

Preservação de Fotografias Nato-digitais

Millard Schisler

Cadernos Técnicos de Conservação Fotográfica

Preservação de Fotografias Nato-digitais

Millard Schisler

Cadernos Técnicos de Conservação Fotográfica

Presidente da República

Luiz Inácio Lula da Silva

Ministra da Cultura

Margareth Menezes

FUNDAÇÃO NACIONAL DE ARTES – FUNARTE

Presidenta

Maria Marighella

Diretor Executivo

Leonardo Lessa

Diretora de Projetos

Laís Almeida

Coordenador de Conteúdo, Pesquisa e Formação

Carlos Eduardo Drummond

Coordenadora do Centro de Documentação

Joelma Neris Ismael

Chefe da Divisão de Preservação

Isabel Mendes

Para adquirir nossas publicações, envie e-mail para a Livraria Mário de Andrade: livraria@funarte.gov.br.

Alguns de nossos títulos estão disponíveis para download gratuito:

<https://www.gov.br/funarte/pt-br/assuntos/edicoes-1>

FUNDAÇÃO NACIONAL DE ARTES
funarte

MINISTÉRIO DA
CULTURA

GOVERNO FEDERAL
BRASIL
UNIÃO E RECONSTRUÇÃO

Equipe de Edições

Caroline Cantanhede

Cristiane Marinho

Gilmar Miranda

Júlio Machado

Maria José de Sant'Anna

Rosilene Alves da Rocha

Produção Executiva

Isabel Mendes

Kadiana Raposo

Preparação de originais

Equipe de revisores da Editora CRV

Projeto gráfico de capa

CCOM Funarte

Projeto gráfico de miolo

Equipe de designers da Editora CRV

Diagramação

Equipe de designers da Editora CRV

Revisão

Luís Carlos Morais Filho e equipe de revisores da Editora CRV

Dados Internacionais de Catalogação na Publicação (CIP) FUNARTE/ Coordenação de Documentação

Schisler, Millard.

Preservação de fotografias nato-digitais. / Millard Schisler. – Rio de Janeiro:
FUNARTE, 2023
218 p. (Cadernos Técnicos de Conservação Fotográfica, v. 9)

ISBN 978-65-5845-013-9

1. Fotografia – Técnicas digitais. 2. Preservação digital. I. Título.

CDD 775

Copyright © Funarte

Todos os direitos reservados.

Fundação Nacional de Artes – Funarte

Av. Presidente Vargas, 3.131 – Cidade Nova – CEP: 20210-911

Rio de Janeiro – RJ | livraria@funarte.gov.br

www.funarte.gov.br

Ao longo de quase cinquenta anos, a Funarte tem desempenhado papel fundamental para a memória das artes, especialmente através de seus Centros de Documentação (CEDOC) e Conservação e Preservação Fotográfica (CCPF), além das publicações impressas e digitais — referências na produção de conteúdos sobre os mais diversos campos relacionados às artes no Brasil.

Nesse sentido, o Centro de Conservação e Preservação Fotográfica da Funarte — CCPF — segue vivo em sua missão e compromisso, com a realização deste 9º volume da coleção Cadernos Técnicos de Conservação Fotográfica. A coleção tem se constituído como uma importante bibliografia sobre o tema, através da promoção e difusão dos trabalhos agora desenvolvidos com o apoio financeiro do Edital Bolsa Funarte de Estímulo à Conservação Fotográfica Solange Zúñiga.

Nesta 9ª edição, apresentamos o olhar sensível do fotógrafo, educador e pesquisador, Millard Schisler, que se debruça sobre as fotografias nato-digitais. No curto instante de tempo em que você lê este texto de apresentação, quantas fotografias estão sendo produzidas por smartphones e câmeras digitais em todo o mundo? Uma avalanche de informações e imagens gerando camadas diárias de conteúdos que se sobrepõem numa velocidade vertiginosa. Qual o destino dessas imagens? Quais delas sobreviverão para dar notícias do nosso agora? O que preservar da maresia do tempo? Como preservar e por quê?

A presente publicação marca um momento histórico em que assistimos às produções de imagens e narrativas serem tomadas como expressão do ódio, da desinformação e do ataque à democracia. Se recuarmos no tempo, poderíamos pensar ainda nos corpos, vozes e imagens apagadas ao longo da nossa história. No rumor do agora, este texto revisa ou produz narrativas que abrem e ressignificam o passado e projetam o futuro. É nessa perspectiva que o Caderno, organizado em 8 temas, nos convoca a pensar sobre a potência desses verdadeiros inventários de instantes e narrativas múltiplas. Uma reflexão sobre a preservação de fotografias nato-digitais num século marcado pelo desenvolvimento tecnológico em escala e permanente movimento, com ciclos de criação e obsolescência de tecnologias.

Viva a potência da fotografia de fazer mover o ontem, o hoje e o amanhã!

Maria Marighela
Presidenta da Funarte

Dedico este livro à Magali Pinhatti, minha companheira de vida e de trabalho. Em 1984, ela abriu as portas para a minha caminhada na fotografia ao me contratar como professor na Focus, onde ela era diretora. Nesse meio-tempo, costuramos uma amizade frutífera, e a fotografia sempre foi um dos elos conectores das nossas vidas. Seja nos muitos projetos, como este, seja na criação dos nossos quatro filhos, a Magali sempre me empurra para a frente, me estimula, e me dá coragem.

Agradeço à equipe da Funarte, responsável pela Coleção Cadernos Técnicos de Conservação Fotográfica, por todo o estímulo à produção de conteúdo que trouxeram para esse campo no Brasil.

Aos responsáveis pelo Edital Bolsa Funarte de Estímulo à Conservação Fotográfica Solange Zúñiga, com o qual fui contemplado na edição de 2019. Certamente este texto não seria escrito sem tal apoio.

À Sandra Baruki e à atual chefe da Divisão de Preservação, Isabel Mendes, pelo trabalho contínuo de promover a formação no país.

Espero que o Edital tenha vida longa e que suas próximas edições produzam muitas contribuições e experiências para todos nós.

Sumário

- 10 Introdução – por que pensar esse assunto?
- 1**
- 15 *Photo Finished* – o fim da fotografia
- 2**
- 26 Onde estão as fotografias digitais
- 3**
- 35 O que queremos preservar
- 4**
- 72 Como devo organizar e descrever meus materiais passivos e atuais
- 5**
- 125 Onde vamos guardar

6
170 Programa de preservação digital

7
174 Do digital ao analógico

8
194 Repensar a captura

200 Comentários finais

201 Notas

210 Créditos de figuras e tabelas

Introdução – por que pensar esse assunto?

Para introduzir esse assunto, gostaria de compartilhar de que maneira cheguei a propor esse texto. A minha história pessoal passa pela produção de fotografias, como fotógrafo e educador em fotografia, e por um envolvimento com a preservação fotográfica, que hoje chamamos de analógica, numa era de fim do século XX, quando a fotografia digital era ainda algo quase experimental. Ao ingressar no século XXI, a fotografia digital invadiu de vez as nossas vidas, com menos filmes fotografados e mais imagens produzidas a partir de dispositivos digitais, como câmeras e, principalmente, celulares. O meu envolvimento no mundo digital se iniciou pela digitalização, e posteriormente, pela produção digital com câmeras e celulares, e a percepção é de que há um campo necessário para a preservação das fotografias nato-digitais produzidas por esses dispositivos.

Historicamente, o campo tradicional da preservação, de forma geral, vem a reboque da produção dos objetos. A conservação preventiva é mais recente e procura prever condições em que os objetos possam ter melhor sobrevida. Entretanto, em muitos dos nossos trabalhos de preservação, recebemos os objetos para serem cuidados muito tempo depois de sua produção, e o volume é tão grande que, em muitas de nossas instituições de memória e cultura, há sempre grande volume de materiais passivos para serem processados e preservados. Podemos também encontrar caixas com negativos ou cópias guardadas há décadas em algum lugar esquecido, e, mesmo assim, ainda poderemos recuperá-las.

No meio da produção nato-digital não existe essa mesma possibilidade. A espera de alguns anos ou décadas após a sua produção já pode significar não encontrar mais os materiais originais, ou não poder fazer a leitura desses dados nos dispositivos em que estão guardados. Nesse sentido, é urgente, na preservação de fotografias nato-digitais, como também nos outros objetos nato-digitais, uma ação imediata de preservação. Neste texto, a preocupação específica são as fotografias nato-digitais. A preservação deverá ser planejada no momento em que as imagens digitais são criadas.

A questão é: quem fará esse trabalho? Se essa etapa estiver próxima da produção, precisará ser feita por quem está ligado à produção das imagens. Nesse sentido, não será o atual conservador de fotografias. As ferramentas de trabalho são muito diferentes quando olhamos para a preservação de objetos em comparação à preservação da informação, armazenada em objetos. Talvez o conservador de fotografias possa ampliar o seu leque para cuidar de todas as fotografias ou imagens, independentemente da forma como são produzidas, tanto via matrizes fílmicas quanto por meio de sensores digitais. Isso representa uma ampliação da profissão, ou uma nova área de atuação, pela diversidade de conceitos e ferramentas necessários para realização de todos esses trabalhos. Um conservador de novas mídias, envolvido amplamente com arte contemporânea, audiovisual, arquivos sonoros, entre outros, poderá também cuidar de fotografias nato-digitais. Contudo, se esse for o caso, o atual conservador de fotografias poderia correr o risco de cuidar somente da fotografia dos séculos XIX e XX.

Estamos lidando com uma área em constantes mudanças, e que passa por ciclos de criação e obsolescência de tecnologias. Este texto também terá, em algum momento próximo, uma possível importância somente como reflexão de um momento na história da produção e da preservação de imagens no início do século XXI. Nesse momento de imensa produção imagética, será necessário ter clareza de que nossa missão não pode ser determinada pela avalanche de conteúdo, mas guiada a preservar pedaços dessa produção, filetes pequenos, aqui e ali, escolhidos, colecionados e preservados intencionalmente. Esse filete arquivístico muitas vezes será o que nós poderemos, de forma sustentável, preservar e difundir em longo prazo.

No livro *Theory and Practice of Digital Preservation*, de Trevor Owens (2018), o autor cita um axioma, o nº 14: *Accept and embrace the archival sliver* (temos que aceitar e abraçar o filete arquivístico). Em todas as culturas, nunca se salvou tudo. Ao termos a clareza de que não salvaremos tudo e de que muito do que é produzido é efêmero,

e sumirá com o tempo, nos posicionamos para uma compreensão maior do que é significativo preservar e do que importa para nós ou para nossas instituições. Temos que ter critérios de seleção atrelados a políticas de aquisição e incorporação de objetos a serem colecionados. Desta forma, alocamos esforços e recursos na preservação dos objetos digitais que mais importam. Isso será um filete pequeno do todo. Com o trabalho das pessoas e das instituições colecionando filetes de todos os tipos, poderemos então preservar pedaços selecionados desse universo da produção digital.

The Memory WAKA



Figura 1. Logomarca de Ross Hemera, do Memory Waka Research Group, com a representação da canoa Maori.

Creio que o nosso desafio atual pode ser resumido no conceito e na analogia criados no *Memory Waka Research Group*¹, da Massey University, da Nova Zelândia.

Waka é uma canoa Maori. Waka também possui outros significados, como um receptáculo dos tesouros de um chefe, o waka huia, contendo seus pertences valiosos, incluindo penas premiadas. Waka também se refere a um grupo de pessoas de um

mesmo parentesco. Metaforicamente, o Memory Waka é um recipiente da humanidade, de ideias e cultura. É também, literalmente, uma forma de viagem para pessoas aliadas a uma causa comum, remando na mesma direção.

Assim, o nosso Waka da memória é que deve carregar nossas imagens preciosas para serem preservadas, vistas e valorizadas pelas gerações atuais e futuras.

Silenciamento da história e dos arquivos

Michel-Rolph Trouillot, um reconhecido antropólogo Haitiano, em sua obra de 1995, *Silencing the Past: Power and the Production of History*, traduzida para o português e publicada pela Huya Editorial em 2016, *Silenciando o passado: poder e a produção da história*², produziu um texto marcante sobre o papel que a história tem em silenciar outras narrativas para manter visões enviesadas de momentos históricos ofuscando outras versões possíveis. Trouillot discute que os silêncios ingressam no processo de produção de história em importantes momentos: quando os fatos são criados, e as *fontes*, utilizadas; quando isso é consolidado nos *arquivos*, na recuperação dos fatos pelas *narrativas*, e, no fim, quando todos esses fatores constituem a elaboração da *história*.

Por não serem documentadas, preservadas, narradas, arquivadas, estas versões silenciadas sumirão da consciência coletiva. Esse conceito se aplica, como dito por Trouillot, à construção dos arquivos e das memórias que preservam. Da mesma maneira que pensamos sobre o que está sendo preservado, deve-se pensar, simultaneamente, sobre o que não está sendo preservado – o que poderá ser silenciado, sumir pelas brechas da história.

Ao pensarmos nesses filetes que decidimos preservar, correremos o risco de realçar histórias em detrimento do apagamento de outras? Estas decisões devem estar muito atreladas, e documentadas, ao que consideramos vitais no nosso próprio trabalho, ou, no caso de instituições, ao papel que possuem na sociedade e na cultura descrita pelas suas missões.

Nesse sentido, tenho aqui a intenção de que sejamos os proponentes de uma guarda, uma seleção, uma curadoria, uma preservação, uma divulgação. O barulho da história, não o seu silêncio.

A democratização da produção, mas não da preservação

A fotografia, em sua era fílmica e rica história, sempre foi feita principalmente por aqueles que tiveram acesso ao conhecimento, à tecnologia e aos recursos financeiros para praticá-la. Isso representa uma elite, principalmente composta por homens brancos, o que também criou uma visão totalmente enviesada do mundo e dos outros. A fotografia digital, principalmente pela sua ampla difusão na captura por celular em praticamente todas as classes sociais, culturas e países, abre a possibilidade de as pessoas poderem contar suas próprias histórias, documentarem suas vidas, famílias, encontros, visões de mundo etc. Nesse sentido, a produção recente de imagens nesta era do celular explode e se torna acessível e democrática. O problema é que, salvo a guarda em redes sociais, que citam claramente não ter a função de fazer a preservação digital, a preservação em longo prazo desse conteúdo exigirá conhecimento, tecnologia e recursos financeiros, criando novamente uma barreira de entrada no universo da preservação que torna esta uma tarefa elitista. O nosso desafio será, então, criar soluções coletivas, fortalecer coletivos e instituições, desenvolver sistemas mais simples e eficazes, e ampliar a difusão da educação acerca de como isso pode ser feito para garantir a preservação de todos esses filetes que queremos preservar. Espero contribuir na construção destas soluções.

Um texto que abre a discussão sobre muitos assuntos

Procuro apresentar uma série de assuntos para abrir a discussão sobre o que representa a preservação de nossa produção digital, e, mais especificamente, das nossas fotografias nato-digitais. Não pretendo concluir esse assunto, mesmo porque são todos assuntos em constante evolução. Cada um dos capítulos deste texto deve ser expandido. Será da responsabilidade do leitor se engajar, pesquisar e estudar esses assuntos para se tornar mais uma pessoa a fim de contribuir e agregar ao crescente movimento necessário na jornada da preservação de fotografias digitais, bem como da preservação digital. Muitas das notas estão nesta obra para apontar a vocês caminhos para continuar a pesquisa.

Temos hoje uma vasta quantidade de fontes confiáveis com informações, textos, vídeos e publicações, muitas mencionadas aqui. Espero que esta publicação sirva de estímulo para os leitores adentrarem nesta jornada e fazerem parte de uma crescente comunidade de difusão de conhecimento.

1

Photo Finished – o fim da fotografia

Este texto foi originalmente publicado na *Brasiliana Fotográfica* em novembro de 2018³, e representa uma visão de abertura para as questões de preservação de fotografias nato-digitais. Diversas alterações no original foram feitas nesta versão.

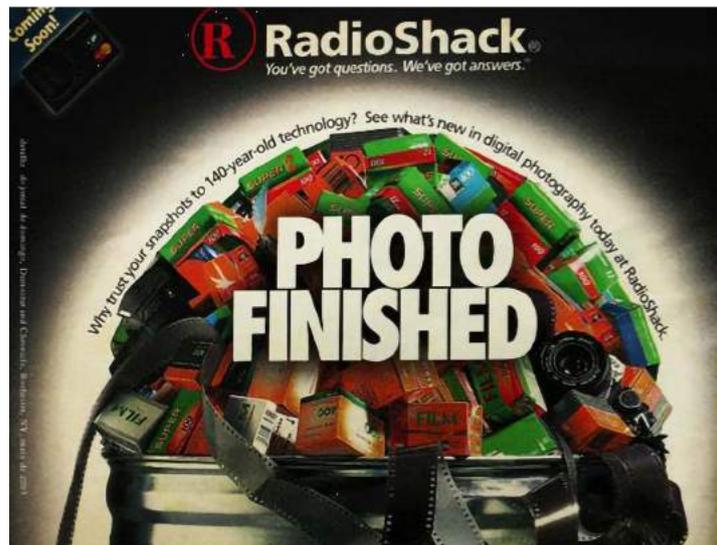


Figura 2. Anúncio da *Radio Shack* no jornal *Democrat and Chronicle*, Rochester, New York, maio de 2003.

Em 2003, vi esse anúncio que faz um trocadilho em inglês entre “fotoacabamento”, *photo finishing*, e *photo finished*, ou seja, “acabou a foto”. Nesse anúncio, temos rolos e caixas de filme jogadas dentro de um cesto de lixo, e um texto em cima de todo esse “lixo” dizendo: “Por que confiar em uma tecnologia velha de 140 anos para as suas fotos?”. Certamente os arquivos nato-digitais realizados naquele ano não devem mais existir, restando somente os “velhos” filmes feitos nesse mesmo ano.

Entretanto, de alguma forma, a fotografia acabou. Grant Romer, um dos historiadores e conservadores de fotografia mais respeitados, afirma, no artigo que escreveu para o Getty Conservation Institute⁴, que

[...] o espichamento e embaçamento da definição de “fotografia” é um resultado direto da evolução da imagem eletrônica [...], e que “[...] os que advogam pelo ‘Progresso’ apontam que a tecnologia fotográfica sempre se transformou desde sua introdução comercial em 1839. Um processo sucumbiu a outro, um após o outro. O que estamos passando agora, dizem, é somente o ‘fechamento de um outro capítulo na história da fotografia’. Mas, é importante considerar que esse pode ser o último capítulo no livro”.

Em 2013, Stuart Jeffries escreve um artigo, *The death of photography: are camera phones destroying an art form?*⁵ (*A morte da fotografia: as câmeras de celulares estão destruindo uma forma de arte?*). Nele, ele conversa com Antonio Olmos, um renomado fotógrafo mexicano que diz: “[...] é muito estranho. A fotografia nunca esteve tão popular, mas está sendo destruída. Nunca houve tanta fotografia sendo feita, mas a fotografia está morrendo [...]”. Stephen Mayes comenta, no artigo *The Next Revolution in Photography is Coming*⁶ (*A próxima revolução da fotografia está chegando*) de 2015, que “É hora de parar de falar de fotografia: não é que a fotografia esteja morta, como muitos dizem, é que ela se foi”. Em entrevista para a France Press em 2016, Sebastião Salgado diz “A fotografia não é imagem. Estamos em um processo de eliminação da fotografia. Hoje temos imagens, mas não fotografias”⁷. Continua, “A fotografia precisa se materializar, precisa ser impressa, vista, tocada...”. Um ano depois, afirma que a fotografia “[...] agora mais do que nunca, tem um longo futuro pela frente”⁸. Sem saudosismos, estas duas frases citadas por Salgado demonstram o nível de confusão desse período de transformação sobre como vemos a fotografia e para onde ela está indo.

Em francês, a fotografia digital é *la photographie numérique*, que designa a fotografia binária, com números. A palavra fotografia vem do grego *phos* ou *photo* (luz) e *graphein* (gravar, escrever, desenhar ou registrar). Por todos os conhecimentos da história da fotografia que temos, podemos atestar que a palavra *photographie* foi cunhada pela primeira vez por Hercule Florence, em 1832, no Brasil, em suas pesquisas com esse processo. Hercule Florence foi pioneiro em muitas áreas, e também o primeiro a produzir fotografias por contato com sais de prata e ouro em 1832 e 1833. Hoje, as fotografias continuam sendo escritas ou gravadas por meio da luz na fotografia binária, pois, de forma analógica, um sensor composto de fotodiodos é exposto com luz, e somente depois esses dados são convertidos para um sinal digital. Em inglês, muitas vezes nos referimos à ideia de *picture* – retrato, foto ou imagem. Podemos também usar fotografia digital, imagem digital, imagem eletrônica, ou infoimagem. Falamos também em fotografia computacional, processando os dados de sensores em combinações algorítmicas, criando novos tipos de imagens em que o autor se confunde com os processos que acabam por montar a imagem. Creio que todos podemos concordar que algo mudou, se transformou, mas o que não mudou é que continuamos a produzir um monte, cada vez mais, disto que temos dificuldade de definir.

Um dos problemas de pensar a fotografia nato-digital, que nasce digitalmente, é que ela se materializa nas nossas telas e em outros dispositivos, mas, em sua essência, é composta de informação ou dados digitais, zeros e uns, interpretados por códigos de *softwares* e armazenados em *hardwares*/dispositivos digitais – são dados digitais, informações digitais e podem existir da mesma forma em muitos lugares distintos e ao mesmo tempo. A fotografia, antes disto, que sempre foi somente fotografia e agora ganhou o adjetivo de analógica, reside nos objetos que lhe deram origem, como negativos ou cópias em papel fotográfico – a imagem pode ser vista sem a necessidade de mediação de *software*, conectores e *hardware*. Por mais que queiramos cuidar das fotografias analógicas, as digitais é que correm mais perigo de sumir devido a essa fragilidade dos sistemas que as compõem. Elas estão por toda a parte e, ao mesmo tempo, em nenhuma parte.



Figura 3. Guarda de objetos *versus* guarda de informação.

A produção cada vez maior e incessante dessa informação digital é gerada principalmente por meio de câmeras de celulares, que representam, de acordo com estudo da InfoTrends, pelo menos 85% das imagens feitas em 2017. Os números são estratosféricos – mais de um trilhão de fotografias produzidas em 2017. As estimativas são de que, desde o início comercial da fotografia, em 1839, até o ano 2000, chegamos a cerca de 100 bilhões de fotografias produzidas. Hoje produzimos facilmente mais do que 10 vezes essa quantidade por ano, com tendências crescentes. Esses números estão além da nossa capacidade de digestão visual. Estima-se que a nossa cultura produziu em torno de 5 trilhões de imagens até 2017, com um acréscimo anual agora de praticamente 1,7 trilhão.

Para tentar entender melhor esses números, fiz estas contas, e você pode fazê-las também. Se você tivesse a tarefa de visualizar estas imagens, uma por segundo, descansando 10 minutos por hora, e fazendo isso 7 horas por dia, 20 dias por mês, durante 11 meses do ano – pelo menos um mês de férias para descansar a vista – você veria 4,62 milhões de imagens nesse primeiro ano. Após um milhão de anos nesse processo, você teria visto quase a totalidade da produção de imagens até

2017, de 4,9 trilhões. Será fundamental deixar um bilhete avisando onde você parou antes de morrer. Contudo, o problema é que só em 2022 foram despejadas mais de um 1,7 trilhão de novas imagens nesta conta.

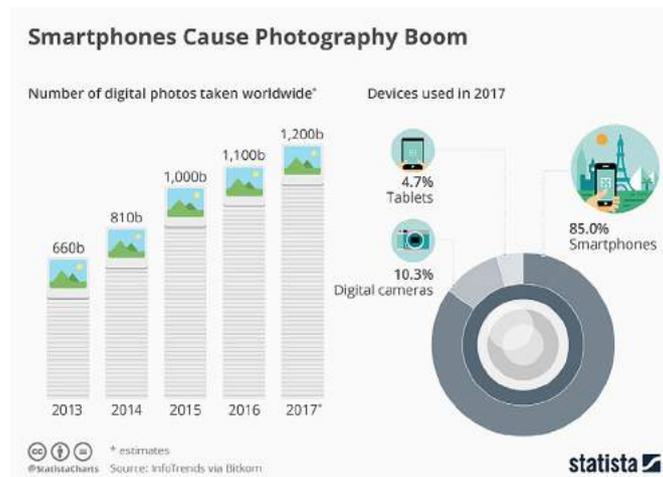


Figura 4. Uso de smartphones, produção de 1,2 trilhão de fotografias/imagens em 2017⁹.



Figura 5. 1 trilhão de fotografias/imagens em 2015 e 4,9 trilhões em 2017¹⁰. De acordo com a Mylio¹¹, a produção de imagens por ano a partir de 2020 é de em torno de 1,4 trilhão.

Creio que todos podemos concordar que a quase totalidade destas imagens não deva ser preservada e pode se perder sem prejuízos para a nossa cultura. Em matéria da CNN¹² de 2015, estima-se que haja em torno de 6,5 bilhões de fotografias de gatos na internet. Existe um efeito efêmero nesta produção – comparo isso ao ditado conhecido, mas com viés para o olhar: “O que entra por um olho, sai pelo outro”. Mal conseguimos lembrar de imagens que vimos dias atrás – já estão soterradas, há muito tempo. Isso me lembra uma fotografia de Howard Cleaves, um dos primeiros ambientalistas, produzida em 1953, no aterro sanitário Fresh Kills, em Staten Island, do lado de New York, com Frank Hauber em cima desse lixo. Todo dia despejamos toneladas de imagens nesse grande lixão visual, que, como um lixão, também contém sua cota de preciosidades – a produção de hoje soterra a produção de ontem, e assim por diante. Mas então, o que preservar? Como devemos direcionar os nossos esforços? O que deve sobreviver, vir à tona, ser resgatado, selecionado?



Figura 6. Frank Hauber dá uma última olhada no decadente *Long Pond* em *Fresh Kills, Staten Island*, New York, todo se enchendo de resíduos e lixo, 1953.

Esse não é um problema novo. Em uma conversa com Grant Romer, alguns anos atrás, discutíamos estas questões da preservação das fotografias digitais. Grant foi para sua biblioteca e trouxe o primeiro dicionário de fotografia de E. J. Wall, de 1902. Ao folhear o dicionário, procurando a palavra *Fading*, “esmaecimento”, ele leu: “Esmaecimento: o pior de todos os males para o qual negativos e cópias boas e interessantes estão sujeitos; mas em outros casos, um benefício para a humanidade”.

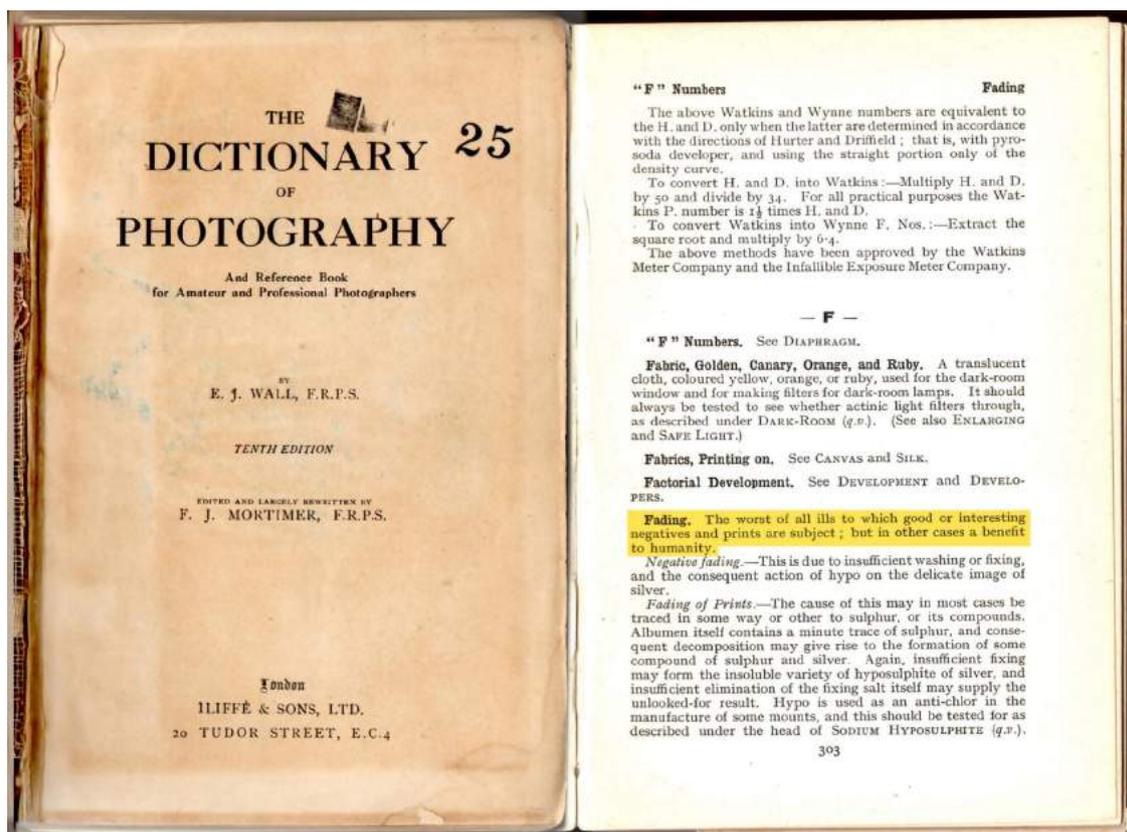


Figura 7. Reprodução de Millard Schisler, *The Dictionary of Photography*, 10ª edição de 1920.

O presente texto surge na época do *boom* de produção de imagens da câmera Brownie e filme de rolo da Kodak – “Você aperta o botão, a gente faz o resto”. Sistemas simples que qualquer um poderia usar como vemos nas propagandas de época. É curioso que, praticamente cem anos depois, também vivemos um outro tipo de *boom* de produção de imagens, a partir de celulares – mas agora você aperta o botão e é você que faz o resto.

Uma outra visão pode vir da entrevista feita pelo Cassiano Viana, em 2016¹³, com a mexicana Mayra Mendoza, então subdiretora da Fototeca Nacional do Instituto Nacional de Antropologia e História (INAH) do México. Mayra nos dá uma luz:

Em primeiro lugar, eu faria uma distinção entre imagem e fotografia. Hoje em dia, todos somos fazedores de imagens, já que uma grande quantidade da população mundial tem acesso a uma câmera, em um dispositivo móvel, em smartphones. Mas isso não quer dizer que todos somos fotógrafos. A fotografia, para além da câmera, implica um ofício e, sobretudo, uma intenção. A função social da imagem se cumpre ao poder compartilhá-la com um público cada vez mais amplo e que produz milhões de imagens diariamente publicadas nas redes sociais. No entanto, em termos de fotografia, apenas algumas transcendem a proposta visual de quem gera essa imagem.

Será que, nesse contexto, podemos pensar em salvar as fotografias e deixar as imagens sucumbirem à efemeridade ou à vida nas redes sociais, preservadas ou não pelos sistemas proprietários onde residem? As que queremos salvar, teremos que preservar ativamente para garantir que possam ser acessadas pelas próximas gerações. Esse universo de imagens está além de nosso alcance, e não poderemos visualizá-lo em sua totalidade para fazermos nossas triagens e decisões daquilo que consideramos importante. Estas imagens fazem parte do *Big Data* e só podem ser processadas e vasculhadas em sua complexidade por sistemas e programas. Um caminho pode ser a criação de algoritmos baseados em critérios estabelecidos para realizar o trabalho de seleção e preservação.

Vivemos em um período de extraordinárias mudanças. A fotografia sempre foi uma ferramenta da classe dominante – com acesso à tecnologia e educação para utilizá-la e com recursos para financiar sua realização. Agora podemos ter registros cotidianos sendo feitos pelos próprios personagens que os vivenciam – não mais necessitamos de fotógrafos de fora. Grupos, culturas, regiões, povos, famílias, todos fazendo sua própria documentação. Esta produção imagética é importante preservar, pois ela conta uma história por uma narrativa do agente e é feita praticamente na sua integridade, de forma digital. Tenho que citar três artigos¹⁴ de muita importância, escritos por duas repórteres de *O Globo* em 2012, Maiá Menezes e Tatiana Farah, com fotografias de Márcia Foletto e Michel Filho. Neles, estas questões são abordadas de forma muito fascinante, com três casos distintos e parecidos ao mesmo tempo, apontando tanto a relevância dessa produção como a sua fragilidade. Importante leitura e reflexão.

Resta entender como faremos o trabalho de peneirar esse universo imagético na busca da guarda do que faz sentido preservar. Deixaremos para cada grupo ou indivíduo a responsabilidade de cuidar de sua própria produção? Cuidar de dados digitais não só exige um grau de conhecimento como também um constante cuidado com recursos humanos e financeiros, mão de obra especializada e compra contínua de equipamentos para realizar esta guarda. A produção pode ter se democratizado, mas o armazenamento em longo prazo, não. A única opção para a grande maioria, que não se enquadra em uma preservação digital intencional, é considerar o YouTube, o Facebook, o Instagram, entre outros, como sendo os repositórios digitais para preservar a sua produção. Por outro lado, muitas imagens digitais acabam em *Dark Data* – material coletado, processado, armazenado e preservado, mas não distribuído, postado ou utilizado para outras finalidades, ou seja, material em estado latente, e com grande risco de não sobreviver. Podemos pensar até em uma analogia com rolos de filmes fotografados, com uma imagem latente gerada pela exposição, mas nunca revelados.

Quando pensamos em preservação em longo prazo, pensamos em 100 anos, ou duas gerações de trabalho ativo. Isso significa que a minha geração irá cuidar e educar/criar uma outra geração que fará o mesmo, assim garantindo pelo menos uns 100 anos de preservação. A próxima geração fará o mesmo, e assim por diante. Teremos as redes sociais de hoje daqui 50 anos?

Todavia, para um programa intencional de preservação de fotografias digitais, teremos que fazer escolhas, seleção, e ter um papel ativo em relação ao que deve ser preservado. Isso certamente passa por um crivo editorial ou institucional, voltado para a missão de quem está fazendo esse processo de seleção e guarda. David Rosenthal comenta, no *post* em seu *blog* de maio de 2012, intitulado *Let's Just Keep Everything Forever in the Cloud*¹⁵ (Vamos manter tudo para sempre na nuvem), que não temos como guardar tudo:

Então vamos ter que jogar coisas fora. Mesmo se acreditarmos que guardar coisas é muito barato, ainda é muito caro. A má notícia é que decidir o que guardar e o que jogar fora não é de graça. Ignorar o problema acarreta o custo de manter os dados; lidar com o problema acarreta o custo de decidir o que jogar fora. Podemos estar em uma situação ruim de não poder arcar com os custos de manter ou jogar fora os dados que geramos. Talvez seja necessário pensar com mais cuidado, antes de mais nada, sobre a geração dos dados. É claro que pensamentos assim também não são sem custo.

É interessante David mencionar sempre a questão do custo. Um outro David, o David Giaretta, cita, em uma de suas publicações de 2008, *Advanced Digital Preservation*¹⁶ (*Preservação digital avançada*), que a preservação digital é algo muito fácil de se fazer, desde que você tenha muito dinheiro para sempre.

A publicação *The Digital Dilemma*, de 2007, traduzida para português pela Cinemateca Brasileira em 2009¹⁷, já aponta para a questão do custo, demonstrando que a guarda de produções cinematográficas em película, filmes em rolo, é em torno de 10 vezes mais barata do que a guarda de uma mesma produção realizada inteiramente de forma digital. Guardar objetos é muito mais simples e barato do que guardar informação! Infelizmente, muitas instituições no Brasil e no mundo ainda batalham para obter recursos a fim de conseguir cuidar de forma adequada de suas coleções de objetos. Enquanto isso, o empilhamento de dados digitais continua ano a ano.

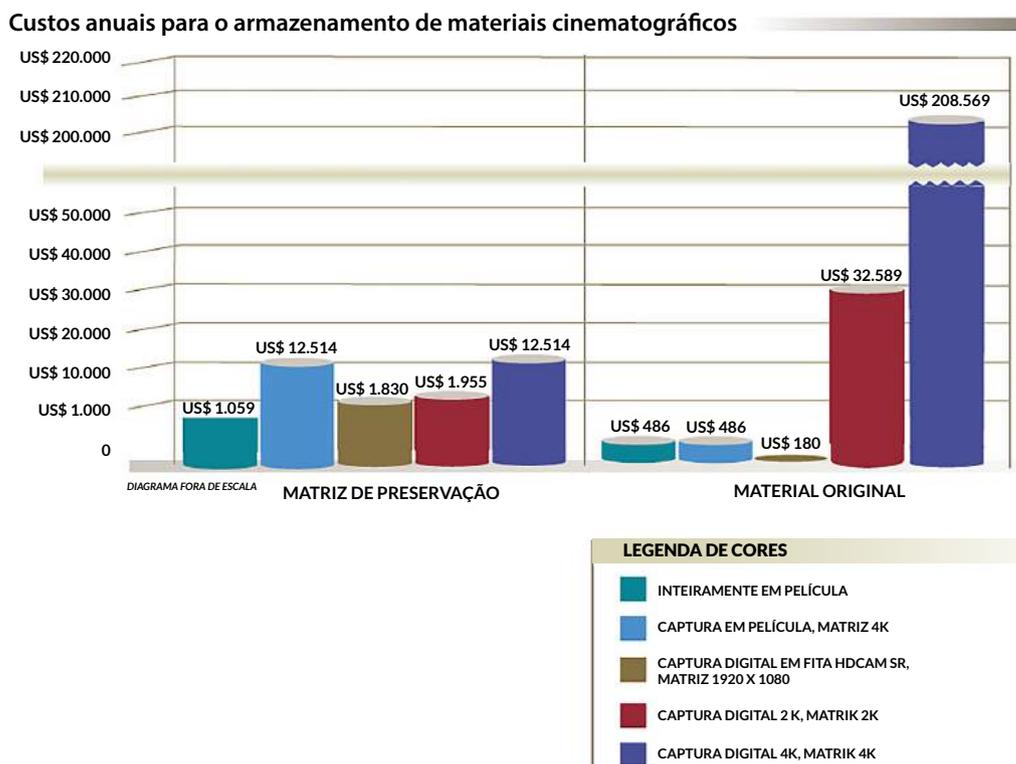


Figura 8. Gráfico na página 44 da publicação *O Dilema Digital*. Olhando somente para as matrizes de preservação, comparamos o custo de armazenamento de um filme produzido inteiramente em película de US\$ 1.059 com um filme produzido em digital 4K de US\$ 12.514. A situação fica mais grave quando pensamos em preservar todos os materiais originais da produção.

O Instituto Americano de Conservação (AIC), publicou um livro¹⁸ em 2011 sobre fotografia digital e documentação em conservação, e, em um dos capítulos, fala sobre cinco etapas fundamentais para a preservação de fotografias digitais.

- 1 Usar formatos de arquivos sustentáveis
- 2 Organizar os dados digitais, inserir metadados para descrevê-los
- 3 Fazer os *backups* e a migração dos dados
- 4 Manter processos contínuos de verificação da integridade dos arquivos
- 5 Imprimir material selecionado/editado em processos mais permanentes

Esses cinco itens representam uma versão simplificada e resumida do processo. Entendemos hoje que a preservação digital não é garantida somente com *backups*, e guardar não resolve se não houver acesso. Temos que pensar na difusão. Fotografias digitais catalogadas devem ser disponibilizadas em sistemas de banco de imagens para consulta e pesquisa. Esse uso é o que dará sentido e vida às coleções, e propiciará sua inserção no cotidiano das pessoas. Será o combustível para a própria manutenção e ampliação desses acervos digitais. Se não for disponibilizado e usado de forma eficiente, um arquivo vira um arquivo morto. Será complicado procurar formas de financiamento para a manutenção de um arquivo morto.

É muito interessante ver o quinto item sobre impressão como forma de preservação de fotografias digitais. Da mesma forma que encontramos cópias em coleções e, na maioria dos casos, não temos os negativos que os geraram, também aumentam as chances de preservação dos arquivos digitais quando materiais editados são impressos em suportes estáveis que possam ser armazenados em acervos climatizados. Publicar livros também é uma forma de consolidar um trabalho de milhares de imagens. A impressão não substitui a preservação das fotografias digitais, é um complemento a esta preservação. Com a participação cada vez maior de pessoas atuando na preservação de dados digitais e de fotografias nato-digitais, teremos mais chances de garantir a sobrevivência do que estamos produzindo hoje.

2

Onde estão as fotografias digitais

As fotografias nato-digitais estão por toda parte. Elas podem ter uma fonte de origem da sua criação, mas logo passam a existir como dados digitais, informação binária, em dispositivos de tudo que é tipo. São enviadas por e-mail, armazenadas em discos rígidos, *pendrives*, copiadas em dispositivos como discos ópticos para enviar para alguém, enviadas e postadas em dezenas de redes sociais, enviadas para armazenamento na nuvem, e enfim acabam por existir em inúmeros lugares, de forma repetida, mas também em diversos formatos de arquivo, níveis de compressão e qualidade de imagem. Muitos lidam com as imagens nas redes sociais e nuvens como se esses lugares tivessem a função de cuidar, gerenciar e preservar estas imagens. Entretanto, como descrito claramente nos termos de serviço, mesmo com a proposta de fazer exatamente isso, não são responsáveis por eventuais perdas e danos aos seus materiais. O risco, portanto, é de ter estas imagens preservadas por todos esses lugares, mas que eventualmente acabem por não existir em lugar algum.

Na imagem a seguir, vemos exemplos de lugares que podem ter imagens digitais: computadores antigos, obsoletos, fora de uso contendo discos rígidos

que podem ter conteúdos digitais que não foram migrados e, portanto, preservados. Computadores novos com dados digitais que só existem neles e sem os *backups* necessários. Telefones celulares de diversos tipos e idades, que podem ter imagens digitais que não foram copiadas para sistemas de nuvem. Cartões de memória dentro de câmeras digitais e cartões de memória fora destas câmeras espalhadas em diversos lugares. Discos ópticos, como CDs e DVDs, que podem conter fotografias digitais originais, e que precisam ser vistos para fazer a migração dos dados destas mídias já entrando em obsolescência; disquetes de diversos tamanhos e idades, com materiais originais que podem não ter sido migrados para outras mídias. *Pendrives* de todos os formatos e tamanhos em inúmeros lugares também precisam ser verificados. Os discos rígidos externos de diversas idades e capacidades talvez sejam mais comuns como dispositivos de *backup*, muitas vezes desordenados e desorganizados. Por fim, imagens que podem existir somente nas redes sociais ou aplicativos de armazenamento na nuvem.



Figura 9. Onde estão as fotografias nato-digitais. Capturas de tela de buscas na internet, compiladas pelo autor.

Quando temos fotografias digitais espalhadas por tantos lugares, fica praticamente impossível pensar em um sistema que gerencie o *backup* desses arquivos de

forma segura e automática. Para que seja possível ter cópias, necessitamos primeiro definir uma matriz, que será a origem das cópias de *backup*.

A primeira ação é fazer um grande inventário de onde estão as suas fotografias digitais. Isso vale em nível individual, como também para instituições, que podem ter seus dados digitais também em inúmeros locais. Precisamos determinar e mapear quais são os dispositivos e lugares em que estas imagens estão guardadas. Será necessário fazer esse levantamento e agrupar todos esses conteúdos digitais para então aplicar ferramentas de *software* a fim de descobrir duplicatas e/ou derivadas de arquivos que acabaram sendo armazenados em lugares diversos. Muitas vezes, na tentativa de ter *backups*, a mesma imagem é copiada em muitos lugares diferentes. Esse mapeamento também serve para entender o volume de dados digitais totais que uma matriz terá que ter. Assim é possível planejar a compra de um disco rígido ou de sistemas como matriz e os discos/sistemas necessários para os devidos *backups*. É necessário ter a compreensão aproximada desses números. Uma previsão, após a verificação de todos os dispositivos de guarda, de um total de 1 TB, 10 TB, 100 TB ou 1 PB de dados digitais, trará propostas de soluções muito diferentes. O movimento aqui deve ser para a criação de uma matriz de todos os seus arquivos digitais. À medida que esta matriz vai sendo construída, será implantado um novo sistema de *backups* a partir dessa nova matriz.

Porém, a previsão de espaço necessário para o passivo deve levar em conta a produção anual crescente. Ter 1 TB de dados, e previsão de produção anual de 1 TB/ano significa ter em torno de 6 TB de dados digitais em 5 anos. Como o planejamento de um sistema de armazenamento deve ser feito para períodos em torno de 5 anos, temos já que prever, nesse caso citado, algo em torno de 8-10 TB de espaço total para os próximos 5-6 anos. Como regra básica, os discos rígidos magnéticos e de estado sólido têm que trabalhar com capacidade máxima com aproximadamente 20% (com mínimo de 15% e máximo de 25%) de espaço livre para que possam funcionar com eficiência e diminuição de riscos de problemas nesses dispositivos.

A segunda ação, após entender onde estão todas as imagens, é realizar o processo de limpeza desses dados, com eliminação de duplicatas e outros materiais indesejados, com o uso de *softwares* especializados, e planejamento para a

criação de uma nova matriz que irá servir de base para todos os seus *backups*. Estas etapas de trabalho serão vistas adiante no capítulo 4, “Como devo organizar e descrever meus materiais passivos e atuais”.

O passivo precisa ser organizado para que possa ser mantido dentro de um fluxo de preservação. Porém, o ideal é planejar e organizar um novo sistema para a produção atual primeiro, e, após uma nova estrutura implementada e em funcionamento, o passivo passará a ser organizado e incorporado, o que veremos mais adiante.

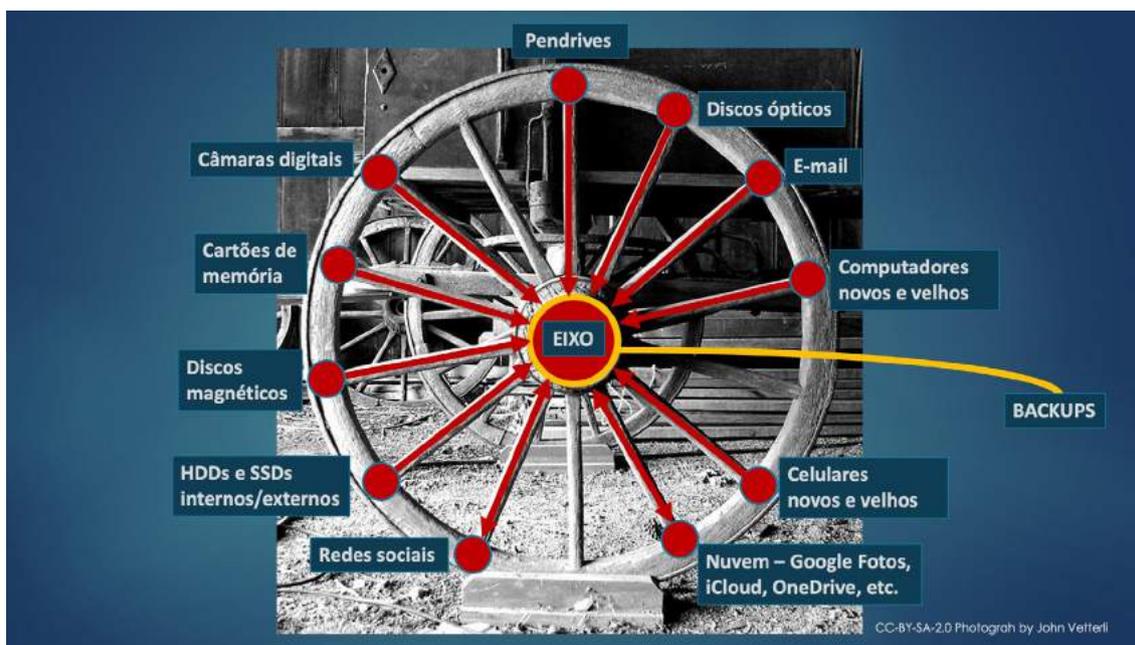


Figura 10. Imagem criada pelo autor a partir de imagem da Creative Commons.

Na imagem acima, o EIXO é a nossa matriz digital, onde todos os dados digitais devem estar, e não esparramados em todos esses dispositivos diversos. A partir dessa matriz, podemos criar, manter e sincronizar os nossos *backups*. Esta também é a hora de entender os objetos que estão mais em risco de obsolescência, discos rígidos mais antigos e frágeis, discos ópticos, e de buscar conhecer todos os formatos de arquivos digitais que estão nestas mídias para planejar os formatos de arquivos digitais que já possam fazer parte de um processo de migração (também conhecido como normalização) para formatos mais seguros de preservação digital em longo prazo.

As redes sociais como áreas de guarda

A crescente produção de fotografias via celulares permite o *upload* dessas imagens diretamente para aplicativos na nuvem (Google Photos e iCloud, por exemplo) e o compartilhamento via os diversos aplicativos na internet (redes sociais como Facebook, Instagram e WhatsApp). Há certa sensação coletiva de que as imagens passam a ser preservadas e acessadas nesses sites e nuvens. Em muitos casos, como já discutimos, a imagem acaba por existir somente nesses *sites*, e o original que gerou esta imagem de difusão já não existe mais. Nesse caso, teremos o que pode ser considerado um “novo original”, já que o arquivo pai/mãe que lhe deu origem não existe mais, e esse novo original nas redes será um arquivo com menor tamanho e qualidade em relação ao seu precursor. Sem um sistema de preservação consciente das versões que deverão ser preservadas, pode ser possível acabar somente com as versões postadas em redes sociais como os representantes sobreviventes. Temos então questões importantes de preservação que devem ser discutidas pensando nos arquivos que desejamos manter – muitas vezes as únicas versões que temos para preservar. No caso dos aplicativos das redes sociais, eles são construídos com a prerrogativa de comprimir ao máximo as imagens recebidas e remover informações de metadados embutidos (feitos intencionalmente), bem como os metadados técnicos de captura EXIF (Exchangeable Image File Format), como dados de GPS e data de criação, entre outros.

A seguir vemos alguns exemplos.

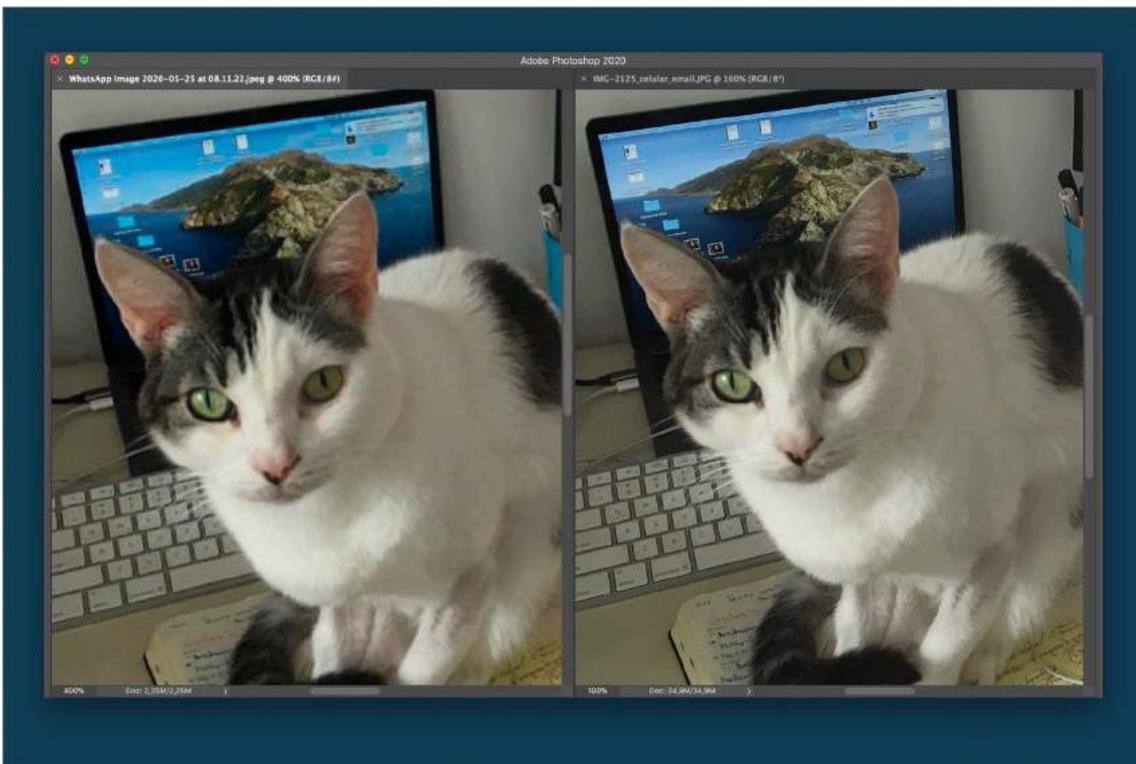


Figura 11. A imagem original, produzida pelo celular iPhone 6S, foi feita pela Joanna Americano com seu gato Momô.

A imagem JPEG à esquerda foi enviada do celular por WhatsApp, e depois baixada do WhatsApp, com tamanho aberto de 2,25 MB. A mesma imagem foi enviada por e-mail, à direita, e tamanho aberto de 34,9 MB (mais de 15 vezes maior). Portanto, isso mostra a perda de qualidade pela compressão e diminuição da resolução aplicada pelo WhatsApp. A imagem da esquerda está com uma ampliação de 400% para atingir o tamanho de 100% da imagem à direita.

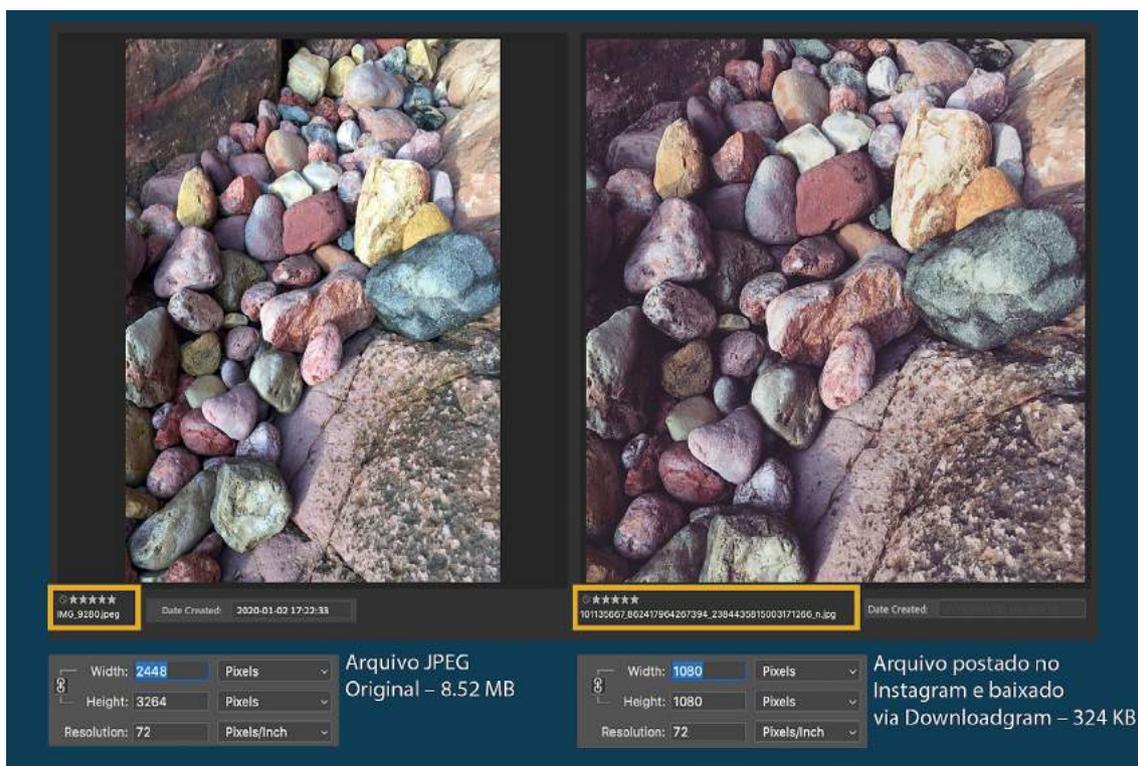


Figura 12. Comparativo entre imagem no tamanho original e mesma imagem baixada do Instagram.

No exemplo acima, a imagem original do autor, à esquerda, possui uma resolução de 2448 x 3264 *pixels*, e tamanho de 8.52 MB. A imagem foi inserida no Instagram e depois baixada desse aplicativo usando *Downloadgram*¹⁹. O arquivo resultante tem um tamanho de 324 KB (26 vezes menor), com muita compressão de dados e diminuição da resolução para 1080 x 1080 *pixels*, troca do nome do arquivo original, remoção da data de criação e dos outros dados EXIF, como GPS (não visto aqui nesta imagem). Podemos ver também uma alteração nas cores da imagem vindo do Instagram.

Devemos considerar, nesse exemplo, a imagem do Instagram como uma derivada. A decisão deve ser sobre a preservação ou não destas duas, uma sendo a imagem original, e a outra, a imagem que o público viu. Preservar a imagem de como foi vista e utilizada tem a sua importância, tanto quanto a preservação do original. Nesse caso, a imagem da experiência do usuário do Instagram não é a imagem de *upload* para o aplicativo, mas a imagem de *download*, como vista aqui. Podemos ter muitas derivadas a partir de uma mesma matriz e teremos que pensar em como lidar com isso. Serão sempre decisões sobre o que guardar e sobre como gerenciar a nomenclatura e guarda do arquivo original e seus versionamentos. Em termos de preservação, corremos o risco de fazer a salvaguarda somente destas imagens

que são distribuídas nestas redes sociais, que são versões com muitas perdas em relação às imagens originais que as geraram.

Aplicativos nas nuvens como áreas de guarda

A nuvem como área de armazenamento será discutida mais adiante, mas, neste tópico, estamos pensando especificamente nas soluções automatizadas em que as fotografias de celulares são armazenadas diretamente e automaticamente na nuvem, como no caso do *iCloud* para os celulares da Apple (iPhones), e do *Google Photos*, tanto para os celulares Android como para os iPhones.

Em quase todas as ocasiões, clicamos o nosso aceite na instalação de *softwares*, sistemas e aplicativos, e passamos direto pelos termos de serviço (*TOS – Terms of Service*) sem sequer ler com cuidado quais são esses termos.

As redes sociais e, especificamente nesse momento, a Google e a Apple, com os seus serviços de armazenamento de fotografias de celulares, citam claramente nos seus Termos de Serviço que não dão a garantia e não se responsabilizam por eventuais falhas ou perdas em seus sistemas, conforme vemos a seguir. Ambos usarão cuidados razoáveis na prestação desse serviço.

