pesquisa, dentre as quais: arquivos e bibliotecas, nacionais e universitárias, e organizações sem fins lucrativos, distribuídas na América do Norte, Europa, Oceania, América do Sul e Ásia.

Trata-se de uma pesquisa de natureza qualitativa, com finalidade exploratória e descritiva. Um dos procedimentos metodológicos utilizados foi a pesquisa bibliográfica, sendo utilizadas as bases de dados *Web of Science*, *Scopus* e Portal de Periódicos CAPES, com os seguintes critérios: a) Idioma: Inglês; b)

Termos: *ethic, ethical, web archive, web archiving*. Sendo recuperadas 72 fontes bibliográficas e selecionadas 36 para leitura mais aprofundada.

Outro procedimento metodológico utilizado foi a pesquisa documental. O corpus da pesquisa foi constituído pelos documentos “termos de uso” e “políticas de privacidade” coletados de 19 plataformas de iniciativas de arquivamento da *web*.

Para a análise dos dados foram utilizadas as técnicas de análise documental e de conteúdo. A análise documental compreende a estrutura e o teor textual de um documento; cada tipo de documento tem uma estrutura própria e, a partir da análise da estrutura é que se propõe identificar conceitos para expor o assunto do documento (NASCIMENTO, 2009). A análise de conteúdo, segundo Santos (2012), se constitui em uma modalidade de interpretação de texto, uma das características essenciais dessa análise é a utilização de categorias, que são levadas para o material empírico com o objetivo de redução do material.

2.2 Organização e análise dos dados

A análise de conteúdo, preconizada por Bardin (2016), se compõe em três etapas; na etapa de pré-análise foram formadas as categorias iniciais a partir da seleção do material e leitura flutuante das plataformas das iniciativas de arquivamento da *web*, dos documentos termos de uso e das políticas de privacidade, resultando na formação de 23 categorias.

A etapa de exploração do material consiste na descrição analítica relacionada ao material textual coletado, submetido a um estudo mais aprofundado orientado pelo referencial teórico (MOZZATO; GRZYBOVSKI, 2011). Após a sistematização dos dados, os documentos foram transferidos para o *software* NVIVO (versão 1.3).

Primeiramente, foi identificado o tipo/descrição da instituição, divididas em Arquivos, Bibliotecas e Universidades, e sua localização - 11 instituições na América do Norte, sendo nove nos Estados Unidos e duas no Canadá; na Europa, são cinco instituições, três no Reino Unido, uma na Croácia e uma em Portugal; na Oceania, uma instituição, na Austrália; na América do Sul, uma instituição, no Chile, e na Ásia, uma instituição, em Singapura.

Quanto à categorização que apresenta o Objetivo/Missão das instituições, a nuvem de palavras construída no NVIVO (FIGURA 2) demonstra que a palavra de maior ocorrência nessa categoria é “Preservar”, que ocorre 14 vezes, a palavra “Acesso” tem uma ocorrência de quatro vezes.

Figura 2 – Nuvem de palavras a partir do objetivo/missão da instituição



Fonte: Elaborado pela autora.

O número de instituições que arquivam a *web* vem aumentando, mas os esforços são mais centrados na coleta (OGDEN; HALFORD; CARR, 2017). Contudo, segundo Pennock (2013), algumas instituições são legalmente obrigadas a capturar e arquivar conteúdo da *web*, e a perda dessas informações é uma responsabilidade organizacional e também uma responsabilidade social.

Quanto à identificação do método de coleta das instituições, as 19 instituições referem abordagem seletiva (evento ou temático), e 6 delas também referem abordagem ampla. Segundo Glanville (2010), a abordagem na coleta de conteúdo influencia no acesso e uso da informação. Uma coleta ampla garante uma cobertura maior de material, mas limita o acesso ao material arquivado.

Foi categorizado também os recursos arquivados pelas instituições, que colocam desafios particulares para as instituições, e também podem influenciar no acesso e uso da informação. A abordagem seletiva (evento) se concentra em *sites* que abordam eventos nacionais ou internacionais importantes; a abordagem seletiva (temático), envolve a seleção de recursos da *web* com base em vários conjuntos de critérios pré-definidos, por exemplo, recursos pertencentes a uma organização. Os conteúdos mais citados pelas instituições foram os *sites* governamentais, mencionado por oito instituições, e os domínios e mídias sociais, mencionado por 5 instituições.

Foi identificado o uso de ferramentas ou serviços utilizados pelas instituições, 14 instituições referem o uso de ferramentas ou serviços. A ferramenta/serviço mais citado é o *Archive-It*, referenciado por sete instituições; é um serviço de arquivamento da *web* baseado em assinatura. Esses serviços são cada vez mais utilizados pelas instituições que preferem não estabelecer e manter sua própria infraestrutura técnica de arquivamento da *web*.

