REQUISITOS MÍNIMOS
DE PRESERVAÇÃO PARA WEBSITES E MÍDIAS SOCIAIS

Conarq
CONSELHO NACIONAL DE ARQUIVOS

Requisitos mínimos
de preservação para websites e mídias sociais

Rio de Janeiro | 2023

Praça da República, 173 ● Rio de Janeiro ● RJ ● 20211-350
e-mail: conarq@an.gov.br

Presidente da República
Luiz Inácio Lula da Silva

Ministra da Gestão e Inovação em Serviços Públicos
Esther Dweck

Presidente do Conselho Nacional de Arquivos
Ana Flávia Magalhães Pinto

Secretário executivo do Conselho Nacional de Arquivos
Alex Pereira de Holanda

Diretora de Processamento Técnico, Preservação e Acesso ao Acervo
Diana Santos Souza

Coordenadora-geral de Acesso e Difusão Documental
Daiana Ribeiro Dantas Martins

Coordenadora de Pesquisa e Difusão do Acervo
Leticia dos Santos Grativol

Revisão
José Claudio Mattar
Mariana Simões

Diagramação e ilustração da capa
Alzira Reis

Projeto gráfico da capa
Mariana Machado Laplace

Dados Internacionais de Catalogação-na-Publicação (CIP)
(Biblioteca Maria Beatriz Nascimento – Arquivo Nacional)

Conselho Nacional de Arquivos (Brasil)

Requisitos mínimos de preservação para websites e mídias sociais. [recurso eletrônico]. / Conselho Nacional de Arquivos – Dados eletrônicos (1 arquivo : 233 kb). – Rio de Janeiro: Arquivo Nacional, 2023.

Formato: PDF

Requisitos do sistema: Adobe Acrobat Reader

Modo de acesso: World Wide Web

1. Preservação digital – Sites da Web. 2. Preservação digital - Mídia social. 3. Arquivamento. I. Título.

CDD 025.84

Ficha catalográfica elaborada por Natália Marques de Souza (CRB7/5223)

Câmara Técnica Consultiva Preservação de Websites e Mídias Sociais
Carolina de Oliveira (coordenadora)
Érika Maria Nunes Sampaio
Gabriela Ayres Ferreira Terrada
Jonas Ferrigolo Melo
Moisés Rockembach

Contribuições de especialistas
Nossos agradecimentos a Ana Javes, Daniela Francescutti,
Mauricio Augusto Cabral Ramos Junior e Ricardo Basílio

Sumário

Introdução 5

1 Âmbito de aplicação e público-alvo 8

2 Objetivo do documento 8

3 Etapas da preservação de websites e mídias sociais 8

3.1 Seleção 10

3.1.1 Definição do escopo do arquivo da web 10

3.1.2 Identificação da comunidade de usuários (público-alvo) 12

3.1.3 Arquivabilidade de websites 12

3.1.4 Priorização dos elementos 13

3.2. Definição de metadados 14

3.3 Captura 15

3.4 Controle de qualidade 16

3.5 Armazenamento 17

3.6 Acesso 17

Glossário 19

Referências 21

INTRODUÇÃO

O comportamento do cidadão diante da facilidade que as tecnologias de informação e comunicação (TICs) proporcionam para obtenção e transmissão de informações exige das entidades públicas e privadas responsáveis pela salvaguarda do patrimônio arquivístico brasileiro ações adequadas às novas formas e ambientes de registro dessas informações.

Dentre as TICs existentes, neste documento serão abordados os websites – conjunto de páginas na internet interligadas por hipertexto, com conteúdo e configuração dinâmica, acessíveis através de um endereço – e as mídias sociais,[1] que são conteúdos criados e compartilhados pelos usuários, em plataformas de redes sociais.

Como principal razão para a preservação de websites e mídias sociais institucionais destaca-se a importância dos serviços on-line como fonte para o exercício da cidadania. No que se refere à esfera privada, os proprietários dos websites e das contas em mídias sociais têm autonomia para definir seus próprios procedimentos de preservação e acesso a esses objetos digitais. Entretanto, ao se considerar a política nacional de arquivos públicos e privados, disposta pela lei n. 8.159, de 8 de janeiro de 1991, e regulamentos, pessoas físicas e jurídicas de direito privado, detentoras de arquivos, podem integrar o Sistema Nacional de Arquivos (Sinar) mediante acordo com o Conselho Nacional de Arquivos (Conarq), assim como conjuntos de documentos arquivísticos dessa natureza serem declarados de interesse público e social.

Além disso, a transitoriedade desses conteúdos, devido à possibilidade do autor ou detentor do website ou da mídia social de apagá-los ou modificá-los em um curto espaço de tempo, e a falta de manutenção, alteração das URLs/domínios ou links quebrados colocam em risco o patrimônio digital nacional.