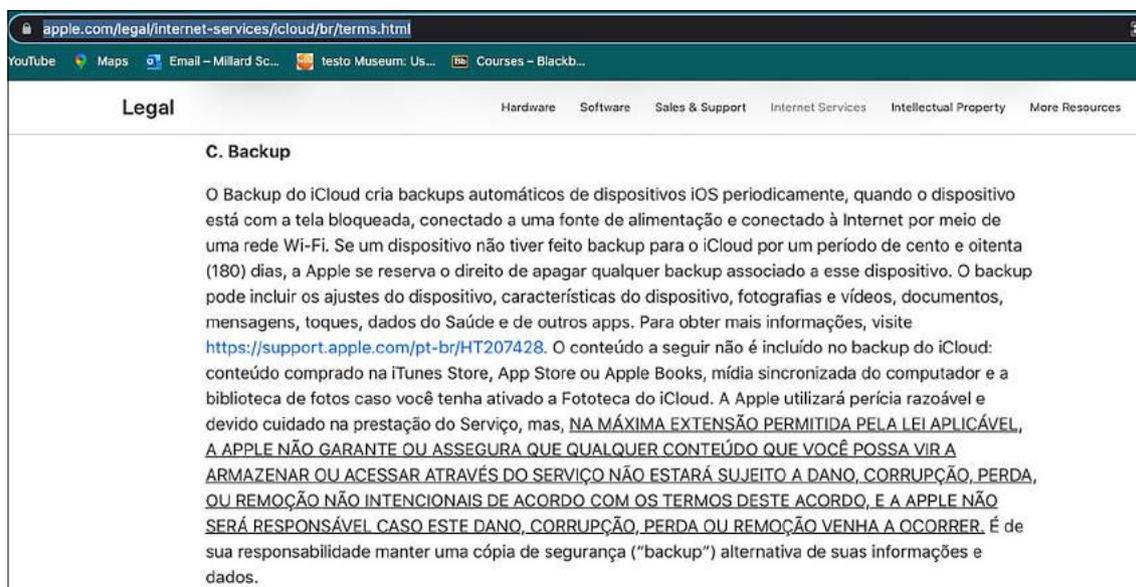


Figura 13. Descrição do site da Apple das regras de *backup* para o iCloud. Fonte: <https://www.apple.com/legal/internet-services/icloud/br/terms.html>.

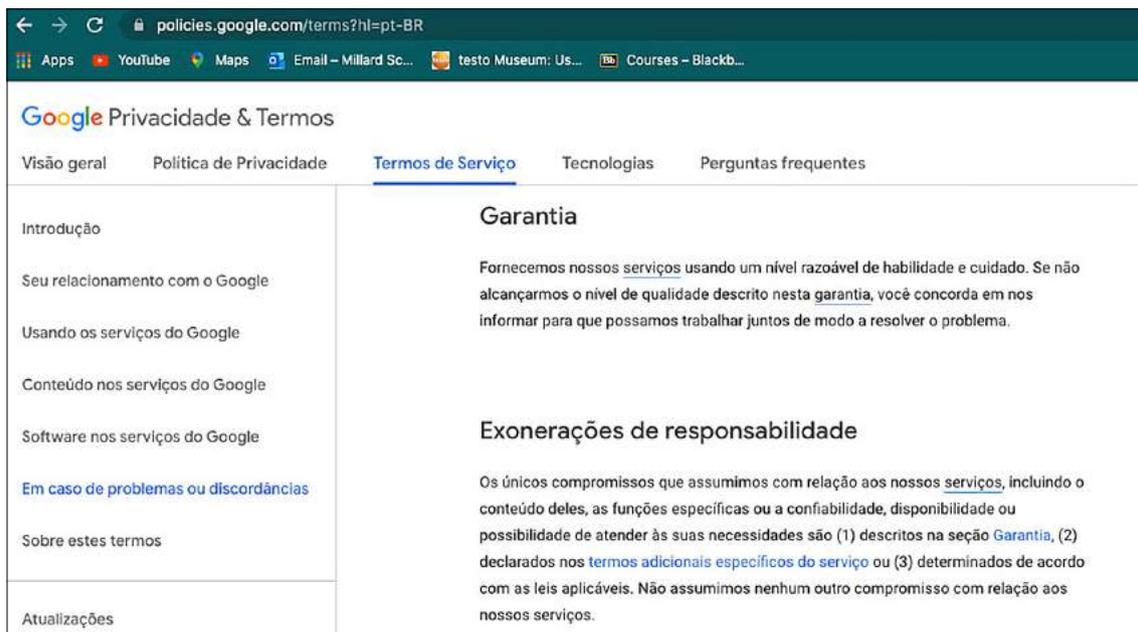


Figura 14. Tela com descrição de termos de serviço do site de busca Google. Disponível em: <https://policies.google.com/terms?hl=pt-BR>.

A dúvida que fica é que, mesmo sem estas garantias, estas empresas terão mais condições de providenciar esta guarda do que uma grande maioria das pessoas que não terão o conhecimento, o tempo, as condições para aquisição e monitoramento de estruturas de preservação em nível pessoal, conforme estamos discutindo neste texto. Novamente, a questão da democratização da produção e a elitização da guarda.

A Google vem investindo em *Machine Learning* (aprendizado de máquina) para identificar e organizar fotografias usando dados de reconhecimento das imagens coordenadas com dados EXIF de data e localização, também utilizado pela Apple. Aproveitar esses sistemas para fornecer uma forma de navegar seu conteúdo de imagens pode ser interessante. Há tutoriais na internet de como fazer o *download* de todo o seu conteúdo para que seja armazenada uma cópia em um sistema fora da nuvem como uma garantia de *backup* de seu conteúdo.

3

O que queremos preservar

É necessário entender o que compõe uma fotografia digital, porque queremos armazenar a melhor versão e a mais próxima da captura digital. Nem sempre temos essas opções, porque as imagens já podem chegar para serem armazenadas em um formato específico e com muita compressão, como já foi apresentado, e estas podem ser as únicas versões referentes àquela produção. Nesse sentido, são os melhores originais (ou os únicos) para preservar. Podemos tecer comparações com a fotografia analógica, relacionando o negativo e as ampliações em papel feitas a partir dele. O negativo filmico é o que está mais próximo da captura inicial. A ampliação é uma derivada produzida posteriormente. É comum, em acervos fotográficos, existirem somente ampliações em papel, sem referências ao paradeiro do negativo original. Vivemos situações semelhantes com a fotografia digital. Derivadas, tratadas, com compressão, enviadas pela internet, representam os materiais de difusão de uma imagem. Elas podem acabar por ser as sobreviventes, com a perda dos “negativos digitais” que lhe deram origem. Decidir o que queremos preservar nesse universo digital é sempre uma etapa, que será também norteadada pelas condições de recursos, humanos e materiais, para esta preservação. Nem sempre temos essas opções, porque as imagens já chegam para serem armazenadas em formatos com compressão, e

já não se encontra mais a imagem de origem. Podemos ter uma versão enviada pelas redes sociais, com perdas, e esta imagem pode ser a única sobrevivente e virar um “novo original”, pois a imagem que deu origem a esta derivada se perdeu.

Nossa intenção deve ser procurar entender os arquivos com os quais estamos lidando e buscar pela preservação das matrizes, como em um negativo original de câmara, e produzir os arquivos necessários para difusão na internet. Vamos começar com os formatos de arquivos – que representam a forma como os dados digitais são codificados e interpretados. Com a evolução nas plataformas e *softwares*, esse universo é dinâmico. Como exemplo, os formatos de arquivos como o Photo CD, da Kodak, do começo de 2000 (.pcd), são considerados obsoletos. Existem parâmetros bem definidos hoje sobre os tipos de formatos de arquivos que são mais recomendados, estáveis, abertos e tendem a prometer uma vida mais longa do que outros. Um dos documentos mais importantes, lançado a cada dois anos, é o RFS 2021-2022, *Recommended Formats Statement*²⁰ (Documento de recomendação de formatos produzido pela *Library of Congress*, EUA).

ii. Photographs – Digital		
	Preferred	Acceptable
A. Faithful representation of the work	<ul style="list-style-type: none"> • Equal in quality to the published version, best edition or master copy • In the same format as the master copy 	
B. Technical Characteristics	<ul style="list-style-type: none"> • Highest resolution available, not rescaled or interpolated • Highest bit depth available, 16 bits per channel if available • Embedded color profile or specified color space used in published version • Uncompressed • Unlayered 	<ul style="list-style-type: none"> • Lossless compression or lower compression ratios • Discrete wavelet transform (DWT) preferred to discrete cosine transform (DCT) • Layered, if supported by preferred or acceptable format
C. Formats	<ul style="list-style-type: none"> • TIFF (*.tif) • JPEG2000 (*.jp2) • PNG (*.png) • JPEG/JFIF (*.jpg) • BMP (*.bmp) 	<ul style="list-style-type: none"> • Photoshop (*.psd, *.psb) • JPEG 2000 Part 2 (*.jpf, *.jpx) • Digital Negative DNG (*.dng) • Proprietary Camera Raw formats (*.nef, *.crw) • GIF (*.gif)
D. Metadata	<ol style="list-style-type: none"> 1. As supported by format <ol style="list-style-type: none"> a. Title b. Creator c. Creation Date d. Place of publication e. Publisher/producer/distributor f. Contact information 2. Include if available: <ol style="list-style-type: none"> a. Common embedded schema (e.g., IPTC) b. Language of work c. Other relevant identifiers (e.g., PLUS ID, DOI, LCCN, etc.) d. Subject descriptors e. Abstracts f. Key or reference to each data field and technical production information (e.g. EXIF metadata from digital camera) 	Metadata provided separately in external text of XML-based file

Figura 15. Nesse gráfico do *Recommended Formats Statement*, na seção de fotografias digitais, vemos as principais recomendações preferidas e aceitáveis: A. Uma representação fiel do trabalho; B. Características técnicas; C. Formatos e D. Metadados.

Para cada tipo de arquivo, como texto, imagem, vídeo, sonoro, entre outros, há uma recomendação de formatos *preferidos*, mais sustentáveis, e outros formatos considerados *aceitáveis*. Estas decisões são elaboradas baseadas nos 7 fatores de sustentabilidade de formatos digitais:

- 1 Divulgação de seu código e especificações – valorização de formatos não proprietários, documentação completa e aberta;
- 2 Adoção pelo mercado – amplamente utilizado por criadores, usuários e indústria;
- 3 Transparência – disponibilidade de inúmeras ferramentas capazes de ler e escrever para esse formato;
- 4 Autodocumentação – metadados técnicos, administrativos e descritivos embutidos no próprio formato;
- 5 Formatos com um mínimo de dependências externas – o formato depende de um tipo especial de *hardware*, sistema operacional ou *software*;
- 6 Impacto de patentes – existem patentes que podem tornar o formato mais arriscado para longevidade.
- 7 Mecanismos de proteções técnicas – há algum tipo de mecanismo de compressão ou criptografia.

O texto *Sustainability of Digital Formats*²¹ (Sustentabilidade dos formatos digitais) da Biblioteca do Congresso explica com mais detalhe esses 7 itens.

Neste capítulo, o intuito é fazer um panorama de informações mais técnicas sobre fotografias digitais. Com uma clareza maior sobre estas questões, podemos fazer escolhas conscientes do que queremos preservar.

Formatos de arquivo de imagens digitais: RAW, DNG, TIFF, JPEG, JPEG 2000, PNG, PDF

Quando salvamos um arquivo digital de uma imagem, os *bits* (abreviação de *binary digit* – dígito binário) são organizados de forma que possam ser posteriormente codificados ou lidos. O método ou fórmula para o empilhamento desses *bits* para que os programas possam ler esse código é o que chamamos de formato de

arquivo. Estabelecemos então que um arquivo de imagem digital para ser visualizado necessita de um programa ou aplicativo específico que por sua vez necessita ser instalado em um sistema operacional de um determinado computador. Assim, a visualização de uma imagem digital é inteiramente dependente de programas e sistemas operacionais. Sem esses decodificadores os *bits* não têm significado e não formam imagens.

Com as mudanças das tecnologias, formas e capacidade de produção de imagens digitais, os formatos de arquivo vão sendo alterados para uma adequação a estas mudanças. Surgem novas funções, novos programas, novas plataformas. Às vezes as variações são pequenas, mas podem ser suficientes para dificultar que um formato de arquivo seja lido pelos sistemas que os liam. É de suma importância pensar na obsolescência dos sistemas de leitura dos formatos de arquivo em projetos de preservação digital, caso haja necessidade de que seus produtos tenham longevidade.

Os formatos de arquivo para imagens digitais que temos hoje já estão bem consolidados, mas podemos considerá-los tão permanentes quanto o *pixel* e os sistemas atuais de decodificação. Esse texto certamente terá um caráter histórico como documentação de uma era em que as imagens eram feitas com *pixels*. Uma das funções principais referentes ao que queremos preservar é pensar em formatos com ampla adoção e com seus códigos documentados, conhecidos e abertos, mas com mira sempre em formatos que podem se tornar obsoletos e necessitar de migração para formatos mais correntes. Esse é um dos desafios da preservação digital – ela exige constante monitoramento e ações para manter os dados digitais acessíveis/visíveis.

Formatos renderizados ou não renderizados

Hoje podemos ter formatos de imagem renderizados ou não. Um arquivo de imagem renderizado é um formato processado e pronto para ser visualizado em editores de imagens, que abrem arquivos de mapa de *bits* (arquivos compostos de *bits*). Esses são os formatos como TIFF, JPEG, e PNG, entre outros. Um arquivo de imagem não renderizado é um arquivo não processado, RAW ou bruto, que precisa de um processo de conversão (conhecido como renderização) para gerar imagens acessíveis pelos editores de mapa de *bits*. Esses são os formatos como RAW e DNG.

Existem muitos formatos de arquivos para imagens digitais, mas discutiremos neste texto os formatos relevantes a um fluxo de trabalho valorizando a preservação digital dos arquivos gerados dentro do escopo do documento de formatos recomendados da *Library of Congress*.

Considerando um fluxo de preservação de fotografias digitais, um arquivo não renderizado pode ser preservado como uma matriz, ou documento original da captura (RAW ou DNG), e os arquivos derivados, como TIFF ou JPEG, podem ser gerados sob demanda. Em um fluxo, por exemplo, JPEGs de baixa resolução para internet podem ser gerados de todos os arquivos RAW. Arquivos TIFFs serão gerados somente para as fotografias que serão tratadas e impressas para exposições ou livros. Como um arquivo RAW tem somente as informações de luz capturadas pelo sensor por baixo de um filtro mosaico de cores RGB, ele possui somente um canal. Uma imagem renderizada, como um TIFF, possui três canais, RGB (vermelho, verde e azul). Portanto, podemos ter números de tamanho de arquivo conforme a tabela a seguir:

Name	Date Modified	Size	Kind
D800e_36mp.dng	12:28	35,5 MB	Digital Negative file
D800e_36mp.NEF	02/08/16	75,6 MB	Nikon Electronic Format file
JPEG 0.jpg	12:24	527 KB	JPEG image
JPEG 4.jpg	12:23	965 KB	JPEG image
JPEG 8.jpg	12:21	2,3 MB	JPEG image
JPEG 10.jpg	12:23	4,3 MB	JPEG image
JPEG 12.jpg	12:22	14,7 MB	JPEG image
JPEG 2000.jpg	12:24	497 KB	JPEG 2000 image
TIFF 8bpc.tif	12:19	108,5 MB	TIFF image
TIFF 16bpc.tif	12:19	216,9 MB	TIFF image
TIFF cmyk 8bpc.tif	12:20	145,2 MB	TIFF image

Figura 16. Nesta tabela podemos ter uma visão de tamanhos de arquivos diferentes a partir de um arquivo original NEF de uma câmera D800e de 36 *megapixels*. Vemos os diferentes tamanhos de arquivos TIFF RGB, com 8 ou 16 *bits* por canal, e com TIFF CMYK, com quatro canais. Temos também as versões com compressão JPEG, e um arquivo DNG.

Arquivos RAW

Arquivos RAW, “cru” em inglês, são imagens em escala de cinza que registram somente os valores de luminância para cada *pixel*, onde cada fotodiodo representa

um *pixel*. Esses valores são processados para gerar uma imagem colorida pela montagem dos valores de luminância vistos através de um filtro Bayer, um mosaico com quadrados verdes, vermelhos e azuis colocado em cima dos fotodiodos. O processo de renderização de uma imagem gera arquivos que podem ser vistos em editores de imagens como o Photoshop. O trabalho de montar e interpretar as informações de luminância visto através do filtro Bayer é o trabalho de um *software* de conversão de arquivos RAW.

A conversão de um arquivo RAW pode ser feita posteriormente à sua captura, e, se esse arquivo RAW for preservado, poderemos voltar inúmeras vezes a ele para processá-lo. Nesse sentido, o arquivo RAW passou a existir como um negativo digital ou matriz digital, gerador de arquivos derivativos. O arquivo RAW não pode ser considerado um formato de arquivo de imagem digital porque cada câmera de cada fabricante, com suas funções específicas, produz um arquivo RAW levemente diferente, já que nesse arquivo estão todas as informações que aquela câmera específica é capaz de registrar. Os fabricantes dos equipamentos detêm a “chave” para abrir esses arquivos, e nesse sentido esses arquivos são considerados proprietários e listados como aceitáveis para a preservação digital no gráfico citado da *Recommended Formats Statement*. Não temos acesso à documentação pública de como esses arquivos organizam os dados da imagem. Esta chave decodificadora é enviada às diversas empresas que vendem *softwares* de conversão de arquivos RAW, para que os usuários possam abrir esses arquivos. O maior problema de pensar em arquivos RAW como matrizes de preservação é que os fabricantes terão a tarefa de reescrever os *plug-ins* (chaves) que abrem esses arquivos para cada atualização de sistema operacional ou aplicativo. Em 2030, um determinado fabricante precisará ter funcionários dedicados para escrever “chaves” para abrir arquivos RAWs de equipamentos feitos em 2020, para que estes possam ser lidos pelos sistemas de 2030. Sabemos que isso não acontecerá, pois, mesmo hoje, nos sistemas atuais, já não conseguimos ler com facilidade alguns dos primeiros formatos de arquivos RAW. Também não temos certeza da longevidade dos fabricantes – já vimos empresas, sistemas, *softwares* e *hardwares* aparecerem e desaparecerem nesse curto espaço de tempo de algumas décadas.

Podemos pensar em algumas soluções: primeiro, obrigar os fabricantes de equipamentos e de seus respectivos RAW a documentar pública e abertamente os

seus métodos proprietários e as *chaves* (chamamos de chaves, mas são os *softwares* escritos pelo fabricante que tornam seus arquivos RAW proprietários visíveis para certos produtos de mercado e sistemas operacionais), algo que provavelmente não irá acontecer. A documentação continua sendo proprietária, mas os fabricantes produzem e disponibilizam as chaves. Segundo, fortalecer um formato RAW universal, público e documentado abertamente, como o DNG; e terceiro, manter o arquivo proprietário RAW, mas estar atento para o momento necessário de migrá-lo/convertê-lo para um formato de arquivo aberto.

Quando a imagem é capturada em TIFF ou JPEG, o arquivo é processado pelo equipamento digitalizador, câmera ou escâner, que converte as informações brutas de luminância em *pixels* conforme os ajustes predeterminados ou pré-ajustados por esses sistemas de digitalização, e salva essas informações em um formato de arquivo, TIFF ou JPEG, jogando fora os dados luminosos capturados pelo sensor. Nesses formatos de arquivo, já processados ou renderizados, não existe a possibilidade de voltar a acessar esses dados brutos do pré-processamento, provenientes do sensor.

Captura em celular com RAW e JPEG

Nestas duas imagens a seguir, podemos ver as diferenças do que representa uma captura em um celular no modo RAW, via uso de aplicativos no próprio celular, e no modo padrão de arquivo processado. A captura em RAW permite registrar e salvar informações que são perdidas no processo de compressão de dados na geração de um arquivo JPEG.

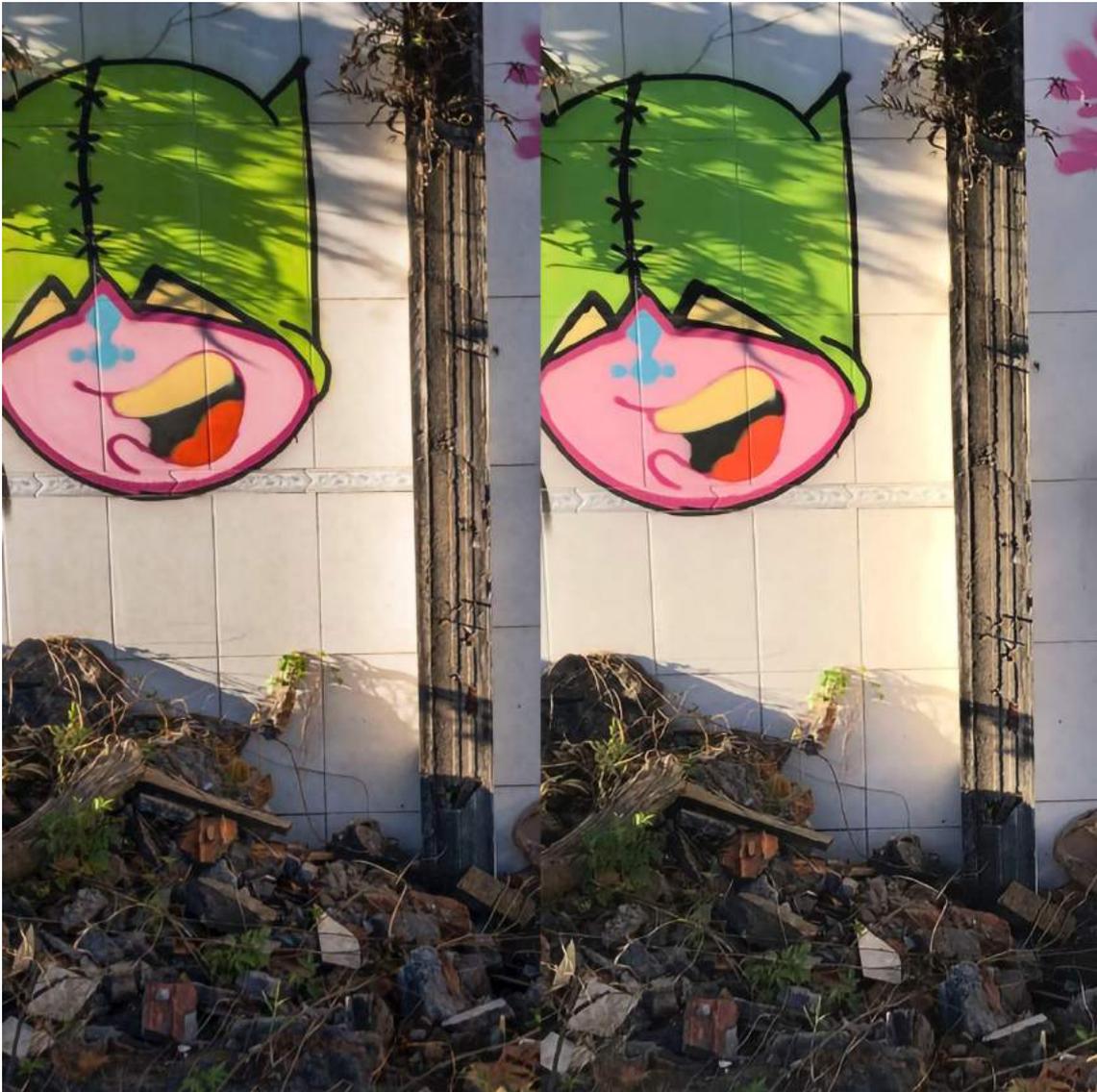


Figura 17. A imagem da esquerda, arquivo RAW, tem muito mais informações, principalmente na região das altas luzes, que estão “estouradas”, atingindo o branco máximo, nível 255, na versão JPEG à direita.

Name	^	Dat...dified	Size	Kind
 IMG_0329.dng		25/05/20	10,8 MB	Digital Negative file
 IMG_0329.jpeg		25/05/20	3,9 MB	JPEG image

Figura 18. Estas duas imagens também apresentam uma boa diferença de tamanho, 10,8 MB para o arquivo DNG, e 3,9 MB para o JPEG.

Digital Negative (DNG)

Pensando no problema de centenas de arquivos RAW provenientes de inúmeras máquinas de diversos fabricantes, a Adobe criou, em 2004, um formato de arquivo aberto conhecido como DNG, ou negativo digital. Esse formato é livre de *royalties*, licenciado gratuitamente para qualquer desenvolvedor e, como extensão do TIFF 6.0, é compatível com o TIFF/EP (*Tag Image File Format/Electronic Photography Format*) ISO 12243-2. Esse formato ainda está em vias de virar uma norma ISO. Todos os tipos de arquivos RAW podem ser convertidos via aplicativos gratuitos, como o *DNG converter*, para um único formato universal, o DNG. O DNG continua representando os dados brutos capturados pelo sensor, sem processamento.

DNG é um recipiente que contém todos os dados proprietários do arquivo RAW e informações adicionais – os metadados originais, as instruções de processamento do arquivo e quaisquer metadados adicionados ao DNG são arquivados no próprio arquivo, em seu cabeçalho. Os arquivos proprietários RAW não permitem alterações e quaisquer mudanças realizadas neles são gravadas em um arquivo adicional de texto, escrito em xml (*por exemplo: nomedoarquivo.nef e nomedoarquivo.xmp*) e salvo na mesma pasta, que documenta estas informações, já que não podem ser escritas/gravadas nesse arquivo proprietário como o NEF da Nikon. Esse arquivo adicional é conhecido como *sidecar*, pela sua semelhança ao assento lateral adicional em motocicletas.



Figura 19. Arquivo original NEF da câmera digital D800e da Nikon. Ao ser aberta e alterada a imagem no arquivo RAW, gera-se um arquivo XMP Sidecar que conterà estas alterações, que não podem ser escritas no arquivo proprietário da Nikon.

Se esse arquivo *xmp* for descartado, as informações inseridas nos metadados ou processamentos realizados via aplicativos *PIEware*, como Lightroom, serão perdidas.

O PIEware, *Parametric Image Editing*, é uma classe de *software* de edição de imagens não destrutiva, na qual o *software* de edição não altera os arquivos originais, mas, em vez disso, registra as alterações nas imagens como conjuntos de instruções ou parâmetros. Em arquivos RAW, estas alterações são salvas em arquivos.xmp com o mesmo nome original, conhecidos como arquivos *sidecar*. Em *softwares* de catalogação, como Adobe Lightroom, alterações em arquivos como TIFF e JPEG, já processados, são armazenadas na base de dados do Lightroom e podem ser exportadas para os metadados da imagem. Diversos *softwares* ajustam imagens dessa forma além do Lightroom, como o Adobe Camera Raw (ACR), o Aperture, o Bibble e o Capture One.

Se o fluxo de trabalho de um projeto de digitalização contemplar o uso de câmeras, os arquivos RAW devem ser convertidos para DNG. Isso pode ser feito no início do processo de trabalho ou após renomearem-se os arquivos, aplicarem-se algumas alterações na imagem ou adicionar-se texto nos campos de metadados. Isso dependerá do fluxo de trabalho escolhido. Se o projeto contemplar escâneres, os arquivos já serão processados em um formato de arquivo universal como TIFF, e não teremos as opções de acessar os dados brutos da captura.

Existem ainda algumas questões não resolvidas com o formato DNG. Alguns parâmetros de arquivos RAW proprietários, como o registro do ponto de foco utilizado durante a captura da imagem, e possivelmente outros, presentes nos arquivos originais de câmera, não estão presentes nos seus respectivos arquivos convertidos para DNG. Questões também relativas ao processo de conversão e ao gerenciamento de cores, entre outros, ainda requerem estudos e pesquisas em sua avaliação. O formato DNG ainda não é norma ISO, mas há um movimento para defini-lo como tal. De qualquer forma, a especificação da Biblioteca do Congresso sobre os formatos de arquivos preferidos e aceitáveis coloca o RAW proprietário e o DNG como formatos aceitáveis. Será da responsabilidade do detentor desses arquivos RAW entender o momento no futuro próximo em que esses arquivos poderão se tornar obsoletos. Por exemplo, se Nikon, Canon, Adobe ou outro fabricante não existirem mais, e com mudanças também nos sistemas operacionais, as chances de haver suporte para esses arquivos RAWs, proprietários ou não, se tornam mais remotas. Esse cuidado de migração de formatos em obsolescência ou obsoletos para os formatos sustentáveis deverá ser tomado com todos os formatos de arquivo.

TIFF

Os termos *Tagged Image File Format* e *Tag Image File Format* foram os subtítulos utilizados nas primeiras especificações dos arquivos TIFF. A partir da especificação TIFF 6.0, em 1992, ficou conhecido somente como TIFF. O formato de arquivo TIFF foi criado originalmente, no meio da década de 1980, para que os fabricantes de escâneres de mesa acordassem na utilização de um formato comum de arquivo em vez de cada fabricante tentar desenvolver e impor seu formato proprietário. A primeira versão da especificação TIFF foi publicada pela *Aldus Corporation* em 1986, e era um formato de imagem binário (capaz de processar somente imagens de 1 *bit*, dois valores possíveis para cada *pixel*). Rapidamente evoluiu para o que é hoje, capaz de processar imagens em alta resolução e grande profundidade de *bits*.

O formato de arquivo TIFF revisão 6.0 sem compressão é considerado um dos melhores formatos para acervos, já que é um formato de arquivo padrão, bem documentado, com uma trajetória de uso de mais de duas décadas e que deverá se manter ainda por um bom tempo. Como também é um formato amplamente utilizado, será fácil pensar na sua migração para formatos futuros.

Arquivos TIFF podem ser salvos com compressão LZW, ZIP e JPEG. Somente a compressão JPEG é com perdas. Para utilização como um arquivo de preservação, não é recomendado ter compressão, pois cria-se mais um fator, o da descompressão, para migrações futuras. Como os arquivos TIFFs são renderizados, são muito maiores que os arquivos não processados (RAW e DNG).

JPEG

O comitê do *Joint Photographic Experts Group*²² foi formado em 1986 e, desde então, vem estabelecendo os padrões para os arquivos JPEG, JPEG 2000 e JPEG XR. Um arquivo JPEG é produzido pela compressão com perda de dados (*lossy compression*), e por isso possui um tamanho menor de arquivo, que facilita a sua guarda e transmissão via internet. Por ser um arquivo com compressão, muitas informações do arquivo original, ou da captura original, são descartadas, e a qualidade final de imagem desse arquivo acaba sendo inferior à do arquivo que lhe deu origem. Como esses arquivos são descomprimidos e comprimidos a cada sessão de

abertura, alterações, salvamento e fechamento, eles devem ser considerados como arquivos de saída, finalizados, e não arquivos de trabalho. Para trabalhar em um arquivo JPEG, ele deve ser aberto e salvo como TIFF antes de realizar alterações, podendo depois ser finalizado, convertido novamente a um arquivo JPEG. Dentro de um projeto de digitalização, o arquivo JPEG deve ser considerado como um arquivo derivado final gerado para distribuição a partir de uma matriz, que deverá ser arquivada. Não temos como pensar em arquivos grandes, como TIFF, para difusão via e-mail e internet.

Como as matrizes em DNG e TIFF podem ser dezenas de vezes maiores do que os JPEGs gerados a partir deles, sempre é difícil definir o que deverá ser armazenado sem considerar os recursos institucionais, volume de material, condições de conservação dos originais e possibilidades de recaptura. Não podemos afirmar categoricamente que tanto os arquivos TIFF como os JPEGs gerados a partir desses TIFFs devem ser armazenados. Em um jornal, por exemplo, onde dezenas de fotógrafos produzem milhares de fotografias quase que diariamente, e que não podem ser recapturadas, o arquivamento dos originais em RAW ou DNG seria o ideal para garantir a qualidade e integridade da captura do original e sua preservação em longo prazo, mas pode ser impraticável ou financeiramente proibitivo, forçando uma guarda somente dos arquivos em JPEG. Como o TIFF pode ser de 3 a 6 vezes maior (dependendo se for 8 *bits* por canal ou 16 *bits* por canal) do que os arquivos RAW, isso terá um impacto ainda maior no custo de longo prazo da preservação.

JPEG2000

O JPEG 2000 é um padrão de codificação e compressão de imagem criado pelo comitê do *Joint Photographic Experts Group*, em 2000, para substituir o padrão de compressão do JPEG, a Transformada Discreta de Cosseno (ou DCT da sigla em inglês para *Discrete Cosine Transform*), por um novo método baseado em Transformada Discreta de Wavelet (DWT). A Transformada Discreta de Wavelet, ao contrário da DCT, não divide a imagem em blocos de 64 *pixels* (8x8); ela é aplicada em toda a imagem, adquirindo mais informações dos *pixels* em cada aplicação e, dessa forma, melhorando a sua eficiência.

Outras vantagens do JPEG 2000 são a possibilidade de salvar imagens em 16 *bits* por canal e fazer a decodificação do mesmo arquivo em diferentes resoluções de saída. O JPEG 2000 pode ser um formato do futuro, mas ainda não teve a repercussão e a implementação esperada, bem como o apoio da indústria e de usuários. Mesmo assim, seu uso está se ampliando. O JPEG 2000 é amplamente usado como formato de entrega para servidores IIF²³ (*International Image Interoperability Framework*), devido à sua descompactação inteligente. A compactação sem perdas torna possível usar o mesmo arquivo como um formato de preservação de longo prazo.

PDF, PNG e JPEG XR

Não usuais como o TIFF, o JPEG e o DNG, esses formatos podem vir a ser interessantes como padrões de armazenamento para preservação. O PDF, muito comum para textos, também pode ser gerado a partir de imagens com seus metadados embutidos, que podem ser facilmente vistos por leitores de PDF. O PNG é um formato que aceita transparências, é de código aberto e pode gerar imagens de 16 *bits* por canal, mas não é muito utilizado pelos acervos ou indústria como um todo. O JPEG XR, originalmente conhecido como Windows Media Photo, depois renomeado para HD Photo Microsoft, é agora um formato padrão ISO. O JPEG XR oferece uma compressão mais eficiente com menos artefatos que um JPEG padrão, tem suporte para imagens de 16 bpc e grande escala dinâmica. Pode ser um formato adotado pelos fabricantes de câmeras digitais. O formato adotado pela Apple em 2017 para seus iPhones o HEIF²⁴ (*High Efficiency Image File Format*) é um exemplo de um formato novo com adoção limitada, ainda sem a mesma ampla adoção do JPEG.

Derivadas e versionamentos

Uma outra forma de pensar nos diversos formatos de arquivos é por uma divisão de tarefas. Nem sempre o arquivo original de captura será o arquivo final de armazenamento. Temos versões de trabalho, onde fazemos tratamentos; versões de entrega e as versões de armazenamento para longevidade.

Captura – a maioria das câmeras digitais simples capturam as imagens, e o conversor A/D (analógico/digital) já as processa e as comprime no formato JPEG. Não existem outras possibilidades. Câmeras mais profissionais permitem gerar arquivos RAW, TIFF ou JPEG. Escâneres de mesa, filmes e outros formatos capturam a imagem e ela é processada ou renderizada quase sempre no formato de arquivo TIFF, podendo ser também em PDF ou JPEG.

Trabalho – podemos fazer alterações nos arquivos digitais de imagens de todos os formatos, mas isso não deve ser feito diretamente nos arquivos JPEG, com o risco de degradação da qualidade de imagem. Trabalhamos em arquivos não processados via *softwares* PIEware²⁵ (*Parametric Image Editing* – edição de imagem paramétrica), que edita arquivos RAW ou DNG (Lightroom, Bibble Labs etc.); ou usamos *softwares* de edição de imagens em mapa de *bits*, como Photoshop, XnView, Gimp, para os arquivos processados/renderizados como TIFFs.

Entrega – os arquivos que podem ser distribuídos para materiais impressos, internet etc. devem ser em formatos processados e finalizados, como TIFF e JPEG. Arquivos com camadas de trabalho, proprietários como o PSD da Adobe, e arquivos RAW e DNG não devem ser arquivos de entrega, pois necessitam de *software* especializado e processamento para serem vistos.

Arquivamento – os formatos de arquivo de preservação mais utilizados hoje são o TIFF e o DNG, como arquivos mestres de alta resolução, e o JPEG como as versões em baixa resolução para difusão. Menos comuns são os formatos JPEG 2000 e PDF. Por outro lado, se os originais foram gerados nas câmeras digitais ou celulares, como JPEG, deverão ser armazenados nesse formato como matriz.

As decisões sobre o que guardar podem significar preservar as imagens digitais originais da captura (RAW/DNG), arquivos tratados de trabalho (TIFF) e versões utilizadas para difusão na internet (JPEG). As versões de trabalho e difusão podem ser somente de uma seleção de imagens originais e feitas somente sob a demanda de seu uso.

Profundidade de cor ou *bits* (níveis, 1 bit, 8 bits, 16 bits)

Uma das decisões que temos que tomar é sobre a profundidade de cor das imagens digitais capturadas e geradas e quantos *bits* terão as imagens finais. Os sistemas atuais de captura podem gerar imagens de 8 ou 16 *bits* por canal. O padrão, hoje, para uso de imagens coloridas RGB ou CMYK (RGB – vermelho, verde e azul,

e CMYK – ciano, magenta, amarelo e preto) na internet, publicações e impressões é de 8 *bits* por canal, mas gerar imagens em 16 *bits* por canal, com o dobro do tamanho de arquivo, pode produzir matrizes de melhor qualidade.

1, 4, 8 e 16 bits

Nas primeiras imagens digitais o pixel era composto por um *bit* (abreviação de *binary digit* – dígito binário). Esse *bit*, ou unidade de informação, apresenta duas possibilidades, como um circuito eletrônico – ligado (1) ou desligado (0). Em uma imagem, o resultado de uma captura em 1 *bit* é produzir somente áreas pretas ou brancas – as duas possibilidades desse *bit*. Conseguimos até produzir a ilusão de uma imagem em tons de cinza pela variação da densidade dos pontos pretos – áreas mais claras com poucos pontos pretos, áreas escuras com maior densidade de pontos pretos, áreas cinza com 50% de pontos pretos. Isso é muito semelhante a imagens em tons de cinza produzidas por prata metálica em papel fotográfico.

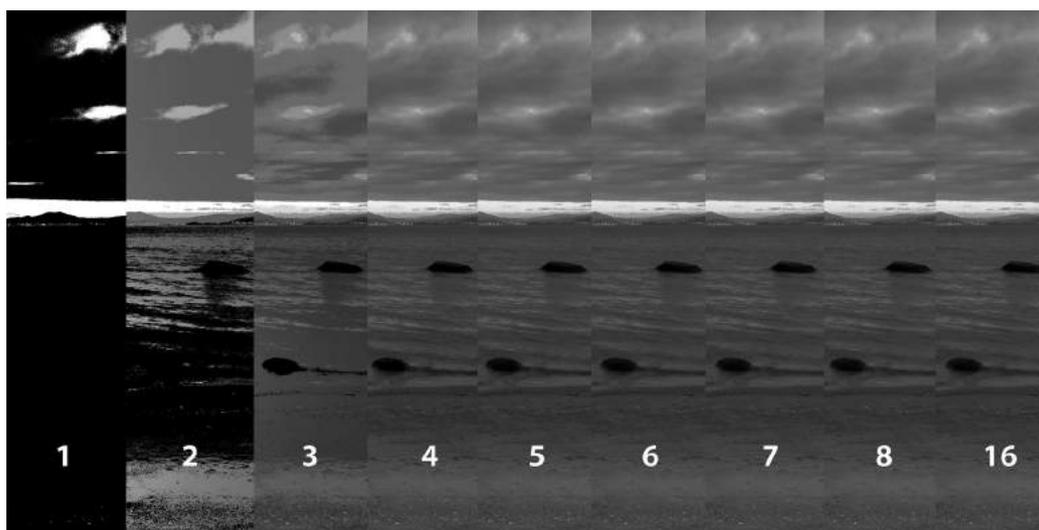


Figura 20. A imagem acima foi convertida para *grayscale* (escala de cinza), em um canal somente, para visualizar melhor as diferenças na comparação de *bits* por canal.

1 *bit* de informação por pixel → ligado x desligado → $1^2 = 2$ tons, preto ou branco

2 *bits* por *pixel* → $2^2 = 4$ tons

4 *bits* por *pixel* → $2^4 = 16$ tons

8 *bits* por *pixel* → $2^8 = 256$ tons (do 0, preto puro ao 255, branco puro)

16 *bits* por *pixel* → $2^{16} = 65.536$ tons

As diferenças visuais são grandes nos recortes de imagem de 1 a 5 *bits* por canal, mas nos de 6 a 16 *bits* por canal são mais sutis e parecem quase iguais. No entanto, para quem trabalha com tratamento de imagens digitais, ter um arquivo de 16 bpc, com uma quantidade enorme de tons ou passagens do preto ao branco, é o que há de melhor na captura da informação feita por dispositivos digitais. Mesmo sendo um arquivo com tamanho grande (pesado, como falamos no jargão), irá gerar o melhor resultado na qualidade da imagem. Contudo, também temos a necessidade de ter arquivos menores para outras finalidades. As redes sociais irão jogar fora muita informação de tonalidades e cores nas imagens para reduzir o seu tamanho. Em uma visualização inicial, a maior parte dos usuários não irá ver essas diferenças em uma tela de celular ou computador, o que é muito parecido com a nossa experiência ao visualizar as transições do 6 ao 16 bpc.

Os primeiros monitores de computadores eram também de 1 *bit*, nas cores âmbar e verde. Logo vieram imagens em 4 *bits*. Como cada unidade pode ter duas possibilidades (0 ou 1), temos então 16 possibilidades de cinza (2^4 – dois, elevado à quarta potência). O próximo desenvolvimento foi de *pixels* com profundidade de cor de 8 *bits*, gerando 256 possibilidades de cinza (2^8). Estas “possibilidades” se traduzem em níveis de cinza nos aplicativos que tratam imagens, como Photoshop. Um *pixel* pode ter um valor de 0, para preto, 128 para cinza médio e 255 para branco, com todos os outros valores intermediários. No espaço de cor RGB, temos um canal para cada cor primária, composto de 8 *bits* (8 bpc – *bits* por canal). A imagem colorida RGB possuirá, então, 24 *bits*, e mais de 16 milhões de cores ($256^3 = 256 \times 256 \times 256 = 16.777.216$ cores). No jargão diário, é comum fazer referência a uma imagem de três canais (RGB, 24 *bits*) como sendo uma imagem em 8 *bits*, e imagens de 16 *bits* para arquivos RGB de 16 *bits* por canal e 48 *bits* totais. Este caderno técnico fará referência sempre ao número de *bits* por canal.

Os sistemas atuais de captura e digitalização permitem registrar e salvar imagens com até 16 *bits* por canal, ou com um total de 48 *bits* para imagens em RGB. Um canal de 16 *bits* pode render 65.536 tonalidades de cinza (2^{16}), e a imagem composta RGB terá mais de 280 trilhões de cores (65.536^3).

Apesar de ser complexo mensurar o que o olho humano consegue enxergar, existe um consenso que não conseguimos nem mesmo distinguir todas as cores de uma imagem colorida de 8 *bits* por canal (24 *bits*). A maior parte das pessoas não consegue perceber e diferenciar mais de 10 milhões de cores. E esse número se reduz muito mais ainda quando pensamos em imagens impressas.

8 ou 16 bits?

Podemos pensar na profundidade de *bits* como uma escada. Quando temos 256 níveis entre o pé e o topo dessa escada (8 bpc), cada degrau é muito mais pronunciado quando comparado à mesma escada com 65.536 degraus (16 bpc). Uma matriz digital de 16 bpc (48 *bits*) tem o dobro do tamanho de arquivo em relação à mesma versão em 8 bpc (24 *bits*). Nesse sentido, 10 *terabytes* de arquivos em 48 *bits* ocuparão 5 *terabytes* se convertidos em 24 *bits*. O mapeamento final de quanto espaço será necessário para todo o material digital produzido poderá ser um dos fatores na escolha de manter os arquivos matrizes em 16 ou 8 bpc.

Uma imagem digital em 16 bpc pode ter mais ajustes e correções sem apresentar perdas nas transições tonais, conhecidas como *banding* ou “posterização”. Esse defeito visual que se apresenta como falhas nas passagens de tons, ocorre em uma imagem quando não há dados suficientes para manter uma aparência de uma tonalidade contínua. É como se alguns degraus da escada estivessem faltando. Um arquivo de 8 bpc, como tem menos tons de cinza para serem trabalhados, não tem a mesma flexibilidade de um arquivo de 16 bpc.

Alguns acervos e conservadores acreditam que mesmo que a maior parte dos sistemas de saída – internet, publicações, *displays* etc. – não tenham suporte para imagens em 16 bpc, vale a pena guardar o registro na melhor qualidade possível, pensando inclusive nos sistemas futuros que virão. Esta decisão não deve ser isolada, mas baseada principalmente no estado de conservação dos materiais que estão sendo digitalizados e na possibilidade de acessá-los no futuro. Quando estamos conversando sobre arquivos nato-digitais, temos que tomar esta decisão na hora de sua preservação. Uma imagem preservada em 16 bpc sempre pode ser reduzida para 8 bpc, mas o contrário, de 8 para 16, quando é feito, não faz sentido. Não gera qualquer ganho de qualidade de imagem, somente no tamanho do arquivo que dobra.

Quando a captura é feita com câmeras digitais, o arquivo nato-digital RAW, não processado, pode ser arquivado no formato DNG como uma matriz. A partir desse DNG, quaisquer arquivos podem ser gerados, e a qualquer momento. Portanto, arquivos TIFF em 8 ou 16 bpc podem ser gerados, bem como JPEGs para difusão. O DNG poderá ser arquivado e acessado se houver necessidade no futuro da produção de uma imagem derivada para finalidades específicas. Esta abordagem é

muito interessante, porque o DNG é muito menor, por ser um arquivo de um canal somente, não processado, e armazenado como um negativo digital – “ampliado” ou processado somente sob demanda. Considerando os custos constantes de armazenamento, a pegada de carbono que representa todo o ecossistema digital de custos e gastos de energia para sua manutenção, faz muito sentido pensar em formas mais sustentáveis de trabalhar. Uma matriz DNG e uma versão de tudo em JPEG para difusão, e a utilização de arquivos processados em alta somente sob demanda pode ser uma abordagem que leva tudo isso em conta.

Resolução

Sempre que possível, queremos preservar a imagem com maior resolução, pois ela nos dá mais possibilidades de usos atuais e futuros. Quando uma imagem é enviada pelas redes sociais, como já vimos neste texto, a versão baixada dessas redes será sempre de menor resolução, mais compressão e, muitas vezes, com alteração de cor e eliminação de quaisquer metadados sobre essa captura (GPS por exemplo capturado em fotografias de celular). O ideal seria preservar sempre a imagem original, antes de entrar nas redes sociais. Preservar a versão baixada das redes pode ser uma opção adicional. Isso pode não ser possível em muitas situações, e teremos que nos contentar com a imagem, nem sempre melhor, que conseguimos obter.

A resolução tem que ser vista em conjunto com a compressão da imagem. Por exemplo, um arquivo TIFF pode ter 4724 x 3153 *pixels*. Quando exibido a 300 *pixels* por polegada, terá um tamanho de 26,7 x 40 cm. Esse mesmo arquivo TIFF pode ser salvo em JPEG com o mínimo de compressão (JPEG 12), ter uma perda de informação, mas continuar com a mesma resolução em *pixels*. O problema dos arquivos *lossy* (com perdas, como no caso do JPEG) é que o arquivo sofre compressão e remoção de dados ao ser salvo, quando é compactado, e, quando aberto, é descompactado, criando novamente os mesmos *pixels*, mas a partir dos cálculos dos algoritmos, que reconstroem a imagem de forma artificial. A imagem tem o mesmo tamanho, mas a informação já não é a mesma. Em resolução de tela, como visto aqui, as diferenças são mínimas nesta imagem, mas dá para sentir a diferença nos tamanhos de arquivo.

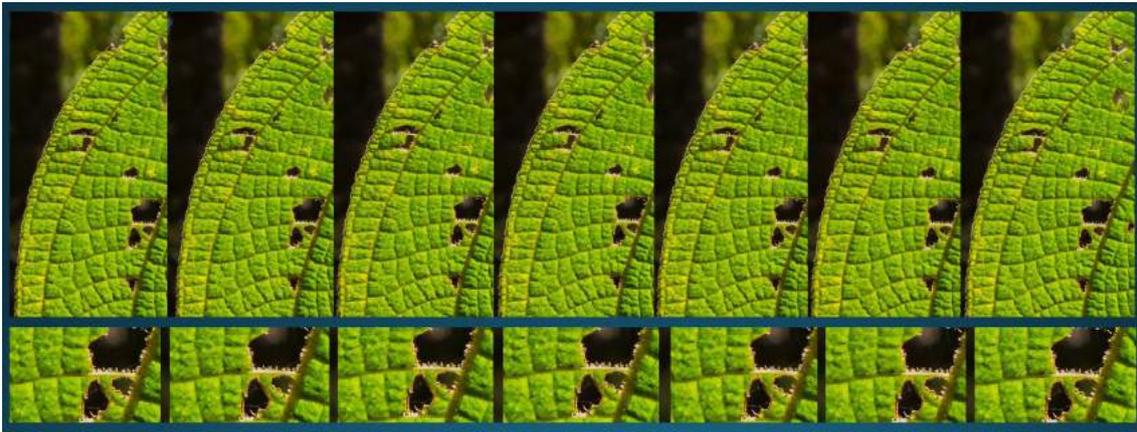


Figura 21. Arquivo RAW de 18,1 MB. Todas as imagens têm a mesma resolução em *pixels*, o que muda é a compressão.

TIFF 16 bpc	TIFF 8 bpc	JPEG 12	JPEG 8	JPEG 4	JPEG 0	JPEG 2000
81,32 MB	40,67 MB	9,37 MB	1,87 MB	795 KB	398 KB	1,4 MB

A interpolação em arquivos digitais, de que falaremos mais adiante, é quando adicionamos ou retiramos *pixels* de maneira artificial, via *software*, para aumentar ou diminuir a resolução original, ou nativa, da imagem de origem. O ideal é sempre preservar uma matriz sem os efeitos da interpolação.

Modos de cor e espaço de cor

Quando falamos de uma imagem RGB, escala de cinza ou CMYK, estamos falando em modos de cor, e isso representa quantos canais teremos para cada imagem. RGB com três canais, CMYK com 4, e escala de cinza (*grayscale*), com um canal.



Figura 22. Principais Modos de Cor. Estamos deixando de fora os modos de cor menos utilizados como LAB Color, Duotone, Mapa de *bits*, Cor indexada e Multicanais.

Slide Kodachrome de foto da minha irmã Debora produzido por meu pai, c. 1958.

O espaço de cor representa o leque de cores que um dispositivo (qualquer tipo de tela/display, impressora, gráfica etc.) é capaz de representar, e, nos casos dos arquivos digitais, quantas cores esse arquivo pode representar. Podemos comparar isso também ao tamanho do estojo de lápis de cor, e à abrangência destas cores, de pouco a muito saturadas.



Figura 23. Quantas cores no seu sistema? Fotografia de Mark McKinney.

A representação das cores é tridimensional, como o desenho de uma árvore de cores, proposto pelo artista, professor e pesquisador Alfred Henry Munsell²⁶ no início do século XX, considerando as três propriedades da cor. Valor (Luminância/*Lightness*), croma ou saturação (*chroma, saturation*) e matiz (*hue*). Assim, o tronco da árvore, ou seu eixo, representa a luminância, o que tem luz, mas não tem cor. O preto nas raízes e o topo atingindo o branco puro, em uma escala de cinza. Os galhos ou braços da árvore representam cada cor ou croma. Quanto mais perto do eixo, menos saturadas as cores, e, à medida que se afastam do centro, vão ficando cada vez mais saturadas. O matiz representa cada direcionamento. Como a árvore é um modelo em 360 graus, temos os diversos matizes, espalhados nos graus desse círculo, como vemos nas imagens a seguir.

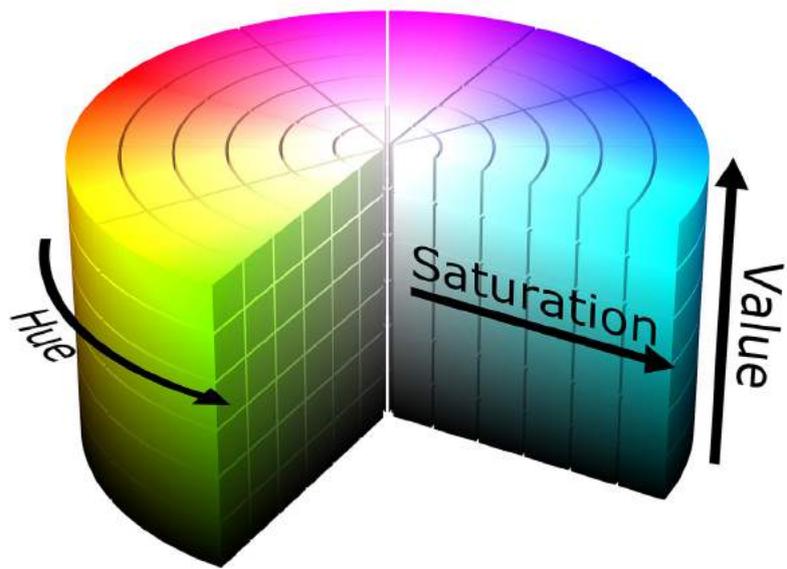


Figura 24. Nesta imagem vemos o matiz, a saturação e o valor.

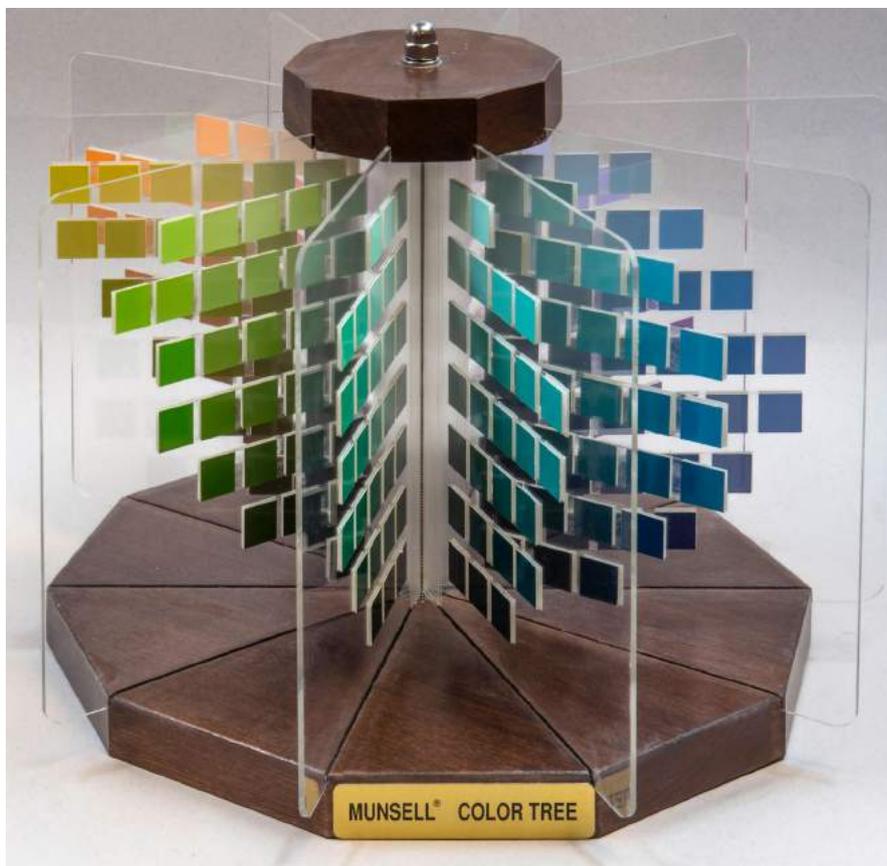


Figura 25. O modelo da árvore de cores do sistema Munsell.

Perfil de cor e gama de cores (*color gamut*)

Considerando esse modelo tridimensional, gama de cores, ou *gamut*, é a paleta de cores que uma determinada tecnologia ou processo é capaz de reproduzir, e representa todo o interior desse espaço de cor. O perfil representa o desenho dos limites das bordas desse espaço. Cada dispositivo terá uma resposta diferente ao procurar representar as cores.

Expressões como “fora do *gamut*” significam que podemos ter uma imagem de uma ave maravilhosa com cores incríveis, mas que, na hora de fazer um livro, com conversão da imagem de RGB para CMYK, muitas cores estarão fora do *gamut* (espaço de cor e seu perfil) da gráfica. A combinação de tinta e papel não consegue reproduzir a saturação e o brilho dessa cor que está no arquivo original digital. Nesse sentido, será necessário resolver esse impasse ao mapear a cor para a cor mais próxima que o sistema consegue reproduzir.



Figura 26. Crejoá (*Cotinga maculata*), Porto Seguro, Bahia. Fotografia de Ciro Albano. 1º lugar, melhor fotografia, Avistar 2010.

Esta linda imagem, gentilmente cedida pelo Ciro Albano para esse texto, é um exemplo claro desse problema. A captura foi processada para um arquivo RGB com espaço de cor Adobe RGB (1998), mas a impressão em catálogo será em CMYK, e no espaço de cor da gráfica de FOGRA 39. Na imagem a seguir, vemos opções no Photoshop para visualizar uma previsão das alterações que devem ser esperadas nesse processo de conversão de RGB para CMYK.

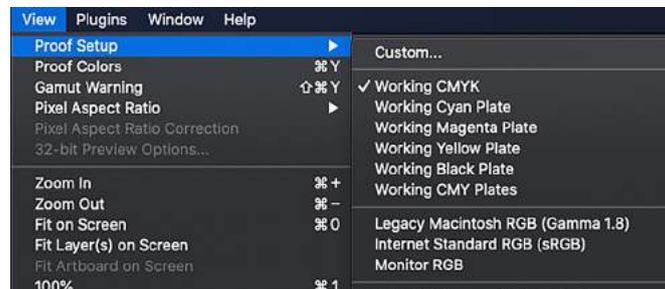


Figura 27. No Adobe Photoshop temos o menu *View* (Visualizar), com as opções de Configurações de Prova (*Proof Setup*), Prova de Cores (*Proof Color*) e Aviso de Gamut (*Gamut Warning*).



Figura 28. Imagem com cores realçadas em vermelho pela função Aviso de Gamut do *software* Adobe Photoshop.

Ao selecionarmos Aviso de Gamut, conseguimos ver na cor de destaque (aqui sendo vermelho, cor que podemos determinar nas configurações de preferências do Photoshop) todas as cores que estão fora do *gamut*, ou seja, não existem no espaço de cor de destino, determinado pela opção de Configurações de Prova – na imagem a seguir, CMYK de trabalho (*Working CMYK*). Nesse momento, o *software* entende que a imagem de origem é RGB com Adobe RGB (1998), e o destino é um processo que utiliza CMYK de trabalho.