A categorização seguinte tratou do tipo de acesso, 16 instituições referem acesso aberto aos arquivos da *web*, mas, dependendo do tipo de informação, poderá haver restrições. O acesso local é utilizado quando a instituição depende de licença dos editores para que o acesso seja aberto; sem a autorização, as informações somente podem ser acessadas nas dependências da instituição; quando o recurso arquivado envolve questões de privacidade, ou quando há restrições estabelecidas pela legislação. Segundo Wickner (2019), deve-se considerar quando é melhor desvincular a preservação e acesso para respeitar ou proteger os sujeitos. A importância de preservar não se traduz necessariamente em acesso público irrestrito, e os danos podem ser mitigados se houver consentimento dos sujeitos.

Com relação à categorização dos termos de uso, a descrição do serviço configura-se em um dos elementos comuns a todos os documentos. No documento, deve estar descrito o serviço e como ele é oferecido; 18 instituições trazem a descrição do serviço no documento, e três ainda informam que o acesso ao serviço oferecido é gratuito. A Tabela 1 mostra como as instituições descrevem seus serviços.

Tabela 1 - Descrição do serviço nos documentos Termos de Uso

Descrição do serviço	Número de Instituições
Apoio ao trabalho científico, à pesquisa, ao conhecimento, educação e ensino	10
Acesso ao patrimônio documental, à cultura, à memória e história	3
Acesso às informações governamentais, programas e serviços do governo	2
Coleta de <i>sites</i> , mídias sociais e conteúdos disponíveis publicamente na <i>web</i>	2
Funções administrativas	1
Serviços/informações acadêmicas	1

Fonte: Elaborado pela autora.

A categoria seguinte refere-se à legislação, normas e regulamentos referenciados nos termos de uso e está relacionada com a proteção legal. Das 16 instituições que citam alguma lei ou regulamento, 13 referenciam às Leis de

Direitos Autorais, considerando nesse total a Lei de Direitos Autorais do Milênio Digital, lei dos Estados Unidos, que permite aos provedores de serviços *online* isenção de responsabilidade por violação desses direitos. O *fair use* (uso justo), mencionado 2 vezes, também é um conceito da legislação dos Estados Unidos, que permite o uso de material protegido por direitos autorais para fins acadêmicos, por exemplo. Importante ressaltar que a proteção legal também está relacionada com a declaração de direitos autorais nos documentos jurídicos; 15 instituições apresentam esta declaração no documento, e 6 instituições mencionam licenças, sendo que 2 delas fazem referência às licenças *Creative Commons*.

A próxima categoria se refere à descrição de uso do serviço; 16 instituições descrevem em seus documentos como deve ser o uso do serviço; 10 informam que o uso sem violação de direitos autorais ou outra lei de propriedade intelectual deve ser garantido pelo usuário; 9 referem que o conteúdo não pode ser utilizado para fins comerciais, modificados, sem autorização, disposição legal em contrário ou acordo de licença, e 6 referem que a instituição deve ser citada como fonte.

Outra categoria aborda as penalidades no descumprimento das orientações de uso dos conteúdos; 9 instituições mencionam aplicação de penalidades em seus termos de uso. Dos tipos de penalidades mencionados, 17% referem que os custos e taxas com advogado será por conta do usuário, 25% referem indenização e isenção de responsabilidade da instituição, 25% referem punição civil e criminal e 33% referem cancelamento de acesso e remoção da conta do usuário.

Foi categorizado também a limitação de responsabilidade da instituição, das 11 instituições que trazem essa informação em seus documentos, 5% referem que não são responsáveis pela suspensão ou cancelamento do serviço, 6% referem que o editor é o responsável pela precisão dos dados dos recursos arquivados, 11% referem que o conteúdo pode ser regido por leis locais, nacionais, internacionais e que a responsabilidade é do usuário, 22% referem que não são responsáveis por perdas ou danos que possam vir a ocorrer no acesso e uso do serviço ou coleção e 56% referem que não são responsáveis pelo conteúdo acessado e o uso que é dado a ele, incluindo *links* para *sites* externos.

Na categoria que trata das garantias oferecidas ao usuário, embora 13 instituições referenciem “garantias” em seus termos de uso, nenhuma oferece qualquer tipo de garantia, 5% referem que não garantem ou que tenham qualquer representação sobre o uso e resultado do uso dos conteúdos, 14% referem que não garantem que o acesso às coleções seja ininterrupto, 33% não oferecem garantia de qualquer tipo para o uso ou reprodução dos itens de suas coleções e

48% referem que não garantem a completude, exatidão, confiabilidade e atualidade dos conteúdos.