---

1 Ver: InterPARES. General Study 09. Disponível em: http://www.interpares.org/ip3/display_file.cfm?doc=ip3_canada_gs09_final_report.pdf.

Estudos demonstram que cerca de 80% das páginas web não se encontram disponíveis na sua forma original após um ano da sua publicação, assim como 11% dos recursos de mídia social, como os postados no Twitter (Ntoulas; Cho; Olston, 2004; Costa; Gomes; Silva, 2016).

Considerando o risco de perda deste patrimônio para a posteridade, devido à rápida obsolescência de hardware e software, além de incertezas sobre recursos e orçamento, responsabilidades institucionais, métodos para manutenção e preservação, e falta de legislação para dar suporte aos processos de trabalho, torna-se desafiadora a preservação desses objetos complexos (Unesco, 2004, p. 80). É necessária a criação de mecanismos para viabilização da preservação dos conteúdos de websites e mídias sociais, visando o acesso e uso como artefato cultural de valor permanente.

O Conarq, através da Resolução n. 13, de 9 de fevereiro de 2001, já salientava a preocupação com a construção de websites para instituições arquivísticas, "recomendando o backup sistemático, por meio de arquivamento eletrônico ou impresso, de forma a garantir a segurança das informações, além do arquivamento das páginas das versões anteriores do website" (Conarq, 2001, p. 9). Ressalta-se que o backup é uma cópia de segurança das informações digitais no caso de perda ou destruição do original, diferentemente do arquivamento de websites e de mídias sociais, que possibilita versionamentos, busca retrospectiva, acesso ao cidadão, reuso de dados, entre outras funções.

Os domínios da web nacionais são recursos que se conectam a milhões de páginas da web, arquivos estes que são publicados, atualizados ou excluídos diariamente. Grande parte desse patrimônio, composto por objetos digitais complexos, deveria ser preservado para o futuro, como, por exemplo, as páginas web, dados brutos de pesquisas, documentos governamentais, arquivos digitais privados de organizações ou indivíduos. Portanto, coletar o patrimônio digital, por meio de diferentes canais e plataformas, demanda esforços e recursos significativos. Dessa forma, é essencial que as instituições nacionais de referência tenham um papel de liderança, estabelecendo políticas e sistemas de coleta e gerenciamento de objetos digitais da web ou liderando redes colaborativas para a adoção de modelos comuns de seleção e preservação (Unesco, 2016, p. 5).

A preservação de websites e mídias sociais vem ao encontro de iniciativas internacionais e nacionais sobre esses objetos digitais (Rockembach, 2018) que têm sido adotadas por pessoas jurídicas de direito público interno, por pessoas jurídicas de direito privado ou por pessoas naturais para a comunicação de suas atividades, assim como para interação com seu público-alvo.

Outros desafios são adicionados às plataformas de mídias sociais, adotadas tanto na esfera pública quanto na privada, por serem de propriedade de empresas. Todas elas possuem termos de uso cujas condições são aceitas pelo usuário no momento da criação de sua conta. Devido a isso, para salvar o conteúdo e a forma das páginas de cada conta é preciso respeitar as cláusulas de tais termos.

Assim como os websites, o uso de mídias sociais pelas organizações demanda a preservação dos perfis e páginas institucionais, a fim de garantir o acesso em longo prazo, independentemente da manutenção do desenvolvedor da rede social.

Destaca-se, ainda, que o arquivamento da web não precisa ser o único método de preservação do conteúdo disponibilizado em websites e mídias sociais. Fotografias, notícias, *releases*, imagens e *cards*, por exemplo, podem ser preservados no contexto do website, por meio do arquivamento da web, mas também no contexto de sua função primária. Isso significa que os procedimentos de gestão documental devem ser considerados em conjunto com as rotinas de arquivamento da web, ainda que essa ação possa causar redundância no armazenamento, muitas vezes necessária para a efetiva preservação digital.

É recomendada, ainda, ao leitor a adoção de outras resoluções do Conarq que fornecem diretrizes para a produção, manutenção, preservação e presunção de autenticidade de documentos digitais. Com esse arcabouço técnico-teórico, a implementação destes requisitos mínimos para preservação de websites e mídias sociais será facilitada.

Por fim, para avançar nas práticas inclusivistas por meio da acessibilidade atitudinal, é importante que os órgãos públicos adotem sempre o recurso de texto alternativo ao postarem como mídias sociais imagens e vídeos, uma vez que se trata de um recurso disponível em todos os aplicativos de redes sociais.

1 Âmbito de aplicação e público-alvo

Administração pública federal, dos estados e dos municípios que mantêm em sua estrutura administrativa instituições arquivísticas, integrantes do Sistema Nacional de Arquivos (Sinar), e demais responsáveis pela custódia de acervos documentais, públicos ou privados, considerados de valor permanente.