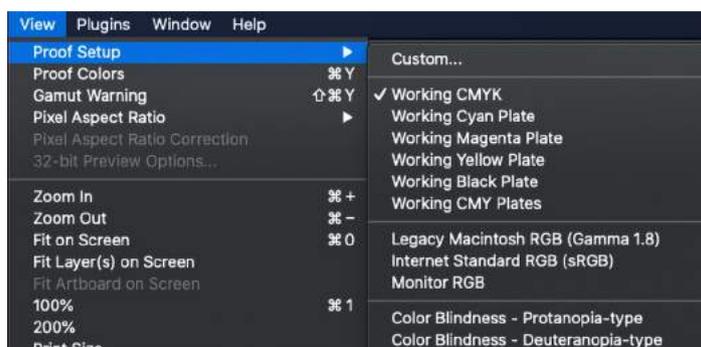


Figura 29. Em Configurações de Prova, posso selecionar *Custom* (Personalizado) para escolher o perfil de cor correto que será utilizado pela gráfica onde o trabalho será impresso.

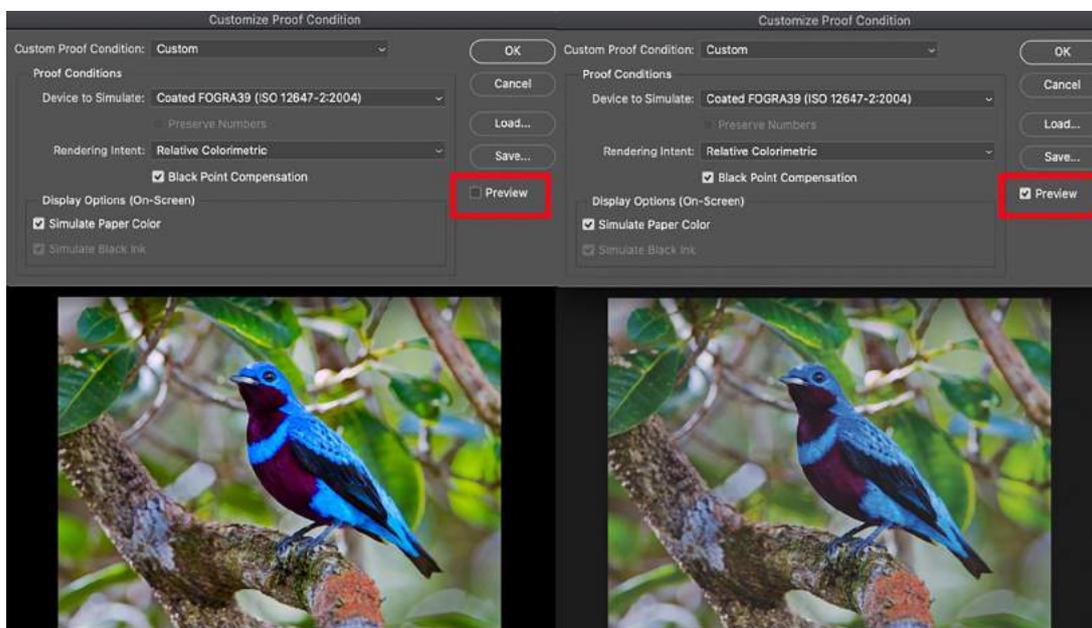


Figura 30. Comparativo de imagens com caixa de seleção *Preview* marcada e desmarcada no *software* Adobe Photoshop para exibir as cores fora do *gamut*.

A imagem da esquerda não tem a caixa de seleção *Preview* (pré-visualização) selecionada, portanto, continua como sendo a imagem RGB com perfil Adobe RGB (1998). A imagem da direita tem esta caixa selecionada. Nesse caso estamos tendo uma visualização prévia pelo *software* de como a imagem aparecerá no dispositivo que estamos simulando. Nesse caso, é uma imagem CMYK com perfil FOGRA 39 Revestido, com finalidade de renderização para Colorimétrico relativo, compensação de ponto de preto, simulação da cor do papel e da tinta preta. Neste exemplo, fica mais fácil entender que as cores vibrantes do Azul/Ciano da ave, estão fora do *gamut*, e que o sistema de impressão não consegue reproduzir a saturação e o brilho destas cores. Quando uma cor está fora do *gamut* para o seu destino, ela será mapeada para a cor mais próxima dentro de espaço de cor. O risco, como já discutimos, é de preservar a imagem final do livro, com menos cores saturadas, e perder o original RGB. Portanto, na preservação digital desse trabalho, será vital pensar nos versionamentos gerados e qual imagem é mais próxima da nossa matriz, com suas derivadas.

O mapeamento pode ser feito através de dois processos principais, conhecidos como “finalidade de renderização” (*rendering intent*), que podem ser o colorimétrico relativo ou perceptivo. Pesquise “Gerenciamento de Cores para Imagens Digitais”²⁷ na dissertação de mestrado de Alexandre Cruz Leão, no site da Adobe²⁸, e na internet para entender melhor as questões de cor em imagens digitais.

Entender o histórico da palavra *gamut* ajuda a compreender o conceito. De acordo com o dicionário Merriam-Webster²⁹:

Para obter informações detalhadas sobre a gama, temos que mergulhar no fundo de uma escala musical desenvolvida pelo músico e monge do século 11, Guido de Arezzo. Guido chamou a primeira linha de sua pauta de baixo de gamma e a primeira nota em sua escala de ut, o que significava que gama ut era o termo para uma nota escrita na primeira linha da pauta. Com o tempo, gamma ut sofreu um encurtamento para gamut, mas foi ampliado o seu significado. Ele se expandiu para cobrir todas as notas da escala de Guido, depois todas as notas na extensão de um instrumento e, eventualmente, toda uma extensão de qualquer tipo.

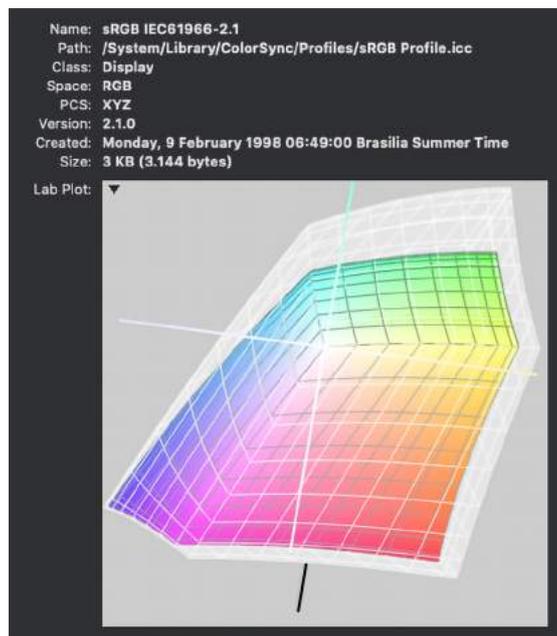


Figura 31. Nesta imagem podemos comparar o espaço de cor, e o seu perfil, do Adobe RGB (1998), em cinza-claro, com o espaço de cor e seu perfil do espaço de cor sRGB, que é menor.

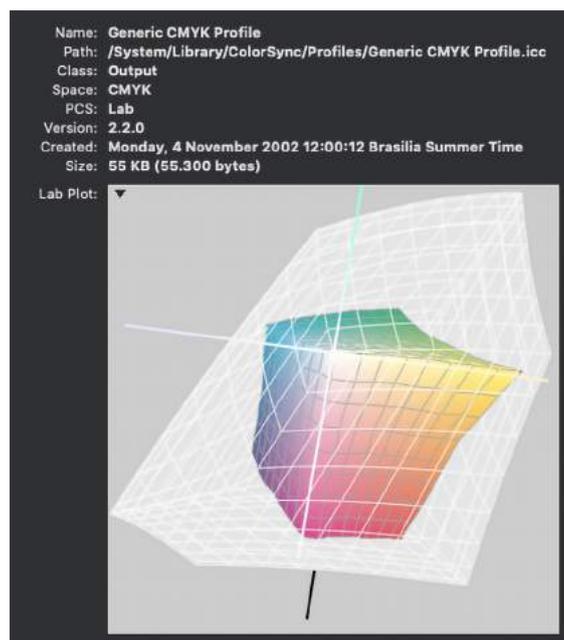


Figura 32. A situação fica ainda pior quando o espaço de cor Adobe RGB (1998) é comparado com o espaço de cor e perfil genérico CMYK. Nesse exemplo, fica claro como temos cores menos saturadas em CMYK. Todas as cores no espaço acinzentado, Adobe RGB (1998), estão fora do *gamut* do espaço CMYK genérico. Na tentativa de representar todas estas cores, a conversão de RGB para CMYK terá que utilizar uma das intenções ou finalidades de renderização.

A importância de entender tudo isso volta para o que queremos preservar em relação ao nosso entendimento de todas as possíveis derivadas.

Em arquivos não processados, ou RAW de câmera, não temos que pensar ainda sobre isso porque estas imagens não estão definidas ainda e existem como informações matemáticas de luminosidade dos fotodiodos que captam a luz, associadas às informações do filtro de Bayer, o mosaico RGB que fica acima desses fotodiodos em um sensor digital. Os sensores de imagem da câmera digital são chips de silício que possuem várias áreas fotossensíveis (*photosites* em inglês) construídas com fotodiodos e organizadas em matrizes dentro das estruturas de chip CCD ou CMOS.

Portanto, uma das vantagens de um arquivo RAW é que o seu processamento pode ser feito a qualquer momento posterior para gerar um arquivo processado, quando teremos que escolher o modo de cor, espaço de cor, e *bits* por canal para a geração dessa imagem renderizada.

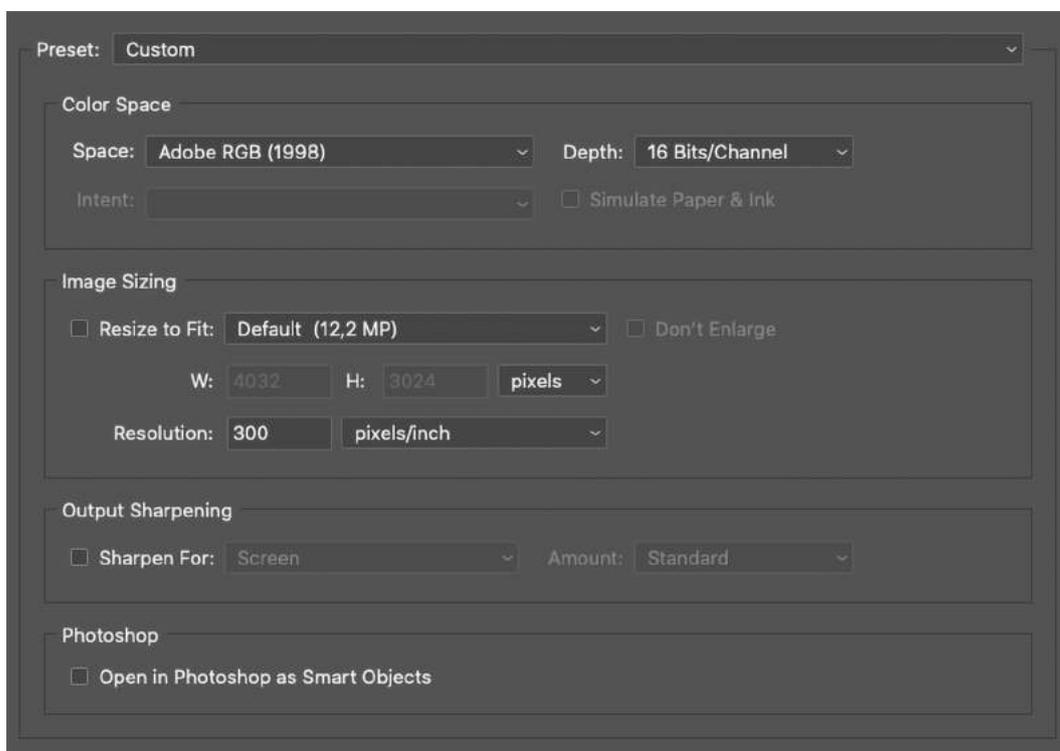


Figura 33. Opções na hora de abrir um arquivo RAW. No *Adobe Camera Raw* podemos escolher o espaço de cor, profundidade de *bits*, tamanho da imagem, aguçamento de saída, e a possibilidade de abrir no Photoshop como *Smart Objects*.

Em imagens já renderizadas, como as de um celular, não configurado para captura em RAW, a imagem pode já ser definida na sua produção como modo de cor RGB e espaço de cor sRGB. Esta será a imagem original desse dispositivo. Ao digitalizar com escâner, também temos que escolher o modo de cor, o espaço de cor e a quantidade de *bits* por canal da imagem renderizada produzida.

O que importa para a preservação da imagem digital é, novamente, como discutimos em resolução, preservar a imagem que tem mais potencial de registrar o máximo de cores da captura original feita pelo dispositivo. Imagens em modo de cor CMYK (ciano, magenta, amarelo e preto) são usadas em publicações gráficas de tinta sobre papel. No processo de conversão de uma imagem RGB para CMYK, ganhamos um canal de cor a mais (arquivo fica maior), mas o espaço de cor acaba sendo bem menor, e muitas cores são eliminadas ao serem mapeadas para a cor mais próxima que cabe no perfil desse espaço de cor. Assim, o modo de cor RGB, com espaço de cor Profoto ou Adobe RGB 1998, é o que preserva a maior quantidade de informações de cores da captura original da maior parte dos dispositivos. As derivadas para internet, em RGB e espaço de cor sRGB, já possuem uma cartela de cores mais reduzida, e ainda mais reduzida quando convertemos para CMYK e impressão de gráfica.

Assim, sempre que possível, queremos preservar a imagem que retém o máximo de informações de cores, que pode ser preservada também com suas derivadas para publicações e internet. Um conjunto composto das imagens original e derivada pode ficar assim: RAW/DNG da câmera, TIFF RGB, Adobe RGB 1998, 16 bpc, tratada com ajustes que refletem a intenção do autor, TIFF CMYK, 8 bpc, com perfil/espaço de cor da gráfica, utilizado para impressões de publicações, e um JPEG, sRGB com 8 bpc. Cada um desses arquivos, da mesma imagem original, tem sua função e especificidade.

Artefatos digitais

Compressão da imagem

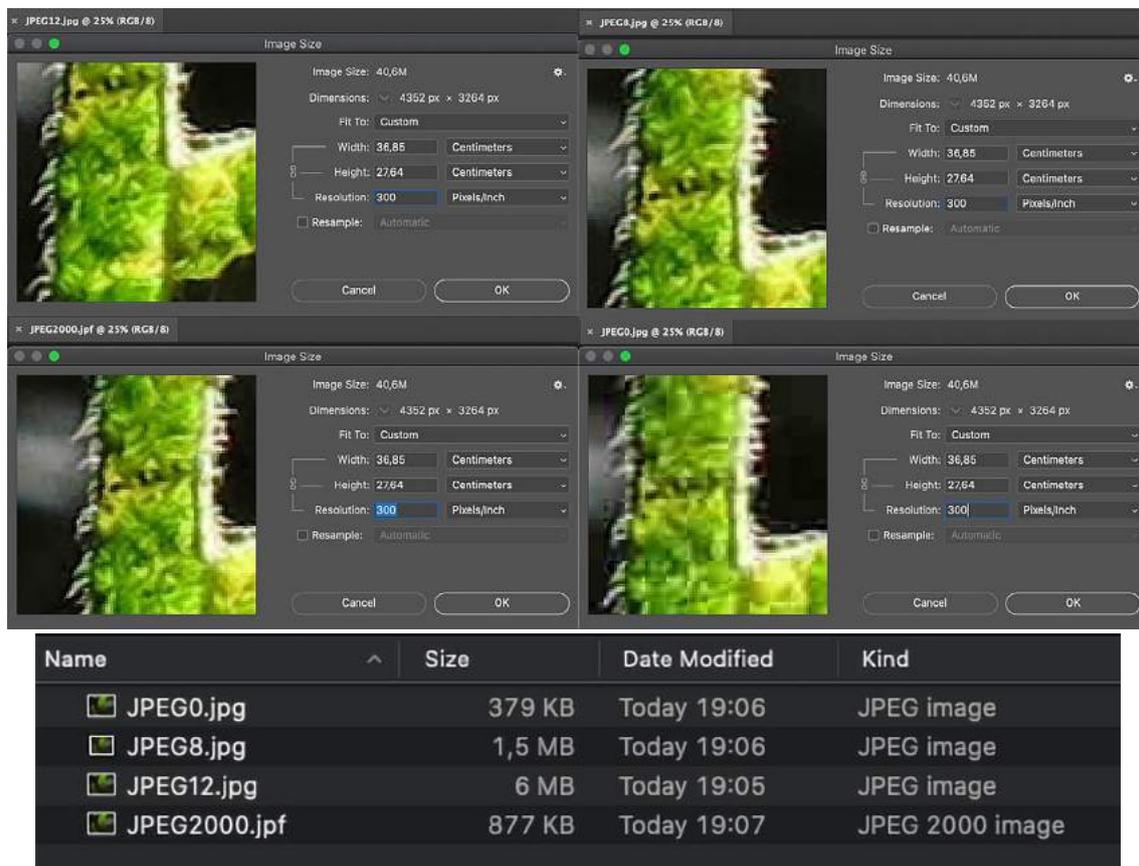
Arquivos matrizes não devem ter qualquer tipo de compressão quando estamos considerando preservá-los para a longevidade. Descomprimir esses arquivos no futuro próximo pode ser um grande desafio. Podemos não ter acesso aos aplicativos de *software* que foram utilizados para fazer a compressão, não encontrar novos sistemas compatíveis para realizar esse trabalho e não ter acesso aos arquivos comprimidos.

A compressão da imagem deve ser feita somente quando são geradas versões de menor resolução utilizadas para difusão e internet. Os melhores formatos de arquivo para isso são o JPEG e JPEG 2000. A compressão pode ser realizada em diversos níveis. Quanto maior a compressão, menor o tamanho do arquivo e pior a qualidade da imagem. Terá que existir um equilíbrio entre tamanho de arquivo e qualidade de imagem que deve ser alcançado de acordo com os objetivos e possibilidades de cada projeto.

Temos compressão com perdas (*lossy*) e sem perdas (*lossless*). A compressão sem perdas, por exemplo, pode ser feita em uma pasta contendo imagens para um formato comprimido, como ZIP (.zip), para envio por e-mail. O destinatário fará a descompressão na outra ponta e não haverá perda na qualidade ou tamanho da imagem. Uma imagem TIFF salva com compressão LZW pode ser depois descomprimida sem esta compressão, e sem perdas. A compressão com perdas é quando reduzimos o tamanho da imagem ao reduzirmos a quantidade de informação contida nela. Esse processo é irreversível – com perdas. O que é eliminado, não poderá ser mais recuperado. O grande ganho é na redução do tamanho do arquivo digital.

Algumas compressões com perda podem providenciar um grande ganho na redução do tamanho do arquivo digital, e com um mínimo de perda na qualidade da imagem. Esse equilíbrio faz sempre parte de um processo de decisão sobre o desejo de guarda e preservação desse conteúdo digital.

O vídeo digital é um ótimo exemplo disto.



Figuras 34 e 35. Nas referências acima, temos as quatro imagens abertas, com a função *Image Size* (tamanho da imagem) no Photoshop. Todas têm exatamente o mesmo tamanho com 300 ppi de resolução. A compressão é que muda, e isso pode ser visto facilmente pelo tamanho da imagem. Como é uma compressão com perdas, o que é removido é irreversível. O aumento é de 300% para realçar a estrutura da imagem. A qualidade da imagem no JPEG12, JPEG8 e JPEG2000 é praticamente idêntica. A qualidade cai tremendamente na versão quadriculada do JPEG0.

Interpolação

Em Matemática, denomina-se interpolação o método que permite construir um novo conjunto de dados a partir de um conjunto discreto de dados pontuais previamente conhecidos. Desta forma, em uma imagem digital, interpolar significa adicionar ou subtrair *pixels* (ou dados pontuais) com base em um conjunto de dados pontuais previamente conhecidos. Podemos também pensar no conceito de reinterpretar. Por exemplo, ao girarmos uma imagem em 90 graus, os *pixels* são simplesmente girados nesse ângulo. Porém, ao girar uma imagem em 3 ou 15 graus,

uma pequena reinterpretação terá que ser feita. No exemplo a seguir, uma linha reta girada em 90 e 15 graus. A resolução não foi alterada, mas a imagem foi interpolada no giro de 15 graus.

Em uma linha reta, as bordas são precisas. A mesma linha, ao ser girada em 45 graus, apresenta o efeito de *aliasing* (imagem à esquerda) ou “escadinha”, porque não é possível girar os *pixels*. O efeito do *anti-aliasing* é uma tentativa de construir uma linha contínua com os *pixels*. Esse é um exemplo da interpolação, ou dos cálculos matemáticos feitos na imagem.

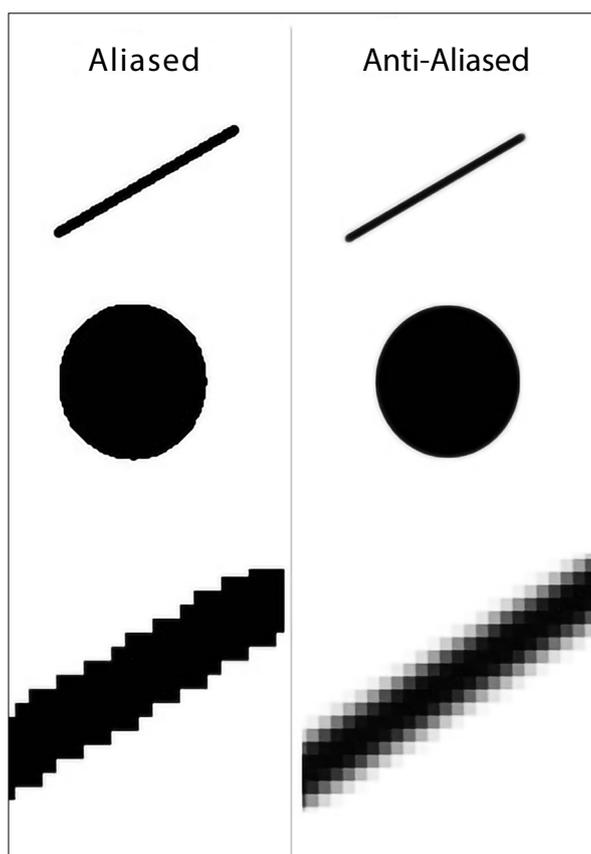


Figura 36. *Aliasing* à esquerda e *anti-aliasing* à direita.

CC BY-SA 3.0 atribuído a MD.

A interpolação também acontece com a captura dos sensores de câmeras digitais que utilizam um filtro Bayer, e em escâneres com sensor CCD, onde as informações de uma linha são lidas por 3 CCDs, com seus respectivos filtros RGB, e depois interpoladas para produzir uma linha em RGB.

Outro conceito principal, que também discutimos na interpolação, é a utilização de um processo para adicionar ou subtrair *pixels* de uma imagem. Quando temos uma imagem com um determinado tamanho e queremos aumentá-la, o *software* analisa os *pixels* e seus arredores e cria novos *pixels*, baseados nos algoritmos selecionados ou especificados pelo *software*. Isso poderá criar uma imagem com mais *pixels* por polegada, mas não significa um real ganho de resolução. Ao capturarmos uma imagem digital que gera uma imagem final no tamanho de 30 x 45 cm com resolução de 300 ppi, é importante frisar que estas dimensões são obtidas na captura nativa do equipamento, e não com artifícios e algoritmos do *software* por meio da interpolação. Gerar uma imagem com captura nativa em 30 x 45 cm com resolução de 300 ppi é significativamente diferente de uma imagem de captura nativa de 12 x 18 cm com 300 ppi, que depois é interpolada para um tamanho de 30 x 45 cm com 300 ppi.



Figura 37. Imagem à esquerda é originalmente de 30x45 cm com 300 ppi. A imagem da direita é uma imagem 12x18 cm, com 300 ppi, interpolada para 30x45 cm, com 300 ppi. A diferença de qualidade é evidente. O processo de interpolação e criação artificial, via algoritmos, de *pixels* é usado nessa imagem para criar a imagem da direita nesta resolução e tamanho.

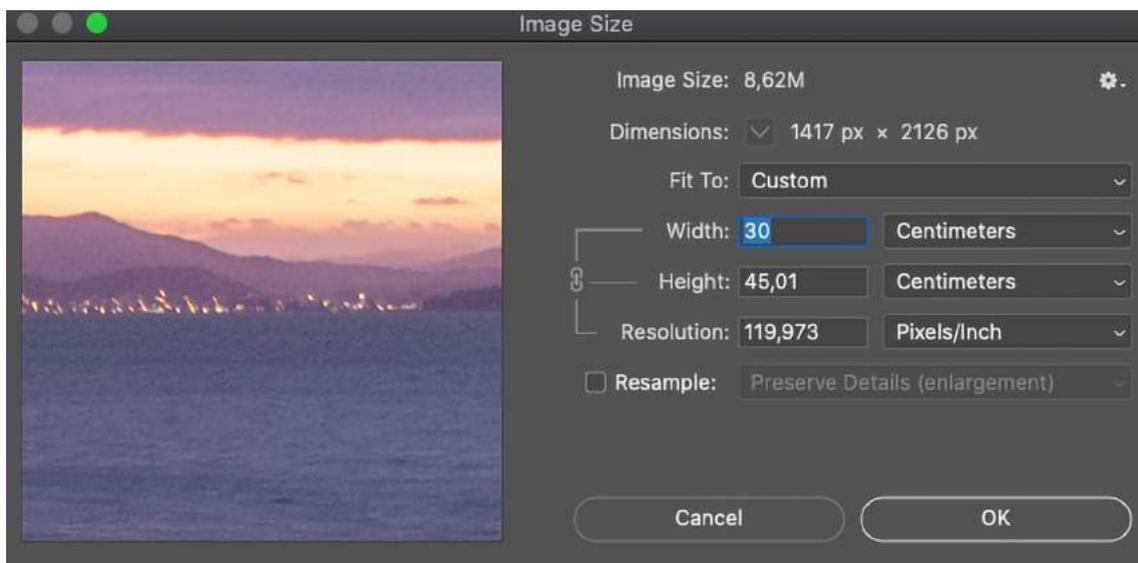


Figura 38. A imagem de 12x18 cm, ao ser ampliada para 30x45 cm, terá 119 *pixels* por polegada. A interpolação, conforme podemos verificar na fotografia a seguir, ao selecionar a opção de *Resample* (reamostragem), consiste em adicionar em torno de 180 *pixels* por polegada para atingir 300 ppi. O tamanho da imagem é alterado de 8,62 MB para 53,9 MB nesse processo.

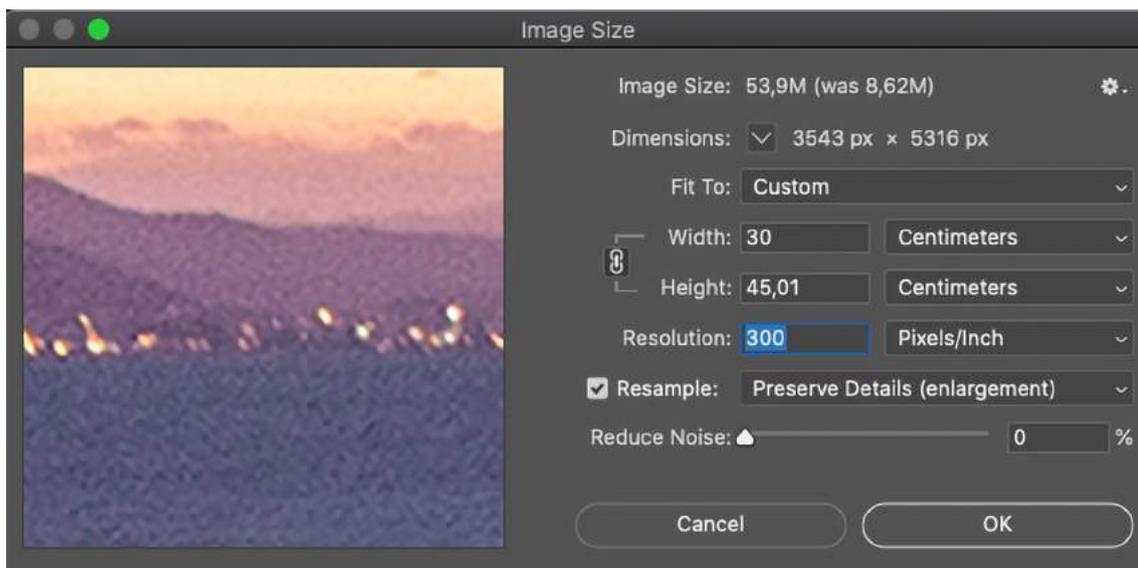


Figura 39. Reprodução de tela do *software* Adobe Photoshop com a função interpolação que permite reduzir ou aumentar o número de *pixels* de uma imagem.

A interpolação é importante e corriqueira na geração de imagens menores, reduzindo nesse caso o número de *pixels* para produzir versões de baixa resolução para uso na internet e e-mail. Adicionar *pixels* via interpolação para tornar a imagem

maior é uma prática usual e pode ser feita dentro de alguns limites razoáveis, que devem sempre ser avaliados para cada projeto e uso, mas não deverá ser parte integrante da prática de captura em um projeto de digitalização. Por exemplo: uma instituição escolheu um método de digitalização que atende a todos os seus usos, mas a imagem tem um tamanho de 30 x 45 a 300 ppi. Com interpolação, seria viável a produção de cartazes com a imagem em 40 x 60 cm a 300 ppi (aumento de 133%), ou, esticando os limites, até 50 x 75 cm a 300 ppi (aumento de 166%). Certamente um sistema com captura nativa de 50 x 75 cm a 300 ppi terá muito mais qualidade, mas também significa que esse sistema de mais resolução gerará arquivos quase três vezes maiores. É importante saber que, caso necessário, pequenos aumentos nas imagens são viáveis por meio de interpolação, para fins específicos e dentro de um padrão de qualidade conforme mencionado anteriormente. A interpolação não é um recurso legítimo para transformar imagens digitalizadas com baixa resolução em imagens de alta resolução. Nas especificações de projetos de digitalização, ou captura digital com câmera, precisa ficar claro qual o papel que terá a interpolação.

Aumento de nitidez/Sharpening

Imagens digitais necessitam de um aumento de nitidez (*sharpening*), ou aguçamento, para contrapor os vários efeitos que ocorrem no nível do sensor de captura – tanto em escâneres como em câmeras digitais. Temos os efeitos da interpolação na conversão das informações analógicas para digitais na geração de uma imagem em RGB e no uso dos filtros *anti-aliasing*, que acabam suavizando a imagem (*Blur filter*) para evitar os efeitos de *moiré* e de *aliasing*. Esse aumento de nitidez pode ser chamado de entrada ou captura, e deve ser conservador o suficiente para repor a resolução original da imagem. Tudo isso é muito subjetivo e visual, e variará também de acordo com o equipamento. Não existem fórmulas prontas, e a pessoa, realizando isso, irá se basear em testes e experiência. Em geral, menos é mais.

Para projetos de digitalização, bem como em arquivos nato-digitais de câmeras, onde a interferência no original deve ser minimizada, o aumento de nitidez deve ser pensado em duas etapas: o *sharpening* de captura ou entrada e o *sharpening* de saída. Fazemos pequenos ajustes na entrada. De acordo com a saída – tipo de material em que será impresso ou projetado e tamanho, faremos um ajuste de nitidez mais específico. Em cada etapa, um pequeno aumento de nitidez é realizado. O maior

problema do aumento de nitidez é ultrapassar o ponto onde a imagem começa a apresentar artefatos, distorções, ruídos e exageros nas diferenças das linhas. Muita comparação, discussão e experiência ajudarão a definir esses critérios. Novamente vivemos a decisão quanto ao que guardar quando pensamos nestas versões. A imagem grande e ajustada com nitidez para um determinado fim pode não ser preservada, desde que a imagem original o seja. Esta imagem original pode gerar novamente esta versão grande. Uma versão grande, com muito ajuste de nitidez não poderá fazer o caminho inverso e produzir a imagem original.

O *sharpening* de captura pode ser realizado diretamente nos arquivos RAW/DNG no caso de câmeras digitais. Em escâneres, isso pode ser realizado via o aplicativo de *software* do escâner, ou mais bem realizado e com mais controles no Photoshop por meio de ações e filtros de aumento de nitidez. O *sharpening* de saída deve ser pensado em relação ao tamanho da saída e o tipo de saída – telas, impressão em papel fotográfico, impressão em livros e revistas e impressões em jato de tinta, entre outras possibilidades.



Figura 40. Exemplo de imagem com ajuste de nitidez no original. Fotografia de Beatriz Franco.



Figura 41. Na versão inferior, a imagem original sem aumento de nitidez. A imagem de cima tem a aplicação do filtro *High Pass* via Photoshop, um dos métodos de *sharpening*. Experiência de trabalho e avaliação dos resultados em tela e impressos ajudam a estabelecer os melhores métodos e intensidade desse tratamento para não criar artefatos ou artificialidades no processo.

No acervo, as matrizes de alta resolução resultantes do processo de digitalização devem ter somente o aumento de nitidez de captura aplicado, e de forma sutil. Quaisquer alterações posteriores, como aumento de nitidez para saída, correções adicionais de cor e contraste, devem ser feitas sob demanda em arquivos duplicados e para esses fins específicos.

O processamento digital de uma imagem pode gerar derivadas de diversos modos de cor, tamanho e resolução, formato de arquivo, profundidade de *bits*, perfil de cor, compressão, entre outros. Conhecer todas estas opções, entender quando são e quando não devem ser utilizadas, nos ajuda a compreender a melhor imagem que queremos ter como matriz de preservação, e quais imagens adicionais preservaremos. Sempre que possível, devemos escolher a imagem mais próxima do original, como um negativo digital, e a mais próxima de seu uso, como imagens finais utilizadas em livros e em páginas na internet.

4

Como devo organizar e descrever meus materiais passivos e atuais

No capítulo 2, foi abordada a necessidade de entender onde estão todas as fotografias digitais. Neste capítulo, faremos uma abordagem de algumas possibilidades de processamento para organizar e descrever os nossos materiais. Não existe uma fórmula única de como fazer isso. Variáveis como o tamanho da coleção, os recursos disponíveis, as formas de trabalhar, os sistemas já implantados e funcionando, determinarão, para cada um, a melhor maneira de fazer isso.

Como discutido no capítulo 2, a prioridade deve sempre ser a de organizar os novos sistemas de trabalho para implantar esses fluxos e garantir que toda a produção atual esteja inserida e funcionando a partir deles. No espírito de preparo para os materiais atuais, será fundamental para estruturar os novos procedimentos olhar para o passivo, refletir sobre o que já foi produzido e sobre como podemos organizá-lo, e compreender todas as etapas necessárias para isso, que serão discutidas neste capítulo.

Portanto, o fluxo proposto neste texto é olhar para o passivo, avaliar e planejar como todos os materiais poderão ser organizados, entender o tamanho em *giga*,

tera ou *petabytes* da nova matriz necessária para os materiais passivos, levando em conta também a previsão da produção atual para os próximos 5–7 anos. O planejamento de nomenclatura, organização de pastas e inserção de metadados será também aplicado para os materiais atuais. Assim, com um plano consolidado, partiremos para organizar a produção atual do dia a dia, e, após os ajustes resultantes do confronto entre nossas teorias e nossas práticas, com tudo funcionando de forma tranquila para os materiais atuais, é o momento de aplicar esses processos em todos os materiais passivos.

A seguir, são realizadas as principais atividades para o planejamento dos materiais passivos e atuais:

Organização de dados passivos

- a) Formas de copiar dados de um lugar para outro a fim de criar matriz de todos os seus dados.
- b) Usar *softwares* para encontrar duplicatas em sua nova matriz.
- c) Usar *softwares* para criar uma lista de formatos de arquivo na busca de materiais que não devem ser preservados, ou que precisam ser migrados para novos formatos.
- d) Organizar estrutura nova de pastas.
- e) Renomear os arquivos conforme necessidade.
- f) Inserção de metadados para descrever os arquivos.

Organização de dados atuais

- g) Construção de sistemas de *backup* e fluxos de trabalho.

Sobre Nobreaks e Antivírus

Aqui temos que abrir uma brecha para inserir dois assuntos de forma sucinta. Esses dois fatores são fundamentais para o gerenciamento de dados digitais e muitas vezes não são discutidos.

Vamos começar pela importância do uso de *nobreaks* (do inglês, sem quebras/interrupções). O *nobreak* é um dispositivo com uma ou mais baterias, que

precisam ser renovadas a cada 1 ou 2 anos, com sensor de queda de energia ou voltagem. No caso de picos ou falha de energia, o *nobreak* mantém o fornecimento contínuo de energia ao sistema, protegendo-o contra falhas. Os sistemas mais econômicos mantêm tudo ligado por pouco tempo, mas é o tempo suficiente para que tudo seja desligado de forma adequada. Esses picos de energia ou falhas são, em muitos casos, responsáveis por danos aos computadores e discos HDD e SSD. É um investimento necessário e contínuo, pela reposição das baterias, para a proteção de dados, e precisa estar dentro de um projeto de preservação de dados digitais.

O uso de um antivírus também é necessário. Na definição da Wikipedia³⁰:

Em informática, um vírus de computador é um *software* malicioso que é desenvolvido por programadores geralmente inescrupulosos. Tal como um vírus biológico, o programa de infecção do sistema faz cópias de si e tenta espalhar para outros computadores e dispositivos de informática.

A maioria das contaminações ocorre por ação do usuário. Um exemplo muito comum se dá por meio para baixar arquivos infectados que são recebidos em anexos de e-mails. A contaminação também pode ocorrer de outras formas: acessando sites de procedência duvidosa ou ainda por meio de arquivos infectados em pendrives, CDs, DVDs ou qualquer outro tipo de dispositivo de armazenamento de dados.

Os vírus de computador têm nomes diversos e são uma forma de *malware* – um termo abrangente, que se refere a vários tipos de “*software* malicioso”, com muitas nomenclaturas que precisam ser pesquisadas, tais como *spyware*, *ransomware*, *worms*, cavalos de Troia, *phishing*, *bots*, *botnet*, *DoS* (*Denial of Service attacks*/ Ataques de negação de serviço e *DDoS* quando é distribuído), entre outros.

Os crimes cibernéticos estão se tornando mais frequentes e lucrativos, com um grande aumento no período da pandemia de covid-19. A única forma de proteção é o uso de *software* antivírus. Existem opções gratuitas e pagas que precisam ser investigadas e analisadas – mas quanto mais robusto for seu sistema, mais importante será profissionalizar esta proteção.

Nota importante: ao incorporar dados digitais provenientes de um dispositivo externo como HDDs, SSDs, discos ópticos etc. para seu sistema, sempre faça uma verificação de antivírus nesses dispositivos antes, bem como utilize um programa de cópiagem com verificação dos dados copiados via *checksum*.

Organização de dados passivos

Formas de copiar dados de um lugar para outro no intuito de criar matriz de todos os seus dados

Talvez muitos de vocês já tenham feito o que eu sempre fazia, que é copiar/colar ou arrastar uma pasta de um lado para o outro. Muitas vezes esses processos eram interrompidos por alguma falha sem termos certeza do que foi copiado e do que não foi. Quando isso acontecia, eu me lembro de ficar um bom tempo tentando entender o que foi copiado e comparando listagens enormes de arquivos para ver o que faltou na tentativa de correção manual. Outras vezes era melhor descartar a cópia feita pela metade e fazê-la novamente. Mesmo copiando sem erros, não podemos garantir que todos os arquivos foram copiados, e de que nenhum arquivo foi copiado com algum erro ou corrupção. Ter erros nesse processo não é comum, mas eles acontecem. Sem um sistema de verificação do que foi copiado, não podemos ter esta certeza. Portanto, temos que sair de processos empíricos e manuais para utilizar ferramentas que nos auxiliem.

O processo que queremos deve nos garantir que o arquivo copiado é idêntico ao arquivo de origem, e que, se houver falhas, poderemos ter, no final do processo, uma lista dos arquivos em que houve problemas. Para isso, precisamos lançar mão de alguns *softwares* que trabalham com estas funcionalidades.

Pesquisa de softwares

Primeiro, um apontamento fundamental sobre o uso de *softwares*. Os *softwares* que serão discutidos neste texto são meros exemplos, dentre muitos outros que não serão abordados, de aplicativos que realizam estas tarefas, e têm a função de mostrar como podemos melhorar os nossos processos com ferramentas de *software*. Contudo, não representam os melhores ou únicos *softwares* para estas tarefas. É necessário pesquisar muito e testar cada um dos *softwares* para entender o que fazem e deixam de fazer. Como novos aplicativos são criados e outros mais antigos deixam de existir, também devemos estar atentos para outras possibilidades de ferramentas. Muitas ferramentas são feitas somente para a plataforma Windows,

outras existem também em macOS e Linux. Não deixe de compartilhar as ferramentas e aplicativos que funcionam para você e as que você já testou e não recomendaria. De forma coletiva, todos nós sairemos ganhando com estas informações.

Linha de comando/ Command Line

Para usuários mais avançados, muitas tarefas podem ser realizadas por linhas de comando disponíveis em Windows, macOS e Linux. Isso envolve uma capacitação em programação fora da esfera de discussão deste texto. Cada vez mais, as instituições de memória cultural, como museus, bibliotecas e arquivos estão se instrumentalizando na esfera da tecnologia e programação. Trazer pessoas com esta *expertise* para trabalhar nos projetos envolvendo a preservação e difusão de dados digitais representa um grande ganho para as instituições de cultura. Neste texto, estamos um pouco mais limitados a trabalhar com aplicativos que possuem uma interface gráfica do usuário (*GUI – Graphical User Interface*), que é a interface visual (*software*) amigável para o usuário, em que realizamos todos os comandos que, por sua vez, executam linhas de programação por debaixo dessa interface.

Checksum

Para o que está sendo discutido neste texto, queremos ferramentas que possam controlar o processo, mas também verificar se os arquivos copiados são idênticos aos de origem. Isso é feito por um *checksum* (soma de verificação), que é o resultado matemático da execução de um algoritmo. Comparar a soma de verificação gerada do arquivo de origem e gerada no mesmo arquivo copiado para outro destino garante que sua cópia seja idêntica e livre de erros. Uma soma de verificação também é às vezes chamada de soma *hash* ou código *hash*.

TeraCopy

Essa ferramenta possui uma versão gratuita para uso pessoal e uma versão Pro que agrega mais algumas funções. TeraCopy está disponível para macOS e Windows. Ele não é um *software* de *backup*, mas um que você pode usar quando

precisar copiar dados de um dispositivo para outro e quiser gerar somas de verificação para ter certeza de que o processo foi 100% à prova de falhas. Use um *software* como esse para copiar uma pasta, ou pastas, de um local para outro.

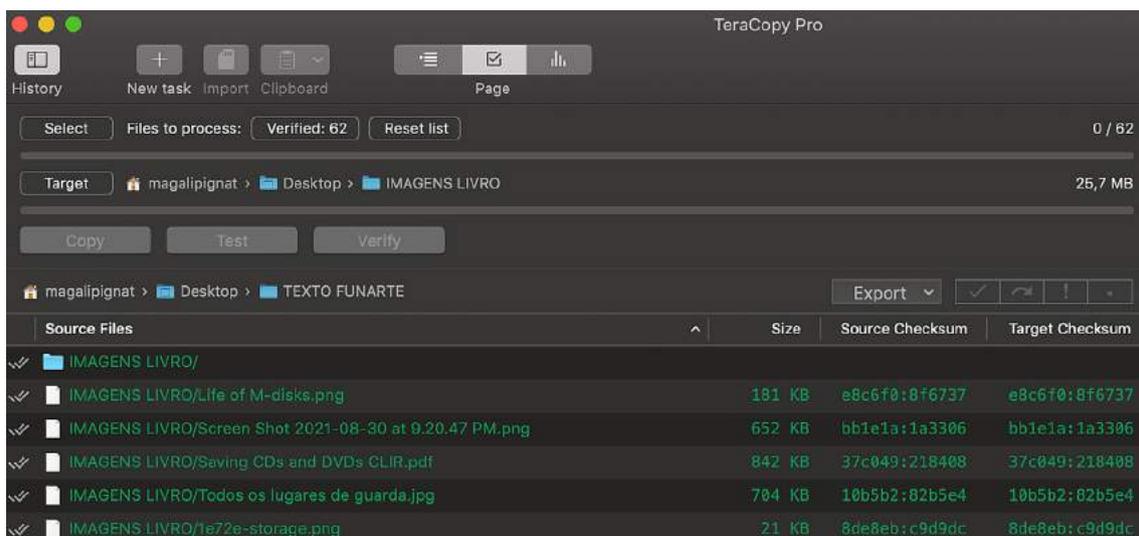


Figura 42. Nesta captura de tela, vemos a pasta *Select* para os arquivos de origem, e *Target*, para os arquivos de destino. Ao selecionar *Copy*, já com a função de *Checksum* selecionada, podemos ver os códigos *hash* de *Source* (origem) e *Target* (destino). Se a soma de verificação falhar, o *software* exibirá o texto no *Target Checksum* em vermelho.

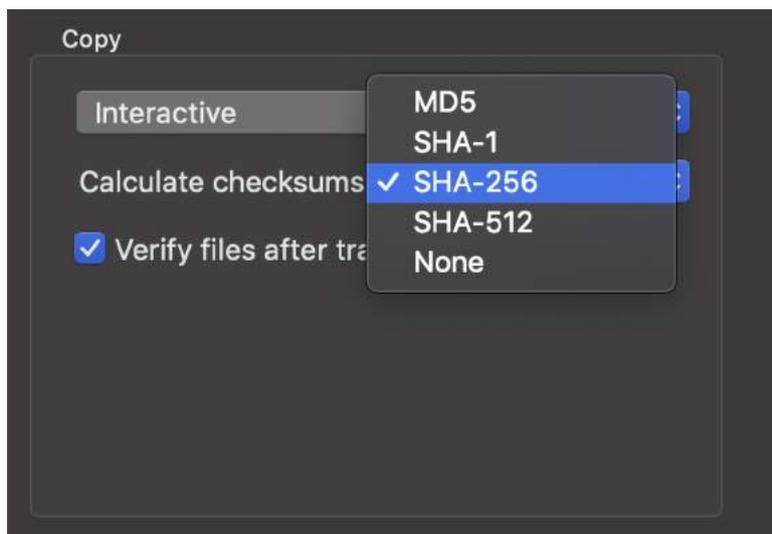


Figura 43. Nesta opção do *software*, definimos qual a soma de verificação que queremos calcular para os arquivos de origem, e selecionamos também para verificar os arquivos de destino após a transferência. Nesse caso, foi utilizado o SHA-256, um *hash* mais complexo em vez de MD5.

Esse *software* tem uma opção bem interessante: a possibilidade de exportar os dados em uma planilha CSV, que pode ser aberta em Excel.

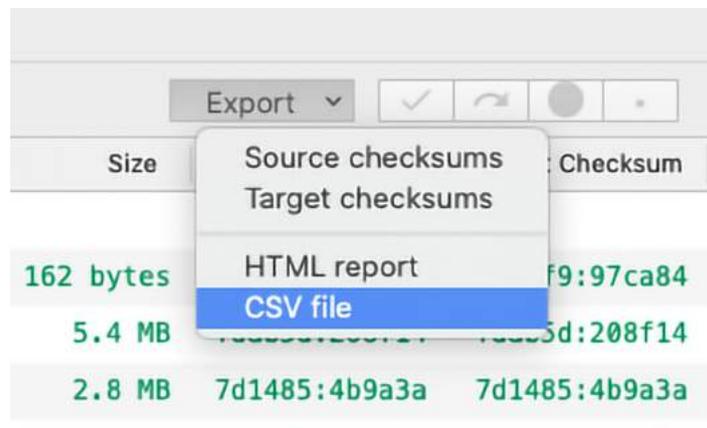


Figura 44. Exportação dos dados em arquivo CSV (*comma separated values*/valores separados por vírgulas, que é um formato padrão para base de dados, e aberto em Excel).

	A	B	C	D	E
1	'Filename'	'Status'	'Size'	'SHA-256 (Source)'	'SHA-256 (Target)'
2	Only first 100 files can be exported in the free version. Please purchase TeraCopy Pro to unlock full functio				
3	.DS_Store	Verified	8196	5a02a3a8359fdde4528b	5a02a3a8359fdde452809
4	1911.04252v4.pdf	Verified	2801395	7d14856fc292ab6f9bfe	7d14856fc292ab6f9bfe84
5	FALTA FAZER NO LIVRO.doc	Verified	15557	c4dbf76d9c8a9fc0362f	c4dbf76d9c8a9fc0362fcd
6	.DS_Store	Verified	6148	e6e7057bc5106eff28ce	e6e7057bc5106eff28ce35
7	1e72e-storage.png	Verified	20743	8de8ebf1db38bb405ee	8de8ebf1db38bb405ee0c
8	3-2-1-1-0.png	Verified	120413	f0332cdda0fb906675b5	f0332cdda0fb906675b5b!
9	3-2-1.png	Verified	98200	7c5c5deee22883eb36b	7c5c5deee22883eb36bba
10	3BarChart.jpeg	Verified	82468	aa70984b187e86cecf8	aa70984b187e86cecf8aff

Figura 45. Acima, um pequeno trecho do arquivo CSV aberto em Excel. Na coluna *Status*, podemos atestar que os arquivos foram verificados (*Verified*). Em um processo de cópia de milhares de arquivos, fazer a busca e filtragem pela planilha Excel irá facilitar o processo de pesquisar por arquivos que tiveram falha no processo de verificação via a soma de verificação.

Beyond Compare

Um *software* multiplataforma, para Mac, Windows e Linux, muito utilizado para realizar trabalhos de sincronização e conferência de diferenças entre pastas. A versão Standard e a Pro podem ser instaladas em inúmeras máquinas e plataformas com a mesma licença, desde que seja um único usuário.

Esta ferramenta permite fazer uma série de comparações entre pastas, subpastas e arquivos. A função de sincronizar pastas (*Folder Sync*) faz a cópia e a verificação para garantir que duas pastas sejam idênticas. Isso é sempre feito seguindo regras que podem ser modificadas.

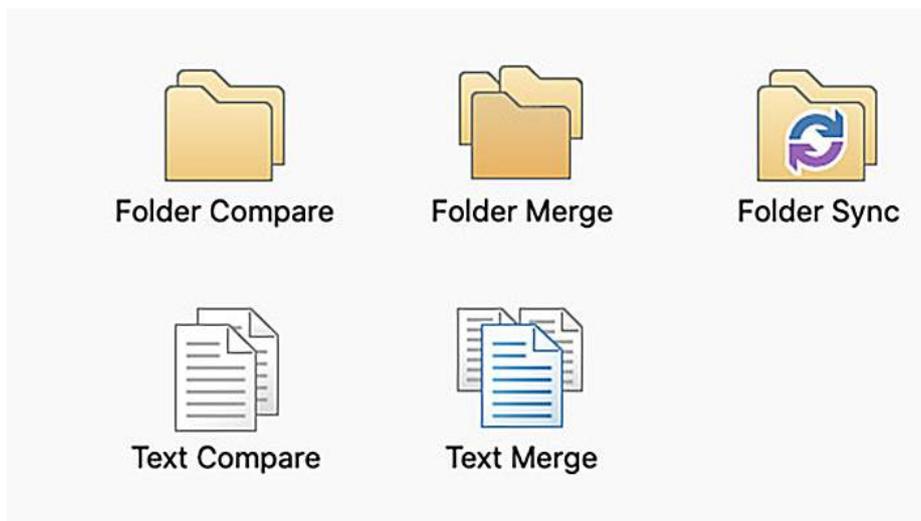


Figura 46. Reprodução de tela do *software* Beyond Compare que permite comparação entre pastas, subpastas e arquivos.

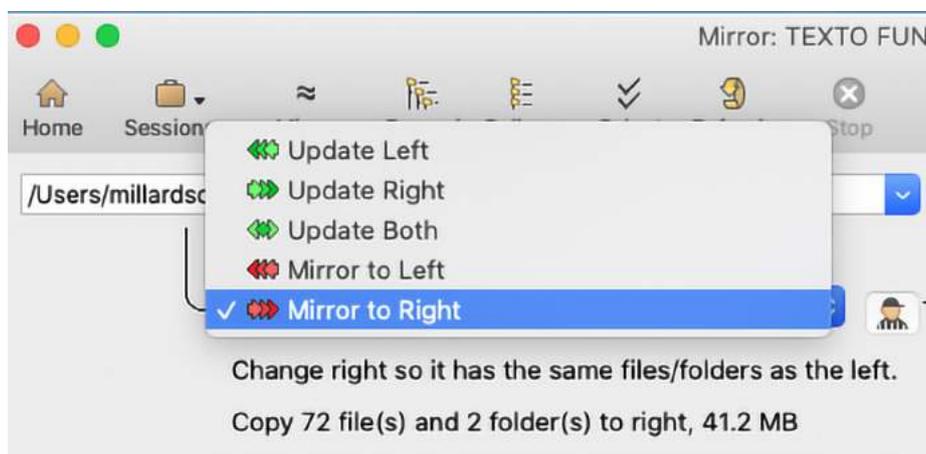


Figura 47. Reprodução de tela do *software* Beyond Compare que permite comparação entre pastas, subpastas e arquivos.

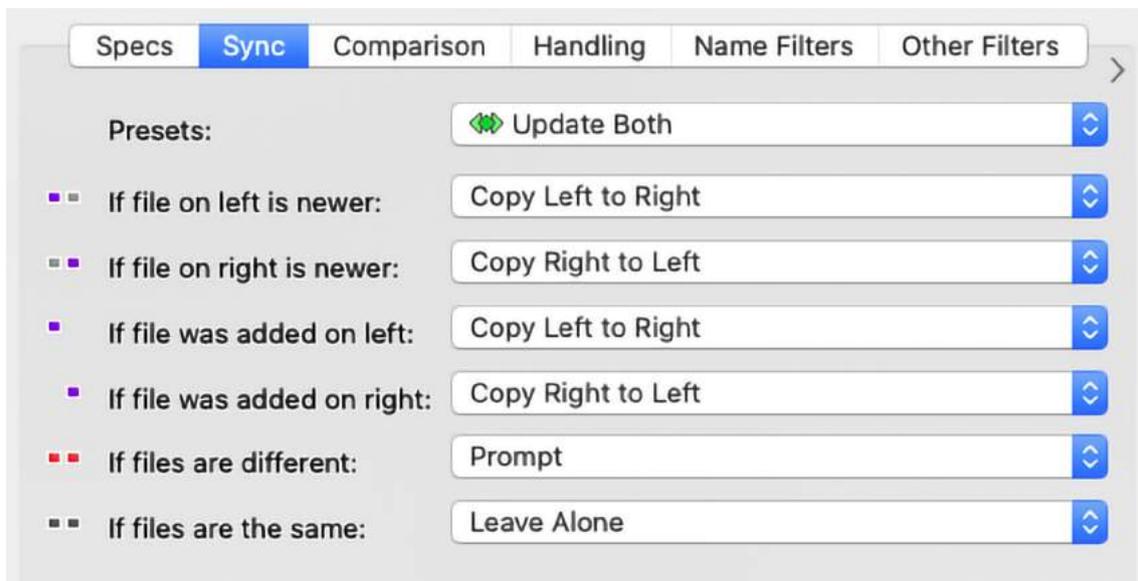


Figura 48. Reprodução de tela do *software* Beyond Compare que permite comparação entre pastas, subpastas e arquivos.

Vemos na imagem acima as opções de atualizar à esquerda (*Update Left*), atualizar à direita (*Update Right*), atualizar ambos os lados (*Update Both*), espelho para a esquerda e para a direita (*Mirror to Left* e *Mirror to Right*). Quase todos os *softwares* de sincronização e *backup* utilizam estas funções básicas, e estamos sempre falando de um lado e outro, esquerda e direita, onde espelho, como a imagem indica, é uma cópia exata, e atualizar é adicionar novos arquivos do original que não existem ainda na cópia.

Para entender bem estas funções é necessário criar pastas covaia com arquivos diversos para realizar testes com estas diversas funções, até entender claramente o que cada uma faz. Só então devemos partir para usar estas funções em pastas de trabalho.

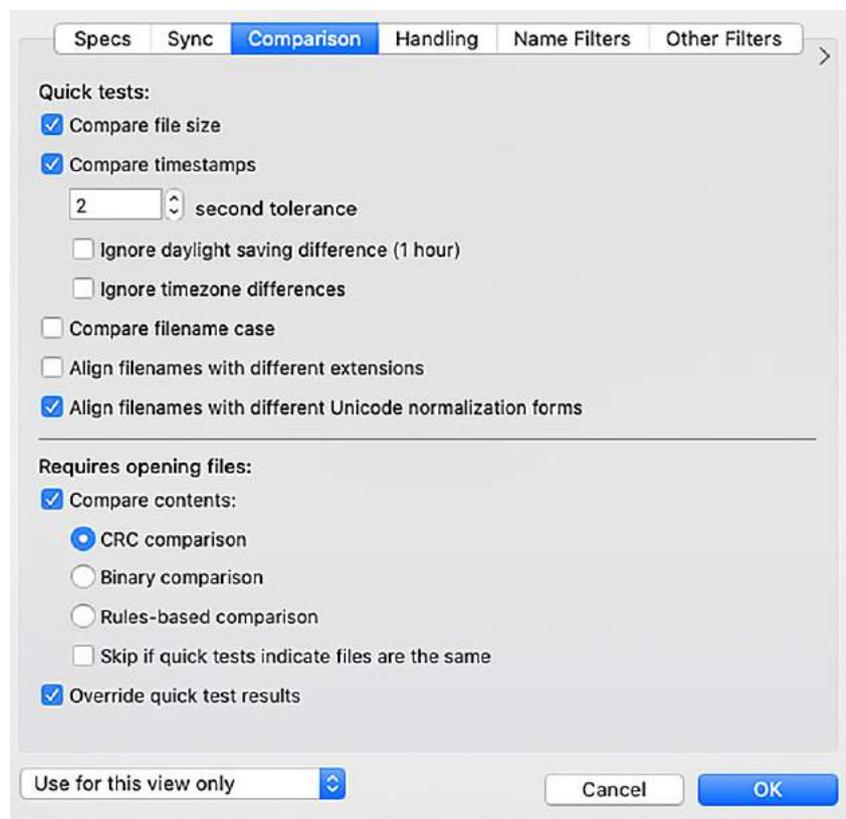


Figura 49. Reprodução de tela do *software* Beyond Compare.

Diversos *softwares* de comparação possuem opções que permitem afunilar como será feita a comparação. Algumas opções do *Beyond Compare* serão abordadas a seguir.

Compare file size (comparar tamanho do arquivo): se o nome do arquivo for igual, mas tiver qualquer diferença de tamanho o *software* entenderá como sendo dois arquivos diferentes.

