

SEÇÃO 6

MEIOS E ESTRATÉGIAS DE ARMAZENAMENTO

MÓDULO 7

Gestão e preservação de documentos digitais

SEÇÃO 6

Meios e estratégias de armazenamento

Adaptação do Arquivo Nacional da Costa Rica

Versão 1, 2024

Este curso foi traduzido e adaptado pela Direção Geral do Arquivo Nacional da Costa Rica, em colaboração com a Seção de Arquivologia da Universidade da Costa Rica, a partir do material original de 2011 da Associação Internacional de Arquivos Francófonos, disponível online no Portal Internacional Arquivístico Francófono. Esclarece-se que podem existir variações em relação ao conteúdo original. Para acessar o material em francês, visite <https://www.piaf-archives.org/se-former/module-7-gestion-et-archivage-des-documents-numeriques>.



Conteúdo

Capítulo 1. Objetivo da seção	5
Capítulo 2. Suportes utilizados para a preservação digital	6
2.1. Gravações digitais e degradação com o passar do tempo.....	6
2.2. Tecnologias existentes	8
2.2.1. Magnético	8
2.2.2. Ótico.....	8
2.2.3. Flash	8
2.2.4. De estado sólido	9
2.2.5. Armazenamento na nuvem.....	9
2.3. Verificação de arquivos com soma de verificação.....	11
2.4. Obsolescência tecnológica.....	12
2.5. Condições ótimas para a vida útil de dispositivos em um centro de dados	14
2.6. Estratégias de preservação digital a longo prazo	15
2.6.1. Seleção de formatos abertos e sustentáveis ao longo do tempo	15
2.6.2. Migração e atualização periódica dos formatos	15
2.6.3. Uso de metadados oportunos e consistentes.....	15
2.6.4. Armazenamento seguro e redundante	15

Capítulo 3. Estratégias de armazenamento	16
3.1. Entidade “armazenamento” do modelo OAIS.....	17
3.2. Abstração da plataforma de hardware	18
3.3. Hierarquia de armazenamento	19
3.4. Características dos serviços de armazenamento.....	21
Capítulo 4. Políticas de armazenamento.....	22
4.1. Aplicação interna.....	22
4.2. Mutualização	22
4.3. Externalização (terceiros arquivistas).....	23
Capítulo 5. Casos de estudo	23
5.1. Projeto de arquivos da UNESCO.....	23
5.2. Projeto de digitalização do patrimônio cultural do Google	23
5.3. Europeia	24
Bibliografia	24

Capítulo 1. Objetivo da seção

Em última análise, será sempre o arquivista o responsável pela preservação da informação. Esta seção 6 é dedicada aos suportes e às estratégias de armazenamento, devendo proporcionar ao arquivista uma visão clara dos possíveis riscos e soluções. Na maioria dos casos, a aplicação das soluções ficará a cargo dos profissionais de informática, mas o arquivista deverá ser capaz de avaliar as vantagens e desvantagens das possíveis soluções e tomar decisões com pleno conhecimento de causa.

A seção é apresentada do ponto de vista da preservação, das sequências de bits e da integridade dessas sequências ao longo do tempo.

As soluções ou as estratégias de armazenamento que sejam adotadas se apoiarão evidentemente nas tecnologias disponíveis neste campo.

Entretanto, não se deve acreditar que se trata de uma questão puramente técnica. As decisões se basearão em uma análise das necessidades e limitações presentes e futuras do Arquivo e de seu contexto.

Estamos em um contexto de revolução digital que nos permite acessar meios de armazenamento mais amplos e acessíveis em comparação com décadas passadas, possibilitando o desenvolvimento das atividades de preservação de dados de forma mais eficaz.

Pergunta-se:

- Onde armazenar os dados? Em qual repositório? Quais meios utilizar? Também deve-se analisar como organizá-los, como preservá-los ao longo do tempo, garantindo que não se perderão e assegurando sua integridade.
- Como sobreviver aos incidentes materiais e ao envelhecimento dos suportes?
- Como garantir sua integridade ao transferir de um meio para outro?
- Como assegurar que os dados não sejam modificados nem lidos de forma não autorizada?

As infraestruturas de armazenamento deverão levar em consideração uma série de requisitos relacionados ao desempenho, aos tempos de acesso aos documentos, à segurança e, ao mesmo tempo, respeitar as limitações de custos.



GLOSSÁRIO

Arquivo: Organização responsável por conservar a informação para que uma comunidade de usuários-alvo possa acessá-la e utilizá-la (glossário do modelo de referência OAIS).

Assinatura digital (impressão digital, resumo ou hash): Resultado de uma função hash aplicada a uma cadeia de caracteres de qualquer comprimento para reduzi-la a um comprimento fixo que a represente. A assinatura digital é um dos elementos que permitem verificar a integridade de um documento, de um fluxo, de um lote, de uma transmissão (comparação de impressões digitais).

Petabyte: Unidade de medida de armazenamento cujo símbolo é representado pelo PB, equivale a 1000 terabytes.

Zettabyte: Unidade de medida de armazenamento cujo símbolo é representado pelo ZB, equivale a 1 milhão de petabytes e a 1 bilhão de terabytes.

Capítulo 2. Suportes utilizados para a preservação digital

Os suportes de preservação digital são os meios ou dispositivos utilizados para armazenar e manter os dados digitais a longo prazo. Entre eles, podemos encontrar discos ópticos, fitas magnéticas, discos rígidos magnéticos (HDD), discos de estado sólido (SSD) e o armazenamento em nuvem.

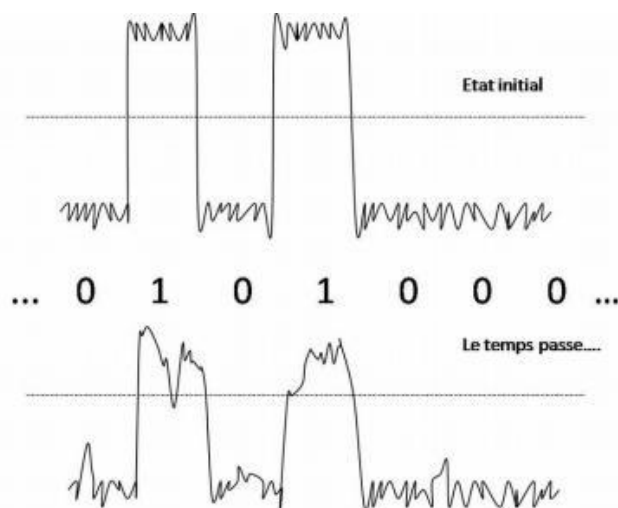
Cada um desses suportes possui características únicas relacionadas à sua longevidade, capacidade de armazenamento, escalabilidade, velocidade de acesso e recuperação dos dados. Deve-se levar em conta, além disso, seu custo e sustentabilidade ao longo do tempo.