Duas categorias tratam das alterações e atualizações dos documentos. Quanto às alterações dos termos, 11 instituições referem alterações no documento e são unânimes em declarar que seus termos de uso podem ser alterados sem aviso prévio, em qualquer momento e que cabe ao usuário revisar periodicamente o documento; 3 instituições mencionam que podem notificar o usuário sobre alterações através de e-mail, declaração ou aviso. Quanto às atualizações, 5 instituições identificam a data da última atualização do documento. Das 5 instituições que apresentavam a data de atualização do documento, 2 informam que o documento foi atualizado em 2020, 1 em 2019, 1 em 2014 e 1 em 2013.

Com relação à categorização das políticas de privacidade, analisando a categoria que trata das leis, regulamentos ou códigos aplicáveis referenciados nos documentos; 11 instituições referenciam em suas políticas de privacidade leis, regulamentos ou códigos; 9 mencionam leis de proteção de dados pessoais e privacidade; a GDPR é mencionada por 2 instituições da Europa, e 2 instituições fazem referência a algum código de ética ou conduta.

Já a categoria que trata do motivo da coleta de dados; 18 instituições descrevem em suas políticas de privacidade o motivo da coleta, sendo que a melhoria nos serviços oferecidos é o motivo pelo qual 13 instituições coletam dados dos usuários; 10 instituições referem a otimização dos *sites* e métricas de uso do *site* como outros motivos para a coleta de dados.

Na categoria que descreve os dados que serão coletados; 18 instituições identificam os dados do usuário que serão coletados, e os mais citados são: *cookies* (15 instituições), informações pessoais (11 instituições) e páginas visualizadas/recursos arquivados (10 instituições).

Na categoria que trata do compartilhamento ou divulgação dos dados dos usuários; 17 instituições informam em suas políticas de privacidade se há compartilhamento/divulgação dos dados, porque esses dados são compartilhados e com quem são compartilhados; 9 instituições mencionam que coletam os dados quando exigido por lei ou ordem judicial, 7 instituições informam que compartilham os dados com fornecedores, consultores, prestadores de serviço e 6 instituições com a equipe/membros da instituição.

Na categoria que trata do monitoramento/controle sobre os dados, 12 instituições descrevem em suas políticas de privacidade como é o tratamento dos dados; 4 instituições informam onde os dados serão armazenados e processados, e se os mesmos podem ser transferidos para outros países, 3 Informam prazos de manutenção dos dados. É importante esclarecer se os dados coletados

permanecerão em um banco de dados ou serão eliminados depois de um determinado tempo. A política de privacidade deve refletir todos os tratamentos de dados pessoais que são feitos pela instituição (REIS, 2016).

Das categorias que tratam as alterações e atualizações dos documentos, na categoria de alterações, 10 instituições se referem às alterações/atualizações no documento, 9 instituições são unânimes em declarar que suas políticas de privacidade podem ser atualizadas e alteradas a qualquer momento, sendo que 4 delas recomendam que o usuário verifique o documento regularmente, 2 instituições declaram que notificarão os usuários através de aviso nas páginas, 1 informa que “pode” notificar o usuário através de e-mail, e 1 informa que o documento é analisado a cada 12 meses para garantir que esteja atualizado. Na categoria de atualizações, das 6 instituições que identificam a data da última atualização do documento, 3 delas o documento foi atualizado em 2020, 1 em 2019, 1 em 2017 e 1 em 2001.

3 CONSIDERAÇÕES FINAIS

As iniciativas de arquivamento da *web* vêm crescendo pelo mundo, e junto cresce o interesse no desenvolvimento de soluções para o acesso e uso das informações coletadas e arquivadas da *web*, de forma ética e de acordo com a legislação. Frequentemente as questões legais são o maior problema não técnico enfrentado pelos arquivos da *web*, questões como o direito legal de fazer cópias do conteúdo, independentemente de ter ou não a permissão do autor, de fornecer acesso aos recursos arquivados, bem como questões que envolvem a privacidade também adquirem um grau de complexidade maior.

Analisar os aspectos éticos e legais pertinentes ao acesso e uso das informações de *websites* arquivados, a partir dos documentos jurídicos disponibilizados nas plataformas digitais de iniciativas internacionais, pode contribuir para os estudos sobre arquivamento da *web* brasileira. As possibilidades de solução estão no trabalho colaborativo e interdisciplinar entre as diferentes áreas, com a formulação de ações voltadas para a educação do usuário, tornando-o capacitado a usufruir das inúmeras possibilidades que o uso do ambiente digital proporciona.

REFERÊNCIAS

- BARDIN, Laurence. **Análise de conteúdo**. São Paulo: Edições 70, 2016. Disponível em: <https://madmunifacs.files.wordpress.com/2016/08/anc3a1lise-de-contec3bado-laurence-bardin.pdf>. Acesso em: 20 jun. 2021.
- BRAGG, Molly *et al.* The Web Archiving Life Cycle Model. **WhitePaper**, p. 1-30, 2013. Disponível em:

http://ait.blog.archive.org/files/2014/04/archiveit_life_cycle_model.pdf. Acesso em: 15 mar. 2020.