Compare timestamps (comparar linha do tempo real): nesse modo de comparação o *software* leva em conta o dia e horário, inclusive os dados UTC (Tempo Universal Coordenado) de fuso horário.

Align filename with different extensions (alinhar nomes de arquivos com extensões diferentes): esta função pode ser muito interessante quando for necessário comparar um conjunto de arquivos com o mesmo nome, mas extensões diferentes, como originais de câmera (.NEF ou .DNG) em relação às versões de difusão (.JPG).

Nas funções de verificação do Beyond Compare, sempre será necessário abrir os arquivos. Se a opção *Skip if quick test results indicate the files are the same* (ignorar

se os testes rápidos indicarem que os arquivos são idênticos) for selecionada, os arquivos não serão abertos.

CRC Comparison (referência da Wikipedia – <https://pt.wikipedia.org/wiki/CRC>): A verificação cíclica de redundância (do inglês, CRC – Cyclic Redundancy Check) é um método de detecção de erros normalmente usada em redes digitais e dispositivos de armazenamento para detectar mudança acidental em cadeias de dados. CRCs são amplamente utilizados e são uma ótima forma de verificação de soma.

Binary Comparison (comparação binária): nesta comparação, mais lenta, os arquivos são abertos e comparados *byte por byte*.

Temos ainda mais opções de filtragem, em que é possível incluir ou excluir tipos de arquivos ou pastas (um arquivo excluído como ***.jpg** indica qualquer nome de arquivo, dado pelo asterisco, com a extensão .jpg. A opção ****** incluirá todos os arquivos – qualquer nome e qualquer extensão).

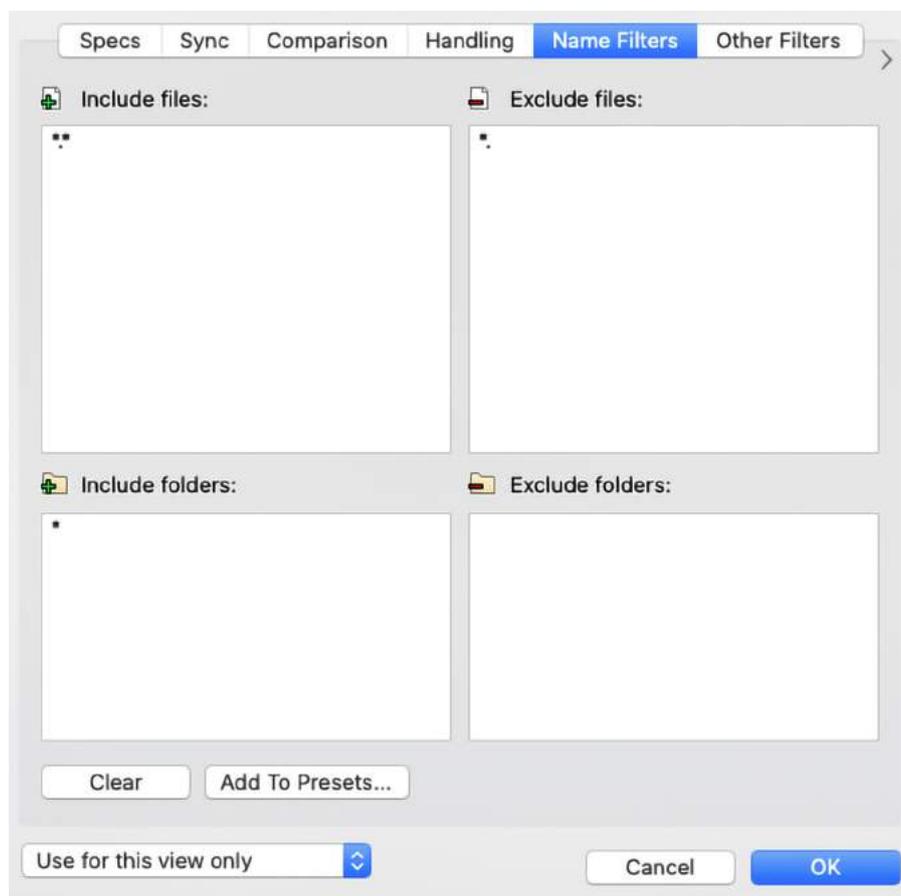
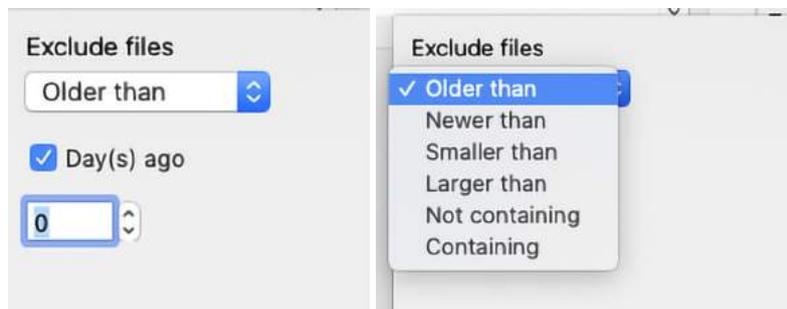


Figura 50. Reprodução de tela do software Beyond Compare com opções de filtragem de arquivo.

Outra opção de filtragem é poder excluir arquivos com base em critérios como “mais velho que”, “mais novo que”, “menor que” e “maior que”.



Figuras 51 e 52. Reprodução de tela do *software* Beyond Compare com opções de filtragem de arquivo baseada em critérios específicos.

Nas duas últimas opções, *Not containing* e *Containing*, de “não conter” ou “conter”, podemos especificar regras de arquivos que estamos procurando ou queremos excluir. Podemos procurar por esses fatores, mas a função mais completa, *Regular expression* (Expressão regular), permite criar expressões como as listadas a seguir.

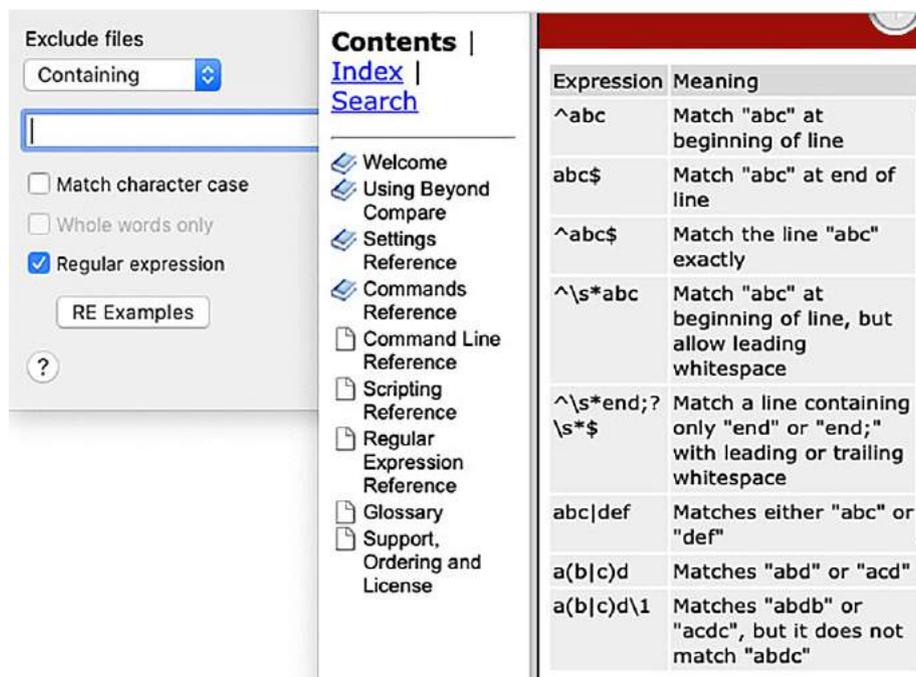


Figura 53. Reprodução de tela do *software* Beyond Compare com opções de filtragem de arquivo baseada em critérios específicos.

Novamente, e reforçando esse conceito, crie pastas para a realização de testes com as diversas opções do *software* em pastas de teste para ter um entendimento melhor de como o *software* funciona antes de começar a aplicá-lo nos seus arquivos definitivos.

Softwares de backup

Softwares que realizam *backups* têm muitas funções de programação para realizar automaticamente o processo de cópia e atualização de seus *backups*. Esses *softwares* são mais indicados para processos de construção e manutenção de fluxos de *backup*, porque as pastas precisam ser configuradas para criar um nome de procedimento de *backup*, incluindo uma origem e um destino. O *software* também pode ser programado, por exemplo, para realizar atualizações no final do dia, uma vez por semana ou continuamente. Em muitos desses *softwares*, temos também a opção de incluir processos de verificação. Não são muito práticos para a realização de cópia de um lugar a outro. *Beyond Compare* é muito mais simples para isso, por exemplo. Existem muitos *softwares* para realização de *backup* e novamente vale a pena baixar e testá-los na busca de um que funcione melhor para seu sistema de trabalho.

Usar softwares para encontrar duplicatas em sua nova matriz

Existem *softwares* que podem encontrar arquivos com o mesmo nome, bem como arquivos semelhantes. Em um processo de organização, é bom fazer esse processo de triagem, porque é comum, como dito anteriormente, na dúvida, fazer mais e mais *backups* das mesmas imagens em lugares diferentes. Como queremos montar uma nova matriz, será necessário juntar todos os nossos materiais. Somente assim é que podemos rodar esses *softwares* na busca de pastas e arquivos repetitivos. Os *softwares* têm também como exibir arquivos com a mesma imagem, mas nomes e formatos de arquivo diferentes.

O número de *softwares* que realizam esse tipo de pesquisa é muito grande, e eles deverão ser testados até que você encontre um que funcione para o que deseja. A versão *Duplicate File Finder Pro* é uma ótima ferramenta, porque analisa todos os tipos de arquivo, não somente de imagem. Há outras ferramentas, como o *Duplicate Photos Fixer Pro*, mas esse *software* só pesquisa arquivos de imagem.

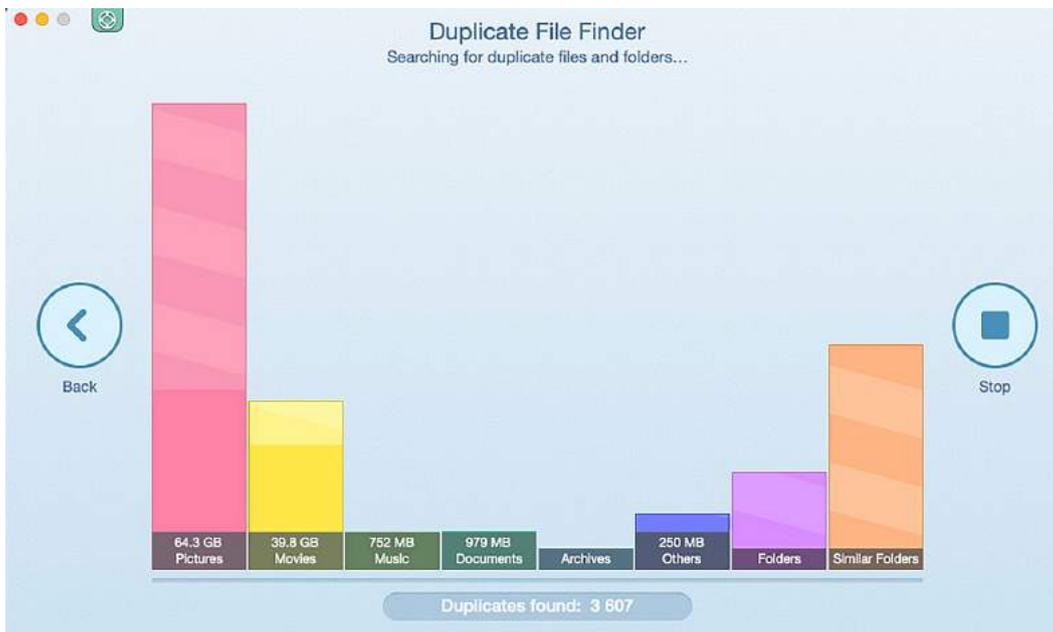


Figura 54. Neste gráfico do *Duplicate File Finder Pro*, podemos ver os tipos de arquivos duplicados em uma grande pasta de arquivos. Tem 3.607 arquivos duplicados. 64,3 GB são fotografias (*Pictures*), e tem 39,8 GB de filmes, pastas e pastas similares. Ao clicar em cada um desses itens, é possível ver e selecionar os arquivos que serão descartados.

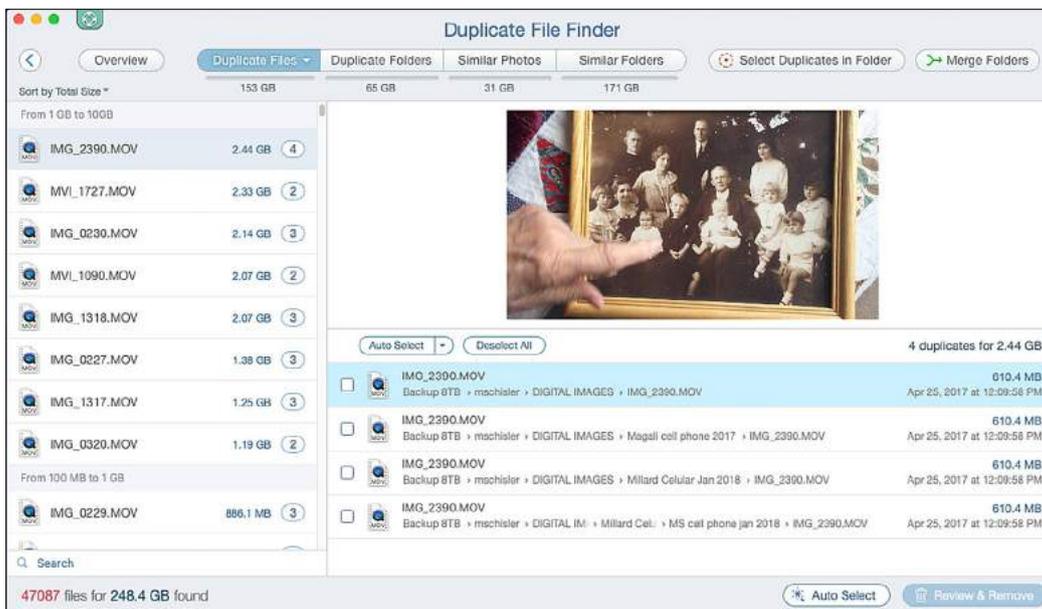


Figura 55. Nesta imagem do *Duplicate File Finder Pro*, temos o mesmo filme IMG_2390.MOV presente em quatro pastas diferentes, mas é exatamente o mesmo arquivo, com a mesma data de origem e tamanho. A seleção para remoção nesse *software*, e em outros, pode ser manual ou por regras de que deverá ser removido.

Nesta imagem vemos que o *Duplicate File Finder Pro* encontrou duas pastas idênticas, uma das vantagens desse aplicativo. Posso analisar o conteúdo de cada pasta, verificar se as duas contêm exatamente a mesma informação e decidir, então, qual manter.

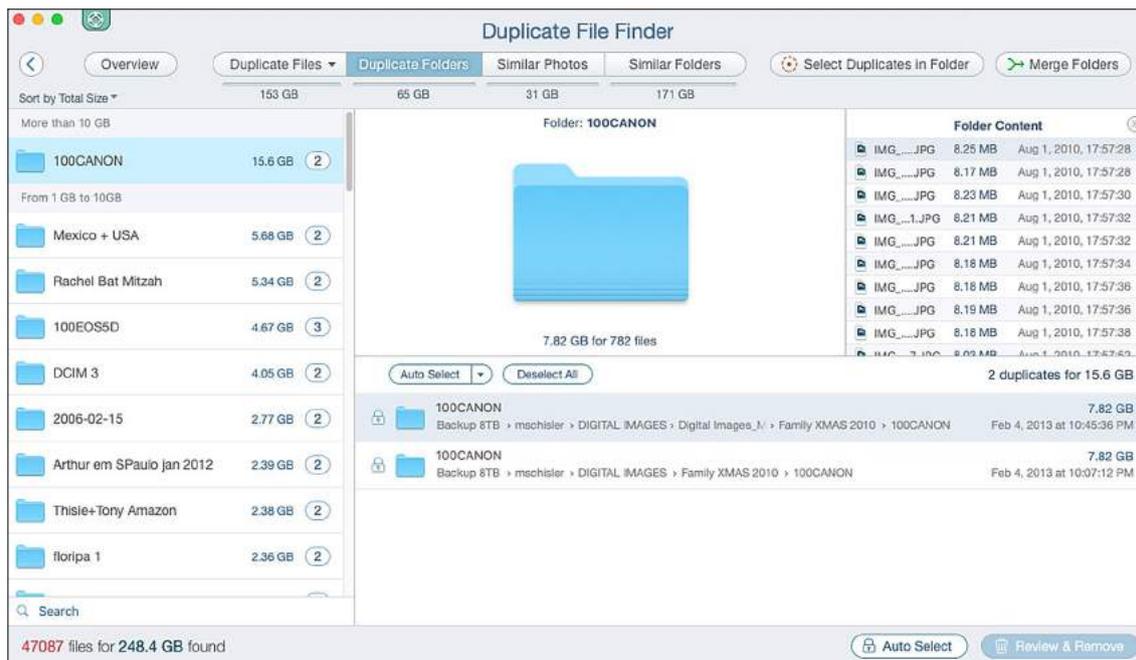


Figura 56. Reprodução de tela do software Duplicate Photos Finder.

No *Duplicate Photos Finder*, a seguir, temos as mesmas imagens, mas com nomenclaturas e tamanhos diferentes. O próprio software faz uma sugestão, com ícone de lixeira, dos arquivos que deveriam ser descartados, preservando a imagem com maior tamanho. Nesse caso, é estranho, e valeria a pena fazer uma pesquisa mais detalhada nestas imagens para entender o que aconteceu antes da tomada de decisão, porque as duas imagens com nome original de câmera/celular (IMG_) são menores, e as outras parecem que passaram por um aplicativo de rede social, onde foram renomeadas.

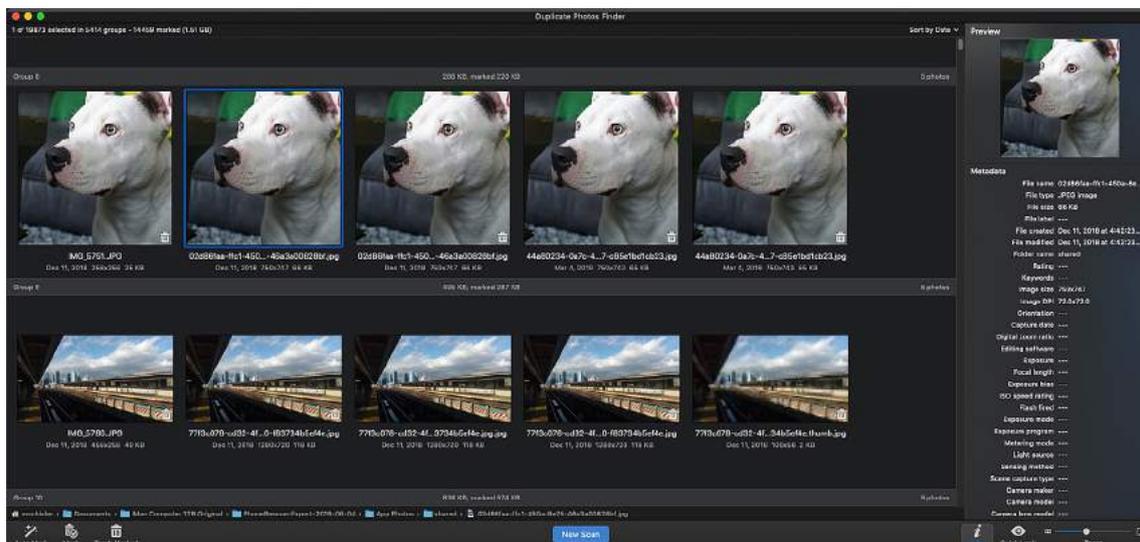


Figura 57. Reprodução de outra tela do *software* Duplicate Photos Finder.

O aplicativo gratuito Anti-Twin escrito por Joerg Rosenthal³¹ é uma ferramenta muito interessante que roda em Windows. Vemos a seguir uma captura de tela da sua interface gráfica com o usuário. Dá para ver pelos itens do menu que o *software* tem diversas opções para *Comparar nomes* (com extensões diferentes, tamanhos idênticos) e coincidências, onde 100% são exatamente iguais, mas, com a redução em coincidências, o *software* encontrará arquivos semelhantes. Em *Comparar conteúdo*, permite-se fazer uma comparação *byte a byte* ou comparar os *pixels* das imagens.

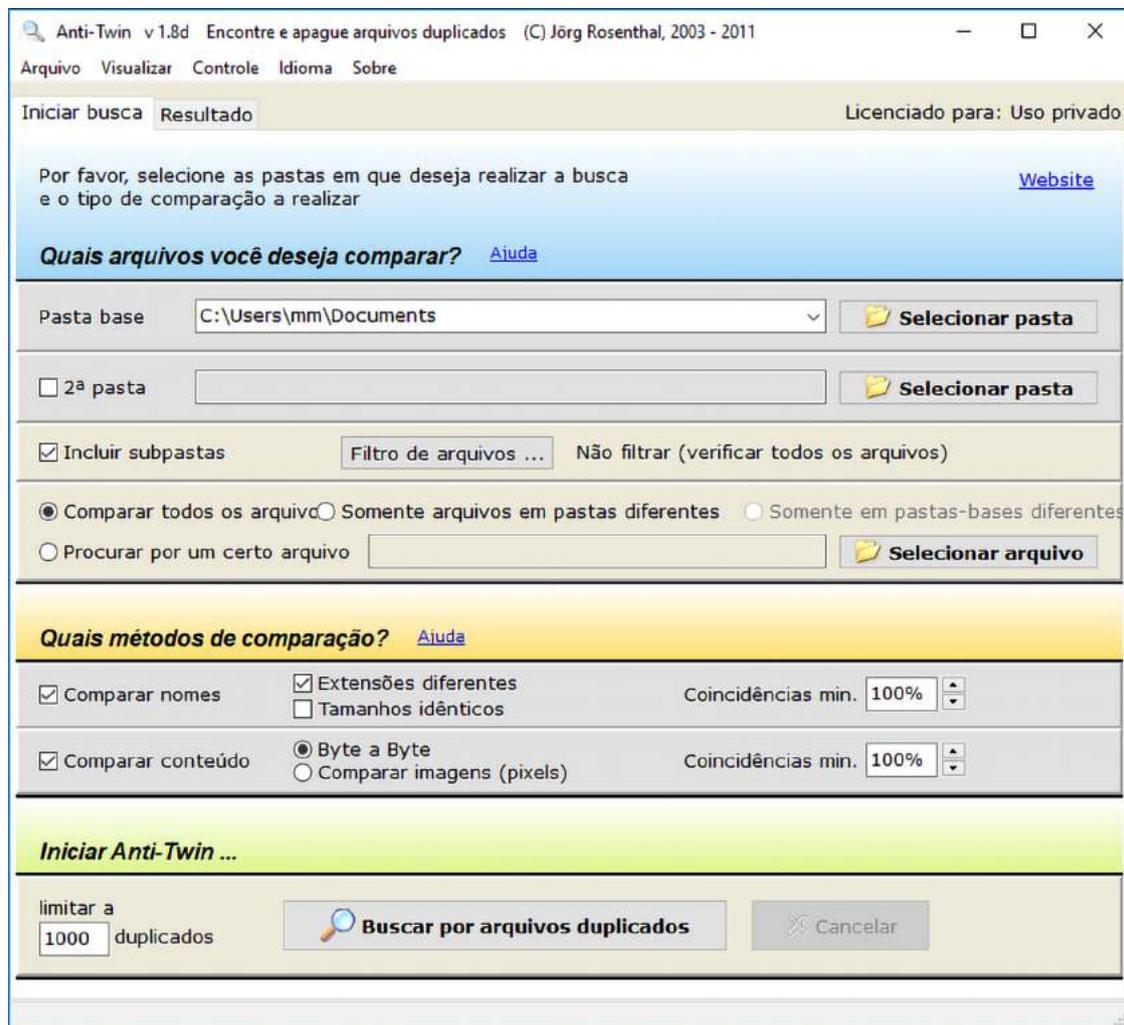


Figura 58. Reprodução de tela do aplicativo Anti-Twin.

Eliminar ou guardar

Até agora estamos usando *softwares* para fazer o processo de limpeza de arquivos duplicados. Eu me lembro de uma palestra que dei em 2008, em que tive um debate com uma pessoa da plateia, que defendia que o armazenamento digital estava ficando cada vez mais barato e que valeria a pena guardar tudo, não jogar coisas fora. Sabemos que o custo de preservar é muitas vezes maior do que o custo em si do armazenamento. Sabemos também que será impossível preservar a produção digital de tudo o que é feito. Porém, será que em nossas próprias coleções vale a pena olhar para o passivo digital e fazer uma triagem?

Estas são decisões que deverão ser tomadas, mas eu realçaria que usar *softwares*, como discutido aqui, para encontrar arquivos duplicados, e removê-los de forma rápida e eficiente é um passo importante. Depois dessa etapa, é de suma importância garantir que os novos fluxos de arquivos digitais, os novos hábitos, não estão repetindo os erros gerados nas práticas anteriores. Sempre será possível voltar ao passivo para processos de organização e limpeza, mas, de forma geral, o custo desse tempo de trabalho é mais elevado do que o custo de guardar esses dados digitais. Veja a tabela comparativa a seguir. Nesse cálculo não está sendo levado em conta o custo de catalogação desses dados. É aí que residem os reais custos e tempo do trabalho. Isso tem que ser uma decisão que envolve analisar com clareza os custos e benefícios de investir tempo nesse material passivo. O mais importante é “estancar a sangria”, resolvendo a produção atual e para a frente, a fim de não gerar um excesso de materiais que depois precisem ser revisitados para decidir o que deverá ou não ser preservado.

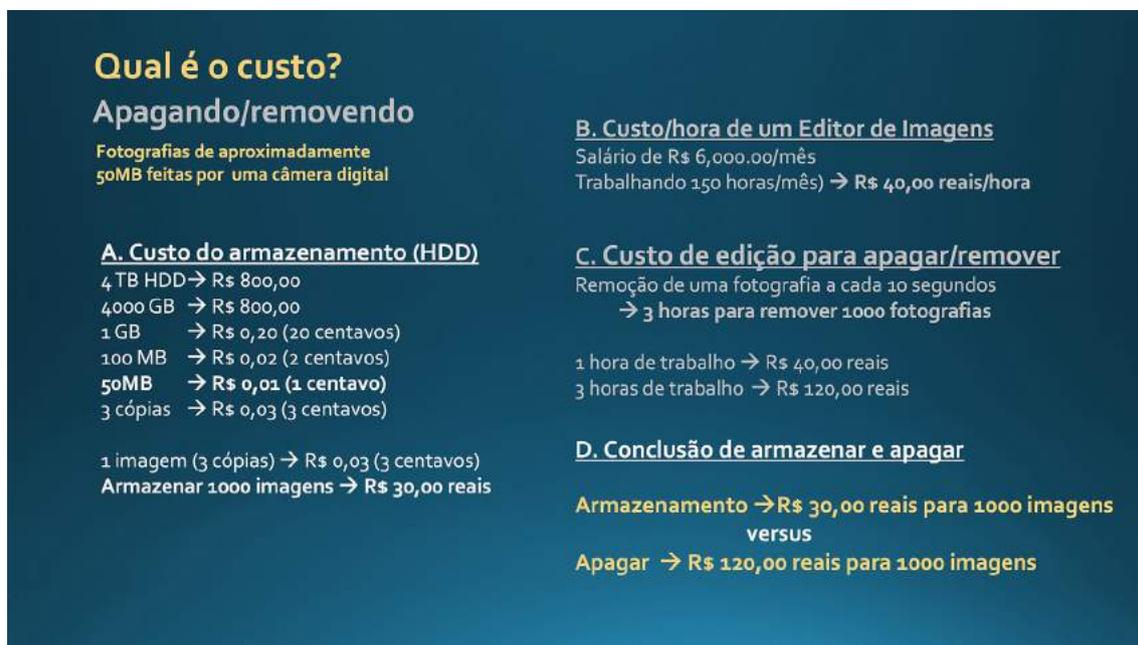


Figura 59. Nesta imagem, podemos comparar o custo somente do armazenamento de imagens de uma câmera digital que produz arquivos RAW de em torno de 50 MB cada. Não estão incluídos nesse exemplo os tempos de alguém que fará a catalogação, um trabalho especializado e que exige *expertise*. O custo do HDD considerado aqui é o de um Western Digital Red, de 3,5 polegadas, no primeiro semestre de 2023.

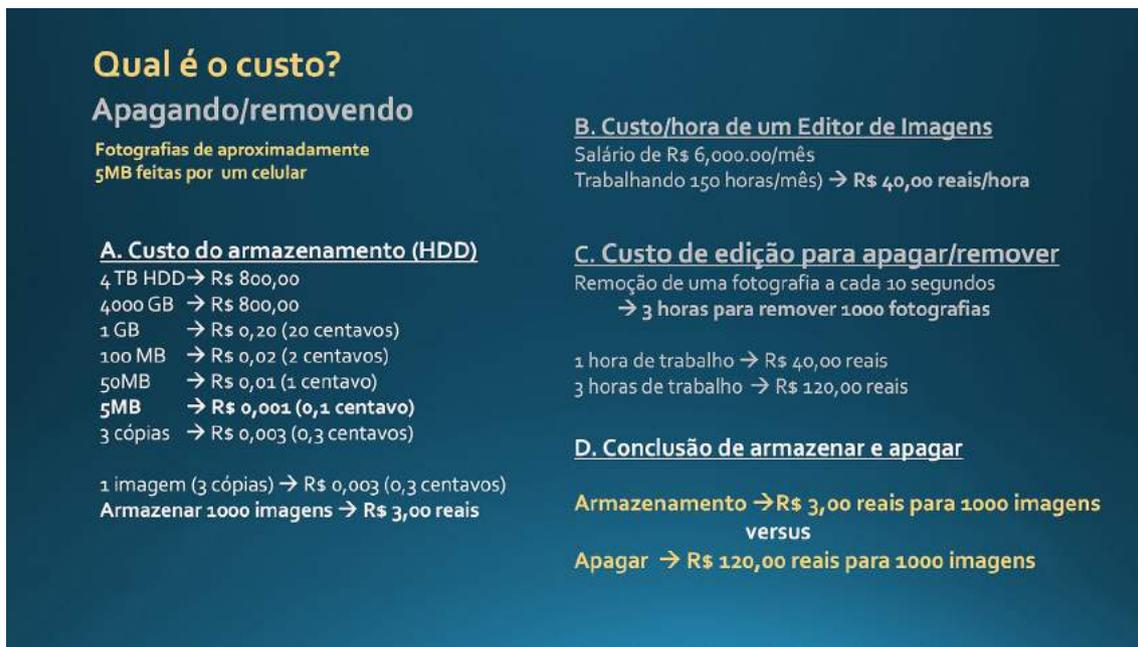


Figura 60. A mesma tabela, mas agora com uma imagem de celular de em torno de 5 MB. O custo de armazenamento é muito menor, considerando o tamanho desse arquivo.

Já discutimos que o armazenamento representa somente a ponta do *iceberg*, e que a manutenção de dados digitais passa pela descrição e por inúmeras outras atividades contínuas de gerenciamento para a preservação digital.

Usar softwares para criar uma lista de formatos de arquivo na busca de materiais que não devem ser preservados, ou que precisam ser migrados para novos formatos

Stefan Trost³² escreveu diversos programas interessantes que são gratuitos, com recomendações de doações para ajudar a mantê-los.

O *FilelistCreator* é uma das muitas ferramentas que descreve todos os arquivos em uma pasta, em subpastas ou em dispositivos, como um HDD. É só arrastar uma pasta ou dispositivo no espaço superior, e o *software* rapidamente processa a listagem.

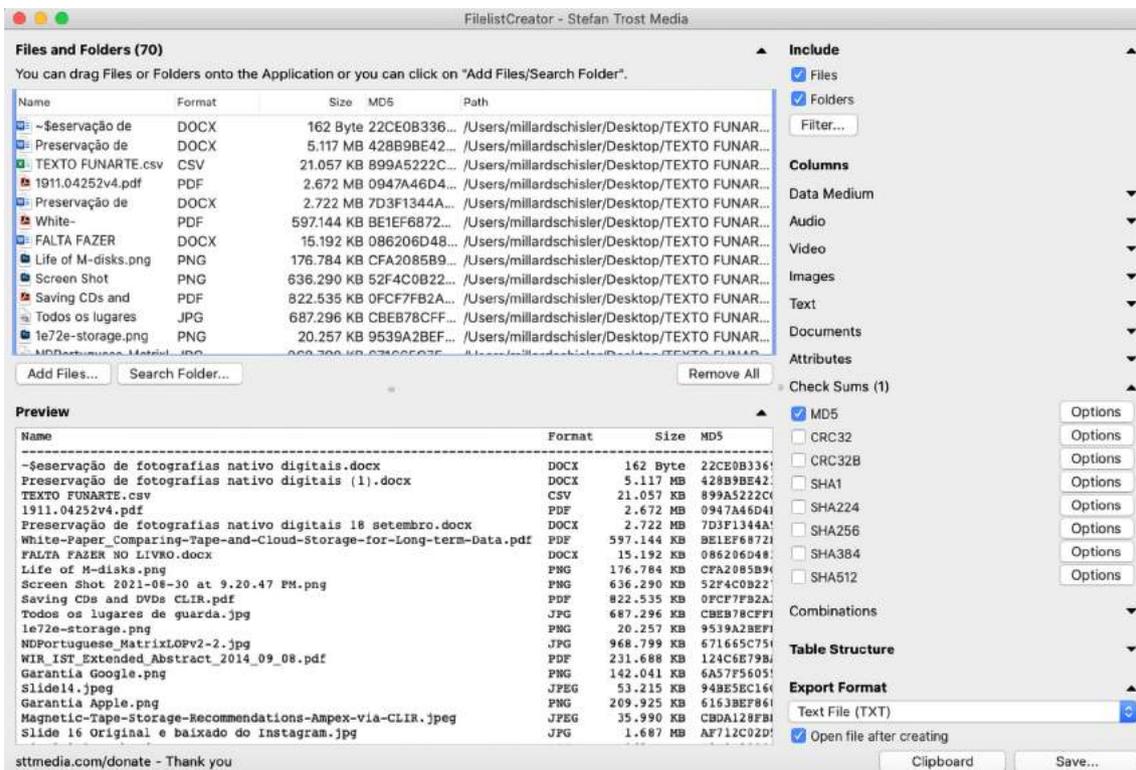


Figura 61. Reprodução de tela do programa Filelist Creator com a prévia de listagem criada.

Esse *software* permite fazer seleções que gerarão colunas na exportação em Excel. Podemos gerar somas de verificação para todos os arquivos, escolhendo diversos algoritmos. Ao copiarmos esses arquivos para outro lugar, podemos verificar se estas somas se alteraram.

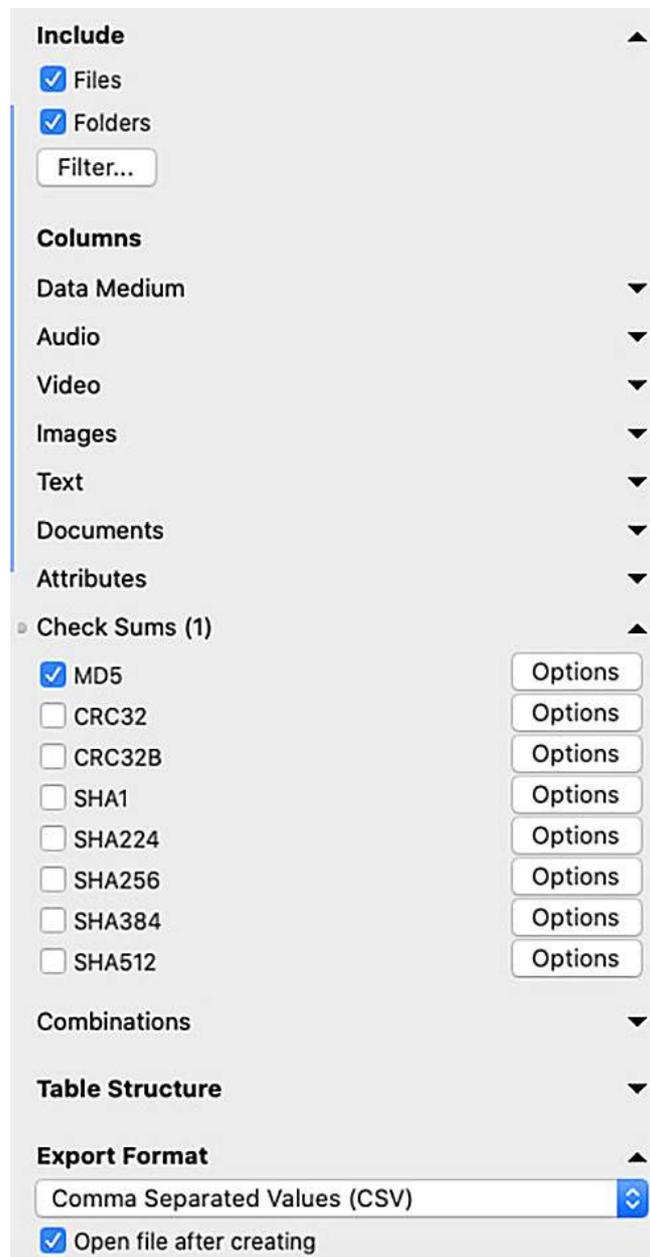


Figura 62. Reprodução de tela do programa Filelist Creator com parâmetros de classificação de arquivos.

Podemos escolher muitas informações para serem analisadas em arquivos de som, vídeo, imagem, texto, entre outros. São opções que nos ajudam a classificar a listagem dos arquivos. Vale a pena explorar com mais detalhes todas estas opções.

Em *Images*, podemos escolher profundidade de *bits* das imagens, por exemplo (*Bit Depth*).

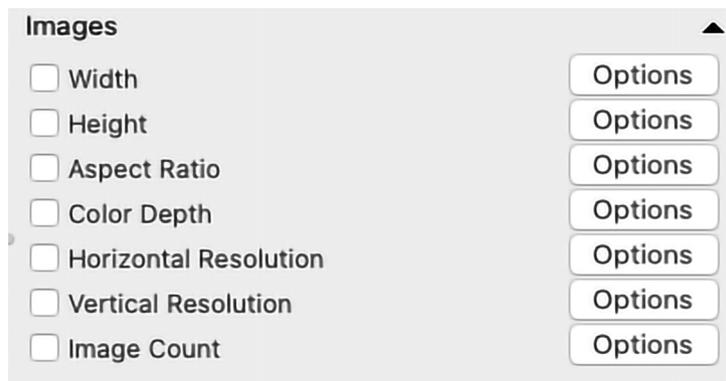


Figura 63. Reprodução de tela do programa Filelist Creator com opções de profundidade de *bits* de imagens.

A listagem completa pode ser exportada para um arquivo.csv, e depois aberta em Excel. Os dados podem ser filtrados e pesquisas feitas, por exemplo, de formatos de arquivo, buscando entender os tipos de arquivos que temos, e se há algum arquivo que deva ser migrado para um formato mais seguro para preservação digital a longo prazo (Normalização).

Name	Format	Size	Color Depth	MD5	Path
~\$eservav8v	DOXX	162 Byte		22CE083369	/Users/millardschisler/Desktop/TEXT0 FUNARTE/~\$eservav8vEo de fotografias nativo digitais.docx
Preservav8v	DOXX	5.117 MB		428B9BE423	/Users/millardschisler/Desktop/TEXT0 FUNARTE/Preservav8vEo de fotografias nativo digitais (1).docx
TEXT0 FUNA	CSV	21.057 KB		899A5222CC	/Users/millardschisler/Desktop/TEXT0 FUNARTE/TEXT0 FUNARTE.csv
1911.04252v	PDF	2.672 MB		0947A46D48	/Users/millardschisler/Desktop/TEXT0 FUNARTE/1911.04252v4.pdf
Preservav8v	DOXX	2.722 MB		7D3F1344A9	/Users/millardschisler/Desktop/TEXT0 FUNARTE/Preservav8vEo de fotografias nativo digitais 18 setembro.docx
White-Paper	PDF	597.144 KB		BE1EF6872D	/Users/millardschisler/Desktop/TEXT0 FUNARTE/White-Paper_Comparing-Tape-and-Cloud-Storage-for-Long-term-Data.pdf
FALTA FAZER	DOXX	15.192 KB		086206D483	/Users/millardschisler/Desktop/TEXT0 FUNARTE/FALTA FAZER NO LIVRO.docx
Life of M-dis	PNG	176.784 KB	32	CFA2085B9C	/Users/millardschisler/Desktop/TEXT0 FUNARTE/IMAGENS LIVRO/Life of M-disks.png
Screen Shot	PNG	636.290 KB	32	52F4C0B227	/Users/millardschisler/Desktop/TEXT0 FUNARTE/IMAGENS LIVRO/Screen Shot 2021-08-30 at 9.20.47 PM.png
Saving CDs a	PDF	822.535 KB		0FCF7FB2A2	/Users/millardschisler/Desktop/TEXT0 FUNARTE/IMAGENS LIVRO/Saving CDs and DVDs CLR.pdf
Todos os lug	JPG	687.296 KB	24	CBEB78CFFB	/Users/millardschisler/Desktop/TEXT0 FUNARTE/IMAGENS LIVRO/Todos os lugares de guarda.jpg
1e72e-stora	PNG	20.257 KB	24	9539A28EFD	/Users/millardschisler/Desktop/TEXT0 FUNARTE/IMAGENS LIVRO/1e72e-storage.png
NDPortugues	JPG	968.799 KB	24	671665C750	/Users/millardschisler/Desktop/TEXT0 FUNARTE/IMAGENS LIVRO/NDPortugues_MatrikLOPv2-2.jpg
WIR_IST_Ext	PDF	231.688 KB		124C6E79BA	/Users/millardschisler/Desktop/TEXT0 FUNARTE/IMAGENS LIVRO/WIR_IST_Extended_Abstract_2014_09_08.pdf
Garantia Goi	PNG	142.041 KB	32	6A57F56059	/Users/millardschisler/Desktop/TEXT0 FUNARTE/IMAGENS LIVRO/Garantia Google.png
Slide14	JPEG	53.215 KB	24	94BE5EC16C	/Users/millardschisler/Desktop/TEXT0 FUNARTE/IMAGENS LIVRO/Slide14.jpeg
Garantia App	PNG	209.925 KB	32	61638EF868	/Users/millardschisler/Desktop/TEXT0 FUNARTE/IMAGENS LIVRO/Garantia Apple.png
Magnetic-Ta	JPEG	35.990 KB	24	CBDA128FB	/Users/millardschisler/Desktop/TEXT0 FUNARTE/IMAGENS LIVRO/Magnetic-Tape-Storage-Recommendations-Ampex-via-CLR.jpeg
Slide 16 Orig	JPG	1.687 MB	24	AF712C02D5	/Users/millardschisler/Desktop/TEXT0 FUNARTE/IMAGENS LIVRO/Slide 16 Original e baixado do Instagram.jpg
~\$y 3-2-1 sui	DOXX	162 Byte		C2E1A90182	/Users/millardschisler/Desktop/TEXT0 FUNARTE/IMAGENS LIVRO/~\$y 3-2-1 sucks.docx

Figura 64. Imagem de listagem de arquivos criada a partir do programa Filelist Creator.

Um *software* como esse, pela criação de listagens, pode também identificar somente pastas, arquivos ou ambos. É uma ferramenta importante para auxiliar na pesquisa de materiais contidos em dispositivos digitais cujos dados serão migrados para uma nova matriz.

Organizar estrutura nova de pastas

Esse deve ser o momento de revisitar a estrutura atual de pastas. Devemos evitar o aninhamento interminável, onde temos uma pasta, uma sequência quase infinita de subpastas, uma dentro da outra, e nomes também grandes para identificar estas pastas. Depois de um número elevado de pastas, chegamos aos arquivos. Veremos a seguir que, em processos de verificação e cópia de arquivos, sempre se leva em conta o nome completo do arquivo, que inclui o nome do trajeto, da unidade de disco, passando por todas as pastas, e chegando ao nome do arquivo. Portanto, na criação de uma matriz nova, esse é um momento de reavaliar a estrutura e organização dos conjuntos e pastas.

Se já não houver, esse também é o momento de criar uma pasta primária que servirá de matriz de todos os dados para facilitar nos processos de *backup* conforme já discutimos. Podemos também criar uma pasta central para todos os projetos em andamento e uma pasta central para tudo o que é permanente ou arquivo. Isso pode facilitar os processos de *backup*, pois o *backup* do arquivo pode ser atualizado com menos frequência que a pasta matriz dos projetos em andamento.

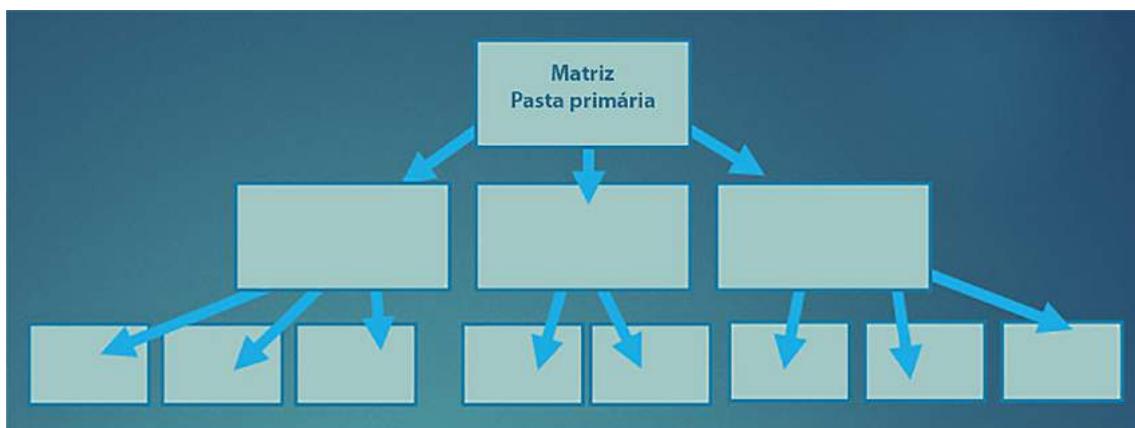


Figura 65. Esquema sugerido para criação de pasta primária como matriz de uma estrutura de arquivos para *backup*.

A diferença de armazenamento e organização das fotografias digitais

O armazenamento descreve como você lida com os arquivos. O conteúdo de uma pasta, os nomes das pastas e as hierarquias utilizadas. A organização, por outro lado, descreve como as imagens são agrupadas de acordo com o conteúdo, uso ou valor. Organização permite que você encontre todas as fotos de um trabalho, projeto ou pessoa, e deve ser fundamentada no conceito de metadados, e não necessariamente de pastas. Na preservação de objetos físicos, o armazenamento e a estrutura organizacional são frequentemente inseparáveis. No mundo digital, tudo isso muda. Uma imagem pode estar fisicamente em uma pasta em sua estrutura de pastas no HDD, mas, em um catálogo de imagens, ela será encontrada por metadados e poderá existir virtualmente em diversas coleções.



Figura 66. A imagem estará em uma pasta organizada por data ou assunto em termos de armazenamento. Ela deve ser encontrada, porém, via metadados descritos em um sistema de catalogação, e poderá existir de forma virtual em diversas coleções, sem ser replicada fisicamente. Imagem do autor.

Conforme o site do dpBestflow³³

Existem dois problemas principais com o uso de pastas para organização. O primeiro é que o trabalho que você pode fazer para organizar as pastas é muito limitado. Deve haver um método organizacional de nível superior, que só pode ser subdividido de forma limitada antes que o sistema se torne muito complicado e quebre. É mais importante dividir por data, cliente, projeto, assunto, classificação ou uso? Além disso, as informações que dependem da estrutura da pasta são muito frágeis. Se você remover uma imagem de uma pasta que designa o que é essa imagem, essas informações de conteúdo podem ser perdidas. Embora a estrutura de pastas possa ser útil na organização de suas imagens, sugerimos que a principal tarefa das pastas seja como uma estrutura de armazenamento.

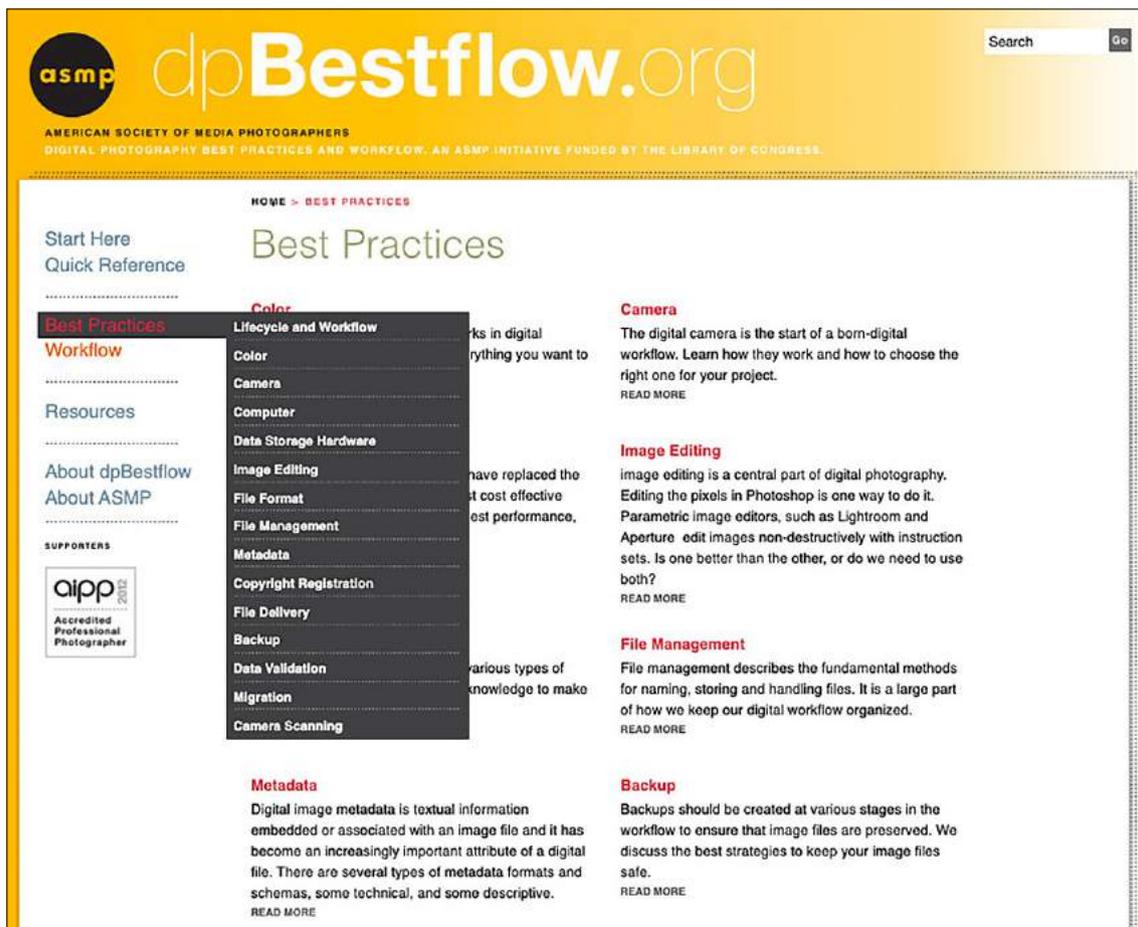


Figura 67. Apesar desse site não ser atualizado desde 2015, e com diversos *links* quebrados, ainda continua uma excelente fonte de informação sobre as melhores práticas na fotografia nato-digital.

Renomear os arquivos conforme necessidade

Quando tratamos de arquivos do nosso passivo, esses podem estar com padrões totalmente divergentes de nomenclatura. Muitas fotografias digitais devem estar ainda com o nome dado pela máquina digital ou celular. Outros arquivos digitais podem ter sido individualmente renomeados com um nome na tentativa de indicar do que se trata aquela imagem ou documento. Mais importante do que renomear os arquivos passivos é garantir que há um novo sistema de nomenclatura para os arquivos presentes, e que há inserção de metadados nos arquivos passivos para que possam ser encontrados. Portanto, é preciso decidir se vale ou não a pena realizar esse trabalho. Muitas vezes os arquivos podem ter sido utilizados em alguma outra

situação, e alterar o seu nome pode quebrar esse vínculo entre o seu nome atual e o nome que foi utilizado.

De qualquer maneira, o processo de renomear sempre deverá ser feito com *softwares* para esta finalidade. Novamente, há muitas ferramentas para realizar esta tarefa, das mais simples a algumas mais complexas. A seguir, vemos o *software* Adobe Bridge na opção *Batch Rename* (renomear em lote). Bridge é um *software* gratuito, mas é necessário ter uma ID da Adobe criada para poder baixar esse *software*. Bridge será muito útil para renomear arquivos, mas principalmente para a inserção de metadados, como veremos adiante.

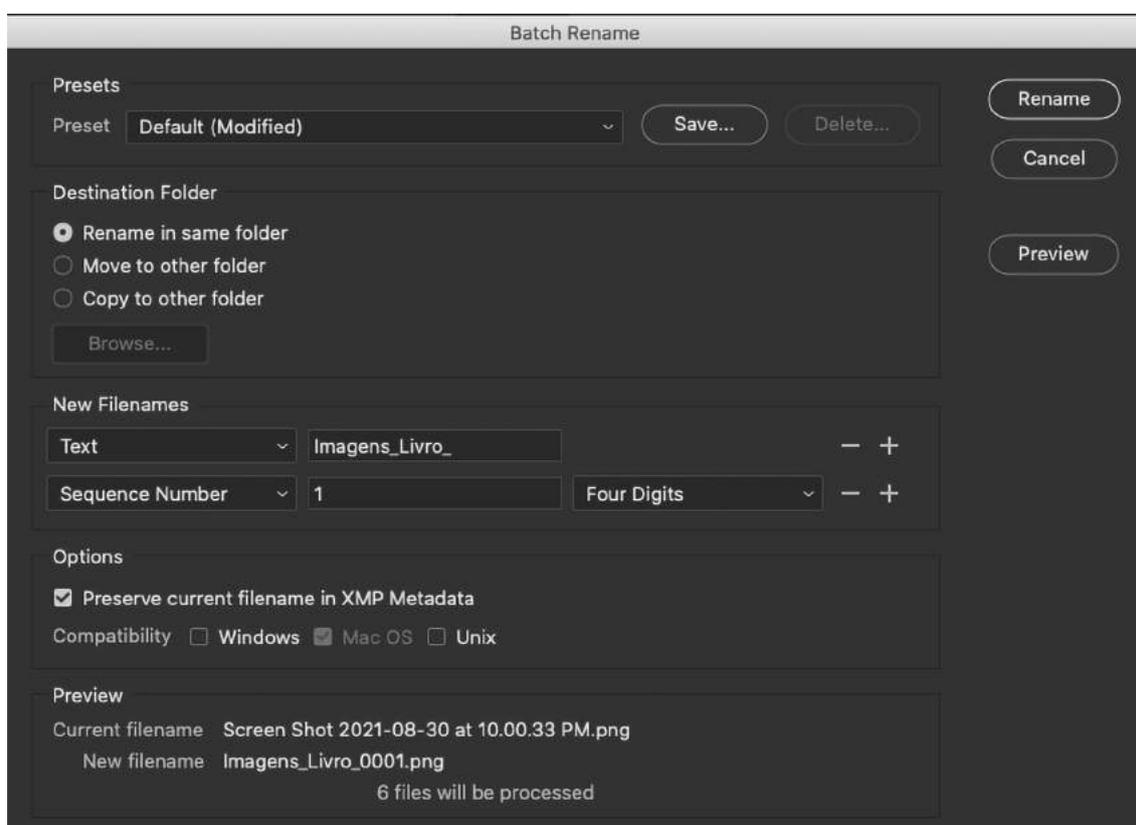


Figura 68. Reprodução de tela do *software* Adobe Bridge.

Nesta tela temos uma opção importante selecionada, que está em *Options* > *Preserve current filename in XMP Metadata* (Preservar nome atual nos metadados XMP). Ao selecionar esta opção, o arquivo será renomeado, mas o nome anterior será mantido dentro de um campo de metadados. Se o arquivo for renomeado novamente, perderemos o nome original, que será substituído pelo segundo nome dado,

a fim de abrir lugar para um terceiro nome. No Bridge, temos a opção de um texto livre, mas também poderemos inserir uma série de outras funções como sequência de números (realçada na Figura 69), sequência de letras, data e hora, entre outros. Podemos também clicar nas caixas de compatibilidade para Windows, macOS e Unix, para garantir não estar usando caracteres que causarão problemas nesses sistemas. Mesmo utilizando macOS ou Windows, muitos processamentos *on-line* passam por redes Unix.

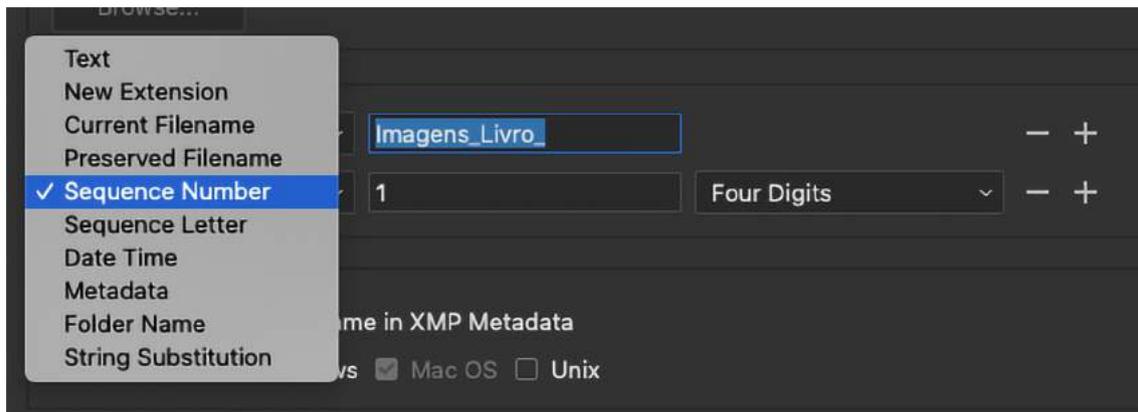


Figura 69. Reprodução de tela com função de uso de sequência de números no *software* Adobe Bridge.

Uma ferramenta muito potente para renomear que permite alterações bem complexas de arquivos e é gratuita para uso individual, é o *Bulk Rename Utility*³⁴, disponível somente para Windows. No site, tem uma seção de Funções, básicas e avançadas, que exibem a complexidade desse *software*.