Outras organizações podem utilizar este documento como referência.

2 Objetivo do documento

A finalidade deste documento é apresentar aos integrantes do Sinar as condições mínimas necessárias para alcançar o objetivo de preservar websites e mídias sociais em longo prazo, assumindo que são objetos digitais complexos de valor secundário.

Ainda que sejam avaliados como de guarda permanente, eventualmente, assume-se a inviabilidade de preservar toda a web e, nesse sentido, o desenvolvimento dos requisitos aqui apresentados considerou uma abordagem sistêmica que consiste em etapas que instrumentalizam a composição de um arquivo da web.

3 Etapas da preservação de websites e mídias sociais

O arquivamento da web enfrenta uma série de desafios devido à sua estrutura complexa, que é materializada pela variedade de formatos e modos de exibição. O leiaute dos websites e mídias sociais varia de domínio para domínio com base nas informações disponibilizadas, ou seja, seu conteúdo; e na sua apresentação, a forma como essa informação é visualizada. Para elucidar, podem-se comparar websites governamentais, de notícias, de comércio eletrônico e redes sociais, que se diferenciam muito em seu conteúdo e estrutura.

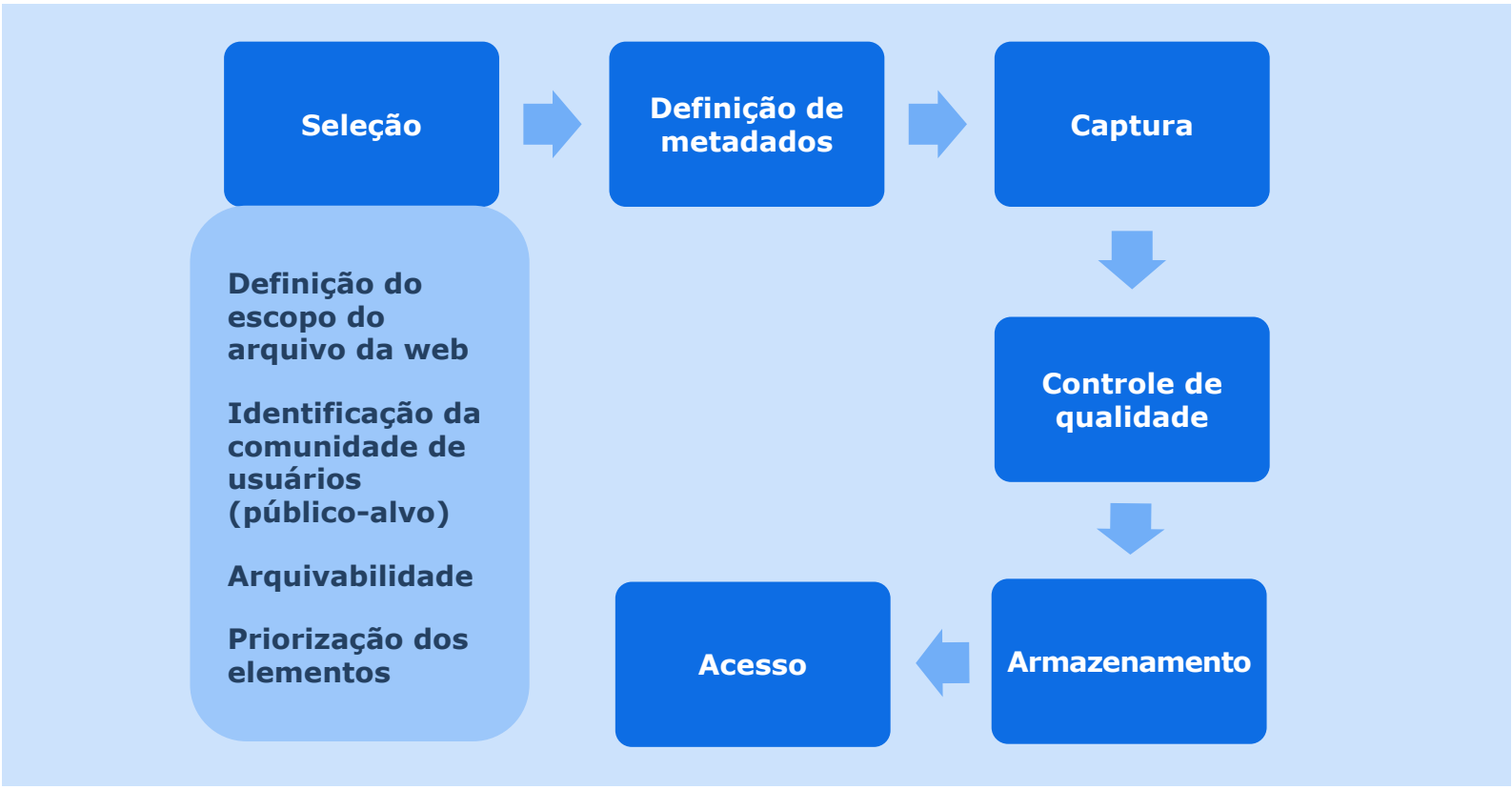
Desse modo, no ímpeto de enfrentar os desafios que o arquivamento da web impõe às instituições e profissionais, este documento apresenta as condições mínimas para a constituição de arquivos da web. Para isso, são propostas seis etapas para a preservação de websites e mídias sociais.