BRAYNER, Aquiles. **Preservação do conteúdo web brasileiro**: mesa-redonda apresentada no 9. Fórum da Internet do Brasil. Manaus, 2019. Disponível em: https://minhaagenda.nic.br/files/apresentacao/arquivo/698/Preservacao_do_Conteudo_Web_Brasileiro.pdf. Acesso em: 04 jan. 2020

GLANVILLE, Lachlan. Web archiving: ethical and legal issues affecting programmes in Australia and the Netherlands. **Australian Library Journal**, Canberra, v. 59, n. 3, p. 128-134, 2010. Disponível em: <https://doi.org/10.1080/00049670.2010.10735999>. Acesso em: 02 abr. 2021.

MORAES, João Antonio de. **O paradigma da complexidade e a ética informacional**. 2018. 170 f. Tese (Doutorado em Filosofia) - Instituto de Filosofia e Ciências Humanas, Universidade Estadual de Campinas, Campinas, 2018.

MOZZATO, Anelise Rebelato; GRZYBOVSKI, Denize. **Análise de Conteúdo como Técnica de Análise de Dados Qualitativos no Campo da Administração**: Potencial e Desafios. **RAC**, Curitiba, v. 15, n. 4, p. 731-747, jul./ago. 2011. Disponível em: <https://www.scielo.br/pdf/rac/v15n4/a10v15n4.pdf>. Acesso em: 22 dez. 2020.

NASCIMENTO, Lúcia Maria Barbosa do. **Análise documental e análise diplomática**: perspectivas de interlocução de procedimentos. 2009. 199 f. Tese (Doutorado em Ciência da Informação) - Faculdade de Filosofia e Ciências da Universidade Estadual Paulista Júlio de Mesquita Filho, Marília, 2009. Disponível em: <https://repositorio.unesp.br/handle/11449/103371>. Acesso em: 13 jun. 2020.

OGDEN, Jessica, HALFORD, Susan, CARR, Leslie. Observing Web Archives: The Case for an Ethnographic Study of Web Archiving. **WebSci** 17: 299-308. 2017. Disponível em: <https://dl.acm.org/doi/pdf/10.1145/3091478.3091506?download=true>. Acesso em 06 abr. 2021.

REIS, Priscila. **Política de privacidade e termos de uso após Marco Civil da Internet**. RMPadvogados, 2016. Disponível em: <http://www.rmpadvogados.com.br/politica-deprivacidade-e-termos-de-uso-apos-marco-civil-da-internet/>. Acesso em: 20 ago. 2020

ROCKEMBACH, Moisés. Inequalities in digital memory: ethical and geographical aspects of web archiving. **International Review of Information Ethics**, Canada, v. 26, p. 138-149, 2017. Disponível em: <https://informationethics.ca/index.php/irie/article/view/286>. Acesso em: 20 abr. 2021.

ROCKEMBACH, Moisés. **Arquivamento da Web**: estudos de caso internacionais e o caso brasileiro. **RDBCI: Rev. Digit. Bibliotecon. Cienc. Inf.**, Campinas, v. 16, n. 1, p. 07-24, 2018. Disponível em: <http://orcid.org/0000-0001-9057-0602>. Acesso em: 15 mai. 2019.

ROCKEMBACH, Moisés. **Preservação do Conteúdo Web Brasileiro**. Mesa-redonda apresentada no 9. Fórum da Internet do Brasil, Manaus, 2019.

ROCKEMBACH, Moisés, PAVÃO, Caterina Marta Groposo. Políticas e tecnologias de preservação digital no arquivamento da web. **RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação**. Brasília: UnB. Vol. 11, n. 1, p. 168-182 2018. Disponível em: <https://periodicos.unb.br/index.php/RICI/article/view/8473>. Acesso em: 20 abr. 2021

SANTOS, Fernanda Marsaro dos. Análise de conteúdo: a visão de Laurence Bardin. **Revista Eletrônica de Educação**, São Carlos, v. 6, n. 1, maio 2012. Resenhas. Disponível em: <http://www.reveduc.ufscar.br/index.php/reveduc/article/view/291/156>. Acesso em: 02 fev. 2020.

WICKNER, Amy. Recognizing Co-Creators in Four Configurations: Critical Questions for Web Archiving. **Journal of Contemporary Archival Studies**, New Haven, v. 6, p. 1-27. 2019. Disponível em: <https://elischolar.library.yale.edu/cgi/viewcontent.cgi?article=1084&context=jcas>. Acesso em: 20 nov. 2020.