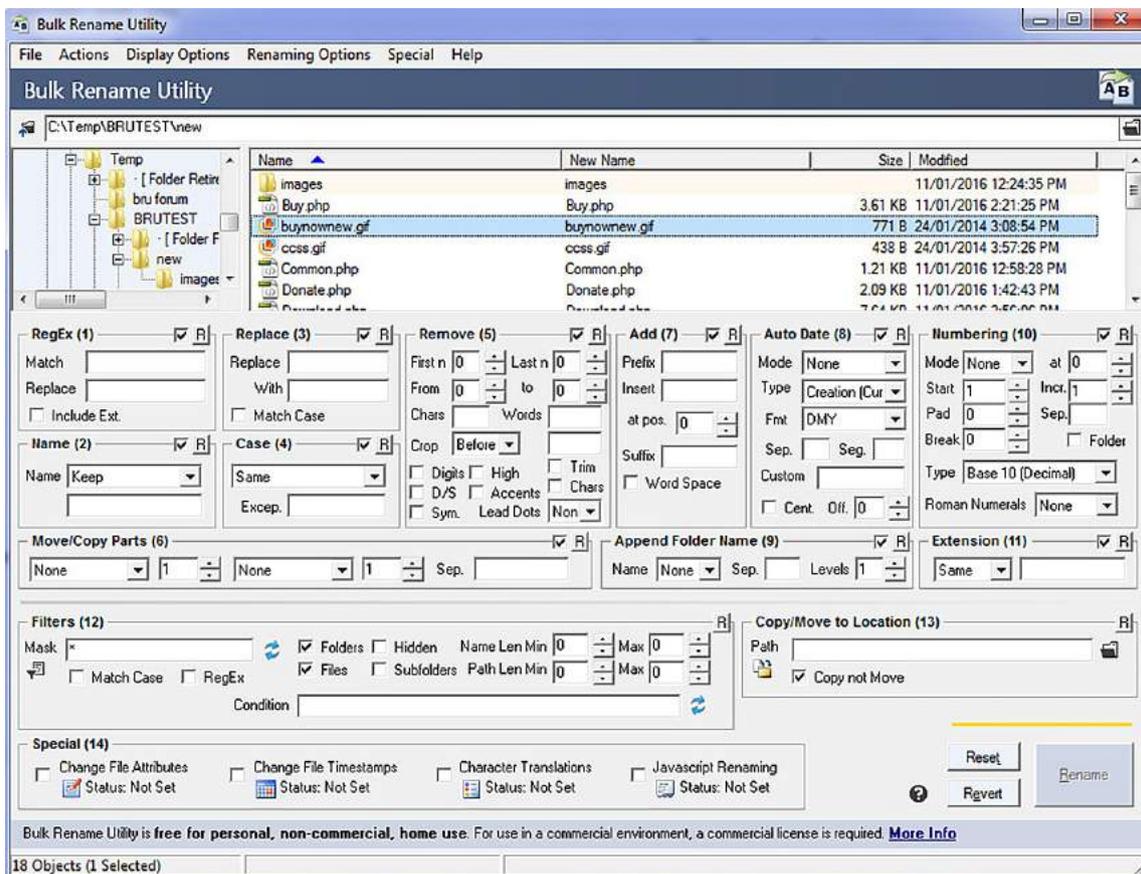


Figura 70. Captura de tela do software *Bulk Rename Utility*.

Como renomear arquivos digitais

Algumas regras importantes para renomear arquivos. Uma das questões importantes sobre a nomenclatura de arquivos e fotografias digitais é a de que o nome não deve ter que refletir o assunto de cada imagem, mas sim um identificador único, individual. As informações sobre o conteúdo deveriam ser inseridas nos metadados.

Existem algumas regras básicas que devem ser observadas:

- Usar caracteres alfabéticos (A-Z ou a-z)
- Usar números (0-9)
- Usar traço/hífen (-) e sublinhados (_), sempre evitando espaços em branco
- Usar o mínimo de caracteres

- Siga as regras básicas do sistema de computador para nomear arquivos
- Use nomes de arquivo exclusivos para cada imagem
- Evite incorporar nomes de trabalhos ou descrições em nomes de arquivos
- Aplique o sistema de nomenclatura de arquivos de forma consistente

Evitar	Exemplo de caracteres	Razões
Separadores de arquivos	: (dois pontos) / (barra) \ (barra invertida)	Você deve evitar o uso de dois pontos e barras nos nomes de arquivos e pastas porque alguns sistemas operacionais e formatos de unidade usam esses caracteres como separadores de diretório. Considere substituir a barra ou dois pontos em um nome de arquivo por sublinhado (<u> </u>) ou traço/hífen.
Símbolos não alfabéticos e não numéricos	¢™\$®	Caracteres não alfanuméricos podem não ser suportados por todos os sistemas de arquivos ou sistemas operacionais, ou podem ser difíceis de trabalhar quando exportados para determinados formatos de arquivo, como EDL, OMF ou XML.
Sinais de pontuação, parênteses, aspas, colchetes e operadores	., [] {} !; "" * ? < >	Esses caracteres são frequentemente reservados para funções especiais em linguagens de script e programação.
Caracteres de espaço em branco, como espaços, tabulações, novas linhas e retornos incorporados		Embora os discos formatados para OS X e macOS suportem espaços em nomes de arquivos, alguns <i>scripts</i> e aplicativos de processamento podem não reconhecer esses caracteres ou tratar seus arquivos de maneira diferente do esperado. Considere substituir espaços por sublinhado (<u> </u>) ou traço/hífen (-).

Tabela 1. Regras para evitar uso de caracteres e sinais especiais em nomes de arquivos.

Outra referência importante que deve ser observada é o número de caracteres utilizados no nome do arquivo. Para os sistemas computacionais, o que é utilizado é o caminho do arquivo, o seu endereço, quando ele é processado. Esse caminho também tem um limite de número de caracteres. São computados todos os componentes desse caminho, do diretório do disco, passando por todas as pastas e incluindo o caractere nulo, utilizado para indicar o final de uma cadeia de caracteres (*strings*). A contagem de todos esses caracteres não deverá exceder 260 (MAX_PATH). `C:\nomepasta1\nomepasta2\nomepasta3\nomepasta4\nomearquivo.ext{nul}` – nesta *string*, temos 67 caracteres. Se na nossa organização as pastas tiverem nomes enormes, e um aninhamento interminável de pastas, com arquivos com nomes enormes, rapidamente chegaremos a esse limite. Esse limite de MAX_PATH é da era antes do Windows 10.

Arquivo
2013_tax_
return.pdf

C:\Pasta1\Caminho realmente longo\Quanta recursão\Tão profundo\Uau\Ainda indo\Vou ficar sem ideias em breve\Tenho problemas organizacionais\Transtorno obsessivo-compulsivo de subdiretórios\ Aqui está uma GUID sem motivo\936DA01F-9ABD-4d9d-80C7-02AF85C822A8\Quase lá\Imposto de renda\2013\ 2013_imposto_renda.pdf

Figura 71. Criado por Ryan Ries, Stack Exchange³⁵, com tradução livre.

O trajeto total em português tem 310 caracteres de comprimento, incluindo todos os espaços e o caractere nulo do final.

O MAX_PATH = 260 caracteres → nome do arquivo + trajeto total.

De maneira bem-humorada, Ryan Ries procura demonstrar a necessidade de nomes sucintos e redução de camadas de pastas e subpastas (aninhamento) na organização dos nossos trabalhos.

No Windows 10, o suporte a nomes de arquivo longos pode ser habilitado, o que permite nomes de arquivos com até 32.767 caracteres (embora você perca alguns caracteres para caracteres obrigatórios que fazem parte do nome). No macOS esse número é de 1.024 caracteres, e em Linux, de 4.096 caracteres. De qualquer modo, trabalhar de forma conservadora, com um sistema enxuto de pastas e nomenclaturas, pode nos auxiliar a não ter problemas com quaisquer programas, redes e sistemas operacionais atuais ou legados, como também ajuda a criar um sistema mais enxuto e organizado.

Identificador único universal (GUID ou UUID)

Um identificador único universal (*Universally Unique Identifier – UUID*) é um número de 128 *bits* usado para identificar informações em sistemas de computação. Também podemos utilizar o termo *identificador único global (Globally Unique Identifier – GUID)*. O GUID válido deve especificar as seguintes condições: → Ser um número de 128 *bits*; gerado automaticamente, ter 36 caracteres (32 caracteres hexadecimais e 4 hifens) de comprimento. Deve ser exibido em cinco grupos separados por hifens (-), conforme esse exemplo de um arquivo JPEG: **123e4567-e89b-12d3-a456-426655440000.jpg**.

Identificadores únicos persistentes (*Persistent Identifiers*)

Diferentemente dos GUIDs ou UUIDs, os identificadores persistentes são formas de identificar um arquivo para um destino permanente. O *Digital Preservation Coalition* – DPC³⁶, tem uma abordagem muito didática de muitos temas voltados para a preservação digital. Sobre os identificadores persistentes, apresenta uma boa definição:

Um identificador persistente é uma referência duradoura a um recurso digital. Normalmente, ele possui dois componentes: um identificador exclusivo; e um serviço que localiza o recurso ao longo do tempo, mesmo quando sua localização muda. O primeiro ajuda a garantir a procedência de um recurso digital (que é o que pretende ser), enquanto o segundo irá garantir que o identificador resolva a busca para a localização atual correta.

Identificadores persistentes visam, assim, resolver o problema da persistência de acesso ao recurso citado, principalmente na literatura acadêmica. Com muita frequência, os endereços da web (links) não levam você ao recurso referenciado que se espera. Isso pode se dar por razões tecnológicas, como falha do servidor, mas as falhas criadas por humanos são mais comuns. As organizações transferem periódicos para novos editores, reorganizam seus sites ou perdem o interesse em conteúdos mais antigos, levando a links quebrados quando você tenta acessar um recurso. Isso é frustrante para os usuários, mas as consequências podem ser graves se o recurso vinculado for essencial por razões legais, médicas ou científicas.

Alguns dos principais identificadores únicos persistentes que deverão ser pesquisados são:

- DOI – Digital Object Identifier
- ORCID – Open Researcher and Contributor ID
- Handle System
- ARK – Archival Resource Key
- PURL – Persistent URL (Uniform Resource Locator)
- URN – Uniform Resource Name

Luis Fernando Sayão, um dos pesquisadores de referência em preservação digital no Brasil, escreveu um texto³⁷ que é uma ótima fonte de introdução, explicação e comparação dos diversos identificadores persistentes principais.

Alguns sistemas de identificadores persistentes são de acesso aberto, como ARK. Outros são feitos via assinaturas anuais de acordo com o número de itens, como DOI. Esse assunto precisa ser estudado com cautela pela instituição para entender os seus objetos digitais, recursos disponíveis, e processo de implementação antes de fazer a adoção de um ou outro sistema.

Inserção de metadados para descrever os arquivos

Metadados são informações (dados) que descrevem outros dados.

Em instituições de patrimônio, os metadados necessários devem ser considerados, como qualquer informação (em forma digital ou física), essenciais para garantir que o material digital a ser preservado seja e permaneça acessível, inteligível e utilizável ao longo do tempo. Os metadados fornecem à instituição as informações necessárias para acessar e preservar o patrimônio digital no futuro.

Guidelines for the selection of digital heritage for long-term preservation, UNESCO/PERSIST Content Task Force³⁸.

Metadados são as informações que agregam valor a outras informações, são os dados do nosso catálogo para os arquivos, dados de localização que permitem o acesso aos arquivos.

Diversos tipos de metadados podem ser produzidos. Alguns podem ser gerados automaticamente por uma máquina ou sistema, e outros, de forma manual, por um indivíduo que insere essas informações para descrever o objeto digital.

Podemos dividir os metadados em cinco grupos de acordo com o clássico texto editado por Murtha Baca do Getty Research Institute, *Introduction to Metadata, Third Edition*³⁹. Na tabela a seguir (Tabela 2 do livro de Murtha Baca), sobre as diferentes categorias de metadados e suas funções temos uma visão básica dos tipos de metadados com uma coluna a mais sobre o uso desses metadados, para finalidades internas ou externas, inseridas por esse autor. Os metadados podem estar embutidos nos arquivos ou em base de dados e/ou sistemas de gerenciamento de coleções ou ativos digitais.

Tipo	Definição	Usos/ interno ou externo
Administrativo	Metadados usados no gerenciamento e administração de coleções e recursos de informação.	São normalmente dados de controle interno dos acervos e não divulgados externamente.
Descritivo	Metadados usados para identificar, autenticar e descrever coleções e recursos de informações confiáveis relacionados.	Esses são os principais dados que descrevem o objeto, e são utilizados por usuários externos para pesquisar os acervos. Podem estar embutidos nos objetos digitais ou em base de dados/sistemas. São normalmente gerados manualmente, como dados catalográficos, inseridos de forma individual ou em lote, mas também podem ser gerados por sistemas de <i>machine learning</i> .
Preservação	Metadados relacionados com a gestão da preservação de acervos e recursos de informação.	Normalmente a documentação é interna e pode estar relacionada ao objeto original, quando falamos de representantes digitais de objetos físicos; e dados de verificação, migração ou atualização em dados digitais como forma de preservação.
Técnico	Metadados relacionados a como um sistema funciona e os dados que gera.	São todos os dados técnicos sobre hardware e software, sistemas, dados da digitalização, entre outros. Podem ser gerados automaticamente (EXIF) ou descritos internamente. Uso interno.
Uso	Metadados relacionados ao nível e tipo de uso de coleções e recursos de informação.	Documentação de uso interno referente ao uso do objeto digital; dados sobre os direitos de uso.

Tabela 2. Reprodução da Tabela 2 do livro *Introduction to Metadata, Third Edition*, de Murtha Baca do Getty Research Institute-GRI.

Esse assunto é muito vasto, e vamos nos concentrar em discutir especificamente o uso de metadados em arquivos fotográficos digitais – produzidos por meio da digitalização, ou nato-digitais.

Início do IPTC – International Press Telecommunications Council

O IPTC foi estabelecido em 1965 por um grupo de organizações de notícias, incluindo a Alliance Européenne des Agences de Presse (EANA), American Newspaper Publishers Association (então ANPA, agora NAA), Fédération Internationale des Editeurs de Journaux (agora WAN-IFRA) e a Agência de notícias da América do Norte (um comitê conjunto da Associated Press, Canadian Press e United Press International) para salvaguardar os interesses de telecomunicações da imprensa mundial⁴⁰.

Estamos falando de uma era antes da internet, quando a comunicação e a transmissão de fotografias eram feitas via linhas telefônicas com aparelhos de telefoto. Nesse interessante artigo do site PetaPixel⁴¹, temos uma visão de como era

a transmissão de fotografias de imprensa com um aparelho da UPI – United Press International. Uma fotografia em preto e branco roda nesse dispositivo, e os dados são enviados por linha telefônica e impressas por outra máquina em um papel especial, que encontramos em muitos arquivos de jornais. Lembrando também que o fotógrafo tinha que revelar o filme e fazer uma ampliação em laboratórios portáteis no intuito de ter uma cópia para a transmissão. Na transmissão, num cabeçalho da fotografia, em uma parte do papel, estarão os dados (metadados) sobre a imagem: autor, data, assunto, descrição, que são as origens analógicas desse conceito e o surgimento do IPTC. Temos também a tecnologia da telefonia nesta época, a dificuldade e instabilidade da realização de ligações internacionais e o custo alto do tempo necessário para transmitir uma fotografia.

No início da década de 1990, o IPTC definiu uma estrutura de arquivo e metadados aplicados a texto, imagens e outros tipos de mídia para criar um padrão de dados que facilitasse a troca internacional entre jornais e agências de notícias – o *Information Interchange Model* (IIM). Esse modelo foi suplantado pela plataforma XMP com *IPTC Core e Extension*, implantado em 2014⁴².

Em 1992, o cientista da computação Tim Berners-Lee estava trabalhando em um pequeno projeto que chamou de *World Wide Web*. Ele pediu ao grupo de rock/humor, *Les Horribles Cernettes*, formado por funcionárias da Organização Europeia de Pesquisa Nuclear da Suíça (ou CERN), uma imagem para contribuir com o seu projeto. Esta foi a primeira fotografia a ser colocada na *World Wide Web*, abrindo a era em que vivemos hoje.



Figura 72. Primeira fotografia a circular na *World Wide Web*, 1992. Fotografia de Silvano de Gennaro.

Plataforma XMP⁴³

A construção dessa plataforma volta ao conceito do cabeçalho onde podemos colocar de forma digital todos os dados sobre o objeto digital em questão.

De acordo com o site da Wikipedia sobre o histórico dessa plataforma⁴⁴, o XMP (*Extensible Metadata Platform*/ Plataforma de metadados extensível) foi introduzido pela Adobe em abril de 2001. Em 2004, a Adobe anunciou sua colaboração com o IPTC. Em março de 2005, foi lançada a especificação da versão 1.0 *IPTC Core Schema for XMP* (esquema de dados do IPTC para XMP). Um conjunto de painéis personalizados para Adobe Photoshop CS pode ser baixado do IPTC.



Figura 73. Uma imagem do autor visualizada por Adobe Bridge, que permite ver os campos EXIF, que estão dentro dessa plataforma XMP. Podemos ver todos os dados técnicos da captura disponíveis por esse equipamento.

Ao abrir esta imagem TEST.tif com um leitor de texto (*TextEdit* no macOS e *Notepad* no Windows), podemos ver um arquivo em cujo início/começo, que conhecemos como o cabeçalho do arquivo, estão armazenados todos os metadados escritos em formato de linguagem XML. XML é uma linguagem muito conhecida e aberta, e segue o formato estruturado de *tags*, onde temos a abertura e o fechamento do *tag* no formato **<Nome do Tag> Texto referente ao conteúdo desse tag </Nome do Tag>**. Nesse exemplo a seguir, na última linha que a captura de tela exibe, temos: <photoshop:ICCProfile>Adobe RGB (1988)</photoshop:ICCProfile>. Vemos também outras informações, como o número de série dessa câmera: 2221203805.

Como exercício, é possível verificar que, ao serem removidos qualquer letra ou símbolo destas *strings* que representam a imagem, e ao se salvar o arquivo de texto, ele já estará corrompido. Ao abri-lo com editor de imagens, se for possível, o arquivo já irá apresentar defeitos imprevisíveis de visualização.

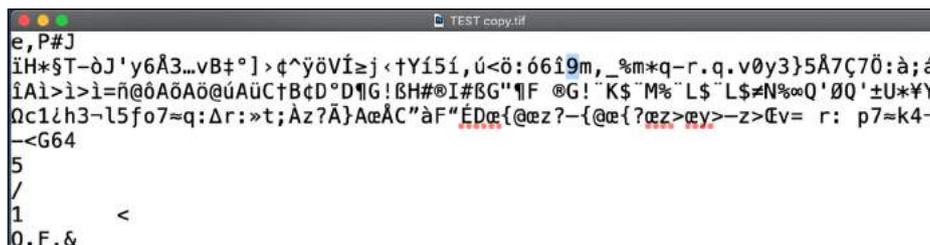


Figura 76. Exemplo de como a área de imagem é representada por um processador de texto.

No meio dessa *string* foi removido o caractere 9. O arquivo foi salvo e, ao ser novamente aberto em editor de imagens, temos a versão a seguir, já com sinais de corrupção.



Figura 77. Exemplo de imagem aberta de arquivo corrompido.

Os resultados são totalmente imprevisíveis, como vemos na imagem anterior, e isso possibilita também uma interessante forma de manipulação aleatória de desconstrução e sem controle no nível de estrutura da imagem digital. Esta experiência demonstra também a fragilidade dos dados digitais. Materiais fílmicos (negativos ou diapositivos) ou cópias fotográficas em papel vão se transformando e degradando lentamente através do tempo. Arquivos digitais, se preservados, manterão suas características sem alterações, dentro de planos de preservação e migração desses dados. Porém, se somente um *bit* falhar (*bit rot*), podemos ter uma ruptura e impossibilidade de leitura da imagem. Em um disco rígido magnético, um ímã pode inverter (*bit flip*) e transformar o valor de 0 em 1, por exemplo. Em inglês é utilizada uma expressão de *brick wall failure* (falha da parede de tijolo). Isso significa que a imagem existe, sem degradação, mas quando há degradação, ela é súbita, a falha é total e instantânea.

Campos IPTC

Os campos são divididos em IPTC CORE e IPTC EXTENSION. O ícone do lápis indica a possibilidade de edição e inserção de dados referentes a cada tópico. No exemplo a seguir, alguns campos foram preenchidos com textos explicativos.

The image shows a screenshot of IPTC metadata fields, divided into two columns: IPTC Core and IPTC Extension. Each field has a pencil icon to its right, indicating it is editable. The IPTC Core section includes fields like Creator, Job Title, Address, City, State/Province, Postal Code, Country, Email(s), Website(s), Headline, Description, Keywords, IPTC Subject Code, Description Writer, Date Created, Intellectual Genre, IPTC Scene Code, Sublocation, City, State/Province, Country, ISO Country Code, Title, Job Identifier, Instructions, Credit Line, Source, Copyright Notice, Copyright Status, and Rights Usage Terms. The IPTC Extension section includes fields like Person Shown, Location in which was created, Location Shown, Name of Featured, Code of Featured, Event, Artwork or Object in the Image, Artwork or Object, Title, Date created, Creator, Source, Source Inventory number, Copyright notice, Artwork or Object, Additional Model Info, Model Age, Minor Model Age, Model Release Status, Model Release Identifier, Image Supplier, Supplier's Image ID, Registry Entry, Max. Available Height, Max. Available Width, Type of Source, Image Creator, Copyright Owner, Licensor, Property Release Status, and Property Release ID.

Figura 78. Esses campos IPTC Core e Extension podem ser preenchidos de forma individual, ou em aplicativos como Bridge (gratuito) e Lightroom. Centenas de imagens podem ser selecionadas, e a informação que é igual para todas elas pode ser inserida somente uma vez, em lote. Também é possível criar modelos (*templates*) para que uma mesma informação possa ser inserida sempre de forma igual. Um modelo para um projeto pode conter todos os dados principais que servem para todos os projetos.



Figura 79. Exemplo de fotografia em tons de cinza. Fotografia de Marcos Issa.

IPTC Core	
Creator	Agencia Argosfoto
Creator: Job Title	
Creator: Address	
Creator: City	Sao Paulo
Creator: State/Province	SP
Creator: Postal Code	
Creator: Country	Brazil
Creator: Phone(s)	55 11 38324600
Creator: Email(s)	agencia@argosfoto.com.br
Creator: Website(s)	www.argosfoto.com.br
Headline	URCA
Description	A Praia Vermelha, na Urca e o Pao de Acucar/ Red Beach in Urca, with the Sugarloaf Mountain, traditional postcard of Rio de Janeiro. "B&W, Brasil, Brazil, CATEGORIES, CVKC, P&B, Rocha, Rock, SOLIDAO, South America, Sugarloaf Mountain, WORLD REGIONS & COUNTRIES, black & white, black and white, bonde, bondinho, brasil, cable car, ceu, granite, granito, land transportation, loneliness, lonely, mar, montanha, morro, mountain, one man only, pao de, pao de acucar, pao-de-acucar, pedra, postal, praia vermelha, preto & branco, quartz, quartzo, red beach, rio, rio de janeiro, sky, soliatrío, sugar loaf, sugar loaf mountain, sugarloaf mountain, teleferico, transportation, um homem,
Keywords	
IPTC Subject Code	
Description Writer	MIS
Date Created	22/05/09
Intellectual Genre	
IPTC Scene Code	
Sublocation	
City	
State/Province	
Country	
ISO Country Code	
Title	Praia Vermelha, Pao de Acucar / Sugarloaf Mountain
Job Identifier	
Instructions	
Credit Line	
Source	Argosfoto
Copyright Notice	© 2009 Marcos Issa/Argosfoto
Copyright Status	Copyrighted
Rights Usage Terms	All Rights Reserved - Todos os Direitos Reservados

Figura 80. Campos preenchidos do IPTC Core da fotografia acima de Marcos Issa, com metadados embutidos, e muitas palavras-chave selecionadas.

Palavras-chave/ Keywords

Dentro da plataforma XMP, temos a opção de criar palavras-chave. Esse não é um campo de texto livre. Criamos uma estrutura de conceitos e ao clicar nestas palavras-chave, com base em um vocabulário controlado para evitar de chamar a mesma coisa com nomes diferentes, ou com grafias diferentes. Em uma coleção, isso é fundamental para se ter certeza de que o nome de um autor é sempre escrito da mesma forma, ou se usamos mar ou oceano, por exemplo. Cada tipo de coleção e/ou instituição deve construir a sua base de vocabulário controlado. David Riecks, ativo no grupo do IPTC, criou um site chamado *Controlled Vocabulary*⁴⁵ (Vocabulário controlado) onde ele traz muitas referências discutindo esse assunto. Antra Silova tem um texto⁴⁶ muito claro sobre como criar o seu próprio vocabulário controlado. Por fim, é importante citar o trabalho que o Getty Research Institute tem feito com seus vocabulários⁴⁷.

Metadados automáticos

Uma câmera digital ou um celular produzirá informações sobre como a imagem foi produzida.

Estas informações estarão presentes no *Exchangeable Image File Format*, mais conhecido como apenas EXIF. Esta é uma especificação que permite que todos os fabricantes de câmeras digitais, celulares ou qualquer outro aparelho com câmera possam gravar todas as informações técnicas utilizadas nessa captura. Esses dados estarão preservados nos campos de metadados EXIF.

Esse é um campo que está se ampliando à medida que mais sensores estão sendo embutidos nos diversos dispositivos de captura de imagens. O avanço de GPS (*Global Positioning System* – Sistema de Posicionamento Global) em câmeras digitais, mas principalmente em praticamente todos os celulares, permite que uma fotografia tenha esta informação de geolocalização. No futuro próximo poderemos ter outros sensores, como altitude, temperatura, umidade relativa do ar, entre outros.

Com o avanço das tecnologias de reconhecimento facial e outras de reconhecimento de imagem por meio de ML (*machine learning*/aprendizado de máquina), começa a ficar mais comum que sistemas passem também a identificar e *taguear*

(do inglês *tag*, como uma etiqueta, adicionar dados/metadados) automaticamente conteúdo nas imagens.



Figura 81. Em um estudo de pesquisadores da Microsoft sobre o desempenho em compreender imagens com o uso de algoritmos e aprendizado de máquina na classificação de imagens da ImageNet⁴⁸, temos a referência do *GT* – *Ground Truth* (a base de verdade da pesquisa) para cada imagem, e as diversas classificações realizadas por esse grupo de pesquisadores. Em 7 de 9 casos, o sistema automatizado conseguiu chegar a um resultado de conteúdo idêntico ao *Ground Truth*.

Os algoritmos não são neutros, como também os conjuntos de imagens utilizados para estas pesquisas não são necessariamente (para não dizer “nada”) inclusivas de todas as realidades sociais, culturais e estéticas. Portanto, existe um movimento emergente discutindo os vieses dos algoritmos, em muitos casos escritos por homens brancos de meia-idade de classe social mais elevada.

Joy Buolamwini é uma ativista e criou a Liga da Justiça Algorítmica⁴⁹. Ela discute (no vídeo *Gender Shades*⁵⁰) as questões dos tons de gênero e dá uma boa visualização sobre a questão do viés nos algoritmos. Em 2019, a revista *Fortune Magazine* referenciou a Joy como “a consciência da revolução de Inteligência Artificial” na sua lista de grandes líderes mundiais. Que venham cada vez mais essas consciências.

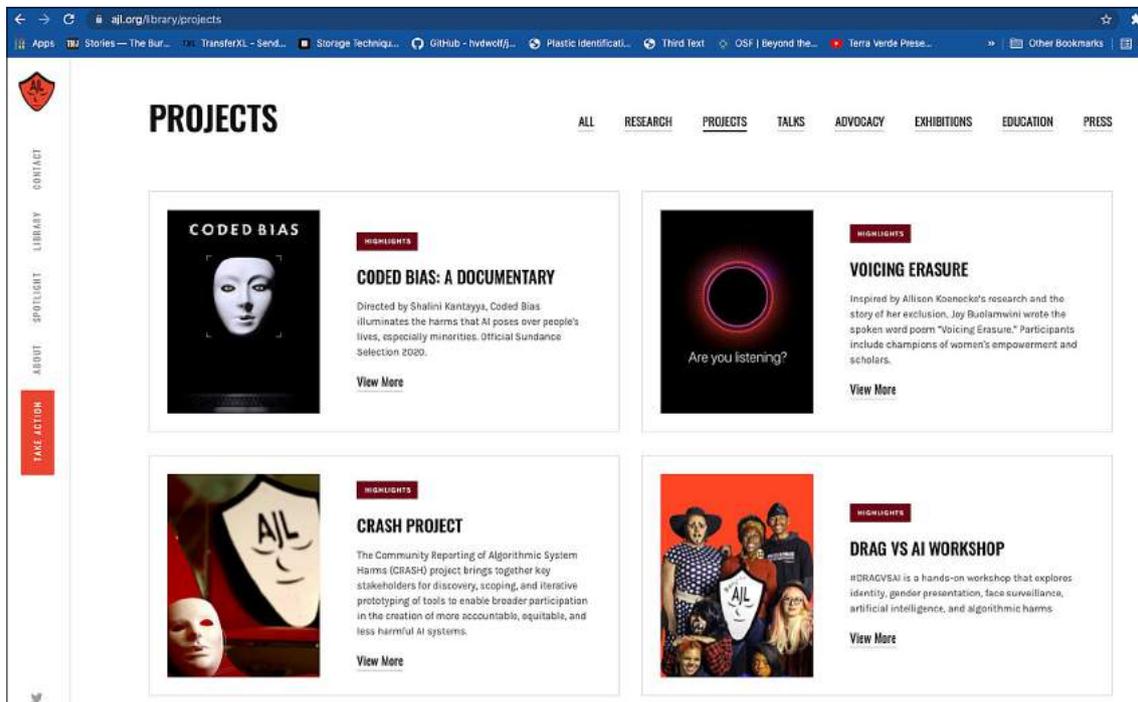


Figura 82. Página de Projetos do site da Liga da Justiça Algorítmica: <https://www.ajl.org/library/projects>.

Conteúdo e contexto

Como a produção mundial ultrapassa 1 trilhão de imagens por ano, fica mais fácil entender que não será possível uma pessoa identificar e descrever cada uma destas imagens. É importante frisar que *conteúdo* e *contexto* são questões muito distintas. Uma imagem como a que vemos a seguir pode ser descrita por uma pessoa ou por sistemas, mas até que ponto saberemos, com qualquer um dos dois métodos, o que esta imagem realmente é e o que significa? O conteúdo pode ser sobre uma fotografia aérea, nuvens, oceano, talvez uma área que pareça ser de areia.



Figura 83. Fotografia da Terra vista do espaço.

O contexto muda ao adicionar mais informações: *NASA, Imagem do espaço, 1965*. Fica ainda mais completo e interessante quando entendemos o que de fato a imagem representa. Esta fotografia do Estreito da Flórida e do Grande Banco das Bahamas foi tirada durante a missão Gemini IV, na órbita nº 19 de 64 órbitas, em 4 de junho de 1965. A tripulação do Gemini IV conduziu experimentos científicos, incluindo fotografias do clima e do terreno da Terra, para o restante de sua missão de quatro dias após a primeira caminhada espacial histórica de Ed White, em 3 de junho. A foto foi feita com uma câmera Hasselblad 500C modificada, lente de 80 mm e suporte de filme de 70 mm. Na missão Gemini IV, cinco carregadores, cada um carregado com aproximadamente 60 exposições de filme Ektachrome MS (SO-217) de 70 mm em uma base estar 2 mil, com sensibilidade de ASA 64, foram transportados. Esses dados são provenientes do Relatório do Gemini IV.

Metadados e acessibilidade

A inclusão e a acessibilidade são assuntos fundamentais para a justiça social. Em outubro de 2021, IPTC anunciou⁵¹ novas propriedades nos metadados das fotografias para torná-las mais acessíveis. Conforme esse anúncio:

Incorporar descrições de imagens para acessibilidade nos metadados das fotos promete ser uma virada de jogo, possibilitando que softwares e sistemas forneçam rotineiramente texto alternativo (*Alternate text* ou *alt text*) com imagens, dando assim ao software leitor de tela a capacidade de ajudar os leitores a visualizar e ouvir as descrições das imagens como elas são lidas em voz alta. Sem descrições acessíveis, as imagens são silenciosas para os milhões que dependem de leitores de tela para acessar totalmente a web.

Metadados inseridos manualmente

A inserção de metadados é uma etapa fundamental para todo o processo de preservação digital, porque não poderemos recuperar as informações e os objetos digitais do nosso acervo se eles não estiverem indexados e catalogados. Da mesma maneira que é feita uma busca na internet para recuperar informações, deixamos de ver muitas referências possíveis por não estarem indexadas, que passam a não existir, mesmo existindo em algum lugar como *dark data*, inacessível.

“Inserido manualmente” significa que uma pessoa qualificada está atuando nesse processo de indexação dessa informação. Um sistema automático pode reconhecer o conteúdo de uma imagem (praia, cidade, pessoas, cachorros, gatos etc.), mas não consegue entender o contexto e significado da imagem, que é derivada do conhecimento do trabalho, do autor, e dos contextos que rodeiam a produção de imagens que estamos indexando.

Com a evolução de *machine learning* a tendência será ter cada vez mais sistemas com a capacidade de ir além da descrição do conteúdo e criar conexões, gerar significados. O tempo dirá como será a evolução disto.

A inserção manual não significa que teremos que indexar cada imagem individualmente. Muitos conjuntos podem ser indexados em lotes, mas são os catálogos que podem e devem conter descrições mais detalhadas sobre os contextos destas imagens. De onde vieram, como chegaram ao acervo, quais são as histórias

relacionadas com estas imagens e quais são os direitos que precisam ser observados, entre outras informações. O aumento de dados sobre uma imagem pode ajudar a facilitar a sua recuperação. A questão que mais nos assombra é sobre quanto tempo queremos ou podemos investir em termos de fazer esse processo de descrição, que poderá também envolver muito tempo de pesquisa.

Mais Produto, Menos Processo/ More Product Less Process – MPLP⁵²

Dois arquivistas norte-americanos, Mark Greene e Dennis Meissner, escreveram sobre esta visão em 2005, pensando principalmente sobre o enorme volume de materiais para processar nos arquivos. Como dinamizar os processos em arquivos para conseguir tornar materiais cada vez mais acessíveis aos usuários e pesquisadores com processos mais enxutos de arranjo, embalagem, catalogação e digitalização? Há um excelente texto, que vale ler para ter um entendimento claro sobre a aplicação prática dessa visão de MPLP, que foi realizada em 2012 pelas Bibliotecas da Universidade da Califórnia⁵³. Na visão dos autores, podemos criar níveis diferentes de processamento de acordo com decisões feitas em todas as áreas de processamento, dedicando mais tempo para as coleções cuja descrição desejamos aprofundar, e descrições em lotes, mais genéricas, para outros materiais.

Voltando para a imagem da NASA sobre a qual discutimos, serão sempre decisões que nos levarão a entender o grau de processamento das coleções, no entendimento de que o processo de descrição, especificamente, pode ser sempre retomado, ampliado e aprofundado conforme o desejo ou a demanda. Teremos níveis mais aprofundados de catalogação para materiais que consideramos mais relevantes, por exemplo.

Metadados embutidos, em base de dados, ou ambos

Os metadados podem estar em uma base de dados simples (como Excel) ou em sistemas de *software* que gerenciam esses dados. Adobe Bridge, é um ótimo *software* gratuito que permite visualizar as imagens em suas pastas, acrescentar metadados e palavras-chave, e renomear os arquivos. As informações estarão

embutidas nas imagens, exceto para as imagens RAW de câmera, que são proprietárias. Qualquer adição de metadados irá gerar um arquivo *sidecar* XMP, que conterá estas informações.

Padrões de metadados

Existem inúmeros padrões de metadados que são utilizados em coleções, além das fotografias. O Dublin Core, um dos mais conhecidos, foi apresentado em 1995 em uma conferência na cidade de Dublin, Ohio, EUA. Esse padrão tem 15 elementos que procuram descrever as principais questões sobre uma imagem. Muitos outros padrões foram desenvolvidos desde esse período para atender a tipos de coleções específicas.

Jenn Riley⁵⁴ fez um excelente estudo, comparando 105 padrões de metadados, que vale a pena ver e pesquisar. O que entendemos hoje como fundamental é que os dados, que são os nossos tesouros, precisam ser construídos de tal forma que possam migrar de um padrão a outro, de um sistema a outro, e que não estejam presos dentro de uma estrutura.

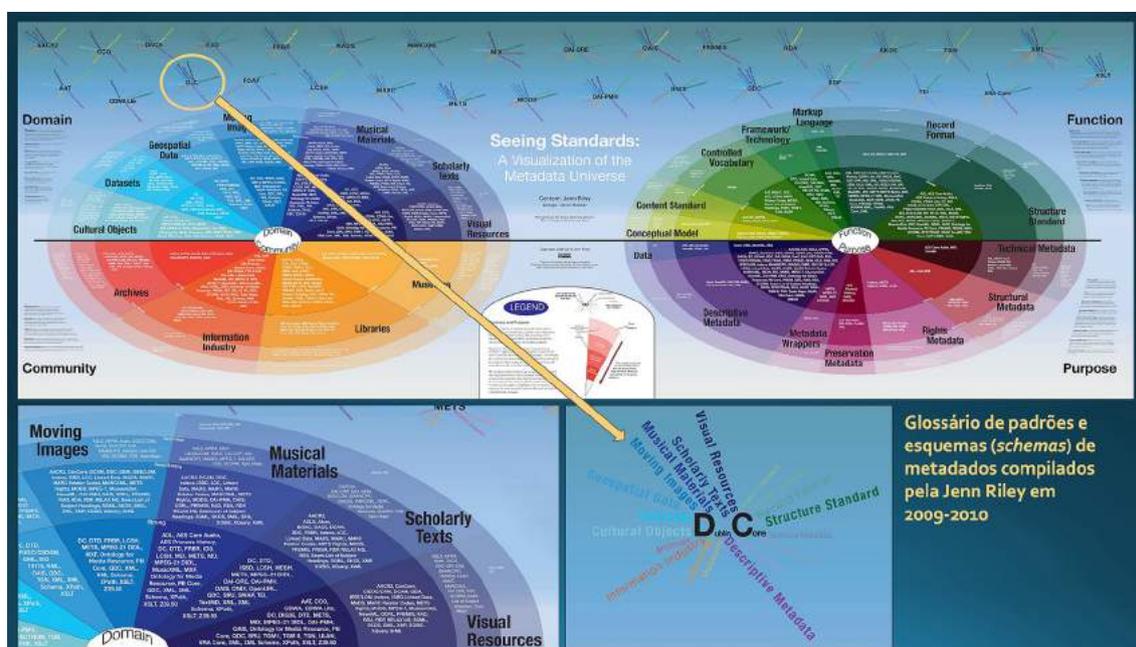


Figura 84. Vendo padrões: uma visualização do universo de metadados, Jenn Riley, 2009-2010. Com destaque o padrão Dublin Core.

O educador e cientista de computação, Andy S. Tanenbaum fez um ótimo comentário sobre padrões: “A melhor coisa sobre os padrões é que existem tantos para escolher”.

Outra questão fundamental com tantos padrões é a possibilidade da interoperabilidade entre eles, para que os dados/metadados possam migrar de um sistema a outro. Em inglês, usamos a expressão “*crosswalk*” (uma faixa de pedestre), que ilustra a ideia de um caminho (um *script* de programação) que faz o *link* entre um sistema e outro.

Organização de dados atuais

Construção de sistemas de backup e fluxos de trabalho

Para qualquer sistema de organização de dados digitais, teremos que ter uma forma de sincronizar os dados da matriz entre os outros diversos dispositivos de guarda. A definição clara desses dispositivos tem que ser feita para na sequência implementar os processos de *backup*.

Separar os **arquivos de trabalho**, que estão em constante mudança, e criar um fluxo específico de *backup* para esses arquivos deve ocorrer separado de um fluxo de *backup* para os **arquivos permanentes**.

Os melhores planos de proteção de dados envolvem uma estratégia sólida de *backup* dos arquivos de trabalho diários (*hot data* ou dados quentes) e uma estratégia de arquivamento dos arquivos permanentes (*cold data*, ou dados frios, porque podem e devem ser armazenados *off-line*, sem energia/desconectados).

Por exemplo, se você faz *backup* dos arquivos de trabalho e permanentes, os seus servidores ativos estarão preenchidos com dados permanentes, que não devem ser modificados, e o seu sistema de *backup* irá fazer cópia desses dados de forma desnecessária e repetidamente – tornando todos os seus *backups* muito maiores do que o necessário.

No site do *dpBestflow*, na seção sobre *backups*⁵⁵, temos estas duas imagens que exemplificam o processo de arquivos de trabalho e os permanentes.

Backups de Arquivos de Trabalho

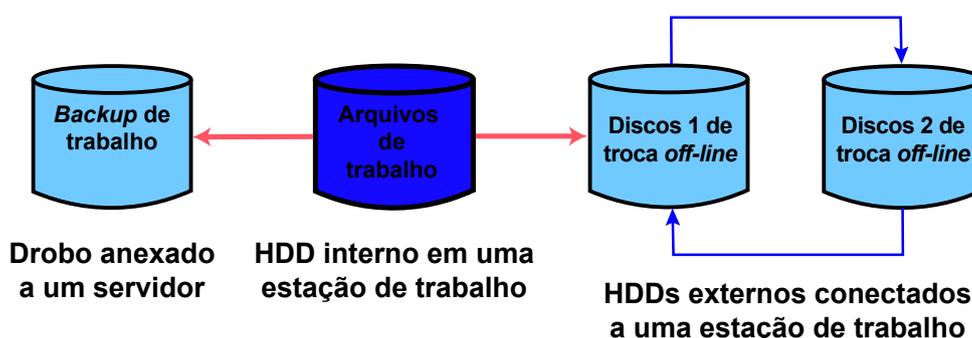


Figura 85. A imagem central, azul-escuro representa o HDD interno, e é a matriz. O *backup* para o HDD à esquerda é feito em um DAS, conectado diretamente ao servidor. A atualização desse *backup* é automática. Os HDDs à direita são conectados de noite para serem sincronizados com todo o trabalho realizado durante o dia. Nesse exemplo, discos de troca (*swapper*) são substituídos *off-site* todos os dias, assim, sempre um disco estará *off-site*. Esses dois discos estão *off-line* e entram *on-line* um de cada vez para fazer a atualização. Ao trazer o Swapper 1 para *backup* na segunda-feira, o Swapper 2 continua *off-site* e *off-line*. Na terça, será a vez do Swapper 2.

Armazenamento Permanente e Backups

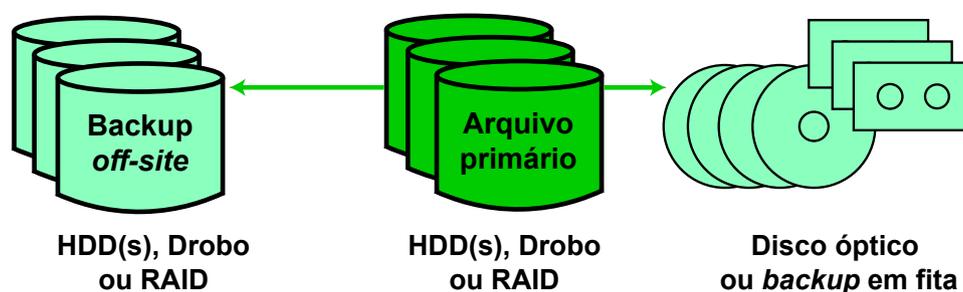


Figura 86. Em uma outra configuração, com outros discos, é construída uma estrutura de arquivamento. O verde-escuro no centro é o sistema de arquivamento primário e permanente, como um RAID da Drobo. À esquerda, há um outro sistema RAID, com a mesma capacidade para fazer o *backup* dessa matriz de arquivamento. À direita, há um *backup* em fita LTO, ou mesmo em nuvem. Todos esses sistemas podem estar *off-line* e entram *on-line* somente para serem atualizados. Esta atualização é mais lenta do que a diária dos arquivos de trabalho.

Nesses dois exemplos anteriores, vemos a necessidade de fazer um desenho do seu sistema e nomear de forma clara o que são cada um dos dispositivos. Esses nomes precisam permitir o entendimento do que cada unidade faz e do que

representa dentro do fluxo de trabalho, que veremos mais adiante. Separar o trabalho do dia a dia do trabalho de arquivamento é uma etapa importante da organização de seus acervos.

Como vamos discutir em armazenamento, podemos usar *softwares* mais simples para gerenciar o *backup* de um acervo digital pessoal ou de uma pequena instituição. À medida que o sistema vai ficando mais complexo, com servidores e centros de dados, existirão *softwares* muito especializados, que precisam ser gerenciados por um *expert* para realizar as tarefas de *backup*, sincronização e verificação de *checksum* dos arquivos.

Neste texto, estamos pensando em soluções de *softwares* simples para indivíduos e pequenas instituições. As principais funções de qualquer tipo de *software* de *backup* passarão por conceitos que observamos a seguir, visto somente da Esquerda para a Direita (mas o inverso também pode ser escolhido, da Direita para a Esquerda) e de modo Bidirecional. Veja com cuidado as definições/regras que cada *software* aponta de como fará o trabalho de cópia. Como dito antes, crie pastas de teste para entender com clareza como é o processo de *backup*.



Backup Esquerda para a Direita

A origem da esquerda será monitorada para alterações e comparada com o destino da direita. Todas as alterações detectadas na origem serão copiadas para o destino. Nota: na primeira vez em que uma operação de *backup* da esquerda para a direita for executada, os arquivos existentes no destino, mas ausentes na origem, **não serão** excluídos do destino.



Sincronização Bidirecional

Ambos os alvos serão monitorados para mudanças. Todas as alterações detectadas serão copiadas para o destino oposto. Observe que na primeira vez que uma sincronização bidirecional é executada, os arquivos existentes em um destino e não no outro **não serão** excluídos do outro destino.



Espelho Esquerda para a Direita

Os arquivos serão copiados apenas da origem ao destino. Se um arquivo na origem/esquerda ou destino/direito for modificado, isso resultará na substituição do arquivo de destino pelo arquivo de origem. Nota: a primeira vez que uma operação *espelhar* da esquerda para a direita for executada, os arquivos existentes no destino, mas ausentes na origem, **serão** excluídos do destino.

Figura 87. Quadro de sinalizações de *softwares* sobre direção de funções de backup.

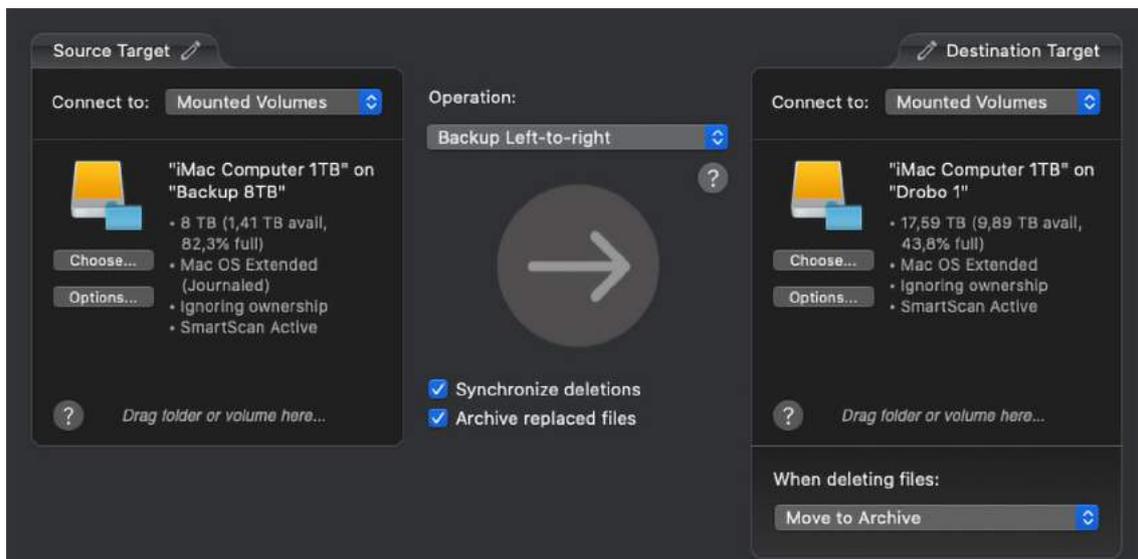


Figura 88. Nesta captura de tela de um *software* de *backup*, vemos a opção de **Backup da Esquerda para a Direita** selecionado. Duas opções estão selecionadas: *Sincronizar os arquivos apagados* (*Synchronize deletions*) apagará arquivos do destino (Direita) se o arquivo na origem (Esquerda) for apagado. Se não se clicar nesta opção ao apagar um arquivo da origem, ele permanecerá no destino até que seja feito um *backup* de *espelho*, e, nesse momento, como o arquivo não existe na origem, ele será apagado do destino. Faça testes com pastas de estudo para entender claramente esses princípios.

Programação do backup

O *backup* pode ser feito manualmente de tempos em tempos, mas corremos o risco de não nos lembrarmos de fazer isso, e infelizmente isso acontece bastante. Portanto, o ideal é programar o *backup* para acontecer de forma automática e durante a sua hora de trabalho, ou quando tudo estiver ligado, e os dispositivos de origem e destino estiverem conectados ao computador.

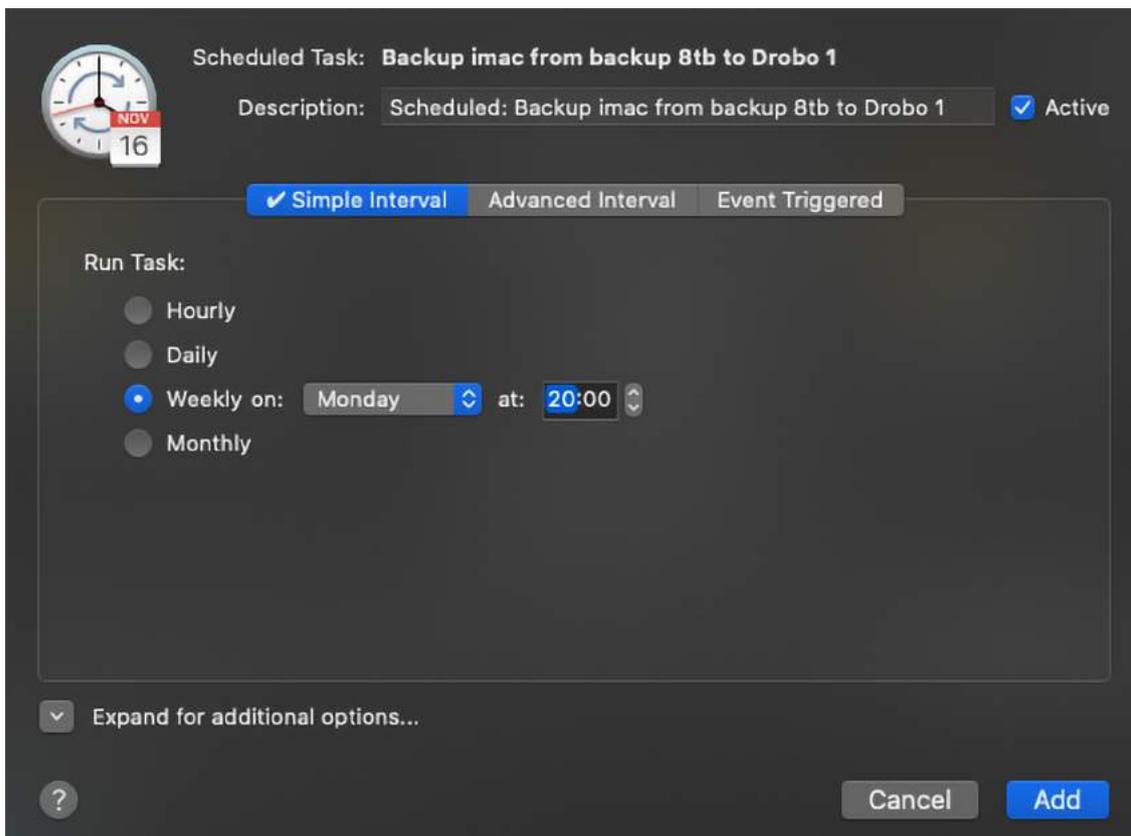


Figura 89. Todos os programas de *backup* oferecem a opção de uma programação. O *backup* pode ser contínuo – sempre que algo é alterado na matriz/origem, é automaticamente copiado para o destino. O *backup* pode ser programado como vemos nesta tela, uma vez por semana, na segunda-feira, às 20h00.

Fluxos de trabalho

É importante definir o conceito do fluxo de trabalho para cada pessoa, pois dependerá do tipo de trabalho feito na captura digital e do esquema de armazenamento, processamento e *backup* utilizado. Definir o fluxo é definir os passos que são tomados. Por exemplo: ao se fotografar com uma câmera digital, o que fazer com o cartão de memória? Onde as imagens são salvas primeiro? Em que momento elas são descritas e renomeadas, e quando os metadados são inseridos? Isso é feito nos arquivos RAW? Depois dessa fase: onde é feito o *backup*? Em qual dispositivo? E assim por diante. Um bom jeito de começar a trabalhar com um fluxo é procurar desenhar em uma folha de papel todas as etapas de trabalho dos seus processos. Nesta folha estarão desenhadas e anotadas a sua câmera, o cartão de memória, o computador, os processos feitos no computador, os *backups*, cada um com um nome

relevante que o identifique, e em passo a passo, com setas que indicam as orientações destas etapas, o que vem antes e depois do que, para deixar, de forma clara, um mapa de trabalho de como tudo deve ser feito. Um *workflow* perfeito deve ser claro e bem escrito, e desenhado a ponto de poder ser seguido por outra pessoa, como um manual de instruções. Esse, inclusive, é um ótimo exercício. Mostre o seu fluxo de trabalho e peça a uma outra pessoa para descrever o que ocorre nele. Se esta pessoa começar a fazer perguntas, será sinal de que não está claro, e é uma chance de fazer correções e adaptações ao fluxo de trabalho. Esse processo deve ser sempre revisitado, em um processo de teoria (o seu desenho) e de prática (o que de fato faz). Com o tempo, o seu desenho passa a ser adaptado a melhores práticas de trabalho, que irão coincidir com o que você faz.



Figura 90. Esta imagem é de uma busca na internet com as palavras *digital photography workflow*, com a visualização das imagens.

Ao analisar inúmeros esquemas de *workflows* diversos, publicados por diversas pessoas, você começará a perceber as falhas que cada um tem. Alguns *workflows* são muito simples, sem detalhes, outros não muito claros na sequência de eventos, e ainda, outros tão complexos que não dá para entender como funcionam. Vale a pena fazer esse exercício de observar *workflows* diferentes e aplicar o seu aprendizado no seu próprio *workflow*.

5

Onde vamos guardar

Guardar e preservar

Diferentemente de guardar negativos, *slides* e cópias em papel, como nos processos fotográficos tradicionais, em que cuidamos de objetos, as nossas fotografias digitais são imagens codificadas em código binário, zeros e uns, interpretadas por um sistema de *software* e armazenadas em algum tipo de mídia conectada a um dispositivo de visualização como um computador, um celular etc.

Portanto, as questões da guarda das fotografias digitais passam por todas as questões de guarda de dados digitais. As nossas fotografias são esses dados tanto quanto outros objetos digitais, como arquivos de vídeo, som, documentos, entre outros. A diferença está nos sistemas de *software* que fazem a construção e a consequente interpretação desses dados. Isso significa que nossas imagens só existem dentro de um ecossistema vinculado a toda essa infraestrutura. Precisamos de um sistema de *software* para construir e interpretar os dados binários. Esse *software* precisa também de uma plataforma, ou sistema operacional, onde ele esteja inserido e operante. Os dados precisam ser armazenados e preservados em algum tipo de mídia que armazene dados digitais.

Nesse sentido, a ideia de **guardar** precisa ser ampliada para um conceito de **preservar**. Eu posso guardar uma coleção de imagens nos formatos de arquivos TIFF e JPEG, por exemplo, em um disco rígido, e colocá-lo em uma prateleira por décadas. Para recuperar estas imagens depois, vou precisar conectar esse disco rígido em algum sistema computacional que possa ler esse disco via algum tipo de conexão compatível com ele. Se isso for bem-sucedido, ainda terei que ter um *software* que possa ler e interpretar esses dados. Por isso, preservar é o melhor conceito que podemos ter. A preservação representa um processo contínuo de monitoramento dos sistemas de guarda e dos formatos de arquivo, na garantia de que estão ainda íntegros e de que ainda existem sistemas de *software* e *hardware* existentes para sua leitura/interpretação. Não podemos armazenar e ignorar (traduzido da frase em inglês *store and ignore*) as nossas fotografias digitais.

Certamente, em algum momento as nossas imagens não serão mais baseadas em *pixels*, e teremos uma fase de transição de migração de imagens construídas com esta tecnologia de *pixels* quadrados para a próxima tecnologia de imagem, construída com outros sistemas de *software* e *hardware*. Temos que estar preparados para estas migrações/transições.

Este texto não é sobre preservação digital, um assunto muito mais amplo, mas temos que adentrar nesse universo um pouco para entender o que significa preservar em médio e longo prazo as nossas imagens.

Na tabela criada pelo NDSA (*National Digital Stewardship Alliance*)⁵⁶ sobre os níveis de preservação digital, temos cinco áreas funcionais: Armazenamento, Integridade, Controle, Metadados e Conteúdo. Estas áreas são divididas em quatro níveis de aprofundamento. O nível mais básico é o Nível 1 – conheça o seu conteúdo. Veja a Tabela 3 com a tradução em português⁵⁷ por Laura Rezende. Nesta coluna, temos importantes informações, sobre como ter duas cópias completas em localizações distintas, mas, no nível 4 de armazenamento já se diz ser necessário ter 3 cópias em localizações geográficas com diferentes ameaças de desastres. Não podemos ter cópias em dois locais diferentes, ambos com riscos de enchentes, por exemplo.