A escolha do suporte a ser utilizado deve ser acompanhada de boas práticas de gestão e manutenção para garantir a integridade e acessibilidade a longo prazo dos dados armazenados, sem nenhum impedimento que coloque em risco a recuperação da informação.

2.1. Gravações digitais e degradação com o passar do tempo

A gravação de informação digital em um suporte, com o objetivo de memorizar ou conservar permanentemente essa informação, resulta na gravação nesse suporte de sequências de 0 e 1. Essa gravação é feita em uma forma codificada, e a codificação pode depender da tecnologia utilizada.

A realidade física implica a gravação de valores discretos no suporte, e essa gravação geralmente se degradará de forma progressiva ao longo do tempo.



Estado inicial / Tempo decorrido

Degradação do sinal registrado ao longo do tempo. Elaborado por PIAF

Na era digital, o armazenamento e a preservação dos dados experimentaram uma evolução significativa. As gravações digitais substituíram em grande parte sua contraparte analógica, oferecendo, entre muitos aspectos de melhoria, maior flexibilidade, qualidade e preservação ao longo do tempo. Apesar de suas vantagens, as gravações digitais estão longe de ser perfeitas. Como parte de seus desafios, enfrentam as condições do passar do tempo, por isso devemos levar em conta uma série de fatores que podem influenciar essa degradação:

- **Perda de metadados:** Os metadados básicos dos arquivos digitais, tais como a data de criação, localização, hora, entre outros, podem ser perdidos, dificultando a identificação e seus dados de referência.
- **Compressão e codificação:** A compressão e a codificação podem afetar a qualidade original dos objetos digitais, de acordo com os algoritmos de compressão utilizados, é possível obter um resultado de degradação e perda de qualidade maior do que o esperado para sua preservação em longo prazo.
- **Erros de armazenamento:** As gravações digitais podem ser corrompidas por erros de armazenamento em setores danificados de discos rígidos, erros de escrita e leitura em diversos meios de armazenamento, ou problemas de conexão durante a transferência na nuvem são outros exemplos.
- **Degradação física do suporte:** Embora os meios digitais não estejam sujeitos à degradação física dos meios analógicos, como as fitas magnéticas, ainda podem se deteriorar com o tempo. Por exemplo, memórias USB podem sofrer danos ao serem expostas à luz ultravioleta, à umidade ou a mudanças extremas de temperatura.
- **Obsolescência tecnológica:** Os formatos e dispositivos de reprodução refletem a rapidez com que podem se tornar obsoletos em um curto período de tempo. Por exemplo, mídias físicas como CDs ou DVDs podem se deteriorar fisicamente e dependem de um aparelho em bom estado para serem lidos — um dispositivo que recentemente tem sido eliminado dos equipamentos pessoais atuais.

2.2. Tecnologias existentes

Existem vários tipos de tecnologias para o suporte do armazenamento de dados, com suas distintas qualidades, entre eles estão:

2.2.1. Magnético

- **Fitas Magnéticas:** As tecnologias magnéticas são as mais antigas. Embora tenham surgido diversas opções para a gravação de dados, essa tecnologia continua vigente atualmente. Trata-se de uma fita revestida com material magnético que é enrolada em um carretel para armazenar dados de forma sequencial. A partir dos anos 1990, popularizou-se o uso das fitas LTO (Linear Tape-Open), que atualmente estão na sua versão número 9, com capacidade de armazenamento de até 45 TB por cartucho em formato comprimido.
- **Discos Rígidos (HDD):** Desde a sua existência em 1960 se tornou uma das principais tecnologias para a gravação de dados. Inicialmente, apresentava custo elevado e menor capacidade, mas atualmente é o recurso preferido dos centros de dados em todo o mundo, tendo se consolidado como uma solução de alta confiabilidade e escalabilidade.

Os HDD são conhecidos como discos rígidos tradicionais, estes dispositivos utilizam discos magnéticos giratórios para escrever e ler informação de maneira persistente. Esse dispositivo se destacava por ter uma grande capacidade de armazenamento, podendo alcançar até 20 TB. Essa tecnologia possui uma longa vida útil, suportando um grande número de ciclos de escrita e leitura antes de apresentar falhas. Além disso, os HDDs oferecem alta capacidade de armazenamento a um custo menor, em comparação com outros serviços ou tecnologias.

2.2.2. Óptico

- **Digital Versatil Disc (DVD):** Disco óptico que utiliza a tecnologia a laser para a gravação e leitura de dados em uma camada refletiva de plástico, embora, devido à crescente popularidade de tecnologias mais acessíveis e com maior capacidade, o uso do DVD tenha diminuído significativamente nos últimos anos, ele ainda continua sendo utilizado em certos setores tratando-se, no entanto, de uma alternativa atualmente em declínio.
- **Blue-ray Disc:** Similar ao DVD, o Blue-ray utiliza um laser de cor azul-violeta para a gravação e leitura de dados em uma camada de plástico de alta densidade, permitindo uma capacidade maior que seu antecessor, uma unidade padrão pode atingir 50 GB, e até 100 GB em suas versões de 4 camadas. Desde o seu nascimento em 2005, enfrenta grandes desafios por ser o principal sucessor dos discos ópticos, em um cenário de transição para mídias removíveis USB e armazenamento em nuvem, entretanto, ainda possui um nicho de mercado voltado para a qualidade de áudio e vídeo.

2.2.3. Flash

- **Memória USB:** A memória USB é um tipo de tecnologia para gravação de dados não volátil que pode ser utilizada em grande variedade de dispositivos eletrônicos, destacou-se amplamente por suas características de acesso rápido, durabilidade e baixo consumo de energia, sendo capaz inclusive de transmitir dados entre smartphones. Sua popularidade no uso pessoal fez com que se tornasse a tecnologia mais utilizada nos últimos anos, graças às suas vantagens em relação às mídias ópticas, à resistência a agentes como poeira ou arranhões, além de sua ampla capacidade, que atualmente pode chegar a 4 TB.

2.2.4. De estado sólido

- **Unidades de Estado Sólido (SSD):** As unidades SSD (Solid State Drive) são dispositivos de armazenamento que utilizam um tipo de chips de memória flash NAND, utilizado para a escrita e leitura de dados de maneira persistente, sua principal diferença em comparação aos discos rígidos tradicionais é que estes não utilizam discos magnéticos giratórios, dando como resultado tempos de acesso muito mais rápidos, assim como taxas de transferência mais altas. Alguns referem-se erroneamente a estes dispositivos como “discos”, entretanto, a diferença de seus predecessores, seus dados já não são armazenados em superfícies cilíndricas.

Entre as suas vantagens destaca-se:

- **Rendimento:** Maior capacidade de acesso e transferência de dados.
- **Eficiência energética:** Possuem um menor consumo que os discos rígidos tradicionais, aspecto que beneficia aos dispositivos portáteis tais como laptops ou tablets.
- **Confiabilidade:** Por não possuírem partes móveis em sua estrutura, os SSDs são menos suscetíveis a danos causados por impactos e vibrações.
- **Tamanho Compacto:** Devido ao seu design sem partes móveis, os SS são mais compactos e leves que os discos rígidos tradicionais.