Khan e Rahman (2019, p. 72) dizem que "um processo de preservação eficaz é aquele que leva a um arquivo da web bem-organizado e de fácil gerenciamento e atende aos requisitos designados da comunidade". Assim, essas etapas podem ajudar a entender os desafios da preservação e as atividades que compreendem o arquivamento da web. Essa estrutura é uma maneira acessível de analisar, projetar, implementar e avaliar o arquivo da web com clareza, compondo um processo eficaz de preservação digital.

Além disso, a preservação digital de websites e mídias sociais demanda uma atuação em rede e interdisciplinar, especialmente quanto ao emprego de recursos de infraestrutura e de pessoal, assegurando a sustentabilidade dessa atividade em longo prazo.

A seguir são elencadas as etapas basilares para os procedimentos de arquivamento da web (Figura 1), definidas tendo como referência as pesquisas *The web archiving life cycle model* (Bragg et al., 2013) e *A systematic approach towards* web *preservation* (Khan; Rahman, 2019), e o processo de arquivamento da web (Rockembach, 2021).

Figura 1 - Etapas da preservação de websites e mídias sociais



3.1 Seleção

Neste documento, entende-se o termo seleção[2] como o processo de escolha do escopo de captura dos websites e mídias sociais que deverão ser preservados por uma instituição arquivística. Esse processo tem que ser baseado em critérios claros visando conciliar a necessidade de informação do público-alvo e os recursos disponíveis para a preservação desses objetos digitais complexos.

A definição de critérios de seleção ajudará a determinar quais os conteúdos do website e da mídia social que precisarão ser capturados, considerando as prioridades, a finalidade e o escopo do arquivo da web.

O processo de seleção é composto por quatro atividades que auxiliarão na definição dos objetos digitais a serem preservados: (i) identificação do escopo do arquivo da web; (ii) definição dos usuários em potencial; (iii) verificação dos níveis de arquivabilidade; e (iv) identificação dos elementos dos websites e das mídias sociais.

Os websites e mídias sociais selecionados, após aplicação dos parâmetros especificados a seguir serão considerados de guarda permanente.

3.1.1 Definição do escopo do arquivo da web

As informações nos websites e mídias sociais são tratadas e apresentadas de diferentes maneiras. Além disso, seu leiaute geral muda de um domínio para outro, o que torna impraticável desenvolver uma sistemática única para preservar todas as páginas web em longo prazo.

O escopo de um arquivo da web é determinado pela definição do objeto digital a ser coletado – o web*site* ou a mídia social. Trata-se de uma questão complexa, pois do ponto de vista do usuário, uma página web é a imagem exibida ao colocar um endereço de URL (*uniform resource locator)* em um navegador. Ainda que essa definição operacional seja

---

2 Para os arquivos da web, natos digitais com valor secundário, são mantidas as versões dos websites em formato Warc.

necessária, não é suficiente, pois um arquivo também deve garantir que os objetos digitais sejam presumidos como autênticos, fazendo com que sejam incluídos seu contexto e seus metadados, e garantindo que se trata do documento original. O escopo do arquivo da web poderá ser definido considerando algumas abordagens:

Centrado no website e perfil de rede social

O arquivamento centrado no web*site* refere-se à escolha de um determinado website em sua totalidade. Isso significa que a preservação será de todo o website, desde sua página principal (*home page*) até o último nível de profundidade de URL.

Essa abordagem, aplicada às redes sociais, diz respeito à definição de um perfil específico como escopo de captura que terá seu conteúdo preservado.

Centrado no conteúdo ou tópico

Essa abordagem tem se tornado cada vez mais popular em razão da natureza efêmera das páginas web. Impulsionado por necessidades diretas de pesquisa, este recurso tem suprido a falta de uma prática sistêmica de preservação de websites e mídias sociais, permitindo que sejam criadas coleções específicas. Como exemplo, cita-se a cobertura de processos eleitorais, temáticas como a da pandemia de Covid-19, Jogos Olímpicos e quaisquer outros conteúdos ou tópicos de interesse na preservação. Instituições poderão definir, por exemplo, que preservarão apenas URLs de seus websites e mídias sociais que tratem de um conteúdo ou tópico específico.