Níveis de Preservação Digital (NDSA, 2019)

<https://osf.io/OGZ98/>

TRADUÇÃO DE LAURA VILELA RODRIGUES REZENDE

Área funcional	Nível 1 (conheça o seu conteúdo)	Nível 2 (proteja o seu conteúdo)	Nível 3 (supervisione o seu conteúdo)	Nível 4 (sustente o seu conteúdo)
Armazenamento	Ter duas cópias completas em localização distinta.	Ter três cópias completas, com no mínimo uma cópia em localização geográfica distinta.	Ter no mínimo uma cópia em uma localização geográfica com riscos de desastres distintos em relação ao local das outras cópias.	Ter no mínimo três cópias em localizações geográficas com diferentes ameaças de desastres.
	Documentar todos os suportes de armazenamento do conteúdo.	Documentar seu(s) sistema(s) e suportes de armazenamento, indicando os recursos e dependências requeridas para seu funcionamento.	Ter no mínimo uma cópia em um tipo de suporte de armazenamento distinto.	Maximizar a diversidade de armazenamento, visando evitar pontos concretos de falha.
	Inserir o conteúdo em um suporte de armazenamento estável.		Monitorar a obsolescência de armazenamento e suportes.	Ter um plano e executar ações para tratar da obsolescência de <i>hardware</i> , <i>software</i> e suportes de armazenamento.
Integridade	Verificar a informação de integridade caso tenha sido disponibilizada juntamente com o conteúdo.	Verificar informação de integridade ao mover ou copiar conteúdo.	Verificar informação de integridade de conteúdo em intervalos fixos.	Verificar a informação de integridade em resposta a situações ou atividades específicas.
	Gerar a informação de integridade caso não tenha sido disponibilizada junto com o conteúdo.	Usar dispositivos com escrita bloqueada quando se utilizam suportes originais.	Documentar processos de verificação de informação de integridade e seus resultados.	Substituir ou corrigir conteúdos corrompidos quando necessário.
	Escanear vírus em todo o conteúdo; isolar o conteúdo em quarentena caso seja necessário.	Fazer um <i>backup</i> da informação de integridade e armazenar a cópia em uma localização distinta do conteúdo.	Realizar auditorias de informação de integridade sob demanda.	
Controle	Determinar os agentes humanos e de <i>software</i> que devem ter autorização para ler, gravar, mover e eliminar conteúdo.	Documentar e pôr em prática quais os agentes humanos e de <i>softwares</i> estão autorizados a ler, gravar, mover e eliminar conteúdo.	Manter registros de ações (logs) e identificar quais os agentes humanos e de <i>software</i> realizaram ações junto ao conteúdo.	Realizar revisões periódicas de ações/registros de acesso.
Metadados	Criar inventário de conteúdo de documentar sua localização atual de armazenamento.	Armazenar metadados suficientes, visando identificar o conteúdo (poderiam estar incluídas combinações de metadados administrativos, técnicos, descritivos, de preservação e estruturais).	Determinar quais padrões de metadados serão aplicados.	Registrar ações de preservação associadas ao conteúdo incluindo quando ocorrerem.
	Fazer um <i>backup</i> do inventário e armazenar no mínimo uma cópia separada do conteúdo.		Encontrar e corrigir problemas nos metadados para cumprir os padrões por eles especificados.	Implementar padrões de metadados escolhidos.
Conteúdo	Documentar os formatos de arquivos e outras características essenciais de conteúdo, incluindo quando e como foram identificadas.	Verificar formatos de arquivos e outras características essenciais de conteúdo.	Monitorar a obsolescência e as mudanças nas tecnologias das quais o conteúdo seja dependente.	Realizar migrações, normalizações, emulação e atividades similares que garantam o acesso ao conteúdo.
		Estabelecer relações/vínculos com os criadores de conteúdo, visando incentivar uma escolha sustentável de arquivos (formatos abertos).		

Tabela 3. Ilustração dos níveis de preservação digital.

Esta tabela aponta para a necessidade de: um estabelecimento dos locais de armazenamento; verificação de integridade ou fixidez para garantir que os dados digitais não tenham sido modificados por corrupção aleatória do estado magnético dos *bits* (*bit rot*); controle do acesso e possíveis alterações desse conteúdo; metadados sobre os objetos e as ações aplicadas a ele; controle do conteúdo, visando à documentação dos formatos; monitoração dos riscos de obsolescência; e planejamento e realização de projetos de migração, normalização (migrar formatos de arquivos com risco para formatos de arquivos considerados mais estáveis), ou emulação (um processo em que um *software* ou dispositivo permite a um sistema imitar o funcionamento de outro sistema obsoleto ou em obsolescência), para garantir acesso ao conteúdo.

Todo sistema de armazenamento irá falhar eventualmente

Assumir que tudo falhará em algum momento é uma das características fundamentais da preservação digital. A falha de um dos nossos sistemas e a perda de uma das cópias não deverá ser um desastre, mas simplesmente algo que irá requerer algum trabalho para fazer a recuperação dos dados provenientes de uma outra cópia.

Um bom exercício é olhar para todos os seus sistemas de guarda (*desktop*, *notebooks*, discos rígidos diversos etc.) e imaginar a perda de um desses sistemas. Se isso lhe causar terror completo pela perda de dados, o seu sistema definitivamente não é seguro.

Portanto, o aprendizado mais importante de todos aqui é compreender que não existe uma marca ou um sistema melhor que o outro. Sim, são diferentes, e podem ter vantagens e desvantagens, e vale a pena estudar e pesquisar o que pode ser melhor para você em um determinado momento. Porém, esse debate do melhor é infrutífero. Todas as marcas ou sistemas falharão em algum momento. Temos que avançar o debate necessário de entender a necessidade de ter muitas cópias, que garantirá a longevidade de nossos dados.

Principais ameaças para a perda de dados

- **Falha do dispositivo** → Todos os dispositivos de armazenamento digital eventualmente falharão.
- **Vírus** → Os vírus podem se propagar silenciosamente de um dispositivo de armazenamento para outro e, em seguida, atacar para destruir os dados. O armazenamento de gravação única, como o disco óptico, oferece a melhor proteção contra vírus.
- **Software malicioso** → Seu computador ou servidor pode ser exposto a *softwares* maliciosos ou *malwares*, de *hackers* mirando em você, sua empresa ou instituição.
- **Falhas de volume e diretório** → Os sistemas de guarda possuem um “mapa” de onde os arquivos estão armazenados, semelhante a um índice. Se esse sistema for corrompido, o computador pode não conseguir localizar os arquivos em uma unidade.

- **Corrupção de transferência** → Sempre que os dados são transferidos de um dispositivo para outro, existe a possibilidade de corrupção. Isso pode ser devido a problemas com a RAM, a unidade, os conectores, a placa de ponte, a rede ou os cabos.
- **Queda de raio/ pico de tensão** → O excesso de tensão de uma queda de raio ou de um transformador da empresa de energia queimado pode fritar seu computador em um piscar de olhos.
- **Roubo** → A proteção contra roubo inclui medidas de segurança, como alarmes ou cofre, mas é mais bem obtida com armazenamento externo.
- **Danos por fogo ou água** → As imagens digitais podem ser destruídas por danos por fogo ou água.
- **Erro humano** → Uma das causas mais comuns de perda de dados é o simples erro humano. *Backups off-line* que não são atualizados imediatamente são uma parte valiosa da proteção contra erro humano.

É por isso que temos pelo menos 3 cópias do mesmo conteúdo, em dois sistemas diferentes, com uma cópia *off-line* e outra *off-site*. Nesse sentido, para uso individual, ter um sistema com HDDs mesclado com SSDs pode ser uma solução. Não estamos colocando todos os nossos ovos na mesma cesta. Os nossos sistemas terão que ser revistos com periodicidade, que varia de ciclos de em torno de 3–5 anos. O aumento da nossa produção de dados, bem como a idade dos componentes de guarda, faz com que atualizemos gradativamente (*refreshing* é termo utilizado em inglês) os nossos sistemas, com compra de novos HDDs e/ou equipamentos com maior capacidade de armazenamento.

Na preservação de fotografia analógica para longa permanência, o prazo sempre foi pensado como sendo de pelo menos 100 anos. É muito complexo planejar algo “para sempre”. Sempre entendi isso como sendo a missão de garantir que durante o nosso ciclo de vida produtiva, em torno de 50 anos, temos que cuidar desses materiais e garantir a passagem dessa tarefa para uma outra geração que fará o mesmo por mais 50 anos, e assim por diante. No meio digital, podemos pensar da mesma forma, e isso terá que ser feito nos mesmos 50 anos, mas com ciclos em torno de 5 em 5 anos. Preservamos durante esse ciclo de 5 anos, e depois mais 5 e assim por diante. É dessa forma, pensando nesses curtos prazos, que podemos planejar a preservação em longo prazo.

Armazenamento 3-2-1 e suas variantes

Um dos primeiros conceitos sobre o armazenamento digital, considerando principalmente a questão da falha dos discos rígidos, foi o de 3-2-1. Temos que ter 3 versões, uma matriz de produção e duas cópias, em 2 mídias diferentes, pensando em não depositar toda a nossa confiança em um único sistema. Esse conceito foi desenvolvido pensando em mídias magnéticas e ópticas (CD, DVD, *blu-ray*) por exemplo, e o 1 representa uma mídia *off-site*, em outro lugar geográfico. Esse modelo tem novas versões que veremos a seguir.

O conceito de 2 mídias como foi inicialmente pensado, voltado para o magnético e o óptico, é complexo de ser utilizado hoje. Guardar dados nestas mídias ópticas não é indicado, principalmente pela baixa capacidade de armazenamento desses dispositivos (4,7 GB para um DVD de camada única e 50 GB para um disco *blu-ray*). Portanto, mantendo os 2, ele representa hoje 2 sistemas diferentes, por exemplo: um disco rígido externo e uma cópia na nuvem, ou um HDD externo e um sistema de armazenamento tipo RAID NAS ou DAS.



Figura 91. A versão clássica do 3-2-1⁵⁸.

A seguir algumas opções de variantes no tema:

- **3-2-1-1** → 3 cópias, 2 sistemas diferentes, 1 cópia *off-site* e 1 cópia *off-line*. É importante ter uma cópia totalmente desconectada de redes e internet, evitando os riscos de contaminação por vírus ou *ransomware* (quando seu sistema é invadido, seus dados são criptografados e um resgate é solicitado para que sejam liberados), o que ocorre com frequência quando todas as cópias estão *on-line*. Uma cópia *off-site* (em outro lugar geográfico) pode estar também *off-line*, cumprindo assim as duas funções.
- **3-2-1-1-0** → 3 cópias, 2 sistemas diferentes, 1 cópia *off-site*, 1 cópia *off-line*, e o 0 representa zero erros de DR (*Disaster Recovery*, testes de recuperação de dados a partir dos *backups*), recuperação de dados de um *backup*. O DR é feito com simulações que demonstram a integridade da cópia de *backup*.

Estratégia 3-2-1-1-0

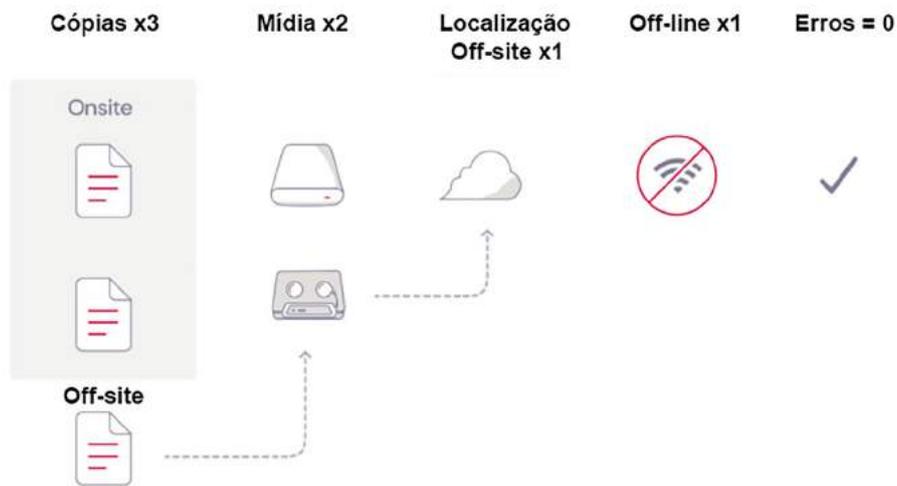


Figura 92. Uma versão mais elaborada e segura da versão 3-2-1.

- **3-1-2** → 3 cópias, 1 tipo de mídia (disco) e 2 cópias *off-site*. Pode ser uma cópia na nuvem e uma cópia em discos rígidos em outro lugar geográfico. Esta versão também poderia ser 3-1-2-1-0, se tivesse uma das cópias *off-line* e zero erros de DR.

Sobre os termos *On-line*, *Nearline*, *Off-site* e *Off-line*

- **On-line** → quando um sistema de armazenamento está conectado ao computador de forma direta ou por meio de uma rede.
- **Nearline** → quando um sistema está próximo, ao lado, mas fica desconectado do computador. Ele entra *on-line* somente para fazer cópias de *backup* e depois volta para o estado de desconectado.
- **Off-site** → quando temos um sistema de armazenamento em outro lugar geográfico, mas ele pode estar *on-line*, *nearline* ou *off-line* nesse lugar.
- **Off-line** → um sistema de armazenamento não conectado. Fitas magnéticas ficam em armazenamento *off-line* e são conectadas somente para a recuperação ou gravação de dados. Chamamos também *off-line* como *cold storage/dark storage* (armazenamento frio ou escuro), pois não dependem de energia enquanto desconectados. Um HDD externo desconectado também está *off-line*.

Um bom exercício para cada pessoa que possui dados digitais é mapear todas as suas unidades de guarda/armazenamento e entender como se classificam. Quantas cópias de fato possui? Existem cópias *off-line*? Alguma cópia em outro local geográfico? Algum teste de recuperação de dados foi realizado? Ao analisar as respostas a estas perguntas, começa a ficar mais claro se há um sistema mais seguro de *backup* sendo utilizado.

Armazenamento como uma atividade dinâmica e temporária

Vários sistemas de armazenamento serão apresentados neste texto, e o mais importante é entender que esses sistemas estão em constante evolução. Dentro de uma década, muitos sistemas utilizados hoje já estarão superados pela sua limitação de capacidade ou pela sua obsolescência e desuso. Os gastos serão também temporários, e sistemas novos precisam ser constantemente adquiridos para manter os dados digitais em sistemas correntes. Isso gera uma preocupação com os custos constantes necessários para manter esses acervos e deve ser um fator essencial de um projeto de geração de dados digitais para garantir sua preservação e sustentabilidade em médio e longo prazo.

Aumento da capacidade de armazenamento

O primeiro computador com disco rígido foi o RAMAC 305 da IBM em 1965. O disco rígido era composto de 50 discos com uma capacidade de 3,75 MB (hoje o tamanho aproximado de uma imagem digital em formato JPEG).



Figura 93. Imagem da esquerda, IBM_305_RAMAC, US Army Red River Arsenal, domínio público; Imagens da direita superior e do meio, domínio público; Imagem inferior da direita, IBM_305_RAMAC, DingirXul CC-BY-SA-2.5, no site do *Computer History Museum*.

Rapidamente, o custo por *gigabyte* foi caindo a partir da década de 1980, com custo médio de mais de US\$ 400.000,00 por *gigabyte* (GB), para US\$ 11,00 por *gigabyte* no ano 2000, e em torno de US\$ 0,025 (dois centavos e meio de dólar) por *gigabyte* em 2020. Temos um grande ganho também em capacidade de armazenamento com discos rígidos em 2022, de 22 TB cada. A indústria vende o conceito de que o custo diminui e a capacidade aumentou e, portanto, temos a capacidade de guardar tudo o que queremos. Por outro lado, é necessário entender que a nossa demanda também foi alterada, com arquivos muito maiores. Temos câmeras digitais hoje de 30 a 60 MP, ou mais, comparadas às câmeras digitais de 2 a 4 MP do início do século XXI.

É necessário considerar que o armazenamento não é o único custo para os nossos dados digitais. Ao avaliarmos o custo total de propriedade (TCO – *Total Cost of Ownership*), vemos que existem inúmeros outros custos não tão visíveis, conforme vemos abaixo, e fundamentais para a garantia da preservação dos nossos materiais digitais. Lembrando, também, que todas essas atividades requerem recursos humanos qualificados.



Figura 94. Imagem do iceberg para demonstrar as camadas visíveis e invisíveis nas atividades de preservação digital.

Uma visão sobre alguns sistemas de armazenamento

Não é o intuito deste livro descrever em detalhes todos os sistemas de armazenamento, mas principalmente falar deles do ponto de vista de sistemas ideais para a preservação de dados digitais.

Discos ópticos

Os discos ideais para guarda são do tipo WORM – *Write Once Read Many*, gravar uma vez e ler muitas, ou seja, discos não regraváveis. Os discos R são

graváveis/*recordable*, e RW são regraváveis/*rewritable*. Porém, hoje, um dos maiores problemas de guardar dados digitais em discos ópticos é a baixa quantidade de dados que podemos armazenar neles. Outra questão é que necessitamos de todos os componentes para acessar os dados nos discos – *hardware, drivers, softwares*, incluindo sistemas operacionais. O manual *Care and Handling of CDs and DVDs – A Guide for Librarians and Archivists*⁵⁹, escrito por Fred Byers em 2003, coloca bem as questões do cuidado com discos ópticos. Apesar de o disco óptico ter as vantagens de um armazenamento sem uso de energia (*cold storage*), riscos no disco, altas temperaturas, umidade relativa do ar, e presença de ar com poluição por enxofre aumentarão as chances de danos às camadas de corantes nos discos, dificultando sua leitura.

Discos com mais de 5–10 anos de vida podem apresentar problemas de leitura. Os novos computadores e *laptops* já não vêm mais com leitores de discos ópticos.

Entendemos que o disco em si dura mais que a tecnologia ao redor do sistema. Temos sempre que avaliar a tecnologia que rodeia o sistema de armazenamento, e não a unidade de armazenamento em si, para determinar sua longevidade. Se esta opção for considerada como uma das opções de guarda, a sugestão é de fazer a migração dos dados para discos novos a cada 5–10 anos, e adquirir discos ópticos de boa qualidade para fazer esta guarda.

Dois casos interessantes para discutir são o dos discos M-Disc e o do sistema óptico da Sony.

M-Disc

O M-Disc usa uma tecnologia proprietária de materiais como se fosse gravar os dados “em pedra”, com referência direta à longevidade da arte rupestre. Com a promessa de durabilidade de 1.000 anos, fica claro hoje para nós que provavelmente nas próximas décadas já não teremos mais as tecnologias necessárias para ler esses discos, mesmo que estejam em ótimo estado de conservação. Isso reforça o fato de que não é sobre o aparato e a tecnologia em si, mas sobre todo o ecossistema que o rodeia e faz com que funcione.

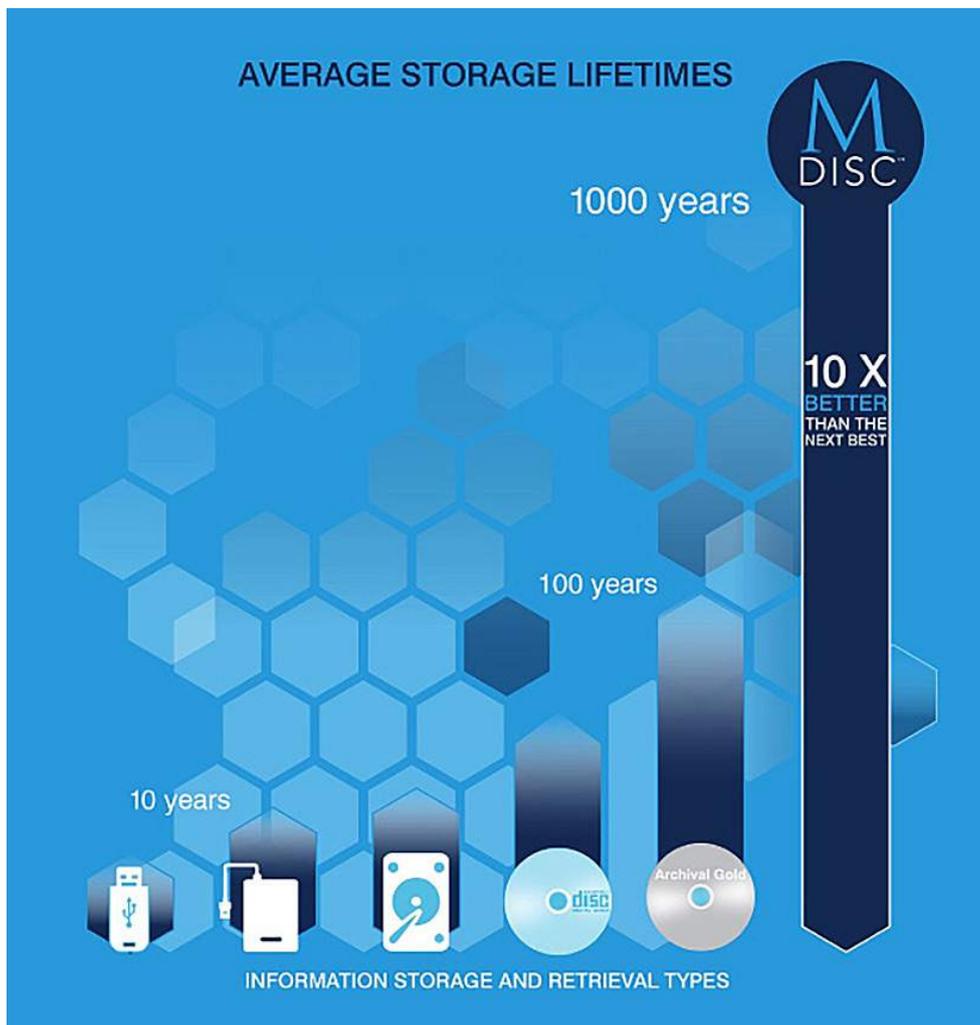


Figura 95. Gráfico de propaganda no site da empresa M-DISC sobre sua longevidade quando comparado a outros sistemas de armazenamento.

Na história sobre o armazenamento, vivemos sempre tentativas de soluções empresariais prometendo longevidade, ou a guarda “para sempre”. A versão realista nos ensina que temos que considerar sistemas de armazenamento como soluções temporárias. O nosso objetivo é preservar “para sempre” os dados digitais permanentes, mas isso será feito, como já foi apresentado, em blocos de 5 a 10 anos por vez, sempre atualizando os nossos sistemas de guarda. Esta realidade poderá ser diferente nas próximas décadas, mas ainda é a nossa realidade atual. Desta forma, é de uma importância fundamental entender e estar preparado para os gastos contínuos que precisam ser feitos em projetos que contemplam o armazenamento digital.

Sony Optical Disk Archive (ODA)



Figura 96. Captura de tela do site da B&H Photo Video, em setembro de 2021.

A Sony lançou um sistema de cartuchos contendo discos ópticos em 2013. Agora na Gen 3 (terceira geração), com guarda de até 5,5 TB para cada cartucho, e com custo em torno de 190 dólares em 2021. Dentro do cartucho há uma série de discos ópticos *blu-ray*. Isso seria uma ótima solução, mas um dos fatores para a garantia de durabilidade de um sistema é a sua adoção. Adoção significa a base de clientes que compraram e usam esse sistema. Com um custo muito elevado da unidade de gravação e leitura, visto anteriormente, e sendo a Sony a única empresa a produzir e vender esse sistema, devemos nos questionar qual a garantia de que este terá uma longa vida de uso. Se não houver adoção, não haverá mercado de consumidores das unidades e dos cartuchos, e ocorrerá sua morte certa, mesmo sendo um bom sistema. Com aumento de adoção ao sistema, teremos mais chances de ele sobreviver em longo prazo. Sem adoção, ficará inviável financeiramente para a Sony, ou para qualquer empresa, manter o produto em sua produção e manutenção. Na revisão final deste texto, em 2023, conforme o site da Sony, todos os dispositivos desse sistema da empresa estão descontinuados.

Discos rígidos magnéticos e estado sólido

Para simplificar a abordagem neste texto, vamos dividir os discos rígidos em duas categorias. Primeiro, temos os HDDs (*Hard Disk Drives*), que normalmente

chamamos de HDs no Brasil. Para não nos confundirmos com a *resolução* HD (*High Definition*), usaremos HDD. Em segundo, temos os SSD (*Solid State Drive*), ou unidade de estado sólido, também conhecido como memória de estado sólido ou memória *flash*. Os HDDs têm uma longa história de produção e desenvolvimento, e continuam mais econômicos em relação ao custo por Terabyte quando comparados à tecnologia SSD. Apesar de as unidades SSD terem sido desenvolvidas na mesma época do desenvolvimento dos discos rígidos magnéticos HDD, elas só entraram para o mercado de consumo na última década. Faremos uma comparação desses dois sistemas com uma discussão maior sobre o que representam para a preservação digital em longo prazo. Esse é um assunto vasto, interessante e com uma rica história, e muita informação na internet.

Comparações entre HDDs e SSDs

Temos que colocar esta discussão, e todas as discussões sobre tecnologia, dentro de uma perspectiva da linha do tempo da história. Este texto certamente estará datado em pouco tempo, bem como novos sistemas e custos aparecerão nos próximos anos. Nesse sentido, o objetivo da abordagem apresentada é criar uma reflexão embasada nas tecnologias atuais, que nos ajudará a pensar sobre estas e sobre as futuras. Existem muitos artigos na internet sobre esse assunto, e o Blog da BackBlaze tem um artigo de 2021, escrito por Andy Klein⁶⁰ bem interessante.

Discos rígidos magnéticos (Hard disk drives – HDDs)

O HDD é uma tecnologia muito testada e implementada há décadas, com alto grau de confiabilidade.

Consiste em elementos móveis, compreendendo um atuador, um braço atuador com cabeça ou cabeçote de leitura e escrita e, o mais importante, disco ou pratos (*platters*) magnéticos, onde serão armazenadas as informações.

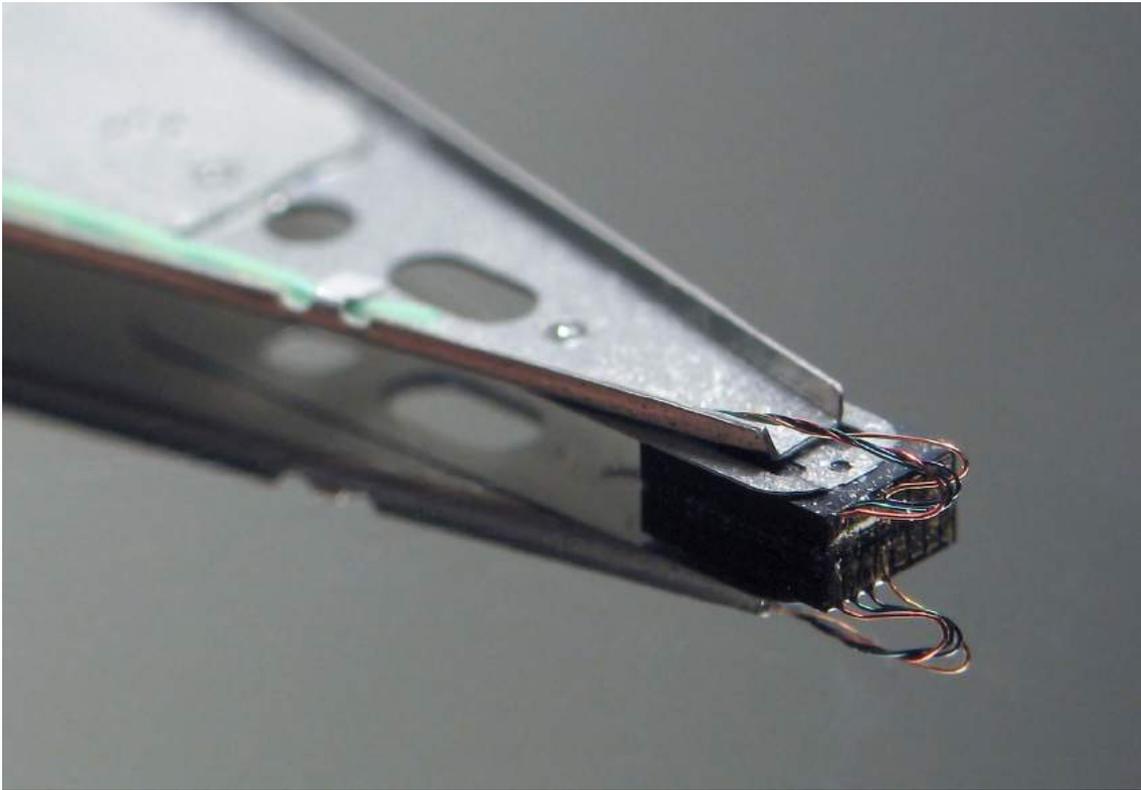


Figura 97. Os dados magnéticos, positivos ou negativos (1s e 0s) são gravados ou lidos com as cabeças de leitura/gravação, que pairam a uma altura de cerca de 5 nm acima do prato, sem tocar sua superfície. Esta distância representa em torno de 14 mil vezes menor que o tamanho de um fio de cabelo, que tem em média 70 microns, ou 70.000 nm (nanômetros).

Esta imagem de Andrew Magill mostra o reflexo do braço atuador visto na superfície do disco (fundo cinza).

Se esta cabeça tocar na superfície do disco, isso causará danos irreversíveis à sua estrutura. Quando a unidade está funcionando, os pratos giram em média na velocidade de 7.200 RPM (rotações por minuto), e a borda do disco de um HDD estará viajando a mais de 120 quilômetros por hora. Quando o HDD recebe um comando de leitura ou gravação, as cabeças procuram as áreas do prato onde os dados devem ser lidos ou gravados.

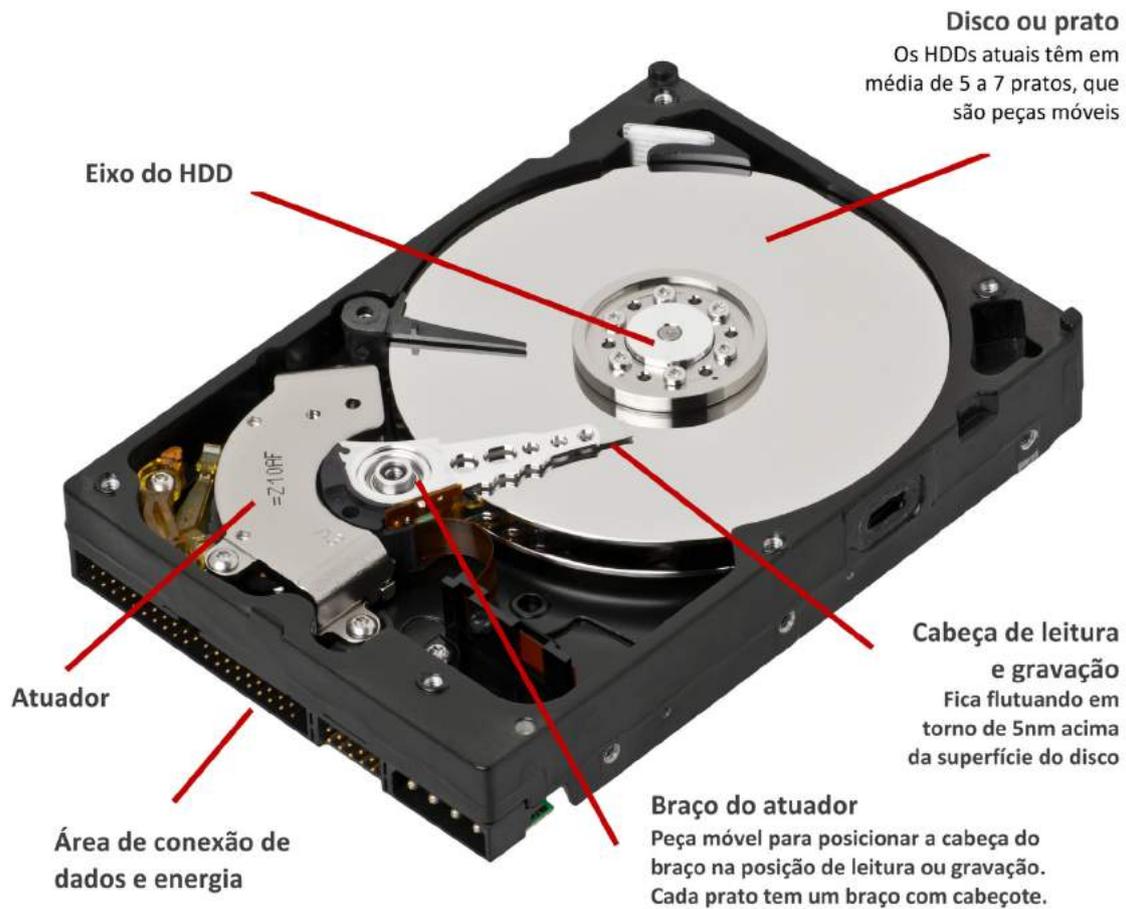


Figura 98. Partes de um disco rígido magnético.

A vida útil desse sistema é algo sempre discutido. O *blog* da empresa BackBlaze é um excelente lugar para buscar artigos importantes sobre armazenamento. Um artigo recente de Andy Klein⁶¹, de outubro 2021, discute a imagem da curva da banheira como símbolo da sobrevida dos HDDs. Discos têm uma tendência de falhar ou no seu início de uso, ou mais para o final da vida, criando esta curva.

O click da morte

Um dos problemas com HDDs é o barulho clássico de HDD clicando seguidamente, como se fosse um metrônomo de música. Esse é o som do braço atuador varrendo de um lado a outro do disco de forma cega, sem encontrar dados, e

esbarrando continuamente no seu ponto de partida. Um disco com esse sinal deve ser aposentado, e deve ser feita uma nova cópia a partir de um *backup*. Caso seja (mas não deveria ser) uma cópia única, desligue-o imediatamente e procure ajuda de uma empresa de recuperação de dados. A tentativa contínua de solucionar o problema poderá causar ainda mais danos se o cabeçote arranhar a superfície do disco nesse processo de varredura contínua.

Drive Failure Over Time: The Bathtub Curve Is Leaking

October 26, 2021 by Andy Klein // 7 Comments

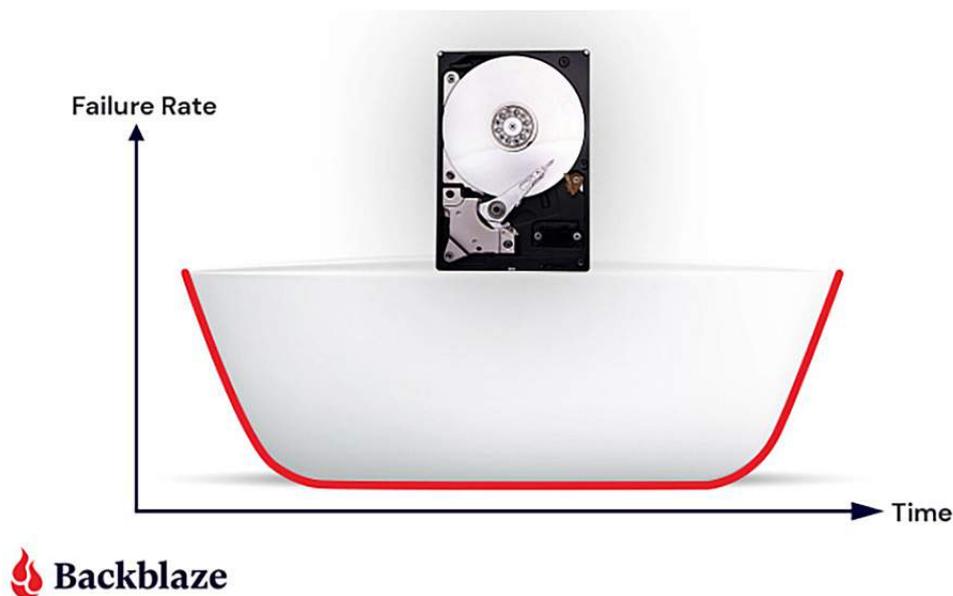


Figura 99. Curva da banheira – taxa de falha de discos HDDs através do tempo. Captura de tela do site da Backblaze, com interessante artigo discutindo a durabilidade de HDDs. Imagem autorizada pela Backblaze.

Nesse artigo, Klein diz que:

Além disso, calculamos que em quatro anos, a expectativa de vida de um disco rígido em nosso sistema seria de cerca de 80%, e prevendo que, em seis anos, a expectativa de vida seria de 50%. Em outras palavras, esperaríamos que um disco rígido que instalamos tivesse 50% de chance de estar vivo após seis anos.

Unidade de estado sólido, SSDs

Os SSDs não utilizam partes móveis para funcionar como os HDDs. Utilizam a memória em *flash*, semelhante à memória de *pendrives* e cartões digitais de câmeras e celulares. As células SSD têm uma vida útil limitada. A previsão é de que uma célula possa chegar a 3 mil ciclos de gravação. A leitura de dados não é o problema, e sim a gravação contínua. Parece pouco, mas em um Notebook com SSD, é bem provável que o disco ultrapasse a vida útil do computador. Mesmo sendo mais robustos que os discos rígidos, continuam sujeitos às mesmas leis de falha que sempre discutimos.

Dados de 2021	HDDs	SSDs
Capacidade de armazenamento	Discos de até 18 TB	Unidades de estado sólido de até 8TB
Velocidade de processamento de leitura	Bem mais lento, valores médios: HDDs: 200 MB/s (O tempo de gravação é sempre mais baixo para os HDDs)	SSDs são muito mais rápidos, mas os SSDs de conexão PCIe são os mais velozes. Vejam os valores médios: SSDs SATA: 560 MB/s SSDs NVMe: 3500 MB/s SSDs PCIe: 7000 MB/s
Custo médio (valores de 2021)	Mais econômico por TB Média de US\$ 30/TB	Custo caindo, mas ainda mais caro Média de US\$ 160/TB
Consumo de energia	Mais alto	Mas baixo
Nível de ruído	Mais alto	Quase nulo
Peças móveis	Partes móveis, discos girando, riscos de danos mecânicos, risco de quedas	Circuito eletrônico, sem peças móveis
Limite de gravações	Os discos rígidos mecânicos têm ciclos ilimitados de gravação e leitura.	<p>“Normalmente, os fabricantes fornecem uma estimativa com os chamados Terabytes gravados (TBW – Terabytes Written) - especialmente quando se trata de SSDs corporativos, mas também para versões de consumidor. Devido ao fato de que, usando o nível de desgaste, os dados serão distribuídos uniformemente por todas as células, essa figura deve dizer quantos dados podem ser realmente gravados no total em todas as células dentro dos chips de armazenamento e ao longo de toda a vida útil.</p> <p>Um valor típico de TBW para um SSD de 250 GB fica entre 60 e 150 Terabytes gravados. Isso significa: Para superar um TBW garantido de 70 TBW, um usuário teria que gravar 190 GB diariamente durante um período de um ano (em outras palavras, para preencher dois terços do SSD com novos dados todos os dias). Em um ambiente de consumidor, isso é altamente improvável!”</p> <p>Tilly Holland, 2020⁶²</p>

Dados de 2021	HDDs	SSDs
Confiabilidade	Os HDDs são considerados bastante confiáveis. Porém, por serem dispositivos de design frágil, são altamente suscetíveis a diversos fatores externos, incluindo danos físicos causados por pequenos choques, quedas ou vibração e danos causados pela água e fogo, umidade excessiva e superaquecimento. Qualquer contato da cabeça de leitura/gravação com o disco causará arranhões que resultam na impossibilidade de restaurar os dados.	A ausência de peças móveis em SSDs (unidades de estado sólido) implica que eles são capazes de resistir aos pequenos acidentes da vida quando comparados aos HDDs. Os maiores problemas são o superaquecimento e os picos de energia, que podem levar não apenas à degradação da unidade, mas também à sua perda completa. E, se deixados sem energia, SSDs começam a perder dados após cerca de um a dois anos em armazenamento, dependendo da temperatura (quanto mais quente pior). Se houver uma falha, a recuperação de dados é mais difícil quando comparado a um HDD.
Preservação digital a longo prazo	Diversos sistemas com discos rígidos magnéticos (HDD) podem ser utilizados para preservação digital	Não é recomendado. Pode ser utilizado como uma das cópias ou uma das mídias, mas certamente não como único sistema/mídia para um fluxo de preservação digital a longo prazo.

Tabela 4. Compilação do autor apresentando algumas comparações de HDDs e SSDs.



Figura 100. Um SSD mSATA Toshiba de 30GB em cima de um HDD interno de notebook de 2,5 polegadas Fujitsu de 320GB.



Figura 101. SSD NVMe 960 Pro da SAMSUNG de 2TB.



Figura 102. SSD PCIe 4.0 980 Pro da Samsung de 1TB. Velocidades de até 7000 MB/s.

Nas publicações da *Memoriav*^{63,64}, há uma recomendação importante sobre armazenamento nas páginas 58 e 59 da publicação *Digital Archiving of Film and Video: Principles and Guidance, Version 1.2*, de novembro de 2019. A seguir, há uma tradução livre dessa tabela.

Categoria	Formatos	Taxa de dados	Área de trabalho	Recomendado para a preservação	Comentários
Dispositivos ópticos	DVD	4–9 Mbit/s	Distribuição	Não recomendado	Suporte não recomendado para arquivamento.
	Blu-ray	Aprox. 36 Mbit/s	Distribuição	Não recomendado	Suporte não recomendado para arquivamento.
Armazenamento com base na TI	M-DISC		Arquivamento	Não recomendado	Mídia não adequada para AV devido à densidade de dados e capacidade de armazenamento. Futuro incerto da produção de dispositivos de leitura.
	ODA (Sony Optical Disk Archive)			Não recomendado	Formato proprietário da Sony, sem experiência de arquivamento.
	HDD			Recomendado condicionalmente	Condições prévias necessárias: múltiplas cópias em locais diferentes, uma seleção de interfaces diferentes, vida útil antecipada de 3 anos.
	RAID			Recomendado	Recomendado desde que haja cópias de backup adicionais em outros sistemas.
	SSD			Não recomendado	O armazenamento em SSD depende de material físico extremamente pequeno que rapidamente atinge seus limites no curso das operações normais ou por meio de influências externas e, conseqüentemente, envelhece muito mal, portanto inadequado para armazenamento de longo prazo
	LTO (7 e 8)			Recomendado	Formato mantido por um consórcio de empresas, e possível para preservação com o padrão de gravação nas fitas LTO-5 LTFS em diante; Fitas LTO-5 e LTO-6 devem ser migradas em breve, as fitas LTO-1 a LTO-4 devem ser migradas imediatamente.
	DLT (Digital Linear Tape)			Não recomendado	Formato desatualizado e sem suporte. Os arquivos armazenados em DLT devem ser migrados para LTO imediatamente.

Tabela 5. Comparação entre diferentes formatos de mídia de armazenamento.

Tabela utilizada com permissão da *Memoriav, Association for the preservation of the audiovisual heritage of Switzerland*.

Diante das discussões apresentadas até aqui, podemos considerar os HDDs mais robustos, mas somente quando em sistemas com mais redundância, como RAIDs, NAS, DAS e SANs, que veremos mais adiante, e mais arriscados quando usados como unidades únicas. Fitas magnéticas LTO também são recomendadas.

HDDs avulsos e em sistemas com diversos discos

Discos rígidos magnéticos são vendidos como discos internos para *desktops* (3,5 polegadas) e *notebooks* (2,5 polegadas), avulsos, ou como discos externos, em um estojo com uma conexão de energia e interface para conectá-lo a um computador. Importante observar que HDDs portáteis, que contêm somente um cabo de conexão, mas não o de força, e tamanhos menores (2,5 polegadas), são normalmente mais frágeis, corrompem com mais facilidade e devem ser considerados meramente como dispositivos de trânsito de informações, e não de preservação. Quase como um *pendrive* de grande capacidade. Na minha experiência, o *click* da morte também é muito mais comum nesse tamanho de HDD.

Discos internos, principalmente HDDs, e o uso crescente de SSDs podem ser adquiridos e colocados em sistemas que comportam uma série de discos, como DAS (*Direct Attached Storage*, armazenamento conectado diretamente a um computador), NAS (*Network Attached Storage*, armazenamento conectado a uma rede) e SAN (*Storage Area Network*, rede de área de armazenamento), que culminam, por fim em grandes *data centers*, que possuem milhares de discos rígidos rodando 24 horas por dia e 7 dias por semana.

Todos esses sistemas têm uma característica em comum. Existe uma redundância interna onde todos os dados digitais gravados são replicados em pelo menos um ou mais discos. Assim, a perda ou falha de um disco não acarretará a perda de dados. O disco falho é simplesmente substituído por um outro com a mesma ou maior capacidade, e os dados são reconstituídos a partir dos outros discos do sistema. Em grandes *data centers*, discos são substituídos constantemente sem haver perda de dados. Do ponto de vista do usuário desses sistemas, há simplesmente uma única área de guarda que se apresenta ao usuário como um único disco. A importância de dividir a discussão sobre esses sistemas nestas quatro categorias tem sentido quando pensamos nos custos iniciais de investimento e o retorno.

O *firmware*, ou a controladora desses sistemas, muitas vezes *software* proprietário, é que controla onde estão todos os dados, os nomes dos arquivos, e a estrutura de pastas desses dados. Se a controladora se corromper, os dados estarão intactos, mas não serão encontrados. A recuperação de dados é inúmeras vezes mais cara do que ter múltiplas cópias, e o processo de recuperação de dados poderá devolver, em muitos casos, arquivos sem a nomenclatura original e a estrutura de pastas,

como também sem a terminação de identificação do arquivo (.tif, .doc, .xls, .pdf etc.). Portanto, qualquer sistema desses, DAS, NAS, SAN, Datacenter, Nuvem etc., deve sempre ser considerado somente uma das cópias de preservação.

É fundamental frisar isso, porque há uma certa sensação de segurança pelo fato de que as falhas de disco não representam perdas, pois os dados estão sempre em outros discos. Esta segurança não é real, pois o *software* que controla os dados pode se corromper. Se isso acontecer, é só formatar o sistema e fazer uma recuperação de dados de um de seus *backups*.

RAID (Redundant Array of Inexpensive Disks)

O RAID representa um conjunto redundante de discos baratos. A ideia é relativamente simples e consiste em ter uma série de discos magnéticos ou de estado sólidos colocados em um suporte para discos formando uma única unidade de disco. Os sistemas podem ser atualizados constantemente ao se substituir discos com menor capacidade por discos de maior capacidade. Isso é fundamental para um sistema que prevê o crescimento ao longo dos anos. Um sistema RAID de 5 discos de 4 TB (total de 20 TB brutos, e em torno de 10 TB de armazenamento em RAID 6) pode virar um sistema com 5 discos de 18 TB (total 90 TB brutos e 45 TB de armazenamento em RAID 6), na medida da necessidade de ampliação de espaço de armazenamento. O RAID pode ser feito via *software*, mas, na maior parte dos casos, é via um sistema de *hardware* com controladora, muitas vezes proprietário.

Existem diversos tipos de RAID, e podemos mencionar alguns dos mais utilizados para preservação:

- Raid 1 – Um sistema de espelhamento no qual um disco é espelhado em outro idêntico. Não haverá perdas de dados com a falha de um dos discos.
- Raid 5 – O sistema é mais complexo, e a paridade, que representa a cópia das informações dos discos, é armazenada de forma alternada nos outros discos do sistema. Assim, a perda de qualquer um dos discos não acarretará perdas, e os dados ainda poderão ser acessados. O disco rígido danificado é substituído e ao acionar um processo de reconstrução (*rebuild*) dos dados, normalmente realizado de forma automática, o sistema de segurança e redundância de dados é reconstruído. Um sistema de Raid 5 precisa ter no mínimo 3 discos.

- Raid 6 – Nesse sistema, temos dupla paridade. Isso significa que os dados são copiados em mais discos, sendo que a perda simultânea de 2 discos não acarretará perda de dados. Em um sistema de dupla redundância, como RAID 6, temos em torno da metade da capacidade total de discos usada para armazenamento, e a outra metade, para a segurança dos dados. Um sistema de Raid 6 precisa ter no mínimo 4 discos.

Neste link da Synology (https://www.synology.com/en-us/support/RAID_calculator), podemos construir um sistema RAID com diversas configurações para entender melhor o funcionamento e a capacidade de armazenamento de diferentes opções de RAID.

DAS

Sistema de armazenamento conectado diretamente a um computador como um HDD externo (*Direct Attached Storage*). Consiste em um contêiner com encaixes para 4, 5 ou 8 HDDs em média. Esse contêiner é conectado diretamente a um computador como um grande HDD externo. Mesmo sendo possível ter uma falha de um HDD, sem haver perda de dados (RAID 5) ou dois HDDs falhando simultaneamente sem perda de dados (RAID 6), esse sistema deve sempre ser considerado somente como uma das três cópias necessárias.

NAS

Sistema de armazenamento conectado à rede (*Network Attached Storage*). Um sistema NAS funciona basicamente da mesma forma que o DAS, porém é conectado a uma rede e pode ser acessado por inúmeras pessoas conectadas a esta mesma rede. Para pequenos grupos e instituições que precisam ter uma área centralizada de guarda, o NAS pode ser uma ótima solução. Esses sistemas normalmente são mais complexos e possuem mais opções de *softwares* que realizam os *backups* e segurança de dados embutidos no dispositivo.

SAN

O *Storage Area Network* é um sistema semelhante ao NAS, mas muito mais complexo, onde temos uma área dedicada de armazenamento conectada aos servidores via um conector de fibra. Um sistema SAN é construído para grandes áreas de armazenamento, com custo inicial elevado de instalação, mas custo reduzido por GB de manutenção após instalação.

A seguir, algumas comparações de custos e benefícios entre esses três sistemas.

	DAS	NAS	SAN
Tipo de armazenamento	Setores	Arquivos compartilhados	Blocos
Transmissão de dados	IDE/SCSI	TCP/IP, Ethernet	Canal de fibra/ <i>Fibre channel</i>
Modo de acesso	Clientes ou servidores	Clientes ou servidores	Servidores
Capacidade	Alguns terabytes até uns 80 TB	Alguns terabytes até uns 120 TB	Sistemas de centenas de terabytes e petabytes
Complexidade	Fácil	Moderado	Complexo
Custo por GB	Alto	Médio	Baixo

Tabela 6. Comparação entre custos e benefícios em diferentes sistemas de armazenamento digital.



Figura 103. Na imagem acima, temos um exemplo de um DAS—*Direct Attached Storage*, armazenamento conectado diretamente a um computador, com 8 baias para discos rígidos de 3,5 polegadas. Esse site permite que você possa colocar HDDs diversos nestas baias, e o sistema vai calculando o espaço de armazenamento que estará disponível. Aqui está selecionada a opção de redundância dupla de disco (*Dual Disk Redundancy*), o RAID 6, que permite a falha simultânea de dois discos. Nesta condição, com 74 TB de discos instalados, com capacidade bruta de 67,4, somente 46,33 TB estão disponíveis para uso. 19,14 TB estão sendo utilizados para proteção.



Figura 104. Sistemas NAS–*Network Attached Storage*, armazenamento conectado à rede, com soluções mais domésticas à esquerda, e mais empresariais à direita, com maior capacidade e possibilidade de expansão com mais unidades.

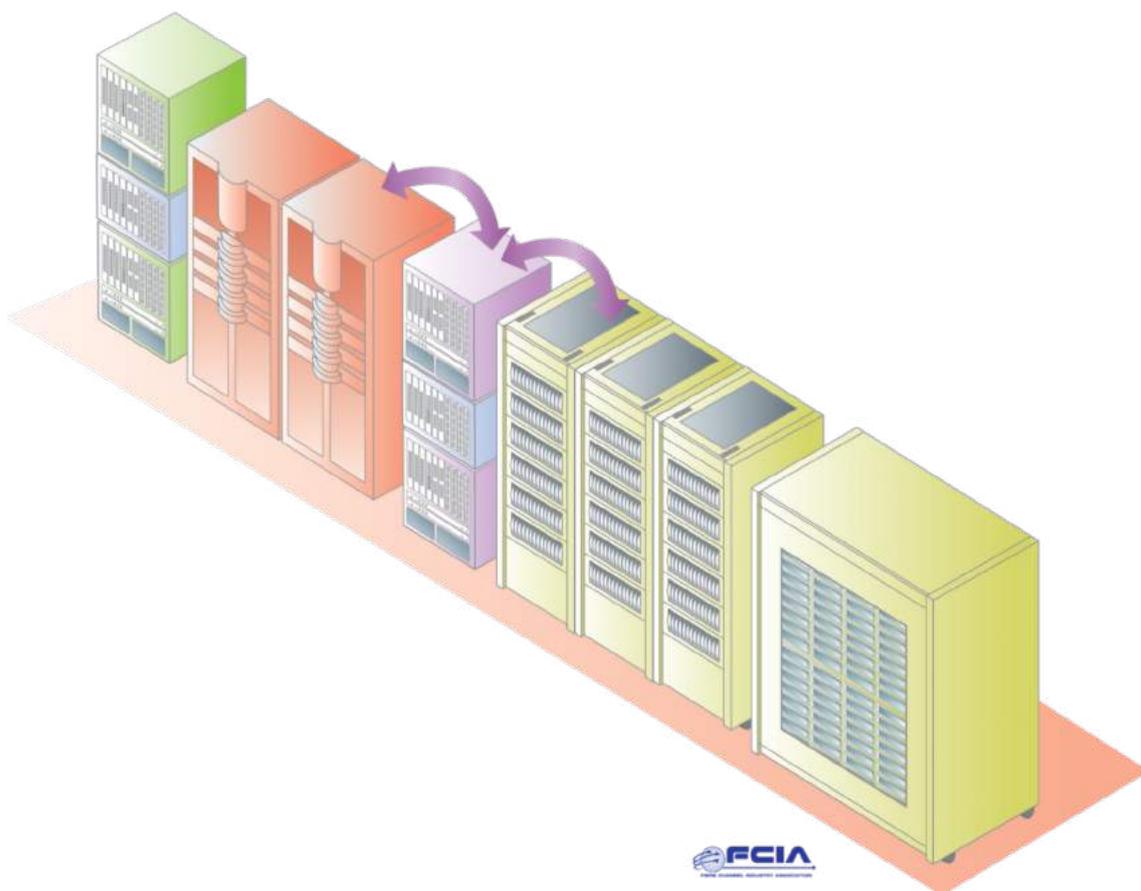


Figura 105. SAN–*Storage Area Network*. Essa rede de área de armazenamento (SAN) conecta os dispositivos de armazenamento amarelos com os servidores laranja por meio dos *switches* de canal de fibra, em roxo.

De sistemas simples a complexos

Como existem discos rígidos HDD externos com capacidade, em 2021, de até 18 TB, podemos ter um sistema de armazenamento com três cópias cada uma em um desses discos externos para quem tem dados em torno de 10 TB com espaço de crescimento. Com um sistema DAS ou NAS, é possível ter sistemas com 5, 8, ou 12 HDDs internos, e, portanto, temos a possibilidade de fazer diversas combinações de discos, podendo começar com um sistema para 5 discos, por exemplo, com 40 TB (5 HDDs de 8 TB cada, lembrando que isso representa 20 TB com Raid 6, em termos de quantidade de dados armazenados). Um sistema assim, de 40 a 150 TB, representa a matriz dos dados, e será necessário ter um outro sistema idêntico como cópia de segurança. A terceira cópia pode estar na nuvem, em *data centers* ou copiada em fitas magnéticas e armazenada *off-line* e *off-site*. Esses sistemas são relativamente fáceis de serem implementados e podem ser gerenciados por pessoas ou equipes semiprofissionais.

Quando queremos ter dados acessados por uma rede de servidores, com velocidade e capacidades acima de 100 TB com perspectiva de aumentos substanciais, podendo chegar a um ou muitos petabytes (PB, equivalente a mil TB), será necessário investir em sistemas complexos, como SAN, com *storage*/armazenamento dedicado e equipe especializada para gerenciamento e manutenção. Nesse momento, também é necessário entender se o armazenamento deve ser feito internamente em centro de dados (*data centers*) ou terceirizado para serviços de processamento e armazenamento na nuvem.

Centro de dados

Um centro de dados (*data center*) é um local que centraliza as operações e equipamentos de TI de uma organização com o objetivo de armazenar, processar e disseminar dados. Os centros de dados são vitais para uma organização, com prioridade máxima de segurança e confiabilidade. No passado, os centros de dados sempre foram áreas físicas, prédios especializados com controle rigoroso de acesso, climatização e acesso à energia, mas hoje, com as diversas opções de nuvens públicas, esse modelo está se modificando. Um centro de dados de uma instituição como o CERN, como na imagem a seguir, funciona como uma nuvem, porém privada, onde somente as pessoas autorizadas da rede do CERN terão acesso aos dados via *firewalls* de segurança.

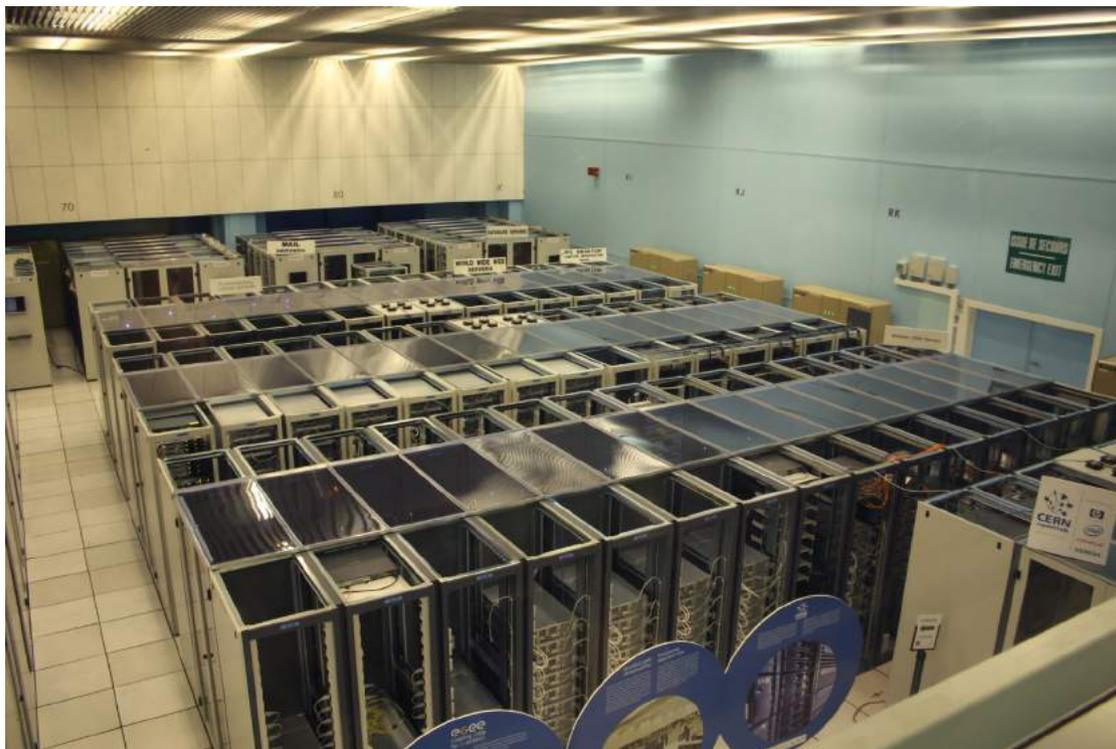


Figura 106. Data Center do CERN em 2010 (*Conseil Européen pour la Recherche Nucléaire*).