Embora os SSD possuam várias vantagens em relação aos discos rígidos tradicionais, eles enfrentam alguns desafios, como limitações na capacidade de armazenamento, que costuma ser inferior à dos discos rígidos tradicionais, além da diferença de custo. Outro aspecto a ser considerado é a degradação desses dispositivos, que possuem um limite no número de ciclos de escrita e leitura.

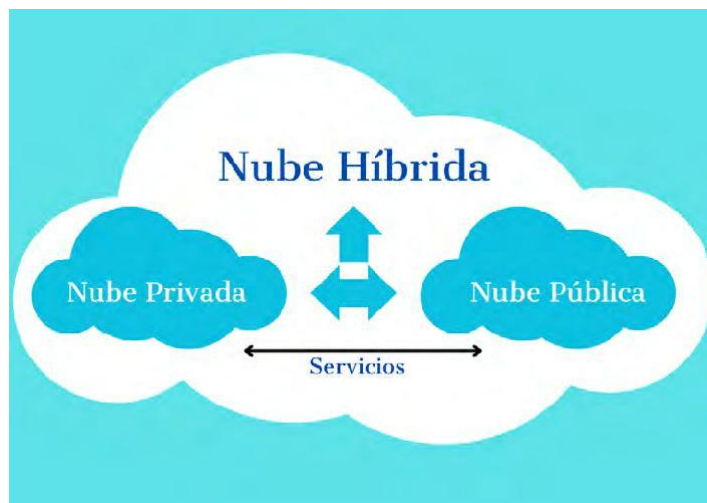
2.2.5. Armazenamento na nuvem

- **Armazenamento baseado na nuvem:** Trata-se de uma tecnologia de gravação de dados que ganhou grande popularidade nos últimos anos. Os dados são armazenados em servidores remotos na internet, de onde podem ser acessados por meio de diversos dispositivos inteligentes, evitando assim a necessidade de guardar os dados em dispositivos locais, como unidades de disco rígido ou unidades flash.

Entre as principais particularidades do serviço de armazenamento na nuvem encontra-se sua acessibilidade, permitindo aos usuários administrar seus dados de qualquer lugar, desde que haja conexão com a internet. A maioria desses serviços permite fazer backup automático da pasta associada à nuvem, além de realizar cópias de segurança e controle de versões, possibilitando a recuperação de arquivos eliminados por engano e a auditoria das atividades realizadas nesse espaço.

Existem três tipos principais de armazenamento na nuvem:

1. **Armazenamento em nuvem privada:** Quando são armazenados dados em um centro de dados privado, utilizada por uma única organização.
2. **Armazenamento em nuvem pública:** Quando são armazenados dados em um centro de dados público compartilhado entre várias organizações, este, por sua vez, é administrado por um fornecedor representante da marca.
3. **Armazenamento em nuvem híbrida:** Quando se faz uma combinação entre o armazenamento em nuvens públicas e privadas, normalmente os serviços são usados de forma integrada.



Representação de tipos de armazenamento em nuvem. Elaboração própria

A revolução digital impulsionou o uso dessa tecnologia nos últimos anos. Segundo um relatório (CyberSecurity-Ventures, 2020), até o ano de 2025 será atingido um armazenamento total de 100 zettabytes de dados na nuvem. Esta pesquisa também evidencia que 50% dos indivíduos utilizavam armazenamento em nuvem, seja para uso pessoal ou profissional, é por isso que as organizações têm voltado sua atenção para o uso do armazenamento em nuvem como parte de sua infraestrutura.

Esse serviço dependerá de uma conexão estável à internet, pois, embora existam serviços de armazenamento em nuvem que realizem sincronização automática das alterações recentes, a quantidade de mudanças pendentes para sincronizar pode aumentar o tempo desse processo, que pode ser limitado conforme a largura de banda da conexão do dispositivo.

Estas tecnologias de armazenamento de dados possuem suas próprias características, sua escolha dependerá do tipo de dado que você precisar armazenar, entre vários fatores, tais como:

- **Confiabilidade:** Capacidade da tecnologia de armazenamento para evitar a perda ou corrupção dos dados.
- **Durabilidade:** Capacidade dos meios para conservar o armazenamento dos dados durante um prolongado período de tempo.
- **Segurança:** A segurança dos dados é crucial; deve-se escolher uma tecnologia de gravação que ofereça proteção adequada contra vazamentos de dados, roubos ou acessos não autorizados.
- **Custo:** As diferentes tecnologias de armazenamento possuem diferentes custos, é importante escolher uma tecnologia que se adapta ao estado atual da nossa organização, mantendo a melhor relação custo-benefício.
- **Compatibilidade:** O tipo de tecnologia a ser utilizada pode ser incompatível com certos formatos ou outras tecnologias. Por exemplo, a migração de dados em a migração de fitas magnéticas para armazenamento em nuvem pode ser complicada devido à incompatibilidade e ao tempo necessário para transferir os dados.
- **Escalabilidade:** As necessidades organizacionais podem variar com o tempo, é importante selecionar uma tecnologia de armazenamento que seja escalável e que seja adaptável às possíveis crescentes necessidades de armazenamento dos dados.

- **Acessibilidade** A acessibilidade é um fator crítico a ser considerado, a tecnologia de armazenamento a ser utilizada deve permitir uma fácil recuperação dos dados quando necessário, sem comprometer sua integridade.



COMPLEMENTO

O meio digital alemão DW publicou uma notícia em fevereiro de 2024 intitulada: Cientistas chineses desenvolvem disco óptico de 1,6 petabytes, fazendo referência ao possível ressurgimento dessa tecnologia de gravação de dados.

<https://www.dw.com/es/cient%C3%ADficos-chinos-desarrollan-disco-%C3%B3ptico-de-16-petabytes/a-68338903>

O site espanhol Hardzone publicou, em julho de 2023, uma reportagem sobre as memórias NAND, com detalhes sobre seu funcionamento e principais características.