Centrado no domínio

O arquivamento centrado no domínio refere-se à escolha de uma determinada extensão de URL. O domínio pode representar uma localidade, tipo de rede ou domínios genéricos. No entanto, é possível distinguir alguns tipos funcionais (.com e .edu, por exemplo) e tipos geográficos (.br e .uk, por exemplo). Os domínios de nível superior geográficos geralmente têm subdivisões funcionais (gov.br e mil.br, por exemplo).

Uma vantagem deste método é que o arquivamento pode acontecer detectando automaticamente websites específicos do domínio escolhido. Destaque-se que essa abordagem se refere apenas ao arquivamento de websites e não se aplica às mídias sociais.

Para este tópico, recomenda-se compreender os conceitos de "coleta intensiva e extensiva" de Masanès (2006).

3.1.2 Identificação da comunidade de usuários (público-alvo)

Conhecer o conjunto de usuários em potencial, ou seja, aqueles que terão interesse (ou necessidade) em acessar o conteúdo arquivado é importante para que se tome a melhor decisão quanto à seleção dos websites e mídias sociais a serem preservados.

Os promotores da iniciativa de arquivamento dos websites e mídias sociais devem identificar esses usuários em potencial do arquivo, suas características e consultas esperadas.

3.1.3 Arquivabilidade de websites

Nesse cenário de complexidade dos websites – em que o desafio do arquivamento se materializa – e na busca de uma solução para compreender as razões pelas quais alguns recursos presentes em websites não são possíveis de arquivamento é que surgiu o conceito de arquivabilidade.

A arquivabilidade é definida como a extensão em que um website atende às condições para a transferência segura de seu conteúdo para um arquivo da web para fins de preservação. Ou seja, é uma noção estabelecida para capturar os aspectos principais de um website, crucial para diagnosticar se ele tem o potencial de ser arquivado com integridade e precisão.

Essa sistemática ajuda a verificar o nível de arquivabilidade de uma página web. Esse nível é calculado considerando seu desenvolvimento, padrões e tecnologias em relação aos padrões estabelecidos pelo consórcio internacional da web, o W3C.[3] Quanto maior a conformidade

---

3 Mais informações podem ser obtidas em www.w3.org/standards.

do website com o estabelecido mundialmente como convenção para seu desenvolvimento, maior será sua responsividade com o arquivamento.

Portanto, para garantir que os websites possam ser arquivados com qualidade, recomenda-se que sejam desenvolvidos considerando os padrões estipulados pelo consórcio W3C. Além disso, algumas práticas simples podem oferecer maiores garantias de arquivabilidade, tais como utilizar URLs "amigáveis" e evitar encurtadores de links. Além de serem boas práticas para a preservação digital, são positivas, também, para melhorar a usabilidade, a segurança e a otimização dos motores de busca.

Para o cálculo da arquivabilidade de um website há ferramentas que fornecem uma abordagem para automatizar o controle de qualidade, avaliando a conveniência de ser arquivado antes de qualquer tentativa de fazê-lo e proporcionando ganhos consideráveis ao reduzir o uso de recursos humanos, computacionais e de rede, e ao não coletar websites não coletáveis.

> De forma demonstrativa, citam-se aqui o método **CLEAR+** e a ferramenta **ArchiveReady** (http://archiveready.com/) que segue o conjunto de facetas de arquivabilidade: acessibilidade, coesão, metadados e conformidade com padrões.
>
> Se for utilizado *script* (tais como JavaScript) na página web, devem ser fornecidas alternativas de HTML simples – isto suporta a acessibilidade para os utilizadores e para o arquivamento. Forneça links estáticos ou "âncoras de página básica" sempre que possível, em vez de URLs geradas dinamicamente. Se a página web inclui conteúdos ricos em mídia e *streaming*, devem-se fornecer alternativas tais como *downloads* progressivos juntamente com conteúdos *streaming*.
>
> Reveja a página web para verificar a acessibilidade de acordo com a norma produzida pelo W3C.

3.1.4 Priorização dos elementos

Para a seleção dos websites e mídias sociais a serem arquivados, outro requisito a ser considerado é a priorização dos elementos que os compõem, definindo aqueles que têm necessidade e condições de serem arquivados.

Os elementos que compõem esses objetos digitais são compreendidos por seu conteúdo e sua estrutura. Podem ser texto, imagem, ícones, vídeos, áudios, rádio web, *banner* rotativo, elementos de acompanhamento em tempo real (horário, previsão do tempo, taxas de câmbio etc.), dentre outros.

A complexidade desses elementos e suas representações podem causar complicações no processo de preservação digital, em razão da inviabilidade de se preservarem todos os recursos utilizados no website e mídias sociais ou até mesmo por sua preservação ser indesejada.

Portanto, a priorização dos elementos é uma fase relevante e requer atenção em dois fatores: o primeiro diz respeito à reutilização potencial dos elementos que compõem o website ou mídia social; e o segundo considera a frequência com a qual o elemento será acessado. Um dos métodos utilizados para medir esses fatores é o MoScoW.