Fitas magnéticas (Linear Tape Open – LTO)

Armazenar dados digitais em fitas magnéticas é algo que vem acontecendo desde os primórdios dos computadores, com inúmeras tecnologias, formatos e empresas, mas a mudança significativa para a preservação de dados vem com a versão LTO-1, em 2000. O conceito foi o de criar uma arquitetura com padrão aberto, indo contra os formatos proprietários de fita magnética disponíveis na época. O consórcio LTO foi criado e é controlado pela Hewlett Packard Enterprise, IBM e Quantum. O cartucho de fitas da LTO-1 tinha uma capacidade de 100 GB. Hoje estamos na versão LTO-9, com capacidade de 18 TB de dados sem compressão e 45 TB com compressão. Em 2010, foi introduzido o LTFS (*Linear Tape File System*), a partir das versões LTO-5, um padrão aberto que separa os metadados do objeto e o próprio objeto, facilitando sua recuperação. Arquivos podem ser encontrados e recuperados como se a fita fosse um disco rígido, navegando pelas pastas na busca de conteúdo.

É interessante ver os discursos da indústria da nuvem vendendo as suas soluções e declarando o fim, ou a morte da fita. Uma busca na internet por *tape is dead* traz mais artigos recentes sobre esse mito e o fato que o armazenamento em fita ainda é a forma mais econômica de guarda de dados digitais. O artigo *Why the future of data storage is (still) magnetic tape: Disk drives are reaching their limits, but magnetic tape just gets better and better*⁶⁵, da IEEE (Instituto de Engenheiros Eletricistas e Eletrônicos), defende que o futuro do armazenamento de dados reside em fitas magnéticas. Com as constantes atualizações de tecnologias, temos sempre que estar revendo esses conceitos. Isso significa interpretar o “futuro” como sendo algo que acontecerá nos próximos 5 ou no máximo 10 anos.

O acesso aos dados em fita é muito lento quando comparado ao acesso rápido de discos rígidos (HDDs) ou memórias semicondutoras (SSDs), mas a fita é pensada como um *backup off-line* e *off-site*, e não para acesso rápido. É construído como o último recurso, depois da falha da matriz e outros *backups* em disco. Fitas precisam ser armazenadas em prateleiras ou em bibliotecas robóticas de fitas, mas conectadas à energia somente quando dados serão escritos ou recuperados da fita. Nesse sentido, a fita é muito eficiente quanto ao uso energético quando comparada às unidades de disco e/ou data centers.

Há o custo energético de manter uma área de guarda com condições climáticas controladas.

No gráfico da Ampex, no texto da CLIR (Council on Library and Information Resources)⁶⁶, vemos que a situação ideal para armazenamento de fitas magnéticas é em torno de 20 C com 40% de umidade relativa do ar.

Vantagens do armazenamento em fitas LTO

Ser um sistema aberto e fazer parte de um consórcio que tem um planejamento constante de novas versões com maior capacidade de armazenamento e velocidade de transmissão de dados (*roadmap*, ou mapa/trajetória da estrada) a cada 2–3 anos é uma garantia de não estar entrando em um sistema sem histórico ou planejamento futuro. A adoção, como discutimos anteriormente, também é um fator, e o armazenamento em fita teve uma queda com o surgimento de armazenamento em nuvem, mas, na década de 2020 a 2030, tem previsão de crescimento de 7,8%⁶⁷.

A segurança dos dados é um dos grandes fatores. Por estar desconectado da rede (*off-line*) e armazenado em um lugar fora da área de trabalho (*off-site*), há o que chamamos de *air gap* (a ausência de uma conexão direta ou indireta entre um computador e a internet, por razões de segurança – a tradução livre da expressão seria “espaço de ar”). Esta característica é muito importante de considerar em um sistema de armazenamento, especialmente em uma era de aumento de taxas de ataques cibernéticos, roubo de dados e *ransomware*. Claro, esse conceito de *air gap* não é exclusividade de fitas e pode ser feito ao deixar um HDD desconectado e armazenado *off-site*.

Desvantagens do armazenamento em fitas LTO

A barreira de entrada maior para uso de fitas LTO é o alto custo inicial do gravador⁶⁸. Esse custo pode ser reduzido se for compartilhado por uma rede maior de organizações. Outra desvantagem é que mesmo a fita tendo uma durabilidade de algumas décadas, o limite maior está na compatibilidade retroativa das unidades de leitura e gravação. Das unidades de fita da geração LTO de 1 a 7, é possível ler fitas de duas gerações anteriores (LTO-6 lerá fitas LTO-4), mas gravam em fitas somente de uma geração anterior (LTO-6 gravará em fitas LTO-5 e LTO-6). A partir da geração LTO-8 e 9, as unidades de leitura e gravação poderão ler e gravar em somente uma versão anterior.

Como temos uma nova geração de fita a cada 2–3 anos, vista no *roadmap*, caminho ou trajetória do consórcio, fitas gravadas em LTO-8 hoje não serão mais lidas em sistemas LTO-10. Nesse sentido, se houver uma quantidade enorme de fitas, também será necessário um processo contínuo de migração dos dados de versões anteriores para novas versões. Se passar muito tempo, a única saída para as fitas LTO-8 é encontrar leitores da série 8 ou 9 para ler os dados, migrar para um servidor (discos rígidos) e, posteriormente, copiar para versão mais atual de fitas ou outro sistema vigente. Ter um leitor funcional, em bom estado, pode ser um problema. Nos dias de hoje, será muito difícil encontrar um leitor de fitas LTO-2 (2003) ou LTO-3 (2005) em ótimo estado de funcionamento. Esta questão é grave quando lidamos com quantidades enormes de dados. A única solução é uma biblioteca de fitas que automatiza esse processo de migração de forma contínua. Para coleções menores, esse processo terá que ser feito manualmente. Novamente vai valer a durabilidade do ecossistema, e não da mídia em si, e nesse sentido a durabilidade das fitas é parecida com a dos HDDs.



Figuras 107 e 108. Gravador e fita LTO-8 à esquerda e *Roadmap* das versões futuras de fitas LTO.

Spectra Stack Calculator

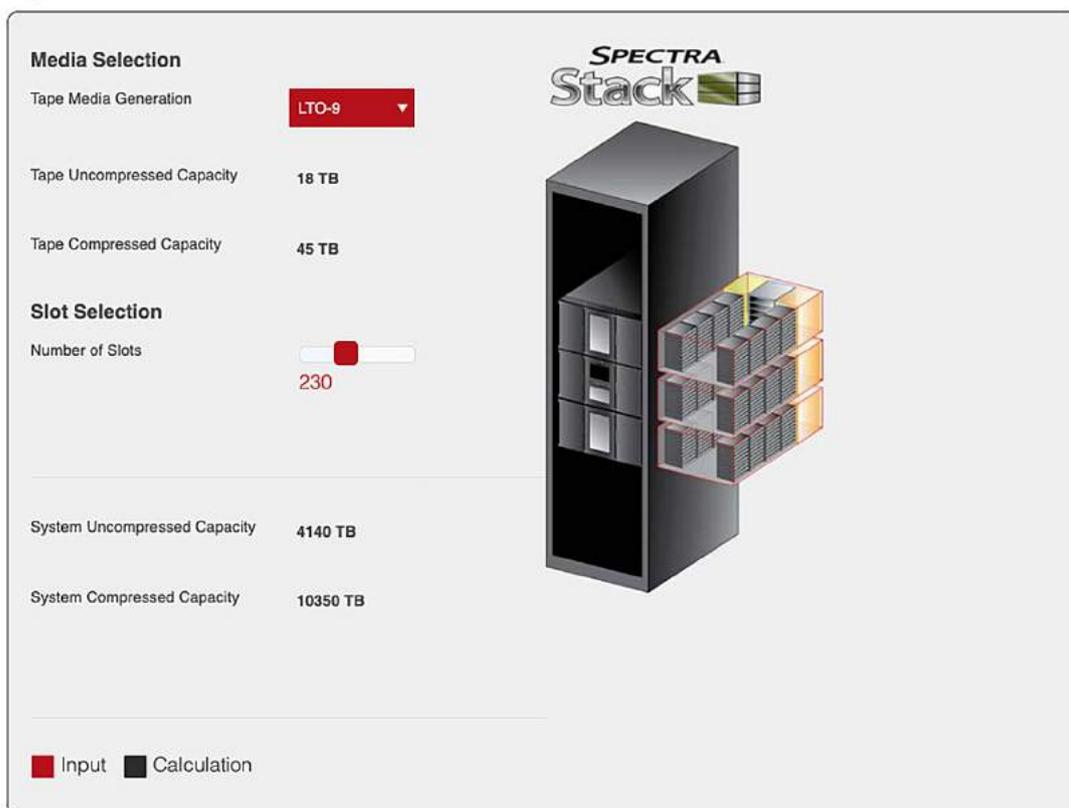


Figura 109. A Spectra, entre outras empresas do mercado, vende soluções de bibliotecas de fita para grandes escalas, chegando ao ExaScale, a maior biblioteca de fitas em 2021 que pode armazenar 1 Exabyte (1 Exabyte é equivalente a 1.000 Petabytes; 1 Petabyte é equivalente a 1.000 Terabytes)⁶⁹.

Armazenamento em Nuvem

Uma descrição bem abrangente do que representa a nuvem é encontrada na Wikipedia⁷⁰:

O armazenamento em nuvem é um modelo de armazenamento de dados de computador no qual os dados digitais são armazenados em pools lógicos, ditos “na nuvem”. O armazenamento físico abrange vários servidores (às vezes em vários locais), e o ambiente físico normalmente pertence e é gerenciado por uma empresa de hospedagem. Esses provedores de armazenamento em nuvem são responsáveis por manter os dados disponíveis e acessíveis, e o ambiente físico protegido e funcionando. Pessoas e organizações compram ou alugam capacidade de armazenamento dos fornecedores para armazenar dados de usuários, organizações ou aplicativos.

A primeira referência à nuvem vem de uma propaganda da AT&T de 1993 sobre o PersonaLink Services. Nesse vídeo, o serviço é comparado à nuvem: *You can think of our electronic meeting place as the cloud* (Você pode pensar em nosso ponto de encontro eletrônico como a nuvem). O artigo *Step back in time: AT&T predicts the cloud in 1993*, escrito por James Bourne em 2014⁷¹ (*AT&T predicts the cloud in 1993*), fala sobre esse fascinante histórico e o vídeo criado por David Hoffman para esse lançamento pode ser visto nesse link no YouTube⁷².

A Chelsey Ferris escreve em 2019⁷³ e dá uma outra abordagem sobre o histórico de armazenamento na nuvem que traz outros elementos, como o início da ARPANET (Advanced Research Projects Agency Network, em português: Rede da Agência para Projetos de Pesquisa Avançada).

A nuvem, como conceito e referência à palavra, tem sido vendida pela indústria como clara, limpa, difusa, transparente, bonita, livre, eterna, natural e, de alguma forma, gratuita. Certamente é também vendida como o futuro e a solução para o gerenciamento e preservação de todos os dados digitais, com custos cada vez mais decrescentes. Vem à minha mente a imagem de uma nuvem como a que vemos na imagem a seguir.



Figura 110. Nuvem lenticular sobre *Harold's Cross* em Dublin, Irlanda, tirada às 11h30, 30 de junho de 2015.

Sabemos que a nuvem é de fato um grande complexo de centro de dados localizados em diversos lugares físicos, construções em solo rígido, com demandas energéticas intensas para rodar discos rígidos 24/7 e alimentar sistemas de resfriamento a fim de dissipar o calor que geram. A nuvem é pesada, próxima, limpa por fora, complexa por dentro, densa, cara, artificial e temporária. Gera nuvens – mas de CO₂.

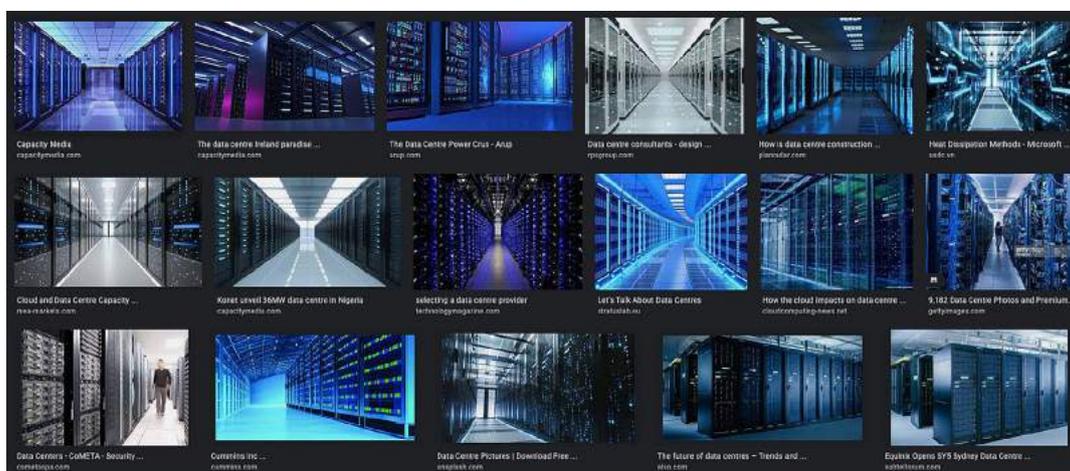


Figura 111. Captura de tela proveniente de pesquisa na internet com a busca de "data center images".

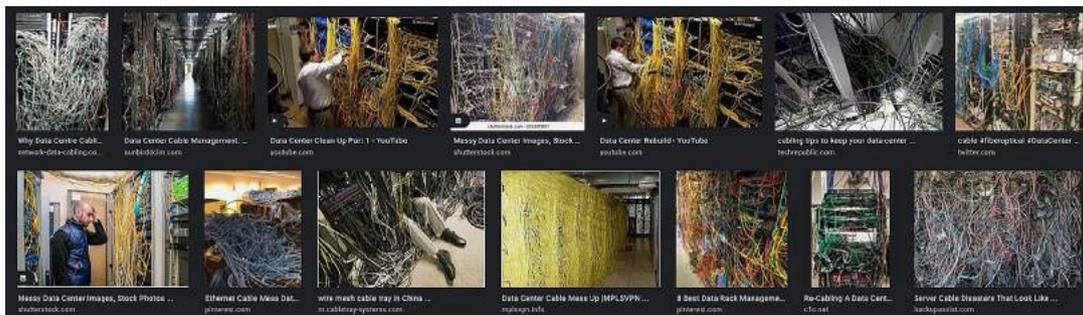


Figura 112. Um tanto exagerado, mas por dentro desses sistemas são milhares de quilômetros de cabeamento para conectar tudo. Captura de tela proveniente de pesquisa na internet com a busca por "data center cable mess".

Custos na nuvem, corrida para custo zero

A AWS – Amazon Web Services é uma das pioneiras no serviço de armazenamento e computação na nuvem, e, no gráfico exibido no artigo de Alex Teu⁷⁴ em maio de 2014 sobre "race to zero", a corrida para zerar o custo do armazenamento na nuvem, vemos o custo por GB (*gigabyte*) das principais empresas do mercado despencando em uma corrida e em uma concorrência para chegar ao custo zero.

Curiosamente, todos os artigos de "race to zero" são em torno de 2014, e, a partir dessa data, os custos de armazenamento em nuvem deram certa estancada em um patamar, sem de fato chegar ao zero, e esta discussão sumiu da internet. Os custos nesse gráfico de 2014 chegam em torno de US\$ 0,03–0,024 (três centavos de dólar por GB).

O artigo *Cloud storage costs: How to get cloud storage bills under control*, de Rene Millman, escrito em 2020⁷⁵ para a Computer Weekly, traz importantes constatações sobre o custo de armazenamento. Dá para perceber que os custos se mantêm praticamente iguais de 2014 a 2020, em torno de 2,5 centavos de dólar por GB.

As entrelinhas dos contratos

Como existem muitos planos e serviços diferentes, que também irão variar de acordo com o local de armazenamento, volume e uso, é necessário fazer uma pesquisa cuidadosa de alguns dos muitos fatores que precisam ser considerados, tais como:

- Como são os contratos? Os SLAs – *Service Level Agreements* / Acordos de Nível de Serviço, precisam ser lidos em minúcia para entender exatamente o que é coberto e como são cobrados os serviços.
- Onde os dados estão sendo armazenados? Lugares diferentes têm preços diferentes.
- Dependendo do pacote, o *upload* (carregamento) pode ser gratuito, mas o *download* (descarregamento) pode ter custos. Isso precisa ser verificado.
- A transferência de arquivos entre diferentes regiões pode ter um custo.
- Como é a estratégia de saída, para retirar todos os seus dados do serviço contratado.

Nuvem – a corrida agora é para zero de emissões

Com o enorme gasto energético dos centros de dados e a urgente necessidade de redução de emissões de CO₂ e gases de efeito estufa para a sobrevivência da nossa existência nesse planeta, a corrida agora é para tecnologias de redução de energia e uso de energias renováveis para a manutenção integral desses centros de dados.

Em um *whitepaper* da DCD (Data Center Dynamics) de Susanna Kass e Alberto Ravagni⁷⁶, eles discutem sobre o *design* e a construção da próxima geração de *data centers* sustentáveis. Não será possível conciliar o avanço da explosão de dados, do *Big Data*, das tecnologias de *Blockchain*, IoT (*Internet of Things*), com a vida nesse planeta sem caminhar para soluções de energias renováveis.

Crescimento da nuvem e saída das empresas

A *International Data Corporation* – IDC⁷⁷ apontou, em setembro de 2021, que as despesas totais de empresas investindo em sistemas de nuvem aumentarão a uma taxa de anual de crescimento composta de 16,9%. Entretanto, mesmo com esse crescimento, muitas empresas, a partir do *boom* inicial de mover tudo para a nuvem em torno de 2011–2012, estão saindo da nuvem pública. O artigo *Lots of Companies Are Leaving the Cloud. Here's Why*⁷⁸, de Ernest Sampera, escrito em 2019, discute algumas das razões.

A questão da segurança é a razão prioritária, com o aumento de *malware* e *ransomware*. O artigo informa que:

Os provedores de nuvem pública oferecem medidas de segurança abrangentes, mas muitos deles dificultam o uso real dessas defesas. A própria natureza da infraestrutura de nuvem pública permite que os dados sejam acessados de qualquer lugar, e os dados frequentemente precisam viajar expostos à Internet aberta, tornando-os vulneráveis a roubo ou infecção por malware. Os provedores de nuvem também gerenciam dados para muitos clientes no mesmo ambiente de nuvem, aumentando a probabilidade de as pessoas erradas receberem acesso ao que deveriam ser dados seguros. Ao mudar para uma solução de nuvem privada, as empresas podem implementar quaisquer protocolos de segurança de sua preferência para garantir a proteção de seus dados.

O controle integral dos dados, a sua centralização, a sua flexibilidade e os custos inicialmente apresentam vantagens para as empresas, por evitarem investimentos em estrutura e recursos humanos para internalizar o armazenamento. Contudo, à medida que os volumes de dados crescem, torna-se mais eficiente e com custo-benefício melhorado para a empresa ou instituição ter seu próprio centro de dados dedicado, saindo da nuvem pública para a nuvem privada, com centro de dados dedicado. Ter sempre os dados disponíveis é outra questão. As empresas de computação em nuvem podem sair do mercado, como já aconteceu com empresas menores, e pode não haver tempo suficiente para remover todos os seus dados armazenados.

Edge e Fog computing

Conceitos que propõem a descentralização dos dados, como *Edge Computing* ou *Fog computing*, estão cada vez mais presentes e comuns nesta discussão, conforme definição da Wikipedia:

Computação em névoa ou em inglês, *fog computing* (também denominada de computação em neblina, computação de borda ou nevoeiro, que deriva da denominação *fogging*), consiste na alocação do poder de processamento mais perto do limite da rede. Portanto, é uma arquitetura de computação descentralizada onde dados, cálculos, comunicações, armazenamentos, medições, aplicações e gerenciamentos são distribuídos no local mais lógico e eficiente: entre a fonte de dados e a nuvem⁷⁹.

O tempo de transferência para *upload* e *download* de dados

Existem inúmeros aplicativos na internet que podem rodar testes de velocidade em sua conexão de internet. Isso serve para calcular as suas velocidades de *upload* e *download*.

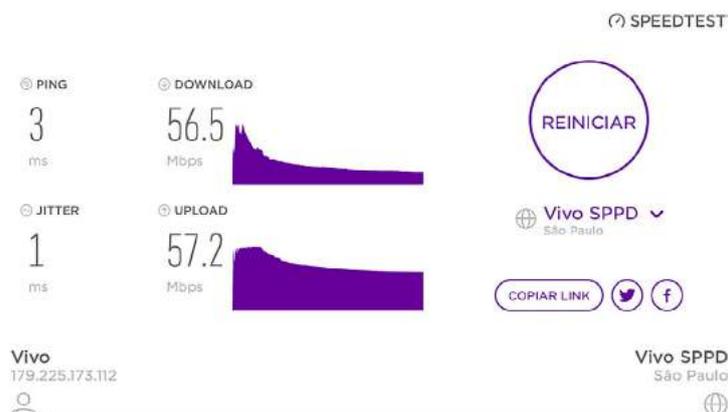


Figura 113. Muitas vezes a velocidade de *upload* é inferior à de *download*. Softwares que testam esta velocidade são úteis para uma previsão de tempos de transferência. A transferência é medida em Mbps (*Megabits* por segundo).

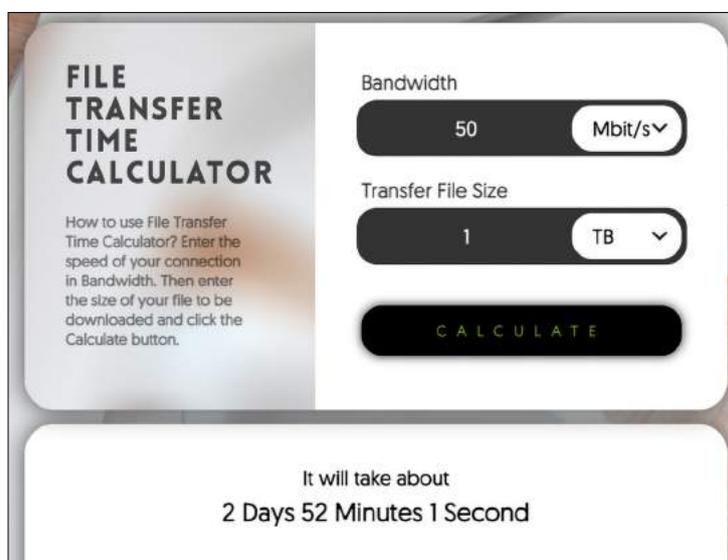


Figura 114. Outros aplicativos que calculam a transferência de arquivos, como este acima, determinam que, na velocidade de 50 Mbps, o tempo para transferência de 1 TB de dados será um pouco acima de 2 dias. Entender esses tempos pode ajudar a entender a disponibilidade de seus dados armazenados na nuvem.

Tecnologias de Blockchain

A guarda de dados em soluções descentralizadas, como a *blockchain*, aponta para um possível futuro de preservação digital, que já é algo sendo bastante discutido nos meios das instituições de GLAM (*galleries, libraries, archives and museums*) como possível solução de armazenamento permanente de suas coleções digitais. O tempo dirá se isso acontecerá. Uma das preocupações da *blockchain* atualmente é o uso intensivo de poder computacional e consumo energético, que gera também preocupação sobre a atual situação mundial de mudanças climáticas e a pegada de carbono destas soluções. Sem condições de vida no planeta, também temos que nos questionar para quem estaremos preservando todos esses dados. Seria muito bom ter um rápido crescimento de uso de energias renováveis.

O custo de armazenamento de dados em *blockchain* hoje é absurdamente mais caro que as formas tradicionais que discutimos neste texto, tornando a questão da acessibilidade à preservação digital com esta tecnologia ainda mais elitista. Como muito do que está sendo discutido neste livro, temos que estar constantemente vigilantes e acompanhando esses desenvolvimentos.

Preocupações para preservação digital a longo prazo

Os benefícios de armazenamento na nuvem são muitos, principalmente pelo fato de uma pessoa, um grupo de pessoas ou uma instituição poder não ter o conhecimento, os recursos e o pessoal para implementar um processo seguro e confiável de armazenamento. Atribuir estas tarefas para uma empresa especializada nisto pode garantir que esta tarefa seja feita de forma mais profissional.

Porém, em termos de preservação digital, a nuvem, quando utilizada como armazenamento, deve ser vista sempre como uma das cópias, um dos sistemas, e não como a única forma de armazenamento, mesmo que utilizando 3 cópias em 3 nuvens distintas. Isso volta aos modelos e variantes do 3-2-1-1-0.

Difusão

O conceito de difusão usado neste texto tem o sentido de fazer com que os nossos acervos se espalhem por múltiplas direções, se tornem públicos e divulgados, sejam acessados, descobertos e conhecidos pela maioria.

Um conceito cada vez mais presente, tanto nas instituições de cultura como em empresas, envolvendo a preservação de dados digitais, é de que se não temos acesso ao que preservamos. Esses dados não poderão ser vistos ou utilizados de alguma forma, e, portanto, acabam por ser um tipo de *dark data* (dados obscurecidos), que são dados coletados e armazenados, mas não utilizados ou acessados para qualquer finalidade. Na maioria dos casos, são dados não descritos (metadados) e, conseqüentemente, não aparecem também em sistemas de buscas.

Podemos também fazer correlação com negativos ou *slides* armazenados e preservados, mas que, se não temos formas de acessá-los, e descobri-los, também se tornam dados obscurecidos. Esses dados existem e podem de fato estar em processos bem concebidos de preservação analógica e *backup* de dados digitais; não serão perdidos, mas também não terão visibilidade.

Os esforços para a preservação do que produzimos fazem sentido somente se esses materiais passam a ser conhecidos e acessados por meio da difusão. Para poder fazer a difusão, temos que ter programas bem consolidados de documentação/catalogação e preservação desses dados digitais. Desta forma, essa preservação existe com a finalidade de dar conhecimento ao que se preserva.

No livro *Theory and Practice of Digital Preservation*, de Trevor Owens (2018), o autor comenta, em um dos 16 axiomas sobre preservação digital:

Acumular e guardar não é preservação. É muito fácil começar a pegar muitos objetos digitais e fazer cópias deles. Isso não é preservação. Para realmente preservar algo, você precisa torná-lo descobrível e acessível, e isso exigirá que você tenha uma abordagem clara e coerente ao desenvolvimento de sua coleção, sua organização, descrição e métodos e abordagens para fornecer o acesso (tradução nossa).

A preservação de dados digitais exigirá um compromisso constante de alocação de recursos humanos e materiais. Assim, fazer a difusão do que estamos guardando também tem um papel vital de ajudar a garantir a sobrevivência desses materiais, o seu

financiamento e, conseqüentemente, a sua preservação. Nesse sentido, o relacionamento entre estas duas atividades, preservação e difusão, é fundamental e fortalece o ciclo em que uma alimenta a outra. Em projetos de digitalização de originais analógicos, gerando novos arquivos digitais, ou na preservação de fotografias nato-digitais, temos que pensar com clareza no ecossistema total, da guarda ao acesso/difusão do conteúdo que preservamos.

Querer nem sempre é poder difundir

Uma das decisões que se deve tomar sobre a difusão envolve entender o que pode ou não ser acessado. Isso abre uma ampla discussão sobre direitos autorais, que não fará parte deste texto. Dependendo do tipo de acervo, pessoal ou institucional, pode haver restrições de diversas ordens. A pessoa ou instituição que deseja fazer a difusão pode não ter os direitos autorais do criador das imagens, como também os direitos de imagem das pessoas retratadas, quando houver pessoas nas imagens. Pode haver cláusulas contratuais no recebimento de materiais de terceiros que restringem o acesso a uma ou mais imagens, como também questões de privacidade que impossibilitam o acesso.

Um excelente manual escrito por Mariana Valente e André Houang⁸⁰ (imagem a seguir), fala sobre o que precisamos saber sobre direitos autorais e as licenças Creative Commons. Com um movimento cada vez maior de acesso aberto, há um crescente desejo de maior compartilhamento de informações e conteúdo aberto. Entretanto, para isso, é necessário entender melhor as questões de direito.

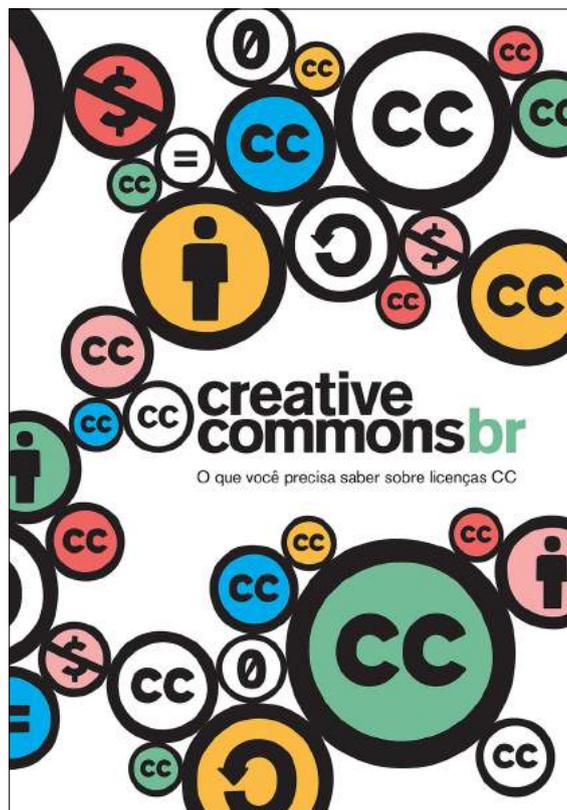


Figura 115. Publicação de Mariana Valente e André Houang, "O que você precisa saber sobre licenças CC".

Sergio Branco contextualiza o entendimento de Domínio Público, conceito jurídico essencial para que instituições de salvaguarda possam empreender a expansão do acesso livre às suas coleções por via digital nesse vídeo⁸¹, gravado durante o evento Abre-te Código, de 2020, do Instituto Goethe de São Paulo.

Como fazer a difusão

Para ter um sistema de acesso *on-line* para difusão, será necessário ter as suas fotografias digitais organizadas com as informações sobre elas em planilhas ou bases de dados que possam ser importadas em sistemas de catalogação, ou com os dados (metadados) embutidos nas imagens.

Existem diversas plataformas onde o material poderá existir, desde *sites* específicos de divulgação do acervo à difusão por meio das inúmeras plataformas das redes sociais, como canais no YouTube, publicações periódicas no Instagram, Facebook, Twitter, entre outras.

Código aberto

Esse é um assunto que vai além desta publicação, mas vale entrar um pouquinho nesse universo. Temos diversos sistemas que possuem código aberto, como Tainacan⁸² para repositórios de acervos digitais em WordPress e desenvolvido no Brasil. Fora do país, temos plataformas como ResourceSpace⁸³ de gerenciamento de bens digitais, ArchivesSpace⁸⁴, Omeka⁸⁵ e Collective Access, entre muitos outros. Uma das grandes vantagens de trabalhar com plataformas abertas é fazer parte de uma comunidade de usuários e usar sistemas que permitem acesso ao código de desenvolvimento, o que permite ampliar e modificar o código-fonte. O fato de não haver custos no licenciamento e no uso desses aplicativos não significa que não haverá custos para fazer o desenvolvimento e a adaptação da plataforma para atender às suas necessidades. Há sempre que considerar o custo de manutenção e o aprimoramento dessas plataformas ao longo dos anos de uso.

Sistemas proprietários

Adquirir um sistema pronto pode ter um custo maior, dependendo do sistema, e muitas vezes ele pode vir já com um contrato anual de manutenção. Dependendo do tamanho da instituição ou do grupo de usuários, pode ser interessante não ter um profissional ou profissionais dedicados internamente à manutenção de uma plataforma e arcar com os seus salários, e repassar esse custo para o próprio sistema proprietário. Isso tem que ser avaliado sempre levando em consideração o custo-benefício para o indivíduo ou instituição.

Algumas premissas sobre esse assunto devem ser consideradas. Sempre teremos que manter qualquer tipo de sistema implantado, seja de código aberto ou proprietário, e necessitaremos de recursos humanos e materiais para manter esses sistemas funcionando. São investimentos contínuos. Como tudo o que estamos discutindo no meio digital, a vida útil de uma solução adotada pode ser em torno de uma década, mas a própria evolução da instituição e das soluções nesta área poderá apontar para outros caminhos. O importante é ter acesso sempre aos seus dados e à sua base de dados para poder movê-los no futuro para outras possibilidades, se necessário ou desejado.

Como diremos a seguir, muitas iniciativas nas áreas de educação, cultura e *marketing* deverão acompanhar esta iniciativa de difusão.

Sustentabilidade da preservação e difusão

Um dos problemas de pensar em sistemas de difusão de coleções é o investimento inicial necessário para implantar e colocar as imagens na internet para acesso e encarar isso como a conclusão de um processo. A etapa intensa de organização e catalogação de suas imagens dá uma sensação de realização, mas, na verdade, colocar as imagens na internet é apenas o começo de um processo contínuo de crescimento, divulgação, ativação do conteúdo e manutenção financeira contínua de toda a estrutura necessária para esse acesso. Assim, pensar na energia, no tempo e nos recursos necessários para esse projeto de difusão antes de entrar é crucial para que os projetos sejam sustentáveis a longo prazo.

Os investimentos para a implementação de um projeto de digitalização ou de organização de fotografias nato-digitais e para a manutenção desses projetos são altos. Deve-se pensar antes e durante o desenvolvimento desses projetos nas ações institucionais voltadas para a utilização desses novos materiais digitais criados. Queremos gerar materiais para serem utilizados.

Para os projetos que pensam na difusão, colocar as imagens em um *site* como esforços para a disponibilização na internet não garante que serão de fato vistos e usados. Ao entrar no *Internet Live Stats*⁸⁶ no final de 2021, temos um pouco mais de 1,9 bilhões de páginas web.

Jonas Palm, ex-diretor do departamento de preservação dos arquivos nacionais da Suécia, escreve sobre isso em seu texto de 2006, *The Digital Black Hole* (O buraco negro digital)⁸⁷. Nesse texto, ele discute a vida de projetos digitais ou de digitalização, e os compara à vida de uma estrela.

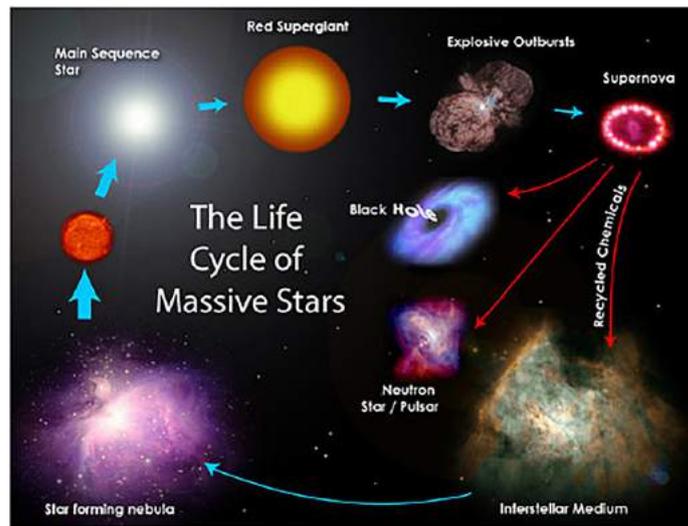


Fig. 1. The Life Cycle of Massive Stars (published on www.star.ucl.ac.uk/groups/hotstar/research.html).

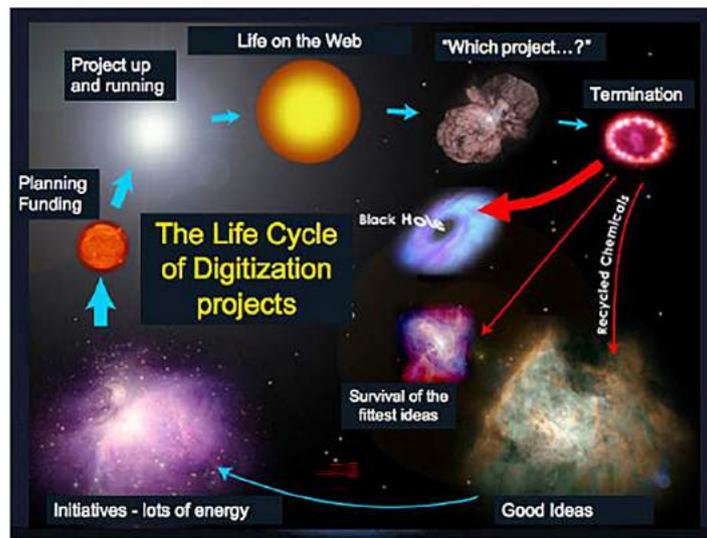


Fig. 2. The Life Cycle of Digitization Projects (modified from Fig. 1 by the author).

Figura 116. Captura de tela do texto de Jonas Palm, *The Digital Black Hole*, 2006.

Normalmente, quando o projeto é lançado ao público, existe uma grande convocação, mas se esse mesmo projeto não tiver ações que o sustentem em médio e longo prazo e que possam garantir a sua renovação e continuidade, ele corre o risco de implosão, como a morte de uma estrela formando um vácuo – um buraco negro, sem deixar rastros.

Entender também qual é o público-alvo que poderá e deverá se interessar por esses materiais também é algo que deve ser pensado. Ações educativas e de *marketing* também são essenciais para a criação e formação de um novo público

usuário. Por exemplo, as visitas a escolas, o treinamento de professores para que conheçam o acervo e suas coleções e a criação de projetos educativos que se utilizam da coleção são algumas, entre muitas, das ações que podem formar um público e fazer com que a instituição tenha um papel na sociedade na qual se insere. Isso não é feito sem planejamento, orçamento e equipes especializadas para a realização destas tarefas. Em muitos casos, os projetos de preservação de dados digitais são orçados e realizados sem pensar nestas etapas. Posteriormente, pode ficar complicado justificar, manter, financiar e realizá-las. Nem todas as instituições têm setores especializados de educação, *marketing* ou de difusão para cuidar desses projetos institucionais.

Outros materiais interessantes, que servem como difusão, são livros, exposições, artigos e seminários, entre outros, que ajudam a divulgar o trabalho. Essas ações são muito importantes, mas momentâneas. Podem gerar um *boom* inicial de acesso e conhecimento que pode não ser duradouro. Em pouco tempo, pode se esvaziar se ações mais concretas não estiverem previstas para o uso em longo prazo destas coleções.

6

Programa de preservação digital

Não podemos pensar na preservação de fotografias digitais sem adentrar no campo da preservação digital, que na era pré-digital, no século XX, era muito diversificado. Um conservador de papel acumulava experiências e *expertises* distintas de um conservador de fotografia, de pintura e de escultura. Algumas ferramentas e técnicas são compartilhadas, mas as áreas de pesquisa e tratamento desses campos e nesses diversos materiais requerem profissionais especificamente treinados. Isso será diferente ao entrar em dados digitais. Temos especificidades em relação aos cuidados de preservação de arquivos digitais de texto, áudio, audiovisual, imagens, e outros tipos de arquivos mais complexos, mas no fim são todos compostos de 0s e 1s, pelo menos nesse momento da nossa história.

Ao cuidarmos de fotografias digitais, cuidaremos desses 0s e 1s e, nesse sentido, poderemos também cuidar desse universo digital como um todo, pois muito do que discutimos neste texto se aplica às fotografias digitais, como também ao universo de todos os dados digitais que produzimos. Assim, temos que avançar no campo da preservação digital se quisermos que nossas fotografias digitais, entre outros documentos digitais, sobrevivam através do tempo.

Não está no escopo deste texto entrar no vasto e importante campo da preservação digital, mas podemos trazer alguns pontos básicos que conectam este texto a esse universo.

Vamos voltar para a tabela criada pelo NDSA (*National Digital Stewardship Alliance*) sobre os níveis de preservação digital⁸⁸, que discutimos no capítulo 5. Essa tabela é uma ótima referência para discutir as melhores práticas e servir como ponto de partida para um projeto maior de preservação digital.

Temos cinco áreas funcionais: Armazenamento, Integridade, Controle, Metadados e Conteúdo, com os 4 níveis de aprofundamento, do básico ao avançado: Nível 1 – conheça o seu conteúdo, Nível 2 – proteja seu conteúdo, Nível 3 – supervisione seu conteúdo e Nível 4 – sustente o seu conteúdo. Essa tabela pode servir de guia para fazer uma avaliação do seu estado atual de preservação digital. Veremos adiante que necessitamos avançar além dessa tabela, mas, se ao menos for possível atingir as metas colocadas nesse tópico, já estaremos fazendo um bom trabalho inicial. Em um texto da OSF – Open Science Framework⁸⁹, demonstra-se como criar critérios de pontuação para cada item. Selecionamos 2 se a organização já alcançou o item, 1 se estiver trabalhando no item e 0 se não iniciou ainda. Ao olhar esta matriz, com suas notas, saberemos melhor as áreas que necessitam de melhorias e implementação de processos de trabalho. Tudo isso tem que ser visto como um processo contínuo. Não compramos ou instalamos processos de preservação digital, eles são construídos continuamente ao longo do tempo. Estaremos sempre em desenvolvimento e evolução em um processo de consolidação destas etapas e outras de preservação digital.

Portanto, temos que garantir um armazenamento seguro, ter certeza da integridade dos dados armazenados, cuidar da segurança e do acesso aos dados, ter metadados sobre todos os dados que estamos armazenando e cuidar do nosso conteúdo (formatos de arquivos) para garantir o acesso a ele. Contudo, está faltando uma etapa fundamental nesta tabela, que é a de difusão e acesso ao conteúdo.

Em 2016, Shira Peltzman⁹⁰, da Biblioteca da UCLA, publicou um texto no *blog The Signal*⁹¹, em que ela aponta a necessidade de um sexto item para esta tabela do NDSA: acesso. Vale a pena ler o texto dela, pois o acesso ao conteúdo preservado é fundamentalmente a razão pela qual estamos preservando e fazendo tudo o que estamos fazendo. Para poder dar acesso a esses conteúdos, estes precisam ser preservados, como também os esforços (e os custos) da preservação existem para

dar esse acesso. Sem poder ver, usar, e integrar o que estamos preservando nas nossas atividades diárias da cultura, de estudo, pesquisa e lazer, entre outros, fica difícil reivindicar e manter os recursos humanos e materiais para a manutenção de coleções como *dark data*.

Na resenha que escrevi em 2020 para a *Revista Eletrônica de Comunicação, Informação e Inovação em Saúde*⁹², comento sobre mais alguns axiomas de Owens. Reforço alguns a seguir.

Trevor Owens⁹³ informa que “*Highly technical definitions of digital preservation are complicit in silencing the past*” (Definições altamente tecnológicas sobre a preservação digital são cúmplices em silenciar o passado). Trevor menciona diversas vezes em seu livro a questão do silenciamento do passado. Quando pensamos na produção digital e no comprometimento necessário para sua preservação, fica fácil entender que há uma relação direta entre acesso à informação, recursos e *know-how* para a realização de um programa de preservação digital. Somente as pessoas ou instituições mais abastadas poderão fazer isso. As outras pessoas e instituições menos abastadas correm o risco de terem suas histórias silenciadas. O entendimento de que esse campo é complexo e requer todos esses recursos acaba por afastar e intimidar as instituições com menos conhecimento, recursos humanos e materiais. Temos que pensar em soluções mais simples, e bons programas de *backup* para garantir a longevidade dos dados digitais como começo para um programa de preservação digital, envolvendo todas as instituições e evitando, assim, o silenciamento destas histórias e desses dados. Um programa de preservação digital começa com pequenos passos e deve evoluir e crescer através do tempo. Mesmo pequenos passos iniciais podem fazer toda a diferença para garantir a sobrevivência de nossos materiais digitais.

Em outro axioma do Owens, vemos que “*Backing up data is not digital preservation*” (Fazer *backup* dos dados não é preservação digital). A preservação digital pensa nos objetos digitais em longo prazo, para que possam continuar sendo descobriáveis e acessíveis no futuro. Isso implica pensar também nos formatos, na migração, na fixidez, e em repositórios digitais seguros, entre outros. Os procedimentos de *backup*, normalmente realizados por equipes de TI, não visam esse alcance. O processo de armazenamento de dados digitais usará procedimentos de *backup*, mas esse será apenas um dos componentes de um universo maior de atividades de preservação digital em longo prazo. Por outro lado, conforme visto no axioma anteriormente, sem

pelo menos a implantação de um bom sistema de *backup*, não poderemos avançar para outros processos de preservação digital. Será necessário investir na educação para implementação de sistemas simples e seguros.

Quando pensamos o termo “em longo prazo”, somente as instituições que atravessam gerações têm condições de garantir a longevidade de todos os materiais da nossa herança cultural – bibliotecas, arquivos, museus, famílias, organizações religiosas, culturais e políticas (governanças em diversas capacidades) etc. É necessário fortalecer essas instituições como forma de garantir a preservação que não conseguimos garantir em nível individual. Não posso contar com os meus filhos e os seus filhos para preservar todo o conteúdo digital que eu estou criando e preservando. Se isso não estiver amplamente organizado e documentado (metadados), será também muito trabalhoso para qualquer instituição receber esse material a fim de garantir a sua preservação.

Os nossos recursos não são ilimitados e, portanto, temos que fazer o melhor uso deles para resolver as situações de maior risco e ameaça de preservação. Owens reforça o fato que a preservação digital não é algo que pode ser realizado e/ou finalizado. É sempre algo em construção – um processo contínuo. Temos que entender os riscos e tomar decisões para resolver as ameaças mais urgentes de preservação, sabendo que não teremos como salvar tudo. Acrescento também que não podemos e nem queremos salvar tudo.

7

Do digital ao analógico

Para iniciar esta discussão, podemos introduzir esse assunto com uma seção do livro publicado pelo AIC – *American Institute of Conservation* em 2011 (2ª edição), *Guide to Digital Photography & Conservation Documentation*⁹⁴ (Guia para a fotografia digital e documentação sobre conservação). Nesse livro são citadas as 5 etapas para preservação de fotografias digitais em longo prazo. A última opção, que ainda não foi discutida neste texto, é a de fazer impressões. Assim, os dados digitais, informação digital ou 0s e 1s, se concretizam em um processo analógico e, portanto, passamos agora a ter um representante físico, tangível, analógico, que passa também a representar o arquivo digital. O texto do AIC propõe um trabalho elaborado de edição e escolha de imagens no intuito de imprimir uma parte do todo, e nunca imprimir o todo. A impressão também não substitui rigorosos processos de preservação dos objetos digitais, mas é uma forma adicional de preservação da obra. A impressão sempre foi vista na fotografia pré-digital como a consolidação final da obra, negativo ou *slide*. Ansel Adams é conhecido por dizer que “*The negative is the equivalent of the composer’s score, and the print the performance*” (O negativo é equivalente à partitura do compositor, e a cópia à sua execução). A fotografia digital pode ser vista em inúmeros dispositivos de

transmissão de luz, como telas de todos os tipos (celulares, *tablets*, projetores, projeções, *notebooks*, *desktops*, *displays* luminosos etc.), mas é somente quando ela é impressa em uma superfície onde veremos a imagem por luz refletida, é que teremos um *print*, uma versão da execução da partitura do compositor. Em português, temos sempre dificuldade de usar algo correlato para *print*. Se é uma ampliação ou cópia, entendemos mais como um processo analógico. Já uma impressão é entendida como um processo digital, jato de tinta, por exemplo, ou impressão em processos gráficos.

Para as pessoas que já trabalharam em arquivos fotográficos, é muito comum encontrar cópias em papel fotográfico, mas, em muitos desses casos, o negativo não se encontra mais. Isso também acontece com fotografias familiares, onde as cópias/*prints* estão preservadas, mas os negativos que lhes deram origem desapareceram. No meio da produção nato-digital, podemos também não ter mais esses “negativos” preservados, como também a versões derivadas impressas em algum tipo de material físico. Esses originais digitais podem até existir, mas em algum lugar incerto ou desconhecido.

5 etapas para preservação digital a longo prazo

AIC – American Institute of Conservation

1. Usar arquivos de imagem sustentáveis
2. Organizar → estrutura de pastas, nomeamento e metadados
3. Backup de seus dados (isto inclui armazenamento, para ter algo para fazer backup!)
4. Verificar seu backup → fixidez
5. Fazer impressões → do digital ao analógico

Figura 117. Quadro com as 5 etapas para preservação digital a longo prazo apresentado pelo Instituto Americano de Conservação (AIC).

Esta seção irá discutir a migração para o analógico como forma adicional de preservação, não nos isentando de trabalhar com os cuidados e preservação também do original digital.

Impressão

O universo das técnicas de impressão de fotografias digitais é muito grande, e não vamos nos ater a uma descrição mais completa, tampouco ao histórico desses processos. Se considerarmos a impressão como uma etapa adicional para ajudar na preservação e na difusão da fotografia digital, devemos procurar processos que sejam mais duradouros, com mais resistência à exposição luminosa, e que possam reproduzir da melhor forma e qualidade as cores ou tonalidades do arquivo digital.

Desta forma, vamos discutir quatro assuntos: a impressão em jato de tinta com pigmentos minerais, a impressão comercial digital, o armazenamento em DNA e outras possibilidades futuras. Certamente esta é uma área que temos que acompanhar pela rápida evolução de sistemas e soluções.

Impressão em jato de tinta com pigmentos minerais

Na década de 1990, explode o universo de impressoras domésticas e comerciais de jato de tinta. A qualidade para o dia a dia era razoável, mas, para fotografias digitais (vindas nesta época principalmente de processos de digitalização de originais analógicos) as imagens resultantes não eram de grande qualidade, como também durabilidade. As cores eram muito fugitivas (do inglês *color fugitive*, ou cores que fogem ou esmaecem com a ação da luz, principalmente de raios UV). Uma impressão poderia ficar quase transparente com o sumiço das cores. O curioso é que a história se repete como os primeiros processos de papéis fotográficos coloridos da Kodak da década de 1950.

A Epson lançou, no ano 2000, duas impressoras com uma nova tecnologia de impressão, utilizando soluções aquosas com micropartículas de pigmentos minerais. A versão menor de mesa, a Epson 2000, e a versão maior, a Epson Stylus Pro 9500. Os pigmentos minerais prometiam trazer durabilidade e resistência ao esmaecimento das cores, ao contrário dos corantes sendo utilizados nas impressoras de jato de tinta. A proposta era criar uma impressora para o mercado de arte (*fine art*), e oferecer uma opção de impressão para fotografias digitais com qualidade de cor e durabilidade. Rapidamente chegaram outras versões com melhoria na qualidade e durabilidade das tintas, e outras empresas com impressoras utilizando pigmentos

minerais que se adicionam à Epson, como a HP e a Canon. Hoje ainda temos os mercados de impressoras jato de tinta, que utilizam corantes, que melhoram muito na qualidade e durabilidade, são de menor custo, mas não têm uma funcionalidade relevante na produção de impressões digitais para longevidade.

Pesquisas sobre a longevidade passaram a ser realizadas principalmente pelo Henry Wilhelm, pela Wilhelm Imaging Research⁹⁵. Esse *site* é um universo de informações dedicadas à preservação de materiais fotográficos, com centenas de artigos e um importante livro que inicia o trabalho de Henry Wilhelm, com contribuições de Carol Brower: *The Permanence and Care of Color Photographs: Traditional and Digital Color Prints, Color Negatives, Slides, and Motion Pictures*, versão original de 1993⁹⁶. Henry continuou suas pesquisas na durabilidade de materiais contemporâneos e, nesse *site*, apresenta estudos com diversos tipos de tintas, corantes e pigmentos minerais, com impressoras e papéis diferentes, por meio de equipamentos de envelhecimento acelerado, onde as condições de temperatura, umidade relativa do ar e luminosidade podem ser controladas para criar condições que simulam o envelhecimento projetado para 50, 100, 200 e 300 anos.

Nesses estudos, as condições são sempre constantes, mas muito diferentes das condições de guarda e uso do dia a dia, com flutuações nesses valores, inclusive quando obras são movimentadas dentro e fora de áreas climatizadas controladas. De qualquer forma, são extremamente válidos como metodologia de comparação de um sistema a outro. As amostras são visualizadas diariamente na busca do *noticeable fading* (esmaecimento perceptível). Quando um observador perceber que a amostra sofreu esmaecimento ao compará-lo à imagem de referência, isso marcará o início da perda de cor. Desta forma os cálculos de meses ou anos se referem ao tempo limite sem esmaecimento perceptível, e em qualquer tempo adicional devemos esperar esmaecimentos. Como um objeto analógico, o esmaecimento se inicia, se torna perceptível, mas a impressão ainda terá muita vida pela frente, pois é um processo de declínio lento.

Os testes não levam em consideração outros fatores importantes, como poluentes atmosféricos, extremamente comuns em centros urbanos. Mesmo com sistemas de guarda com filtragem do ar, temos que considerar a exposição a esses poluentes, como sulfatos, nitratos, amônia, cloretos e VOCs (compostos orgânicos voláteis). Estes, misturados com gases como ozônio, óxidos nítricos de enxofre e monóxido de carbono, podem gerar aerossóis ambientais. Sem entrar na discussão do impacto

negativo à saúde humana desses compostos, sabemos, na conservação, do efeito prejudicial na durabilidade de objetos pelo efeito degradante dos poluentes ambientais. Essa é uma área que necessita de mais pesquisa em relação à longevidade de impressões com pigmentos minerais. Sabemos, desde já, pela experiência, que impressões sem vidro em ambientes urbanos poluídos sofrem esmaecimentos com grande intensidade e velocidade.

No artigo *A Survey of Print Permanence in the 4x6-Inch Consumer Digital Print Market in 2004-2007*, escrito por Henry Wilhelm em 2007⁹⁷, podemos analisar a longevidade das cores de uma pequena impressora doméstica. Na primeira linha onde são utilizadas as tintas Vivera (a linha de pigmento mineral da HP) com papel HP, a previsão é de não ter esmaecimento perceptível em até 25 anos. Na primeira fileira de imagens onde são utilizadas as tintas Vivera (o pigmento mineral da HP) com papel HP, a previsão é de não ter esmaecimento perceptível em até 25 anos. A segunda fileira de imagens é impressa com refil de tintas à base de corantes e papéis da loja *Office Depot*, e a terceira fileira de imagens é impressa com refil de tintas à base de corantes e papéis da loja *Staples*.

Fica evidente nesse artigo o nível de esmaecimento sofrido pelas tintas à base de corantes.

Em outro estudo⁹⁸ sobre os pigmentos minerais Ultrachrome Pro da Epson nas impressoras P10000 e P20000, podemos destacar alguns conceitos. O documento completo deve ser lido para se ter um entendimento mais amplo de todas as questões abordadas aqui, inclusive com as referências de parâmetros para os itens citados no estudo. A longevidade das tintas está diretamente relacionada com o tipo de papel. Papéis diferentes reagem de forma diferente em relação à longevidade. Ter vidro protegendo a obra também aumenta a longevidade das cores, e com vidros que absorvem radiações UV, o tempo de vida da imagem é ainda maior. Exibir sem vidro reduz muito sua sobrevida. Manter temperatura baixa (em torno de 20° C) e umidade relativa do ambiente em torno de 50% ajudará a garantir sua longevidade. Altas temperaturas e UR do ar acelerarão os processos de esmaecimento. Umidades relativas do ar acima de 60-65%, dependendo da temperatura, já podem contribuir para germinação de esporos de fungos nestes materiais. As tabelas continuam discutindo questões como resistência ao ozônio, alta umidade, água e se o papel possui ou não branqueadores ópticos. Os tempos discutidos são sempre os tempos em que um observador não observará mudanças perceptíveis.

Com esses dados, podemos afirmar que as impressões exibidas debaixo de vidro com proteção contra radiação ultravioleta (filtro UV) são mais resistentes do que vidro comum. Exibir sem vidro diminui mais ainda o tempo de longevidade. Novamente, não usar vidro não está levando em consideração a exposição a poluentes atmosféricos. Portanto, a longevidade sem esmaecimento perceptível é uma combinação da tinta pigmentada e do tipo de material em que esta é depositada, em conjunto com fatores muitas vezes difíceis de controlar, como exposição à luz, variações de temperatura e umidade relativa do ar, exposição a elementos poluentes e oxidantes, manuseio, entre outros. Guardar as imagens em um acervo climatizado controlado com baixa umidade relativa do ar e temperatura, com filtragem do ar, aumentará em muito a longevidade dessas impressões, porém isso limitará sua difusão e sua visualização.

A maior parte das impressoras tem pelo menos 3 pretos (um preto para papéis foscos ou brilhantes, um preto mais claro, e um bem mais claro). Na impressão de imagens em preto e branco, podemos usar todas as tintas da impressora (em torno de 8 a 12, de acordo com o modelo), mas, como os pigmentos minerais utilizados nas tintas pretas são extremamente estáveis por serem pigmentos minerais à base de carbono e resistentes ao esmaecimento, podemos imprimir imagens preto e branco somente com estas tintas. Uma parte muito pequena das outras tintas também é utilizada nesse modo. Há um aumento substancial nos números de anos antes do aparecimento de esmaecimento perceptível nesse modo de impressão. Existem também empresas, como a Cone Editions⁹⁹, que produzem uma linha completa de tintas pigmentadas monocromáticas para as impressoras jato de tinta. A impressora é transformada para esta configuração.

Uma das discussões sobre esses processos de impressão é encontrar o equilíbrio entre preservação e difusão, e o que configura a obra. Fazer essas impressões e guardá-las em um acervo climatizado é uma ótima opção para preservação a longo prazo, mas, se queremos difundir, exibir, tornar acessíveis estas impressões, temos que exibi-las, expô-las à luz, o que começa a reduzir o tempo de vida da obra. Na era do filme, os negativos em preto e branco e coloridos são apresentados como obra pelas tiragens de cópias em papel. O *slide* colorido poderia ser utilizado como obra em projeções e audiovisuais, como também em ampliações em processos indiretos (com internegativos) ou diretos (como Cibachrome originalmente, depois adquirido pela Ilford e renomeado Ilfochrome¹⁰⁰).

Como isso se dá com fotografias nato-digitais? Elas podem existir de inúmeras formas e ser impressas, mas nem sempre para a função de preservação. Existem inúmeras tecnologias de impressão e suportes que produzem resultados excelentes, mas não são duráveis. Impressão em tecidos, vidros, acrílicos, metais, lonas e outros materiais com tintas corantes ou tintas curadas (secas) por radiação ultravioleta (conhecidas com impressoras UV, impressoras de tinta UV, ou tinta de cura UV). O resultado de processos de difusão, como exposições ou instalações, com esses tipos de impressão pode ser de grande impacto, mas não de permanência. Portanto, esses materiais poderão ser utilizados como derivadas da obra. São conceitos novos, que exigirão novas abordagens. Dentro de um conceito de preservação, podemos pensar em preservar as matrizes digitais, toda a documentação sobre como essa obra pode e deve ser utilizada e exibida, sempre que possível construída com a participação do autor da obra, para poder replicar o processo de difusão no futuro e sob demanda.