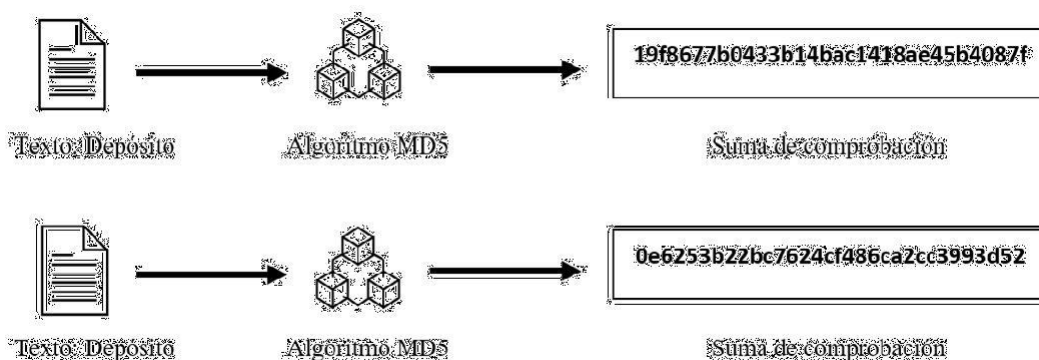
<https://hardzone.es/reportajes/que-es/memoria-nand-flash/>

2.3. Verificação de arquivos com checksum (soma de verificação em português)

A soma de verificação é uma sequência de caracteres usada para referenciar de maneira única um arquivo. Este procedimento é utilizado para verificar se um arquivo conserva sua integridade de dados durante sua transmissão ou se ele foi alterado. Esse valor é gerado de forma que a mínima alteração nos dados do arquivo resulte em uma soma com valor totalmente diferente.

Para realizar uma soma de verificação, deve-se processar o arquivo utilizando um algoritmo adequado, resultando em uma sequência alfanumérica única, comumente chamada de “checksum” ou soma de verificação.

Os algoritmos de soma de comprovação mais comuns são SHA-1, SHA-256, MD5, entre outros, podem ser utilizadas em arquivos por meio de aplicações online, assim como em versões para desktop.



Exemplo de soma de comprovação com resultados diferentes. Elaboração própria a partir de PIAF

Uma vez obtida a soma de verificação do arquivo original, deve-se realizar um segundo processo com o arquivo a ser comparado, ao contrastar ambas as sequências de caracteres, é possível determinar se o arquivo copiado sofreu alguma alteração e, se desejado, investigar a causa dessa discrepância.

As somas de comprovação são utilizadas em cenários como:

- Verificação da integridade de arquivos em um sistema de armazenamento e suas cópias de segurança.
- Verificação da integridade de arquivos durante a transferência de dados (envio por e-mail ou arquivos na nuvem)
- Verificação da integridade de arquivos baixados da internet e distribuição de software.
- Auditorias de segurança a sistemas ou sistema de armazenamento.

Em resumo, a soma de verificação possui uma ampla gama de aplicações para a verificação da integridade e autenticidade dos arquivos. Atualmente trata-se de uma prática comum para garantir a segurança e confiabilidade dos arquivos em um repositório digital.

2.4. Obsolescência tecnológica

A obsolescência dos suportes digitais é um problema de grande índole na indústria da tecnologia, informação e comunicação. Refere-se ao risco ao qual os dados armazenados em um formato específico estão expostos, diante da possibilidade de se tornarem inacessíveis devido à tecnologia atual associada a esse formato.

As causas deste fenômeno podem ser variadas, entre elas podemos destacar:

- **Mudanças nos padrões da indústria:** Os padrões de armazenamento e comunicação, como os tipos de dispositivos, interfaces de conexão e protocolos de transferência, podem variar com o tempo. Se um arquivo depende de um padrão que se torna obsoleto, os dispositivos e softwares modernos podem não ser compatíveis com esse arquivo.
- **Formatos de arquivo obsoletos:** À medida que a tecnologia evolui, é possível que novos formatos de arquivo sejam estabelecidos, com maior qualidade, compressão ou eficiência, fazendo com que os formatos mais antigos se tornem obsoletos ou apresentem dificuldades para a manipulação ou leitura dos dados armazenados.
- **Degradação física dos meios:** Todos os meios de armazenamento têm uma vida útil limitada e podem deteriorar-se ao serem expostos a certos fatores como a exposição ao calor, umidade ou luz ultravioleta. Isto pode resultar em uma perda dos dados e na obsolescência dos meios.
- **Descontinuação de produtos e serviços:** Em algumas ocasiões, os fabricantes deixam de produzir dispositivos de armazenamento e softwares compatíveis com certos tipos de mídias ou formatos digitais. Isso pode fazer com que os formatos se tornem obsoletos e difíceis de manter no futuro.



EXEMPLO

- O formato digital MPEG-2 (Moving Picture Experts Group -2) "oi, em certo momento, o padrão de compressão de vídeo e áudio digital, desenvolvido em 1995 como uma melhoria de seu antecessor, MPEG-1, foi amplamente utilizado para a televisão digital, DVDs e a transmissão ao vivo por meio da internet. Embora este não tenha se tornado totalmente obsoleto, seu uso diminuiu devido à criação de formatos mais modernos como o H.264 (AVC) e H.265 (HEVC).
- O formato Windows Media Audio (WMA), desenvolvido pela Windows, é outro claro exemplo da obsolescência em formatos digitais. Embora tenha sido amplamente utilizado décadas atrás, foi substituído por formatos mais populares e universais, como o MP3 e o AAC.

Com o objetivo de mitigar os riscos associados à obsolescência nos suportes digitais, é importante tomar medidas proativas e estar atualizado com as informações mais recentes sobre o assunto. Realizar tarefas como a migração para formatos ou meios modernos e gerenciar cópias de segurança desses arquivos fazem parte das alternativas recomendadas, além da adoção de formatos padrão para preservação a longo prazo ou de software livre, tais como o PDF/A, TIFF ou MKV.



COMPLEMENTO

Exemplos de formatos digitais de preservação a longo prazo

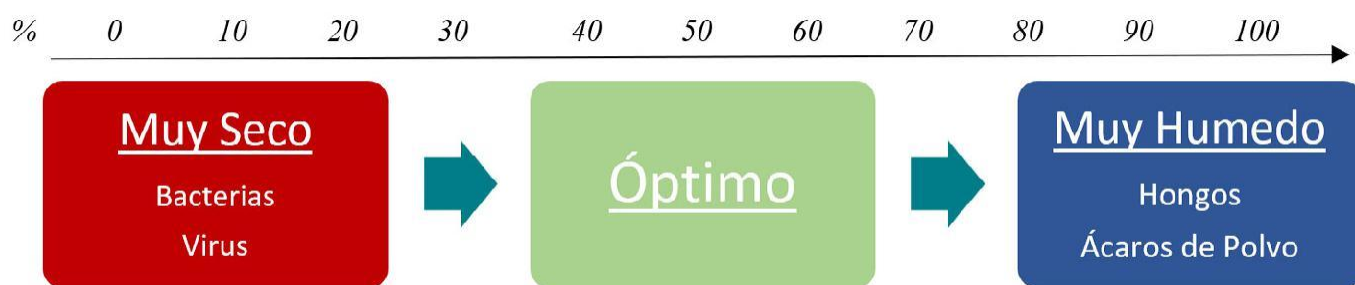
Formato	Descrição	Vantagens	Limitantes
PDF/A	Formato PDF projetado para a preservação a longo prazo de documentos eletrônicos.	<ul style="list-style-type: none"> • Ampla Compatibilidade. • Proteção contra inserção de URL, código Javascript, multimídia, entre outros. 	<ul style="list-style-type: none"> • Dependência de fontes incorporadas no documento. • Limitações com características dinâmicas presentes em alguns PDF padrão
TIFF	Formato de imagem sem perdas, projetado para o armazenamento de grande qualidade.	<ul style="list-style-type: none"> • Conserva a qualidade da imagem original. • Suporta metadados descritivos. 	<ul style="list-style-type: none"> • Problemas de armazenamento derivado do grande tamanho dos arquivos.
FLAC	Formato de compressão de áudio sem perdas para a preservação da qualidade original.	<ul style="list-style-type: none"> • Possui grande compatibilidade com os distintos reprodutores de áudio. • Conserva a qualidade do som original. 	<ul style="list-style-type: none"> • É possível que os arquivos FLAC sejam mais pesados que outros formatos de compressão.
MKV	Formato de compressão utilizado para armazenar áudio, vídeo e legendas.	<ul style="list-style-type: none"> • Suporta metadados descritivos. 	<ul style="list-style-type: none"> • Possível incompatibilidade com alguns reprodutores de mídias.