Método MoSCoW

Trata-se de uma técnica de priorização utilizada em gerenciamento e análise de negócios, gerenciamento de projetos e desenvolvimento de softwares para alcançar um entendimento comum com as partes interessadas sobre a importância que elas atribuem à entrega de cada requisito. Também é conhecido como priorização do MoSCoW ou análise do MoSCoW (Haughey, 2021).

O termo é um acrônimo derivado da primeira letra de cada uma das quatro categorias de priorização: *Must have* (Deve ter), *Should have* (Deveria ter), *Could have* (Poderia ter) e *Won't have* (Não terá).

3.2. Definição de metadados

Os metadados desempenham um papel vital na preservação em longo prazo de objetos digitais e são importantes para se recuperar um objeto específico do arquivo após a preservação. Eles asseguram respostas sobre a integridade do objeto, sua criação, descrição, se sofreu alteração e sua relação com outros objetos.

Metadados são informações estruturadas que descrevem, localizam, gerenciam, recuperam e acessam recursos de informação e são geralmente chamados de “dados sobre dados” ou “informações sobre

informações", mas pode ser mais útil e informativo descrever esses dados como "documentação descritiva e técnica" (Niso, 2017).

Entendido dessa maneira, fica evidente que os metadados precisam suportar uma gama de funções, incluindo, por exemplo, acesso, registro de contextos e proveniência de objetos, documentação de ações e políticas de repositório. Para o propósito deste documento, serão três os tipos de metadados em foco: metadados descritivos, metadados estruturais e metadados administrativos.

Os metadados descritivos constituem um recurso para fins de descoberta e identificação. Podem consistir em elementos para um documento, como título, autor(es), resumo e palavras-chave. Os metadados estruturais descrevem como os objetos compostos são reunidos, por exemplo, como indicação de hierarquia. Os metadados administrativos abarcam os metadados técnicos e os de preservação. Também podem incluir os metadados sobre direito e reprodução. Fornecem informações para facilitar o gerenciamento de recursos digitais, onde constam informações como requisitos de acesso e trilhas de auditoria.

A preservação e o arquivamento de objetos digitais requerem padrões de metadados para rastrear e garantir o acesso a esses objetos. Os elementos de metadados podem ser personalizáveis, inclusive a partir de uma junção de diferentes padrões, a fim de melhor se adequarem à natureza dos objetos digitais a serem preservados pelas instituições custodiadoras.

3.3 Captura

A captura de um website ou mídia social corresponde à recolha desse objeto digital para aplicação dos procedimentos destinados à sua preservação. Ela é baseada em uma lista de entradas ou URLs predefinidas. Essas entradas são chamadas na literatura específica de *seeds* (sementes).

A URL é o endereço global de recursos e documentos na web. Isto significa que tudo em um website ou mídia social pode ser facilmente rastreado e, portanto, arquivado. Por isso, sempre que possível, mantém-se todo o conteúdo sob uma URL de raiz, evitando o uso de encurtadores de link, por exemplo.

Os arquivos da web podem apresentar uma ampla variedade de formatos de arquivos, sejam páginas de conteúdo on-line textual, imagens, vídeos ou arquivos em PDF e outros formatos. Para preservar esse conteúdo, o processo de arquivamento da web usa diferentes formatos de armazenamento contendo metadados e utiliza técnicas de compactação destes dados.

Uma vez que as instituições tenham escolhido quais e quantos websites e mídias sociais serão capturados, as ações com o software de rastreamento são colocadas em funcionamento. Neste processo é definida a frequência em que os rastreamentos ocorrerão.

3.4 Controle de qualidade

Os conteúdos capturados podem ter maior ou menor qualidade, conforme vários quesitos, como a completude dos dados, incluindo textos, imagens e reprodutores de multimídia (som e vídeo) ou na reprodução posterior do conteúdo.

A fidelidade em reproduzir o website ou a mídia social com o mesmo *design* original é importante, porém a captura do conteúdo das URLs é primordial. Por isso, torna-se necessário realizar uma etapa de controle da qualidade, verificando como o rastreador da web foi executado e a reprodutibilidade do conteúdo armazenada.

Realizada a captura dos websites e mídias sociais selecionados, os dados capturados são analisados e sua qualidade e integridade avaliadas por meio de relatórios gerados pelos rastreadores ou nos próprios websites e mídias sociais arquivados.

Ressalta-se que limitações técnicas e legais na captura de um website ou mídia social mostram que uma cópia perfeita nem sempre é alcançada, por isso medidas de garantia de qualidade são necessárias para definir o sucesso do arquivamento, assegurando que os recursos estejam adequados para preservação e uso em longo prazo.

O método ou técnica de controle de qualidade pode ser elaborado conforme a necessidade e exigências da iniciativa que está compondo o arquivo da web. Com relação aos indicadores de qualidade, eles devem ser definidos pelas instituições arquivísticas com base em suas políticas internas.

3.5 Armazenamento

O armazenamento de websites e mídias sociais pode ser compreendido como a guarda desses objetos digitais, sendo necessário para o acesso ao longo do tempo conforme os padrões internacionais de preservação digital.

Os arquivos da web devem ser preservados no formato WebARChive (Warc), que foi desenvolvido pelo International Internet Preservation Consortium (IIPC) tendo sido adotado, em 2009, como extensão padrão para arquivos web, como definido na ISO 28500.[4]

Os arquivos da web precisam manter a autenticidade e a integridade do conteúdo após seu arquivamento mesmo que os requisitos mudem de acordo com o objetivo da coleta e a priorização dos elementos. Em alguns cenários, preservar apenas o conteúdo intelectual é suficiente; em outros, a estrutura dos elementos também precisa ser preservada. Entretanto, reforça-se que a autenticidade e a integridade deverão ser mantidas. E o registro de metadados e de informações complementares apoia a autenticidade dos objetos digitais.

É pertinente salientar que a natureza, qualidades e especificidades dos websites e mídias sociais exigem dos pesquisadores da área revisitação de conceitos e métodos adotados para outros objetos digitais, uma vez que a base teórica de preservação digital é a mesma.

3.6 Acesso

O acesso aos websites e mídias sociais preservados por uma instituição arquivística é fundamental e justifica o investimento por ela realizado com tal finalidade. Para tanto, a definição de uma política de acesso, uso e reutilização é um importante elemento na composição de um arquivo da web.

---

4 Disponível em: https://www.iso.org/standard/68004.html.

Cabe reforçar que o acesso à informação pública deve ser extensivo, de acordo com a Constituição Federal de 1988, a Lei de Acesso à Informação (LAI) e o Marco Civil da Internet, assegurando a liberdade de expressão e impedindo a censura. Casos excepcionais, tais como documentos sigilosos, dados sensíveis ou com proteção de direitos autorais, devem ser tratados conforme previsões em legislação específica.

Além disso, assume-se a possibilidade de arquivos da web acessíveis apenas àqueles retrospectivos (ou acesso atrasado) para evitar concorrência com o website proprietário. Essa prática de acesso atrasado é amplamente utilizada, por exemplo, pelo Arquivo.pt, UK Web Archive e Internet Archive.

A facilidade de recuperação de conteúdo em um arquivo da web está implicada em fatores como o processo de seleção, os metadados e, especialmente, a ferramenta de busca utilizada. O processo de preservação, quando bem definido, é um fator fundamental para que o arquivo da web alcance seu objetivo de disseminar informações preservadas em suas bases.

Para usar todo o potencial dos arquivos da web são necessárias uma interface amigável que facilitará a pesquisa pelo usuário e a utilização de metadados que ajudem a recuperar a partir do texto completo, palavras-chave ou outras informações que, porventura, poderão ser relevantes para a recuperação da informação pretendida pelos usuários.

GLOSSÁRIO

Arquivamento da web

Processo que compreende capturar, armazenar e disponibilizar a informação retrospectiva da World Wide Web para cidadãos, servindo como preservação da memória institucional.

Arquivo da web

Conteúdos publicados na web, sejam websites ou mídias sociais, sobre os quais uma instituição assumiu a responsabilidade e tomou providências para a preservação digital e que mantêm as mesmas características de informação e navegabilidade das páginas web originais.

Autenticidade

Credibilidade de um documento como documento, isto é, a qualidade de um documento ser o que diz ser e de estar livre de adulteração ou qualquer outro tipo de corrupção.

Captura

Etapa do método de preservação dos websites e mídias sociais correspondente ao recolhimento das páginas web, baseada em uma lista de entradas (URLs predefinidas).

Documento digital complexo

Documentos não estáveis que agregam dados, metadados e às vezes serviços em uma única entidade digital lógica.

Elementos da web

Compreendem o conteúdo (textual, visual, multimídia etc.) e a estrutura (leiaute, apresentação, comportamento de navegação) da web.

Encurtador de link

Técnica utilizada na internet para transformar um endereço http em um *link* mais curto. A URL original passa a ser acessada pelo novo *link* com menos caracteres.

Mídias sociais

Conteúdos criados e compartilhados pelos usuários, em plataformas de redes sociais.

Objeto digital

Conjunto de uma ou mais cadeias de bits que registram o conteúdo do objeto e de seus metadados associados.

Perfil de rede social

Ambiente particular em uma rede social utilizado para interação de indivíduos e entidades, bem como para a disseminação de informações.

Rastreador

Ferramenta de software que captura o conteúdo da World Wide Web de forma automatizada.