2.5. Condições ótimas para a vida útil de dispositivos em um centro de dados

A vida útil dos recursos físicos em um centro de dados está diretamente ligada às condições e medidas de proteção que eles têm. Um ambiente deficiente pode reduzir rapidamente a vida útil dos dispositivos, causando danos catastróficos e irreparáveis se não for detectado e corrigido em tempo hábil.

A seguir, são apresentadas algumas recomendações que visam prolongar a vida útil dos dispositivos nos centros de dados:

- **Temperatura:** Manter uma temperatura estável entre 18 e 27 °C; deve-se evitar tanto o superaquecimento quanto o resfriamento excessivo, pois essas condições podem causar danos aos dispositivos.
- **Controle de poeira e outras partículas:** Deve-se minimizar a quantidade de poeira e partículas no ar, com o objetivo de prevenir que eles se acumulem nos dissipadores e em outros componentes dos dispositivos; a instalação de filtros de ar é uma solução eficaz para esse ponto.
- **Alimentação elétrica estável:** Os dispositivos dos centros de dados requerem uma alimentação elétrica estável e de grande fidelidade para funcionar de maneira correta. A instalação de reguladores de voltagem, fontes de alimentação redundantes e UPS de dupla conversão são medidas eficazes para proteger os equipamentos contra quedas de energia e picos de tensão.
- **Umidade relativa:** A umidade relativa evita o acúmulo de carga eletrostática e a corrosão dos componentes elétricos dos dispositivos, assim como a geração de descargas eletrostáticas causadas pela baixa umidade. Recomenda-se manter uma umidade relativa entre 40% e 60% no ambiente do centro de dados, a instalação de um desumidificador é útil para controlar esse parâmetro.



Estados de acordo com o percentual de umidade relativa em um centro de dados. Elaboração própria

- **Monitoramento e manutenção:** Deve-se realizar um monitoramento contínuo tanto das condições ambientais da área quanto do desempenho e dos alertas emitidos pelos equipamentos ou seus componentes.

Seguir essas recomendações contribui para aumentar a vida útil dos dispositivos nos centros de dados, é importante antecipar e lidar com os riscos que possam comprometer a funcionalidade do ambiente, assim como adotar boas práticas, como a aquisição de peças de reposição ou consumíveis de reserva, a fim de aplicar as correções necessárias no menor tempo possível.

2.6. Estratégias de preservação digital a longo prazo

A preservação dos recursos digitais é um processo de extrema importância, não basta ter um plano de criação e gestão de backups das informações, é necessário garantir a acessibilidade e a integridade desses meios ao longo do tempo.

Algumas estratégias recomendadas para alcançar este objetivo são:

2.6.1. Seleção de formatos abertos e sustentáveis ao longo do tempo

Dar prioridade ao uso de formatos de arquivos abertos, que sejam reconhecidos por diversas entidades como amplamente compatíveis e com baixa probabilidade de se tornarem obsoletos em curto prazo. Podem ser considerados como parte desses formatos os observados na tabela: Exemplos de formatos digitais de preservação a longo prazo, do capítulo 2.4.

Quanto aos padrões relacionados a este ponto, a norma ISO 19005 – Gestão de documentos – Formato de arquivo de documento eletrônico para preservação a longo prazo, define as diretrizes para a criação e o uso de formatos de arquivos eletrônicos que são ideais para a preservação de documentação a longo prazo. Esta norma especifica nas suas diferentes seções, o uso do formato PDF/A e as características das suas variantes.

2.6.2. Migração e atualização periódica dos formatos

Devem ser implementados processos de migração e atualização de formatos para garantir que os recursos digitais permaneçam acessíveis à medida que evoluem os novos padrões e tecnologias.

2.6.3. Uso de metadados oportunos e consistentes

Adicionar ao plano de trabalho a aplicação e manutenção de metadados descritivos e técnicos para os recursos digitais gerenciados pela organização. Esses metadados facilitarão a identificação, a busca e o contexto geral dos arquivos digitais ao longo do tempo, contribuindo para sua preservação e leitura.

Como marco básico para a implementação de metadados, o conjunto Dublin Core¹, baseado na norma ISO 15836:2009 – Informação e documentação, nos fornece diretrizes para a implementação de metadados descritivos para recursos digitais, como documentos, imagens, vídeos, sites, entre outros. Esses elementos fornecem informações básicas por meio de etiquetas sobre o recurso, como título, autor, data, tipo de conteúdo, entre outros, facilitando sua busca, gestão e recuperação, componentes fundamentais para sua preservação a longo prazo. (Ver seção 9 dedicada aos Metadados)

2.6.4. Armazenamento seguro e redundante

Recomenda-se o uso de sistemas de armazenamento seguro e redundante para a proteção dos recursos digitais, devendo-se priorizar a garantia dos arquivos contra condições que os exponham a riscos de corrupção de dados, perdas parciais ou totais, além de acessos não autorizados.

1 <https://www.dublincore.org/>

É possível utilizar soluções físicas para armazenamento distribuído, o uso de discos rígidos redundantes e até mesmo serviços de armazenamento em nuvem, que ofereçam as características de segurança e acessibilidade necessárias para os recursos.

A norma ISO 16363:2017-Sistemas de transferência de informação e dados espaciais - Auditoria e certificação de repositórios digitais de confiança, estabelece os requerimentos para auditar e certificar os repositórios digitais confiáveis para a preservação de dados a longo prazo. Essa norma avalia o cumprimento dos padrões necessários que garantam a acessibilidade, autenticidade e integridade dos recursos digitais armazenados em um repositório de longo prazo.

Entre os aspectos que contempla esta norma encontram-se:

- As definições e termos-chave relacionados à preservação dos dados e aos repositórios digitais confiáveis.
- Requisitos gerais para a administração de repositórios digitais, incluindo políticas e procedimentos.
- Requisitos para a gestão de dados e metadados, incluindo a documentação dos recursos digitais e a preservação dos metadados técnicos e descritivos.
- Procedimento de auditoria e certificação para avaliar a conformidade de um repositório com os requisitos estabelecidos pela norma.