Streaming

Tecnologia que transmite dados pela internet, principalmente áudio e vídeo, sem a necessidade de baixar o conteúdo. O arquivo é acessado pelo usuário de forma on-line.

URL

Do inglês *uniform resource locator* ou localizador uniforme de recursos, identifica um recurso em uma rede informática.

Valor secundário

Valor atribuído a um documento em função do interesse que possa ter para a entidade produtora e outros usuários, tendo em vista a sua utilidade para fins diferentes daqueles para os quais foi originalmente produzido.

Warc

Formato de arquivo em modelo *open source*, baseado na ISO 28500:2017, para preservação de arquivos da web em longo prazo.

Websites

Compreendidos como uma coleção de páginas web relacionadas por conteúdo ou domínio e com configuração dinâmica.

REFERÊNCIAS

BRAGG, Molly; HANNA, Kristine; DONOVAN, Lori; HUKILL, Graham; PETERSON, Anna. *The web archiving life cycle model*. WhitePaper. 2013. Disponível em: http://ait.blog.archive.org/files/2014/04/archiveit_life_cycle_model.pdf. Acesso em: 14 jul. 2022.

CONSELHO NACIONAL DE ARQUIVOS (Conarq). Resolução n. 13, de 9 de fevereiro de 2001. Dispõe sobre a implantação de uma política municipal de arquivos, sobre a construção de arquivos e de websites de instituições arquivísticas. Disponível em: https://www.gov.br/conarq/pt-br/legislacaoarquivistica/resolucoes-do-conarq/resolucao-no-13-de-9-de-fevereiro-de-2001. Acesso em: 11 ago. 2021.

COSTA, Miguel; GOMES, Daniel; SILVA, Mário J. The evolution of web archiving. *International Journal on Digital Libraries*, p. 1-15, 2016. Disponível em: https://www.researchgate.net/publication/302777958_The_evolution_of_web_archiving/link/5e1e66c6a6fdcc904f7049ef/download. Acesso em: 22 set. 2021.

HAUGHEY, Duncan. *Moscow Method*. Project Smart website. 2021. Disponível em: https://www.projectsmart.co.uk/moscow-method.php. Acesso em: 27 jul. 2022.

KHAN, Muzammil; RAHMAN, Arif Ur. A systematic approach towards web preservation. *Information Technology and Libraries*, v. 38, n. 1, p. 71-90, 2019. Disponível em: https://doi.org/10.6017/ital.v38i1.10181. Acesso em: 1 ago. 2019.

MASANÈS, Julien. *Web archiving*. Paris: Springer-Verlag; Berlin: Heidelberg, 2006.

NISO. National Information Standards Organization. Understanding metadata: what is metadata, and what is it for?: A Primer. Beltimore: Niso, 2017. Disponível em: https://www.niso.org/publications/understanding-metadata-2017. Acesso em: 27 jul. 2022.

NTOULAS, Alexandros, CHO, Junghoo, OLSTON, Christopher. What's new on the web? The evolution of the web from a search engine perspective. *Proceedings of the 13th international conference on World Wide Web*. WWW2004, May 17-22, 2004, New York, New York, USA. ACM 1-58113-844-X/04/0005 2004. Disponível em: http://cs.brown.edu/courses/csci2531/papers/www04-ntoulas.pdf. Acesso em: 22 set. 2021.

ROCKEMBACH, Moisés. Arquivamento da web: estudos de caso internacionais e o caso brasileiro. *RDBCI*: Revista Digital de Biblioteconomia e Ciência da Informação, v. 16, n. 1, p. 7-24, 2018.

ROCKEMBACH, Moisés. A web brasileira na Covid-19: arquivamento da web e preservação digital. *Liinc em Revista*, v. 17, n. 1, 2021. Disponível em: https://doi.org/10.18617/liinc.v17i1.5713. Acesso em: 17 de jul. 2022.

UNESCO. Carta sobre la preservación del patrimonio digital. In: UNESCO. Records of the General Conference: 32nd session, Paris, 29 september to 17 october 2003. Volume 1: resolutions. [Paris], c2004. p. 74-77. Disponível em: https://unesdoc.unesco.org/ark:/48223/pf0000133171.page=80. Acesso em: 11 ago. 2021.

UNESCO. *Directrices UNESCO/PERSIST sobre selección del patrimonio digital para su conservación a largo plazo*. [*S.l.*], 2016. Disponível em: https://www.ifla.org/files/assets/hq/topics/cultural-heritage/documents/persist-content-guidelines_es.pdf. Acesso em: 11 ago. 2021.

ARQUIVO NACIONAL

MINISTÉRIO DA
GESTÃO E DA INOVAÇÃO
EM SERVIÇOS PÚBLICOS

GOVERNO FEDERAL
BRASIL
UNIÃO E RECONSTRUÇÃO