Em conclusão, ao combinar essas estratégias, é possível desenvolver uma abordagem integral baseada em normativas atuais, que nos ofereça os passos iniciais para a preservação dos recursos digitais a longo prazo, garantindo de forma proativa sua integridade e acessibilidade para consulta pelos usuários.

Capítulo 3. Estratégias de armazenamento

Armazenar é apenas uma das funções do arquivo e não deve ser confundido. Se o armazenamento é um elemento essencial do processo, em nenhum caso constitui o único componente de uma estratégia de sustentabilidade, mas é, sim, a base sobre a qual o arquivo se fundamenta.

O que se espera de um serviço de armazenamento dentro de um arquivo digital?

É uma infraestrutura capaz de garantir a restituição da informação.

- Sem alterações (integridade, autenticidade)
- A longo prazo (migração, obsolescência)
- Em um problema de custo controlado

Em primeiro lugar, vamos revisar com mais detalhes o que o modelo OAIS diz sobre a entidade funcional “Armazenamento”.

3.1. Entidade “armazenamento” do modelo OAIS

A entidade “armazenamento” do modelo OAIS assegura as seguintes funcionalidades:

1. Receber os AIP (pacotes de informação do arquivo)

Registrar em uma área de armazenamento adequada e enviar um certificado de que o armazenamento foi realizado.

2. Gerenciar a hierarquia do armazenamento

Com base nos requisitos de qualidade dos serviços das entidades de pagamento e acesso, e considerando a frequência e o uso previsto do AIP, serão selecionados os suportes adequados que permitam o acesso online, em diferido ou em acesso levemente diferido. Deverá ser verificado previamente que os AIP não foram alterados durante a transferência.

Esta função também fornece estatísticas sobre os meios disponíveis e a capacidade de armazenamento disponível nas diferentes camadas, assim como sobre a utilização dos AIP.

3. Garantir a migração dos suportes

Esta função oferece a possibilidade de reproduzir os AIP ao longo do tempo. Estas são as migrações tecnológicas. Durante as operações, o conteúdo informativo e a informação perene não devem sofrer modificações.

Entretanto, os dados que constituem a informação de empacotamento podem ser alterados, desde que continuem desempenhando a mesma função. A estratégia de migração consistirá em escolher um meio de armazenamento levando em conta as taxas reais e esperadas de erro que caracterizam os diferentes tipos de suportes, seu desempenho e seu custo de aquisição.

4. Controle periódico da integridade da informação confiada

Esta função garante, com uma probabilidade estatisticamente aceitável, que nenhum AIP foi corrompido em uma transferência interna dos dados da entidade “armazenamento”. A informação de integridade assegura, em certa medida, que o conteúdo da informação não foi modificado ao longo do tempo, independentemente de o AIP ter sido transferido de um meio para outro, e independentemente de esse AIP ter sido acessado ou não.

5. Regenera os dados em caso de sinistro

O “plano de recuperação” proporciona um mecanismo para duplicar o conteúdo digital do Arquivo e armazenar a cópia em uma instalação geograficamente distante.

6. Fornece os AIP para a entidade de acesso.

Esta função transmite à Entidade “Acesso” as cópias dos AIP armazenados. Esta função recebe uma solicitação de AIP especificando os AIP requisitados e os registros sobre o tipo de suporte solicitado, ou os transfere para uma área de armazenamento provisória.

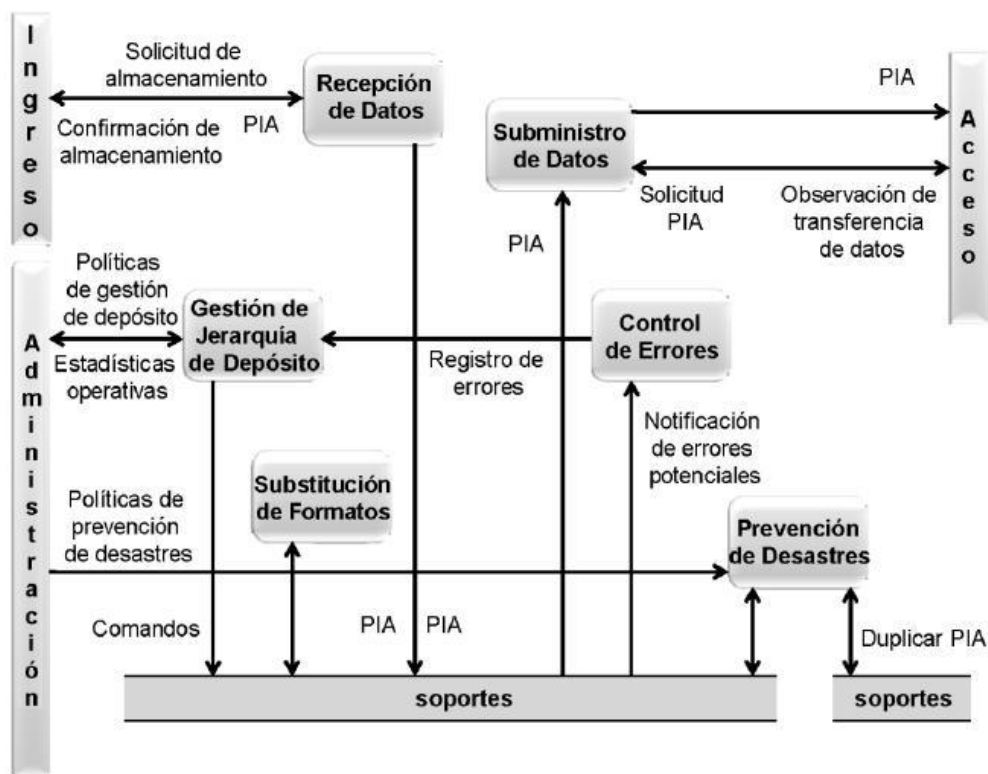


Figura 4-3 – Funciones de la Entidad Funcional Almacenamiento de Archivo

ISO 14721 (2015) p.46. Funções da entidade funcional: armazenamento

Vamos encontrar todas estas características na implementação de um serviço de armazenamento.

3.2. Abstração da plataforma de hardware

O armazenamento assume a forma de um conjunto de meios organizados em sistema. Um “sistema de armazenamento” é um conjunto de software e hardware que proporcionam um serviço plenamente definido.

Sabendo que a plataforma vai evoluir regularmente, já que é constituída por hardware e software que irão mudar com o tempo, será necessário gerenciar operações de monitoramento e renovação dos meios de armazenamento, parece haver um grande interesse em constituir essa plataforma de forma autônoma em comparação com o restante do sistema de arquivo digital.

Essa abordagem pragmática baseia-se principalmente na experiência das plataformas existentes.

Com uma abordagem desse tipo:

- As evoluções da plataforma de armazenamento não devem afetar a organização logística do arquivo: o produtor dos dados e também o arquivo gerenciam uma organização lógica dos dados e a capacidade de transformar esses dados em informação inteligível. Essa organização lógica não corresponde à organização da arquitetura de armazenamento utilizada, organização que pode ser migrada ao longo do tempo em função das tecnologias e dos serviços disponíveis.

- As mudanças na política da hierarquia de armazenamento (classes de serviço) e as migrações da arquitetura base do repositório devem ser transparentes, ou seja, não devem impactar as entidades usuárias do serviço de armazenamento.

Portanto, os usuários (também chamados clientes) do serviço de armazenamento, não necessitam saber sob qual arquitetura de armazenamento seus dados estão protegidos. Entretanto, precisam dispor das garantias do serviço que lhes é oferecido nos termos de sustentabilidade, integridade e acessibilidade.

Trata-se de uma abordagem de serviço. É, por exemplo, o que esperamos de um serviço de entregas, esperamos que nosso pacote chegue ao destinatário, em bom estado e no tempo previsto. Não importam os meios de transporte que o fornecedor do serviço utilize para transportar o pacote. Esse design também facilita a compartilhamento, ou seja, a utilização do serviço de armazenamento por diferentes grupos, departamentos, projetos dentro da organização e até mesmo entre organismos associados.



COMPLEMENTO

A NASA utiliza uma combinação de sistemas de armazenamento em nuvem, sistemas de armazenamento de alto desempenho e sistemas de armazenamento de longo prazo para gerenciar seus dados científicos e de pesquisa de missões espaciais, entre outros meios digitais. Para isso, eles utilizam o Centro de Distribuição da Terra (EOSDIS)² na sigla em inglês, para armazenar e distribuir essas informações.

No artigo intitulado “Earth Science Data in the Cloud: The EOSDIS Cumulus Project” (Blumenfeld, 2017) comenta-se sobre a evolução deste projeto e a transição desses serviços para a plataforma em nuvem, Blumenfeld também menciona o estado atual do armazenamento dos dados coletados, que se estimava em 22 petabytes (PB) até aquela data, com uma expectativa de crescimento para até 247 PB em 2025, evidenciando a enorme quantidade de dados processados e coletados como resultado das operações deste programa da NASA.

3.3. Hierarquia de armazenamento

Levando em consideração que nem todos os recursos digitais têm a mesma importância para a organização e que alguns são consultados com mais frequência do que outros, a hierarquia de armazenamento nos oferece uma abordagem estratégica que possibilita classificar os recursos digitais em diferentes níveis de armazenamento, otimizando a preservação a longo prazo sem comprometer a acessibilidade ou a segurança dos arquivos protegidos.

Sob essa hierarquia, os recursos digitais são organizados em diferentes níveis de armazenamento, desde o mais seguro e duradouro até o de menor custo e com maior suscetibilidade a falhas. Poderíamos optar, na parte superior da hierarquia, por meios físicos de armazenamento de alta durabilidade e qualidade, tais como discos rígidos ou fitas magnéticas. Em contraste, na parte inferior da hierarquia poderíamos colocar o armazenamento em nuvem a longo prazo, que oferece acessibilidade e flexibilidade, mas pode não ser tão durável e normalmente está associado a

² <https://www.earthdata.nasa.gov/>

um custo elevado para sua manutenção a longo prazo.

Alguns elementos disponíveis para a escolha de uma hierarquia de armazenamento são os seguintes:

- **Armazenamento em arquivo morto:** É comumente considerado o nível mais baixo da hierarquia, consistindo no uso de meios físicos de armazenamento de longa duração. Esses meios são utilizados para armazenar recursos digitais que não são acessados com frequência, mas que devem ser conservados por longos períodos de tempo.
- **Armazenamento na nuvem a longo prazo:** Alguns provedores de serviço em nuvem oferecem soluções de armazenamento projetadas especificamente para a preservação a longo prazo dos recursos digitais, com características como criptografia de ponta a ponta, detecção e correção de erros e redundância de dados.
- **Arquivos digitais em repositórios específicos:** Esse nível implica a utilização de repositórios digitais especializados na preservação a longo prazo dos recursos digitais, estes podem ser projetados segundo o esquema do modelo OAIS, como o conhecido software livre archivematica.³ desenvolvido pela empresa Artefactual.
- **Armazenamento em base de dados de preservação digital:** Algumas organizações optam pelo armazenamento dos recursos digitais em bases de dados especializadas, que podem ser desenvolvidas localmente segundo os padrões existentes para a preservação digital a longo prazo.
- **Armazenamento em repositórios de arquivos institucionais:** Em alguns casos, as instituições culturais e estatais podem estabelecer repositórios de arquivos institucionais para a preservação e acesso a longo prazo dos recursos digitais importantes para a organização.



Exemplo de princípio de hierarquia de armazenamento. Elaboração própria a partir de PIAF

É importante considerar que a ordem desses níveis de armazenamento é uma estratégia integral de preservação.

digital a longo prazo. Essa escolha em particular dependerá de diversos fatores, como a natureza dos recursos digitais, os requisitos de preservação, o orçamento disponível e as diferentes regulamentações às quais a organização deve se submeter.

3.4. Características dos serviços de armazenamento

As características de armazenamento de serviços se concentram em atender às necessidades dos usuários e organizações em termos de armazenamento e administração de seus dados; entre as mais básicas, podemos mencionar:

- Capacidade de armazenamento escalável
- Acesso remoto
- Sincronização de arquivos
- Cópias de segurança automática
- Segurança e criptografia

Os serviços de armazenamento geralmente oferecem opções para definir permissões de acesso granulares, permitindo que os proprietários dos arquivos gerenciem quem pode visualizar, editar ou comentar os arquivos compartilhados. Dessa forma, garante-se que a informação sensível esteja protegida e que apenas os usuários desejados possam ter acesso.

A colaboração e o compartilhamento de arquivos são elementos fundamentais normalmente oferecidos por esse serviço, permitindo que os usuários trabalhem conjuntamente em documentos e arquivos de forma remota e em tempo real, adicionem comentários e façam notificações aos participantes. Essa característica permite o gerenciamento de versões e a restauração de arquivos, ferramenta útil para rastrear mudanças, ou recuperar dados perdidos.

O aspecto multiplataforma também é relevante, sendo atualmente primordial o acesso aos dados a partir de diferentes sistemas operacionais, seja em desktops (Windows, macOS, Linux) ou em dispositivos móveis (Android, iOS), permitindo a maior parte de suas funcionalidades de forma transparente.

A integração com aplicativos de terceiros é um recurso complementar que apoia a administração dos recursos dos usuários em suas tarefas diárias; algumas integrações podem incluir:

- Clientes de e-mail
- Ferramentas de colaboração
- Ferramentas de edição e design
- Gerenciadores de senhas
- Ferramentas de análise e visualização de dados

Essas ferramentas permitem que os usuários aproveitem ao máximo seu armazenamento, trabalhando de maneira eficiente e colaborativa de forma integrada.

Capítulo 4. Políticas de armazenamento

Podemos mencionar três políticas de armazenamento:

- A aplicação interna
- A mutualização
- A externalização

4.1. Aplicação interna

A aplicação interna oferece à organização (uma empresa, uma instituição...) o controle completo da plataforma de armazenamento e de sua evolução. Essas liberdades de decisão permitem adaptar o conjunto das disposições e escolhas às necessidades específicas e rigorosas dessa organização.

Em contrapartida, essa aplicação exige recursos humanos e financeiros significativos, além do desenvolvimento de competências especializadas.

Esta aplicação interna apenas é desejável para grandes empresas ou instituições com suficiente recurso para manter e atualizar a plataforma ao longo dos anos.

4.2. Mutualização

A mutualização de um serviço de armazenamento duradouro entre vários sócios oferece vantagens importantes:

- Divisão de custos.
- Importantes economias de escala: o custo médio anual de armazenamento de 1 GB diminui significativamente em função do volume armazenado.
- Uma troca de competências que permita aumentar a segurança e a confiabilidade do armazenamento.

É possível estabelecer um armazenamento mutualizado segundo critérios geográficos ou institucionais, independentemente da lógica do negócio e dos conteúdos.

A virtualização oferece toda uma gama de ferramentas para segmentar a plataforma e obter o máximo aproveitamento dos recursos mutualizados.

Exemplos:

- Possibilidade de considerar, ao nível de um departamento francês, a gestão de todos os dados digitais por uma única entidade (processos ativos e arquivos correntes tratados pelos serviços administrativos e processos arquivados geridos pelos serviços de arquivo).
- Outra possibilidade a nível interno: armazenamento seguro e centralizado dos dados e documentos de pesquisa das universidades.
- Existência nos Estados Unidos do San Diego Supercomputer Center (SDSC), que federou o armazenamento de dados de 40 universidades, incluindo a Universidade da Califórnia.
- O padrão «Digital Library Control Systems o projeto CLARIN (Common Language Resources and Technology Infrastructure)⁵ para os recursos da linguagem seguem nessa direção, mas com um espectro que vai muito mais além

4 <https://www.sdsc.edu/>

5 <https://www.clarin.eu/>

da questão do armazenamento.

O principal risco da mutualização é a dificuldade de encontrar parceiros comprometidos com a proposta.

4.3. Externalização (terceiros arquivadores)

Essa via permite recorrer a agentes com competências e recursos. Facilita a economia de custos de escala, considerando a importância dos custos fixos (pessoal, material). Entretanto, é preciso considerar a questão da continuidade do provedor do serviço e da segurança da rede que permite transmitir ou recuperar dados. Isso pode incluir desde o armazenamento de dados na nuvem, uso de serviços de hospedagem gerenciada, contratação de serviços externos de gestão de dados, entre outros.

A norma ISO 14641 intitulada “Sistemas de gestão de documentos – Aplicação da gestão de documentos eletrônicos – Princípios e requisitos” aborda, em seu conteúdo, as diretrizes de armazenamento e preservação, incluindo a gestão do armazenamento, backups, migração de formatos e medidas de segurança, como a proteção de dados.

Capítulo 5. Casos de estudo

Esses casos têm como objetivo destacar alguns arquivos digitais patrimoniais que conseguiram estabelecer com sucesso repositórios de preservação a longo prazo, servindo como referência para outras instituições e organizações que iniciaram projetos dessa índole:

5.1. Projeto de arquivos da UNESCO⁶

Estabelecido em 1992 pela UNESCO (United Nations Educational, Scientific and Cultural Organization) como uma iniciativa para a preservação e divulgação do patrimônio documental do mundo. Este projeto foca na importância de preservar a diversidade cultural e o patrimônio documental da humanidade, oferecendo acesso em benefício das gerações presentes e futuras. Entre esses recursos estão bibliotecas digitais, arquivos audiovisuais, documentos digitalizados, entre outros materiais de valor cultural e histórico.

5.2. Projeto de digitalização do patrimônio cultural do Google

A empresa Google tem colaborado com organizações como museus, bibliotecas e arquivos ao redor do mundo. Tem como objetivo difundir o patrimônio cultural por meio do público global da internet. Este projeto inclui várias iniciativas e programas que foram desenvolvidos ao longo dos anos; entre eles, podemos encontrar:

- **Google Arts & Culture⁷:** Plataforma online que oferece acesso a um amplo catálogo de conteúdo cultural e patrimonial.
- **Google Street View⁸:** Função do Google Maps que permite aos usuários explorar lugares de interesse cultural e patrimonial por meio de imagens panorâmicas em 360 graus, incluem-se museus, bibliotecas, documentação.

6 <https://unesdoc.unesco.org/archives>

7 <https://artsandculture.google.com/>

8 <https://www.google.com/intl/es/streetview/>

histórica, entre outros.

5.3. Europeana9

É uma plataforma online que oferece acesso a milhões de recursos digitais de museus, bibliotecas, arquivos e galerias da Europa. Fundada em 2008 pela Comissão Europeia, hospeda gratuitamente uma ampla variedade de recursos digitais, incluindo materiais audiovisuais e documentos do patrimônio artístico e cultural da Europa.

Bibliografia

Asociación Española de Normalización y Certificación UNE-ISO 14721. (2015). Sistemas de transferencia de datos e información espaciales Sistema abierto de información de archivo (OAIS) Modelo de referencia. Recuperado de Bases de Datos del SIBDI: AENORMÁS.

BANAT-BERGER F., HUC C., DUPLOUY L., *L'Archivage numérique à long terme, les débuts de la maturité?* (Primera obra de síntesis sobre el archivo digital en lengua francesa) Paris, La Documentation française, 2009

BANAT-BERGER F., HUC C., Module 7 - Gestion et archivage des documents numériques. Portail International Archivistique Francophone. 2011. <https://www.piaf-archives.org/se-former/module-7-gestion-et-archivage-des-documents-numeriques> (Se identifica en el texto como PIAF)

Blumenfeld, J. (18 de Julio de 2017). earthdata.nasa.gov. Obtenido de <https://www.earthdata.nasa.gov/learn/articles/earth-science-data-cloud-eosdis-cumulus-project>

CyberSecurity-Ventures. (2020). cybsecurityventures.com. Obtenido de The 2020 Data Attack Surface Report: <https://cybersecurityventures.com/wp-content/uploads/2020/12/ArcserveDataReport2020.pdf>

IASA. (s.f.). Sistema abierto de archivo de información (OAIS). Obtenido de <https://www.iasa-web.org/tc04-es/618-sistema-abierto-de-archivo-de-informaci%C3%B3n-oais>