Papéis fotográficos analógicos coloridos

Ainda temos alguns laboratórios fazendo impressões em papel fotográfico colorido. Chamamos esse tipo de papel de fotoquímico, porque usa um processo de exposição à luz, hoje realizado principalmente por máquinas que utilizam lasers RGB para realizar a exposição no papel, e porque o papel exposto é revelado em processo químico conhecido como RA-4 da Kodak ou CP48S e CP49E da Fuji.

Na campanha da Kodak Alaris, em sua aba de consumidores, há um sistema *on-line* de envio de arquivos digitais e impressão em papéis fotoquímicos da Kodak. É interessante ver como a Kodak promove a impressão em papel fotográfico verdadeiro. A menção do termo *true* (verdadeiro) é um comentário para inúmeros processos de impressão de fotografias em papel que utilizam processos comerciais de baixa qualidade e, principalmente, baixa durabilidade.

O texto na imagem a seguir: “Faça as memórias durarem! Nuvens se dissipam (com referência ao armazenamento em nuvem), Celulares são perdidos e Discos rígidos se corrompem”, e um texto final afirmando que somente papel fotográfico verdadeiro é o único método comprovado para preservar memórias.

Nós temos que desenvolver processos de preservação de dados digitais em que possamos confiar como forma de preservação das nossas fotografias digitais, porque imprimir tudo o que temos não é a solução.

MAKE MEMORIES LAST!





Clouds drift off



Cell Phones get lost



Hard drives crash



True photo paper is the only proven method to preserve memories



Figura 118. Propaganda da Kodak Alaris dizendo que somente papel fotográfico verdadeiro é o único método comprovado para preservar memórias. Entendemos verdadeiro, neste contexto, como sendo papel fotoquímico.

Kodak Professional HOME FEATURES KEY ADVANTAGES PRODUCTS VIDEOS MARKETING HELP GET SOME Kodak Professional Main Site

Kodak Professional
Enjoy Quality For Life



FAMILY OF PRODUCTS





CONSUMER
MORE FEATURES
MID-LEVEL PROFESSIONAL
MORE PERFORMANCE
PREMIUM PROFESSIONAL

Ektacolor Edge Paper
Royal Digital Paper

Consumidor

Elite Paper
Pro Image Paper
Premier Digital Paper
Premier Digital II Paper
Photo Book Paper

Endura Premier Paper
Endura Metallic Paper
Brilliance Paper

Profissional

Figura 119. A Kodak produz papéis de boa qualidade como a linha profissional Endura, e a versão Edge para consumidores. Toda a linha de papéis tem o selo de True Photo Paper.

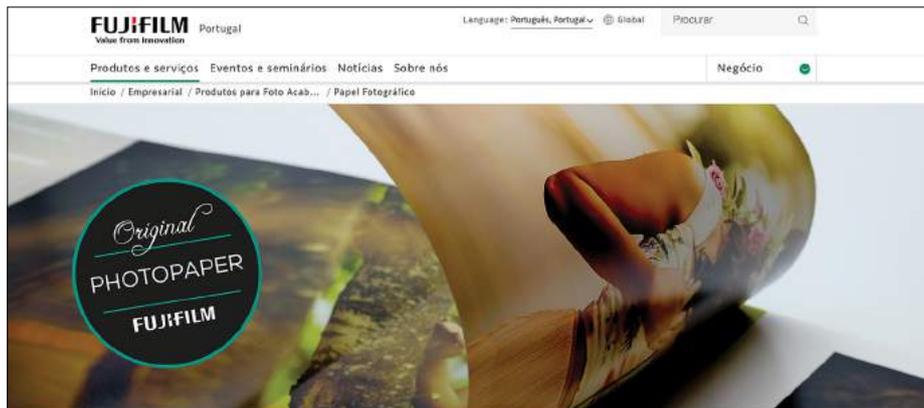


Figura 120. A Fuji produz um excelente papel chamado Crystal Archive, ou Fujicolor CA. O selo utilizado pela Fuji reforça que é papel fotográfico original. Reprodução do site da Fujifilm.

Em uma busca pela internet com muitas propagandas de impressão de lotes de arquivos digitais em papel fotográfico, temos duas empresas que anunciam seus serviços com papel Kodak e Fuji. Ao entrar em contato com estas empresas, a informação é que o papel da Kodak é o Edge e o da Fuji o Crystal Archive.

É necessário fazer perguntas sobre o tipo de papel, o processo utilizado, sobre se a empresa é consistente com o uso e os testes dos materiais, e criar um relacionamento com a empresa para entender como os arquivos digitais são tratados e que tipo de processamento é aplicado aos arquivos. Algumas empresas podem aplicar processos automáticos que podem alterar o arquivo recebido com aumento de contraste, saturação e nitidez. O ideal é solicitar um tratamento neutro, ou sem alterações, para que você possa entender como seus arquivos estão sendo impressos com os seus ajustes.

<p>Compra pelo Mercado livre</p> <ul style="list-style-type: none"> • Qual papel? • Qual processo? • É consistente? • Entender os procedimentos • Travar relacionamento? 	<p>Papel Fuji?</p>	<p>Revelação De 80 Fotos Digitais 15x21 Fosco/brilho Fujifilm</p> <p>R\$ 380</p> <p>Estoque disponível</p> <p>12x R\$ 299 sem juros</p> <p>VISA MESTRO</p> <p>Ver as opções de pagamento</p> <p>Entrega a combinar com o vendedor</p> <p>Envio: São Paulo</p> <p>Ver custos de envio</p> <p>Quantidade: 1 unidade de 1 (10 disponíveis)</p>	<p>Fuji Crystal Archive</p>
	<p>Papel Kodak?</p>	<p>Revelação De 100 Fotos Tamanho 10x15 Cm Papel Kodak Com Frete Grátis</p> <p>R\$ 59⁹⁰</p> <p>Estoque disponível</p> <p>10x R\$ 59⁹⁰ sem juros</p> <p>VISA MESTRO</p> <p>Ver as opções de pagamento</p> <p>Frete grátis</p> <p>Carapina, São Paulo</p> <p>Ver custos de envio</p>	<p>Kodak Edge</p>

Figura 121. Imagem montada pelo autor com produtos vendidos no Mercado Livre.

Por fim, em um dos anúncios da Fuji, notamos uma comparação interessante entre impressão e revelação. A revelação é entendida como processo fotoquímico, molhado, com papel fotográfico tradicional (verdadeiro). A impressão, nesse caso, são os outros processos comerciais de baixa durabilidade que não usam papel fotoquímico.



Figura 122. Exemplo de propaganda de revelação fotográfica.

Impressão comercial digital

A produção de livros também é uma forma de consolidar trabalhos fotográficos. Livros têm vida longa e não dependem de qualquer tipo de tecnologia para serem vistos, e, portanto, nunca serão obsoletos. Porém, como são produzidos normalmente em escalas menores, seu alcance de difusão é mais limitado. Existem muitos projetos de climatização passiva para a construção de bibliotecas sendo feitos hoje com redução do uso de energia, usando iluminação natural, ventilação natural, isolamento térmico e a inércia térmica predial.

Um livro impresso com tiragens mais altas, normalmente acima de 500 a 1.000 exemplares, será mais econômico se for feito em tecnologia *offset*, a forma tradicional de impressão. Os custos para um projeto de impressão *offset* são altos, mas o

custo unitário por livro cai muito quando estamos falando de milhares ou dezenas de milhares de unidades.

A outra forma de impressão é a digital, em que um livro pode ser rapidamente impresso e em pequenas quantidades. A impressão digital criou o conceito da impressão sob demanda, no qual as tiragens são feitas de acordo com a necessidade, pelo fato principal de que o custo unitário da impressão do livro é quase sempre o mesmo, independentemente da tiragem.

No gráfico a seguir vemos uma comparação entre o custo unitário da impressão e a sua quantidade nos dois métodos de impressão. A impressão *offset* é muito cara para baixas quantidades, mas muito mais econômica para grandes quantidades. A impressão digital é quase uma linha reta, com um custo levemente maior para pequenas quantidades. O ponto de equilíbrio entre estas duas tecnologias é algo em torno de 500 ou 1.000 unidades, dependendo do tipo de tecnologia utilizada, mas isso significa que podemos pensar em projetos de livros com tiragens abaixo de 100, algo impensável na era da impressão *offset*. Como forma de fazer a distinção entre impressão digital e um parque gráfico com impressora digital, é mais fácil usar o termo de Gráfica Digital e Gráfica Offset (em inglês a distinção de *digital press* e *offset press*).

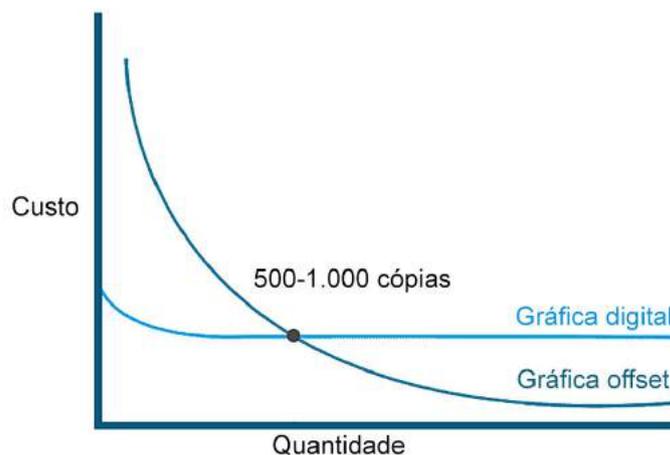


Figura 123. Comparação de custo e quantidade entre impressões *offset* e digitais.

Os custos de impressão *offset* despencam quando as quantidades estão acima de 500-1.000 exemplares. Abaixo de 500 os custos são muito atraentes para impressões digitais.

O número exato varia dependendo da impressora digital concorrente.

Gráfico elaborado pelo autor.

A qualidade de impressão no digital sempre foi abaixo dos resultados obtidos em impressões *offset*. Hoje, somente um olho bem treinado é capaz de discernir diferenças de qualidade entre impressões *offset* e digital.

Duas empresas do setor de impressoras digitais, a HP e a Xerox, produzem impressoras digitais (*digital press*) de grande qualidade.

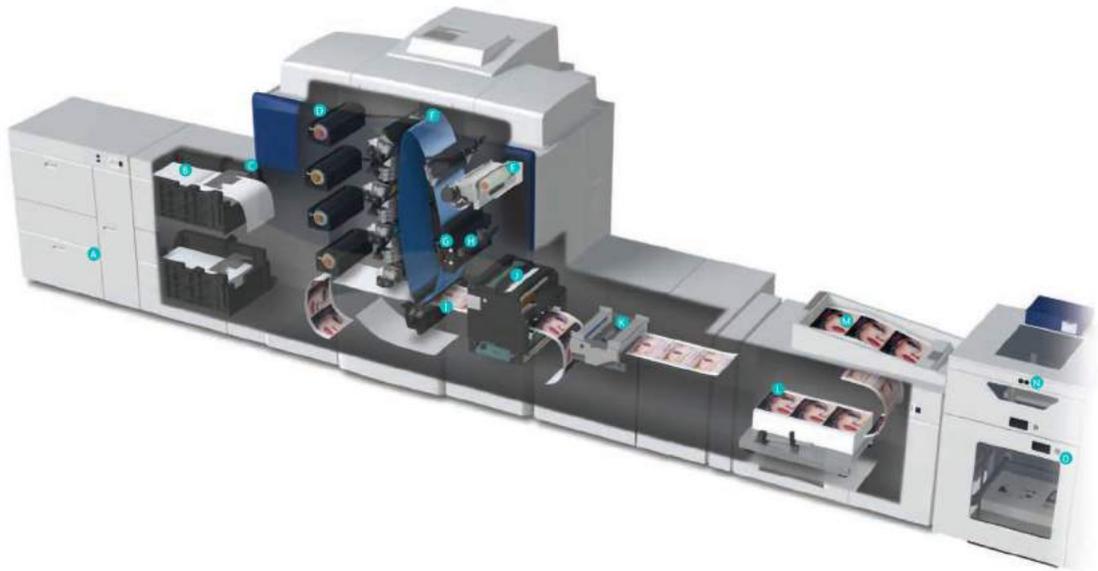


Figura 124. Gráfica digital iGen5 da Xerox.



Figura 125. Gráfica digital HP Indigo 12000 da Hewlett Packard.

Henry Wilhelm, no seu site, tem estudos de durabilidade das cores (antes do esmaecimento perceptível) dos *toners* utilizados na impressão nos dois processos. Tanto a Indigo, da HP¹⁰¹, quanto a iGen5¹⁰², da Xerox atingem acima de 200 anos quando armazenados sem exposição à luz, desde que sejam respeitadas as condições especificadas nos testes. Esses valores são maiores do que os obtidos nos papéis Fuji Crystal Archive e Kodak Edge.

Também podemos utilizar esses equipamentos para imprimir as nossas fotografias. Não é um processo fotográfico/fotoquímico, é um processo de tinta (*toner*) no papel. Existem muitos sistemas *on-line* em que você pode enviar o seu PDF e escolher os modelos, acabamentos, tamanho e quantidade de livros. Conhecer a empresa e entender como lidam com os seus arquivos, qual o perfil de impressão que utilizam e que tipo de tratamento é ou não é feito às imagens o ajudará a prever melhor os resultados de seu trabalho. O ideal é sempre começar com pequenos projetos.

Uma desvantagem das impressoras digitais é que o tamanho máximo do papel é menor. Em gráficas *offset* é algo em torno do A0 (84,1 x 118,9 cm), e na maior parte das impressoras ou gráficas digitais é em torno de 53 x 75 cm. Outra vantagem da impressão digital, não disponível na impressão *offset*, é a possibilidade de utilização de dados variáveis. Os textos e imagens podem estar atrelados a uma base de dados e *script*, e isso significa que cada livro ou impressão poderá ser impresso de forma única, com dados personalizados fornecidos pela base de dados.

Back to film – laser film recorder

Com a evolução digital em cinematografia, muitas das produções passaram a ser inteiramente digitais, mas, ao se finalizar um filme digitalmente, uma das saídas para a produção de cópias de exibição a partir de um negativo original é gravar os dados digitais com lasers RGB (vermelho, verde e azul) em películas fotográficas com posterior revelação fotoquímica. Assim estabelecemos o conceito de voltar para o filme – *back to film*. O resultado é esse material fílmico proveniente dos dados da produção digital. Com o aumento de salas com projetores digitais, o uso de cópias de exibição em película despencou, mas os acervos e grandes estúdios de cinema passaram a usar o *Laser film recorder* para gerar um *backup* da produção digital em formato analógico como forma de preservação de uma matriz da produção.

Como o filme em preto e branco, em base de poliéster bem processado, tem alta durabilidade em acervos climatizados, com maior longevidade em relação a

filmes coloridos, alguns estúdios estão gerando separação da matriz colorida em seus componentes ciano, magenta e amarelo. No futuro, estas matrizes em preto e branco poderão ser utilizadas para gerar uma imagem colorida. Não existem experiências de gravar arquivos de fotografias digitais em película com esse processo, mas certamente seria uma das possibilidades de conversão de um formato digital para analógico, como vem sendo feito com imagens em movimento.

Armazenamento em DNA

O DNA ou ácido desoxirribonucleico, é uma longa molécula que contém nosso código genético único. Como um livro de receitas, ele contém as instruções para fazer todas as proteínas em nossos corpos. Seu genoma? É feito de uma substância química chamada ácido desoxirribonucleico, ou simplesmente DNA¹⁰³.

ACTG – Adenina, citosina, timina e guanina. Essas moléculas, ou letras, são responsáveis pela estrutura do nosso DNA no formato de dupla hélice, descoberta publicada em 1953 pelo inglês Francis Crick e o americano James Watson. Em 1962, a dupla recebeu o Prêmio Nobel de Medicina e Fisiologia, que os tornou dois dos cientistas mais importantes da história moderna.

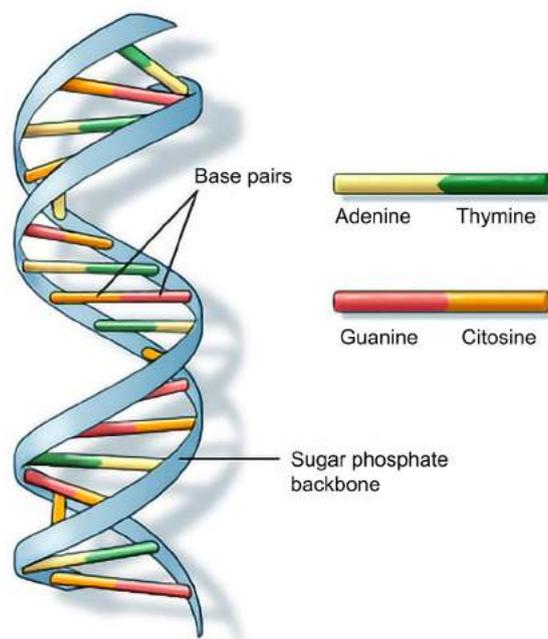


Figura 126. Estrutura do DNA com ACGT.

Em 2012, o bioengenheiro Sri Kosuri e o geneticista George Church, do Instituto Wyss da Harvard University armazenaram, e recuperaram 700 terabytes de dados em um grama de DNA, utilizando as moléculas de ACGT. Eles salvaram 70 bilhões de cópias do próximo livro de Church, um conteúdo de aproximadamente 700 TB.

No modelo de Church, feito no anúncio de sua descoberta em 2012, os dados binários, 0s e 1s, são codificados em sequência de moléculas ACGT. Esta sequência é sintetizada para fitas de DNA, que são depois encapsuladas e armazenadas por tempos indeterminados. Para recuperar a informação, o DNA é liberado do encapsulamento, as fitas de DNA são sequenciadas em moléculas ACGT, que posteriormente são decodificadas em código binário de 0s e 1s. Esse processo de sintetizar e sequenciar DNA representa a maneira de armazenar dados binários em DNA.

O DNA é estável à temperatura ambiente. “*Você pode largá-lo onde quiser, no deserto ou no quintal, e ele estará lá 400.000 anos depois*”, disse Church¹⁰⁴.

Desde a experiência de Church e Kosuri, estamos assistindo a inúmeras outras experiências e mais avanços nesse processo. Hoje já é uma tecnologia viável, mas longe ainda de ser uma tecnologia de *desktop*, acessível e de uso disseminado.

Em um artigo¹⁰⁵ de março de 2021, escrito por Sarah Vitak, a autora diz que:

O universo digital pode adicionar cerca de 175 zetabytes de dados por ano até 2025, de acordo com a empresa de análise de mercado IDC. Isso é 175 com 21 zeros depois (1.000 terabytes equivalem a um petabyte, e 1.000 petabytes equivalem a 1 exabyte, e 1.000 exabytes equivalem a 1 zetabyte). Essa quantidade de informações exigirá enormes centros de dados e vastos recursos de energia para manter. Um pequeno, mas crescente grupo de pesquisadores defende o DNA como um substituto estável e sustentável.

Esses esforços aumentaram em novembro passado, quando uma coalizão de empresas de computação e biotecnologia, incluindo Microsoft, Twist Bioscience, Illumina e Western Digital, anunciaram que estavam formando a DNA Data Storage Alliance (DDSA). A aliança espera “organizar a indústria e pensar em como construir todo o ecossistema para armazenamento de dados de DNA”, diz Xavier Godron, diretor de tecnologia da DNA Script, uma desenvolvedora de sintetizadores de DNA de bancada com sede em Paris e membro da aliança.

Em um interessante artigo de 2015¹⁰⁶, que fala sobre a pesquisa liderada por Robert Grass, um pesquisador do departamento de química e biociências aplicadas

em ETH Zurich apresenta como o armazenamento de informações em longo prazo e sem erros pode ser alcançado, potencialmente por mais de um milhão de anos. Primeiro, eles encapsulam os segmentos de DNA com informações em sílica (vidro) e, segundo, usam um algoritmo para corrigir erros nos dados. O texto científico está disponível na publicação do *Angewandte Chemie*¹⁰⁷.

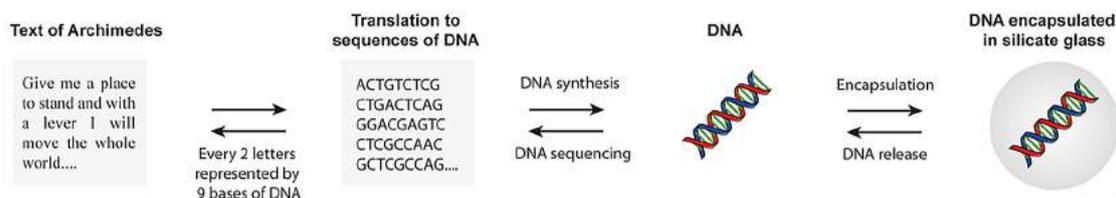


Figura 127. Esquema representativo de como o armazenamento de dados em longo prazo, sem erros, pode ser alcançado.

©Robert Grass, ETH Zurich, Imagem utilizada com permissão de Robert Grass.

Por fim, como tudo o que já discutimos até aqui, a tecnologia de sintetizar e sequenciar DNA deve continuar melhorando. Mais rápido, eficiente, preciso e barato. Todavia, temos que pensar em todo o ecossistema. Em 100 ou 1.000 anos, a recuperação de um código binário após o sequenciamento destas fitas de DNA nos levará a imagens, textos, trechos de áudio ou imagens em movimento, por exemplo, mas, para chegar a esse conteúdo, a sequência de *bits*, ou fluxo de *bits* (*bit stream*) precisa ser interpretada pelos aplicativos de *software* que geraram esses arquivos. Teremos também acesso às interpretações desses *bits* dadas pelas instruções e códigos desses *softwares*, e os sistemas operacionais em que foram projetados para funcionar?

Em um texto visionário de Jeff Rothenberg na *Scientific American* de 1995¹⁰⁸, ele já falava sobre a interpretação de um fluxo de *bits*. Na segunda página de seu texto ele afirma que:

Um arquivo não é um documento em si — apenas descreve um documento que passa a existir quando o arquivo é interpretado pelo programa que o produziu. Sem esse programa (ou *software* equivalente), o documento é um código refém de sua própria codificação. A tentativa e erro podem decodificar o texto pretendido se o documento for uma simples sequência de caracteres. Mas se for complexo, é improvável que tal abordagem de força bruta tenha sucesso. O significado de um arquivo não é inerente aos próprios *bits*, bem como o significado dessa frase é inerente em suas palavras.

Para entender qualquer documento, devemos saber o que o conteúdo significa na linguagem de seu leitor de origem. Infelizmente, o leitor de um arquivo de documento é um programa. Documentos como apresentações de multimídia são impossíveis de ler sem o *software* apropriado: ao contrário de palavras impressas, elas não podem ser apenas “vistas com a luz”.

Outras possibilidades futuras

Algumas tecnologias que prometem virar soluções de armazenamento, mas dependerão de toda uma indústria para virarem produtos que precisarão de adoção para sobreviverem e crescerem. Listo alguns exemplos a seguir para que pesquisem:

- Holografia quântica eletrônica¹⁰⁹.
- Piql¹¹⁰, salvando dados binários em filmes fotográficos com base de poliéster¹¹¹.
- Project Silica da Microsoft¹¹², salvando dados em vidro de quartzo, e outro artigo discutindo esta tecnologia de armazenamento 5D¹¹³.

Quando o pixel não for mais quadrado, e quando não for mais um pixel, e não for mais digital

Quando o *pixel* deixar de ser digital, as nossas imagens serão construídas de forma analógica, com conceitos de biologia e química, como somos construídos.

A primeira imagem digital veio em 1957, quando Russell Kirsch, um engenheiro do National Bureau of Standards, fez uma imagem digital de 5 x 5 cm com uma matriz de 176 x 176 *pixels* digitalizando uma fotografia de seu filho de três meses, Walden.



Figura 128. Uma das primeiras imagens digitalizadas. Produzida pelo NIST (*National Institute of Standards and Technology*) em 1957.

A criança da imagem acima é Walden Kirsch, filho de Russel Kirsch que liderou a equipe responsável pelo desenvolvimento do *scanner* de imagens. Esta imagem¹¹⁴ é um composto de duas digitalizações binárias para produzir níveis aproximados de cinza.



Figura 129. Imagem do primeiro scanner¹¹⁵. Nessa imagem vemos Richard B. Thomas da equipe do Russel Kirsch, no *scanner* SEAC (Standard Electronic Automatic Computer). Nos fundos está todo o console de operações do scanner. Imagem da publicação do Kirsch em 1998 sobre esse projeto realizado na NIST.

A baixa resolução se devia ao fato de que o computador que usavam não era capaz de armazenar mais informações devido ao altíssimo custo de limitações de armazenamento nesta época.

Em 2010, Russell Kirsch pede desculpas por criar uma imagem com *pixels* quadrados, como também questiona a falta de avanço nesta tecnologia e a manutenção, pela indústria da tecnologia, do *pixel* quadrado. Realmente não faz muito sentido as nossas imagens serem feitas por minúsculos quadradinhos. O texto *Why Russel Kirsch Apologized for his Pixel Technology*, de Lee Nourse¹¹⁶, fala sobre o legado do Kirsch e de sua morte em 2020. Certamente teremos um momento em que os *pixels* não serão mais quadrados, como também não serão mais *pixels*, e certamente não serão mais digitais.

Computadores biológicos

Os computadores biológicos são feitos de células vivas. Em vez de carregamento elétrico, esses computadores usam entradas químicas e outras moléculas derivadas biologicamente, como proteínas e DNA, para realizar cálculos computacionais que envolvem armazenamento, recuperação e processamento de dados. Na página da Wikipedia¹¹⁷ sobre o assunto, fica claro que esta é uma área ainda em desenvolvimento com muito potencial para o futuro, mas longe ainda de ser uma realidade atual.

De qualquer maneira, tudo que estamos discutindo aqui sobre *pixels*, formatos de arquivo, *hardware*, *software* etc. são conceitos da nossa atualidade, e parecem ser elementos já consolidados, vieram para ficar. Mas, na realidade, são totalmente transitórios na linha do tempo. Esse tempo nos mostrará grandes mudanças em todos esses assuntos.

8

Repensar a captura

Um pensamento que deve estar na nossa pauta é o que significam os custos e compromissos energéticos necessários para todas as fases de trabalho de produção e preservação de dados digitais, como as nossas fotografias. Repensar a captura, o que produzimos e geramos, faz parte de um importante movimento necessário, que envolve inclusive repensar toda forma de viver, o que necessitamos e o que consumimos, o nosso uso de energia, água e comida. Este texto não poderia discutir a preservação de fotografias nato-digitais se isso também não incorporasse ter para quem preservar isso tudo de que estamos falando – que somos todos nós.

Alguns meses antes de eu finalizar este texto, o *IPCC – Intergovernmental Panel on Climate Change* (Painel Intergovernamental sobre Mudanças Climáticas) publicou a contribuição do Grupo de Trabalho I para o sexto relatório de avaliação *Sixth Assessment Report (AR6)*¹¹⁸. Esse relatório apresenta um cenário crítico para a humanidade, em que já estamos sofrendo consequências da ação humana sobre o planeta. Veja a matéria¹¹⁹ da WRI Brasil (*World Resources International*) para uma leitura sobre esse relatório.

São necessárias ações urgentes em todas as áreas da nossa sociedade. Falamos muitas vezes sobre a preservação em longo prazo, mirando ultrapassar mesmo a

minha geração e de todos os leitores deste texto. Para isso, será necessário existir uma sociedade que usufrua da memória que estamos construindo.

Nesse último capítulo, repensar a captura é fazer uma reflexão acerca do que significa de fato tudo isso que estamos produzindo. Já discutimos de forma clara que não faremos a preservação de tudo, nem temos como realizar isso. Queremos e podemos salvar filetes dessa produção. Será que podemos reduzir essa produção para algo mais perto desses filetes? Para finalizar este texto, apresento algumas ideias para alimentar esta discussão.

O que entra por um olho sai pelo outro

Talvez tudo que estamos produzindo seja efêmero, e devamos acatar isso. Afinal, é fácil não cuidar, e cuidar exigirá ações da nossa parte, que implicam em recursos humanos e materiais, muitas vezes não disponíveis. As camadas novas diárias de imagens vão se sobrepondo às camadas anteriores a ponto de ser quase impossível fazer um resgate do que ficou soterrado. O que entra por um olho sai pelo outro, nada fica, nada permanece.

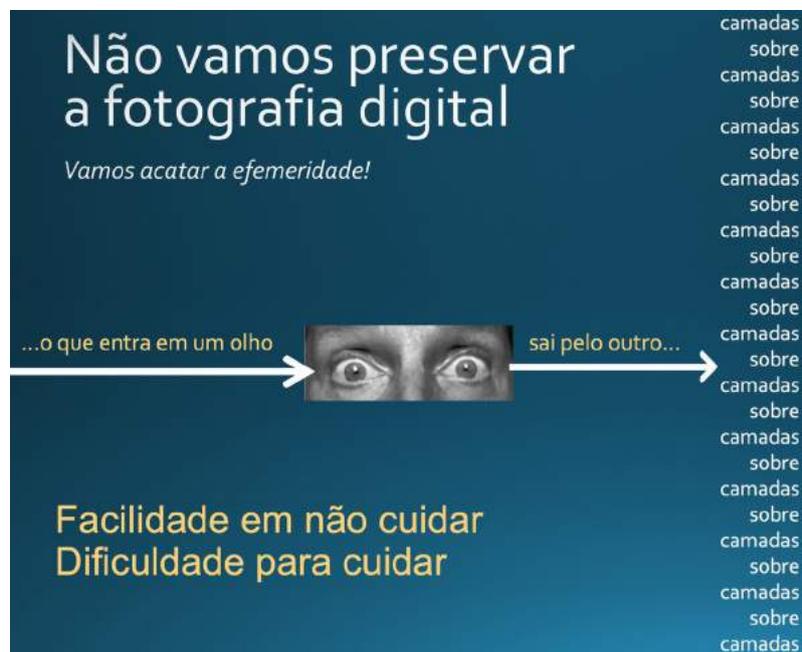


Figura 130. As novas gerações pensam em fazer a preservação de fotografias nato-digitais?

Photo taking impairment effect (efeito de diminuição da memória por tirar fotos)

Linda A. Henkel, psicóloga e professora na Universidade de Fairfield, fez um estudo¹²⁰ em 2013 sobre o impacto de tirar fotografias e seu efeito na memória. No artigo *No Pictures, Please: Taking Photos May Impede Memory for a Museum Tour*¹²¹ da Association for Psychological Science há uma boa descrição do estudo, conforme vemos a seguir:

Para descobrir, ela montou um experimento no Bellarmine Museum of Art da Fairfield University. Os alunos de graduação foram conduzidos a um passeio pelo museu e solicitados a tomar nota de determinados objetos, seja fotografando-os ou simplesmente observando-os. No dia seguinte, sua memória para os objetos foi testada. Os dados mostraram que os participantes foram menos precisos no reconhecimento dos objetos que fotografaram em comparação com aqueles que apenas observaram. Além disso, eles não foram capazes de responder a tantas perguntas sobre os detalhes visuais dos objetos para aqueles objetos que haviam fotografado.

Henkel chama isso de “efeito de diminuição da memória por tirar fotos”: “Quando as pessoas contam com a tecnologia para se lembrar por elas – contando com a câmera para registrar o evento e, portanto, não precisando prestar atenção a tudo sozinhas – isso pode ter um impacto negativo em quão bem elas se lembram de suas experiências”, explica ela.

Linda Henkel argumenta que usamos a câmera como um tipo de *drive* externo de nossa memória. Caberá ao aparelho lembrar-se das coisas por nós, portanto, não precisamos continuar a processar aquele objeto ou imagem. Por isso não interagimos, nem nos envolvemos com as coisas que nos ajudariam a lembrar dele. Também podemos pensar nisto como “nós fotografamos para esquecer”, ou fotografamos tudo para ver depois, mas não teremos tempo depois, nem saberemos direito aonde ir para ver depois. Vale a pena ver este curto filme de François Ferracci, *Lost Memories*¹²² (Memórias Perdidas).

O estudo da Linda e o filme de Ferracci me fazem lembrar de um depoimento que eu li anos atrás, cuja referência, infelizmente, não guardei e não acho mais na internet. Na minha memória, a história é mais ou menos a seguinte: uma mãe, nos EUA, acompanha todos os jogos de futebol (*soccer*) da filha. Ela não perde um jogo, e sempre leva uma filmadora montada em um tripé com baterias e fitas extras para poder

registrar o jogo inteiro. Em um dos jogos, ela teve um problema técnico ao começar a filmagem, e nada funcionou. Desesperada e frustrada, acaba por não conseguir filmar o jogo, restando a simples ação de assistir ao jogo do começo ao fim. Depois, em estado quase de choque, percebe que é a primeira vez em sua vida que vê a sua filha jogar.

Movimentos *Slow Food* e *Slow Photography*¹²³

O movimento *Slow Food* é criado em 1986 na Itália após uma manifestação no local pretendido de um McDonald's na Escadaria Espanhola, em Roma. Um dos fotógrafos do movimento *Slow Photography*, o Kurt Budliger, comenta sobre a relação do movimento *Slow Food* com o *Slow Photography* nesse interessante depoimento¹²⁴:

O movimento do *slow food* tornou-se popular em parte como uma resposta ao aumento do consumo e dependência de nossa cultura de alimentos altamente processados de fazendas industriais. Queremos comodidade, volume e velocidade. No entanto, esse tipo de alimento não nutre nossos corpos tão bem quanto o alimento completo, o que, claro, pode levar a todos os tipos de problemas de saúde. Além disso, perdemos uma faceta quase igualmente importante da comida; a nutrição espiritual e social que vem de preparar e desfrutar de uma refeição com amigos e familiares.

Tim Wu escreve, em 2011 um texto¹²⁵ explicando a importância e a necessidade do movimento do *Slow Photography*.

Não, a verdadeira vítima da fotografia rápida não é a qualidade das fotos em si. A vítima somos nós. Perdemos outra coisa: o lado da experiência, a alegria da fotografia como atividade. E tentar lutar contra essa perda, tratar a fotografia como uma experiência, não um meio para um fim, é a própria definição de fotografia lenta. Definida com mais cuidado, a fotografia lenta é o esforço para inverter a relação usual entre o processo e os resultados. Normalmente, você usa uma câmera porque quer os resultados (as fotos). Na fotografia lenta, a ideia básica é que as próprias fotos – os resultados – são secundários. O objetivo é a experiência de estudar algum objeto cuidadosamente e exercitar a escolha criativa. É isso.

Um exercício interessante de fazer seria “desenterrar” cartões antigos de memória na faixa de 256 MB, 512 MB ou 1 GB e fazer projetos de fotografia com eles. Um cartão de 256 MB não faria mais do que 5 ou 12 fotografias, dependendo da câmera digital sendo utilizada. Estimular a/o fotógrafo/a a realizar um trabalho

com somente um cartão desses, com um limite muito reduzido de imagens, fará com que seja necessário elaborar bem o que será fotografado.



Figura 131. Fotografia do autor segurando um cartão de memória de 256 MB.



Figura 132. Esta imagem produzida pelo fotógrafo italiano Gabriele Chiesa faz uma remissão interessante sobre a evolução da fotografia, do Daguerreótipo, passando por filme e chegando ao digital.

Assim, repensar a captura não significa ter regras, mas ter um espírito inquisitivo sobre todas as questões que discutimos neste texto, para onde estamos indo com a nossa produção atual e pensando em toda a cadeia de sustentabilidade de nossas ações e os nossos desejos de preservação dos filetes que estão dentro da nossa missão.



Figura 133. Fotografia do acervo de família do autor.

Ao finalizar este texto, a nossa filha mais nova nos enviou uma carta contendo somente estas duas fotografias; Ela está comigo em uma, e a na outra, com a Magali, minha companheira de vida. Essas fotografias mexeram muito comigo, porque o ato de segurá-las na mão parece criar uma conexão com as imagens e com as pessoas contidas nelas que me lembra muito o que deveria ter sido, nos primórdios da fotografia, segurar um daguerreótipo com sua imagem na palma da mão. Estas imagens existem em formato nato-digital, e já até cheguei a vê-las um tempo atrás, mas com esta quase-existência digital, elas parecem se evaporar no universo imagético. Ao serem consolidadas neste texto, contudo, mesmo sendo cópias feitas em algum processo qualquer, não permanente, existem agora ao meu lado, onde fazem parte do meu dia a dia à medida em que vão esmaecendo.

Comentários finais

Fico feliz de ter colocado uma série de ideias no papel, algumas que têm uma duração maior no tempo, que transcendem as questões tecnológicas apontadas neste texto, e outras que servirão talvez por alguns anos, mas terão que ser “migradas” para as visões que virão.

Sinto vontade de continuar, pois ainda há muito para falar, pesquisar e escrever. Fica então o desafio, para todos nós, de continuarmos a construir textos com esses e outros temas relacionados, para ampliar e aprofundar os assuntos discutidos nesta obra. Precisamos construir comunidades de pessoas e instituições que possam colaborar nestas tarefas, pois os desafios são muitos e contínuos.

Notas

Todos os *links* foram acessados entre agosto de 2021 e julho de 2022

- ¹ Para mais informações sobre Memory Waka Research Group, ver: <http://memoryconnection.org/about/>.
- ² TROUILLOT, Michel-Rolph. *Silenciando o passado: poder e a produção da história*. Tradução de Sebastião Nascimento. Curitiba: huya, 2016. Disponível em: https://edisciplinas.usp.br/pluginfile.php/5663280/mod_resource/content/1/Trouillot%2C%20Silenciando%20o%20Passado%20-%20completo.pdf.
- ³ SCHISLER, Millard. Photo Finished – o fim da fotografia. *Brasiliiana Fotográfica*, [s. l.], 1 nov. 2018. Disponível em: <https://brasilianafotografica.bn.gov.br/?p=13350>.
- ⁴ ROMER, Grant B. Fundamentals of the Conservation of Photographs. The Getty Conversation, Institute, [s. l.], 2010. Disponível em https://www.getty.edu/conservation/publications_resources/teaching/photo_tn_what_is.pdf.
- ⁵ JEFFRIES, Stuart. The death of photography: are camera phones destroying an art form? *The Guardian*, [s. l.], 13 dez. 2013. Culture. Disponível em: <https://www.theguardian.com/artanddesign/2013/dec/13/death-of-photography-camera-phones>.
- ⁶ MAYES, Stephen. The Next Revolution of Photography is Coming. *Time*, [s. l.] 25 ago. 2015. Disponível em: <https://time.com/4003527/future-of-photography/>.
- ⁷ SEBASTIÃO Salgado prevê fim da fotografia 'em 20 ou 30 anos'. *G1*, [s. l.], 28 out. 2016. Pop & Arte. Disponível em: <http://g1.globo.com/pop-arte/noticia/2016/10/sebastiao-salgado-preve-fim-da-fotografia-em-20-ou-30-anos.html>.

- 8 CAGNETI, Duda. Sebastião Salgado afirma: “a fotografia não está acabando”. *Portal Photos*, [s. l.], 14 fev. 2017. Disponível em: <https://web.archive.org/web/20180331091230/http://photos.com.br/sebastiao-salgado-fotografia-nao-esta-acabando/#expand>.
- 9 CAKEBREAD, Caroline. People will take 1.2 trillion digital photos this year – thanks to smartphones. *Insider*, [s. l.], 31 ago. 2017. Disponível em: <https://www.businessinsider.com/12-trillion-photos-to-be-taken-in-2017-thanks-to-smartphones-chart-2017-8>.
- 10 WITHERILL, Huntington. The trivialization of photography. *The Luminous Landscape*, 9 dez. 2015. Disponível em: <https://luminous-landscape.com/the-trivialization-of-photography/>.
- 11 CARRINGTON, David. How many photos will be taken in 2020? *Mylio*, [s. l.], [2020]. Blog Home. Disponível em: <https://blog.mylio.com/how-many-photos-will-be-taken-in-2020/>.
- 12 YANG, Jeff. Grumpy Cat may be gone, but our obsession with internet cats will never die. *CNN*, [s. l.], 2 abr. 2015. Atualizado em: 17 maio 2019. Opinion. Disponível em: <https://edition.cnn.com/2015/04/02/opinions/yang-internet-cats/index.html>.
- 13 MENDONZA, Mayra. Produzimos imagens, mas não somos todos fotógrafos. Entrevista concedida a Cassiano Vianna. *Itaú Cultural*, São Paulo, 2016. Disponível em: <http://www.itaucultural.org.br/entrevista-com-mayra-mendoza>.
- 14 MENEZES, Maiá; FARAH, Tatiana. Barateamento de celulares e câmeras revolucionam forma como classes C, D e E registram dia a dia. *O Globo*, [s. l.], 4 nov. 2012. Economia. Disponível em: <https://oglobo.globo.com/economia/tecnologia/barateamento-de-celulares-cameras-revoluciona-forma-como-classes-d-e-registram-dia-dia-6629781>.
MENEZES, Maiá; FARAH, Tatiana. Proliferação de imagens esbarra no caos e em problemas de HD. *O Globo*, [s. l.], 4 nov. 2012. Economia. Disponível em: <https://oglobo.globo.com/economia/tecnologia/proliferao-de-imagens-esbarra-no-caos-em-problemas-de-hd-6629859>.
MENEZES, Maiá; FARAH, Tatiana. Com celulares à mão, índios, agricultores e cabeleireiras retratam pela 1ª vez suas famílias. *O Globo*, [s. l.], 4 nov. 2012. Economia. Disponível em: <https://oglobo.globo.com/economia/tecnologia/com-celulares-mao-indios-agricultores-cabeleireiras-retratam-pela-1-vez-suas-familias-6629919>.
- 15 ROSENTHAL, David. Let’s just keep everything forever in the cloud. *DSHR’S Blog*, [s. l.], 14 maio 2012. Disponível em: <https://blog.dshr.org/2012/05/lets-just-keep-everything-forever-in.html>.
- 16 GIARETTA, David. *Advanced digital preservation*. New York: Springer, 2011. Para ter acesso ao artigo, o interessado precisa se cadastrar e solicitar o envio do pdf por meio do link: https://www.researchgate.net/publication/321505986_Advanced_Digital_Preservation.
- 17 SCIENCE AND TECHNOLOGY COUNCIL. *O dilema digital: questões estratégicas na guarda e no acesso a materiais cinematográficos digitais*. Tradução Fernanda Paiva Guimarães. São Paulo: Cinemateca Brasileira, 2010. Disponível em: https://web.archive.org/web/20210805162236/http://cinemateca.org.br/wp-content/uploads/2016/08/Dilema_Digital_1_PTBR.pdf.
- 18 FREY, Franziska *et al.*; WARDA, Jeffrey (ed.). *The AIC guide to digital photography and conservation documentation*. Washington, DC: American Institute for Conservation of Historic and Artistic Works, 2011.
- 19 Aplicativo *Downloadgram*. Disponível em: <https://downloadgram.org/>.
- 20 LIBRARY OF CONGRESS. *Recommended Formats Statement 2021-2022*. Disponível em: https://www.loc.gov/preservation/resources/rfs/RFS%202021-2022_Archival%20Only.pdf.

- 21 LIBRARY OF CONGRESS. *Sustainability of Digital Formats: Planning for Library of Congress Collections*. Última atualização: 5 jan. 2017. Disponível em: <https://www.loc.gov/preservation/digital/formats/sustain/sustain.shtml>.
- 22 JOINT Photographic Group. *Wikipédia, a enciclopédia livre*. Flórida: Wikimedia Foundation, 2022. Disponível em: https://en.wikipedia.org/wiki/Joint_Photographic_Experts_Group.
- 23 Para mais informações sobre International Image Interoperability Framework, ver: <https://iif.io/>
- 24 HIGH Efficient Image File Format. *Wikipédia, a enciclopédia livre*. Flórida: Wikimedia Foundation, 2022. Disponível em: https://en.wikipedia.org/wiki/High_Efficiency_Image_File_Format
- 25 Para mais informações sobre o software PIEware, ver: <https://www.acdsee.com/en/parametric-editing/>.
- 26 Para mais informações sobre Alfred Henry Munsell, ver: <https://munsell.com/about-munsell-color/munsell-color-company-history/albert-h-munsell/>.
- 27 LEÃO, Alexandre Cruz. *Gerenciamento de Cores para Imagens Digitais*. 2005. Dissertação (Mestrado em Artes Visuais) – Universidade Federal de Minas Gerais, Belo Horizonte, 2005. Disponível em: https://repositorio.ufmg.br/bitstream/1843/JSSS-7XGFG3/1/dissertacao_alexandre_leao.pdf.
- 28 ADOBE. *Configurações de gerenciamento de cores para a melhor saída impressa*. Última atualização em 13 jan. 2022. Disponível em: <https://helpx.adobe.com/br/photoshop-elements/kb/color-management-settings-best-print.html>.
- 29 Definição da palavra *Gamut* disponível em: <https://www.merriam-webster.com/dictionary/gamut>.
- 30 VÍRUS de computador. *Wikipédia, a enciclopédia livre*. Flórida: Wikimedia Foundation, 2021. Disponível em: https://pt.wikipedia.org/w/index.php?title=V%C3%ADrus_de_computador&oldid=61775513.
- 31 Aplicativo Anti-twin disponível em: <http://www.joerg-roenthal.com/en/antitwin/>.
- 32 Aplicativo FileListCreator disponível em: <https://www.sttmedia.com/software>.
- 33 Disponível em: <https://dpbestflow.org/links/32>.
- 34 Disponível em: <https://www.bulkrenameutility.co.uk/>.
- 35 RIES, Ryan. Resposta à discussão “Overcoming maximum file path length restrictions in Windows”. *Server Fault*. [s. l.], 25 dez. 2014. Disponível em: <https://serverfault.com/questions/232986/overcoming-maximum-file-path-length-restrictions-in-windows/654756#654756>.
- 36 DIGITAL PRESERVATION COALITION. Persistent identifiers. In: HANDBOOK: Digital Preservation. 2. ed. [S. l.]: Digital Preservation Coalition, 2015. Disponível em: <https://www.dpconline.org/handbook/technical-solutions-and-tools/persistent-identifiers>.
- 37 SAYÃO, Luis Fernando. Interoperabilidade das bibliotecas digitais: o papel dos sistemas de identificadores persistentes – URN, PURL, DOI, Handle System, CrossRef e OpenURL. *Transinformação*, v. 19, n. 1, p. 65-82, 2007. Disponível em: <https://www.scielo.br/j/tinf/a/NTr5XbPG7LG5pWH876MmWVN/?lang=pt>.
- 38 THE UNESCO/PERSIST CONTENT TASK FORCE. *Guidelines for the selection of digital heritage for long-term preservation*. [S. l.]: Unesco, 2021. Disponível em: <https://repository.ifla.org/bitstream/123456789/1863/1/Guidelines%20For%20the%20Selection%20Of%20Digital%20Heritage%20For%20Long-Term%20Preservation%20%E2%80%93%202nd%20Edition.pdf>.

- 39 BACA, Murtha. Intro to Metadata. In: BACA, Murtha. (ed.). *Introduction to Metadata*, 3. ed. Los Angeles: Getty Publications, 2016. Disponível em: <http://www.getty.edu/publications/intrometadata>.
- 40 Para mais informações, ver: <https://iptc.org/about-iptc/>.
- 41 ZHANG, Michael. This is how press photos were transmitted back in the 1970s. *PetaPixel*, [s. l.], 26 jul. 2015. Disponível em: <https://petapixel.com/2015/07/26/this-is-how-press-photos-were-transmitted-back-in-the-1970s/>.
- 42 INTERNATIONAL PRESS TELECOMMUNICATIONS COUNCIL. *IPTC Photo Metadata Standard*. Versão de 2022. Disponível em: <https://iptc.org/standards/photo-metadata/iptc-standard/>
- 43 Para mais informações sobre a Plataforma XMP, ver: <https://www.adobe.com/products/xmp.html>.
- 44 EXTENSIBLE Metadata Platform. *Wikipédia, a enciclopédia livre*. Flórida: Wikimedia Foundation, 2022. Disponível em: https://en.wikipedia.org/wiki/Extensible_Metadata_Platform#History.
- 45 Para mais informações sobre o site Controlled Vocabulary, ver: <http://www.controlledvocabulary.com/cv/>.
- 46 SILOVA, Antra. How to create your own controlled vocabulary. *Databasics*, [s. l.], 24 nov. 2015. Disponível em: <https://www.databasics.com.au/blog/how-to-create-your-own-controlled-vocabulary..>
- 47 Para mais informações sobre os vocabulários controlados desenvolvidos pelo Getty Research Institute, ver: <https://www.getty.edu/research/tools/vocabularies/>.
- 48 HE, K.; ZHANG, X.; REN, S.; SUN, J. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. In: INTERNATIONAL CONFERENCE ON COMPUTER VISION, 7-13 dez. 2015, Santiago. *Proceedings [...]*. Santiago: IEEE, 2015. Disponível em: <https://doi.org/10.1109/ICCV.2015.123>.
- 49 Para mais informações sobre The Algorithmic Justice League (a Liga da Justiça Algorítmica), ver: <https://www.ajl.org/about>.
- 50 GENDER Shades. *The Gender Shades Project*, [s. l.] 2018. Disponível em: <http://gendershades.org/index.html>.
- 51 IPTC Announces new properties in photo metadata to make images more accessible. *IPTC*, [s. l.], 27 out. 2021. Disponível em: <https://iptc.org/news/iptc-announces-new-properties-in-photo-metadata-to-make-images-more-accessible/>.
- 52 GREENE, Mark A.; MEISSNER, Dennis. More Product, Less Process: Revamping Traditional Archival Processing. *American Archivist*, v. 68, n. 2, p. 208-263, 2005. Disponível em: <http://www.archivists.org/prof-education/pre-readings/IMPLP/AA68.2.MeissnerGreene.pdf>.
- 53 BACHLI, Kelley *et al.* *Guidelines for efficient archival processing in the University of California Libraries*. Versão 3.2. 18 set. 2012. Disponível em: https://libraries.universityofcalifornia.edu/wp-content/uploads/2022/06/Efficient_Archival_Processing_Guidelines_v3-1.pdf.
- 54 RILEY, Jean. *Seeing standards: a visualization of the metadata universe*. 2009-2010. Disponível em: <http://jennriley.com/metadatamap/>
- 55 KROGH, Peter. Back up overview. *dpBestflow*. Última atualização: 22 set. 2015. Disponível em: <https://dpbestflow.org/backup/backup-overview#threats;>

- 56 NDSA. Levels of digital preservation. Versão 2.0. NDSA, [s. l.], 2019. Disponível em: <https://ndsa.org/publications/levels-of-digital-preservation/>.
- 57 NDSA. A matriz dos níveis de preservação digital. Tradução de Laura Rezende. [S. l.]: NDSA, 2019. Disponível em: <https://osf.io/3amnu>.
- 58 RABINOV, Natasha. What's the Diff: 3-2-1 vs. 3-2-1-1-0 vs. 4-3-2. *Backblaze*, [s. l.] 21 jul. 2021. Disponível em: <https://www.backblaze.com/blog/whats-the-diff-3-2-1-vs-3-2-1-1-0-vs-4-3-2/>.
- 59 BYERS, Fred R. *Care and Handling of CDs and DVDs – A Guide for Librarians and Archivists*. Washington, DC; Gaithersburg, MD: National Institute of Standards and Technology: Council on Library and Information Resources, 2003. Disponível em: <https://www.clir.org/wp-content/uploads/sites/6/pub121.pdf>.
- 60 KLEIN, Andy. Hard Disk Drive (HDD) vs. Solid-state Drive (SSD): What's the Diff? *Backblaze*, [s. l.], 5 out. 2021. Disponível em: <https://www.backblaze.com/blog/hdd-versus-ssd-whats-the-diff/>.
- 61 KLEIN, Andy. Drive Failure Over Time: The Bathtub Curve Is Leaking. *Backblaze*, [s. l.], 26 out. 2021. Disponível em: <https://www.backblaze.com/blog/drive-failure-over-time-the-bathtub-curve-is-leaking/>.
- 62 HOLLAND, Tilly. SSD Lifespan: How Long do SSDs Last. *Ontrack*, [s. l.], 30 jul. 2021. Disponível em: <https://www.ontrack.com/en-us/blog/how-long-do-ssds-really-last>
- 63 Memoriav é uma organização nacional da Suíça que visa proteger a herança cultural audiovisual para todas as organizações de memória do país. Veja lista de importantes publicações em: <https://memoriav.ch/fr/publikationen/>.
- 64 MEMORIAV. *Digital Archiving of Film and Video: Principles and Guidance*. Versão 1.2. nov. 2019. Disponível em: <https://memoriav.ch/en/dafv/?lang=en>
- 65 LANTZ, Mark. Why the future of data storage is (still) magnetic tape: Disk drives are reaching their limits, but magnetic tape just gets better and better. *IEE Spectrum*, [s. l.], 28 ago. 2018. Disponível em: <https://spectrum.ieee.org/why-the-future-of-data-storage-is-still-magnetic-tape>.
- 66 CLIR. How Can You Prevent Magnetic Tape from Degrading Prematurely? *CLIR*, Alexandria, VA, [20--?]. Disponível em: https://www.clir.org/pubs/reports/pub54/5premature_degrade/.
- 67 ALLIED MARKET RESEARCH. Resumo sobre o relatório Tape Storage Market Global Opportunity Analysis and Industry Forecast, 2020–2030. [S. l.]: Allied Market Research, 2021. Disponível em: <https://www.alliedmarketresearch.com/tape-storage-market>.
- 68 Um gravador/leitor externo LTO-8 custa no Brasil em torno de 35 mil reais, a fita LTO-8 tem um custo em torno de 600 reais, ou R\$ 50 reais/TB para armazenamento de dados sem compressão. Esses são valores aproximados em 2023, mas sofrem constantes atualizações.
- 69 Para mais informações sobre a Tape Library da Spectra Logic, ver: <https://spectralogic.com/products/tape-storage/spectra-stack-tape-library/>.
- 70 CLOUD storage. *Wikipédia, a enciclopédia livre*. Flórida: Wikimedia Foundation, 2022. Disponível em: https://en.wikipedia.org/wiki/Cloud_storage.
- 71 BOURNE, James. Step back in time: AT&T predicts the cloud in 1993. *CloudTech*, [s. l.], 6 maio 2014. Disponível em: <https://www.cloudcomputing-news.net/news/2014/may/06/step-back-time-t-predicts-cloud-1993/>.

- 72 What Is The Cloud – by AT&T. Direção: David Hoffman. [S. l.: s. n.], 2013. 1 vídeo (6 min 47 s). Publicado pelo canal David Hoffman. Disponível: https://www.youtube.com/watch?v=_a7hK6kWttE.
- 73 FERRIS, Chelsey. History of Cloud Storage. *Capacity*, [s. l.], 5 dez. 2019. Disponível em: <https://capacity.com/cloud-storage/history-of-cloud-storage/>.
- 74 TEU, Alex. Cloud Storage Wars: Amazon, Google and Microsoft are Racing to Armageddon. *Linkedin*, [s. l.] 28 maio 2014. Disponível em: <https://www.linkedin.com/pulse/20140528192040-3841600-cloud-storage-wars-amazon-google-and-microsoft-are-racing-to-armageddon>
- 75 MILLMAN, Rene. Cloud storage costs: How to get cloud storage bills under control. *Computer Weekly*, [s. l.], 11 ago. 2020. Disponível em: <https://www.computerweekly.com/feature/Cloud-storage-costs-How-to-get-cloud-storage-bills-under-control>.
- 76 KASS, Susanna; RAVAGNI, Alberto. *Designing and Building the Next Generation of Sustainable Data Centers*. [S. l.]: DCD, 2020. Disponível em: <https://shre.ink/kntu>.
- 77 SHIRER, Michael. IDC Forecasts Worldwide “Whole Cloud” Spending Reach \$1.3 Trillion by 2025. *IDC*, [s. l.], 14 set 2021. Disponível em: <https://www.idc.com/getdoc.jsp?containerId=prUS48208321>.
- 78 SAMPERA, Ernest. Lots of Companies Are Leaving the Cloud. Here's Why. *Vxchnge*, [s. l.], 13 set. 2019. Disponível em: <https://web.archive.org/web/20210302102948/https://www.vxchnge.com/blog/why-companies-are-leaving-the-cloud>.
- 79 FOG computing. *Wikipédia, a enciclopédia livre*. Flórida: Wikimedia Foundation, 2022. Disponível em: https://pt.wikipedia.org/wiki/Fog_computing.
- 80 VALENTE, Mariana; HOUANG, André. *Creative Commons BR: O que você precisa saber sobre licenças CC*. [S. l.]: Equipe CC Brasil, 2020. Disponível em: <https://br.creativecommons.net/wp-content/uploads/sites/30/2021/02/CartilhaCCBrasil.pdf>.
- 81 BRANCO, Sérgio. Abre-te Código – Sérgio Branco, Instituto de Tecnologia e Sociedade. São Paulo: Goethe Institut, 7 ago. 2020. 1 vídeo (1 h 01 m 33 s). Disponível em: <https://www.goethe.de/prj/hyc/pt/bra/pdf/clc/its.html>.
- 82 Para mais informações sobre o Tainacan, ver: <https://tainacan.org/>
- 83 Para mais informações sobre o ResourceSpace, ver: <https://www.resourcespace.com/>.
- 84 Para mais informações sobre o ArchivesSpace, ver: <https://archivesspace.org/>
- 85 Para mais informações sobre o Omeka, ver <https://omeka.org/>.
- 86 Para acessar o Internet Live Stars, ver: <https://www.internetlifestats.com/>.
- 87 PALM, Jonas. *The digital black hole*. Stockholm: Riksarkivet 2006. Disponível em: <https://cdn-kb.avanet.nl/wp-content/uploads/2019/05/31173546/Palm-Black-Hole.pdf>.
- 88 Para mais informações a respeito da tabela sobre os níveis de preservação digital criada pelo NDSA, ver nota 57.
- 89 BAUCOM, Erin *et al.* *Using the Levels of Digital Preservation as an Assessment Tool*. [S. l.]: NDSA, 2019. Disponível em: <https://osf.io/m6j4q/download>.

- ⁹⁰ PELTZMAN, Shira. Expanding NDSA Levels of Preservation. Library of Congress. *The Signal*, [s. l.], 12 abr. 2016. Disponível em: <https://blogs.loc.gov/thesignal/2016/04/expanding-nds-a-levels-of-preservation/>.
- ⁹¹ Para mais informações sobre o blog de preservação digital The Signal, da Library of Congress, ver: <https://blogs.loc.gov/thesignal/>.
- ⁹² SCHISLER, Millard. The theory and craft of digital preservation: comentários sobre os 16 axiomas de preservação digital. *Revista Eletrônica de Comunicação, Informação & Inovação em Saúde*, [s. l.], v. 14, n. 3, 2020. Disponível em: <https://www.reciis.icict.fiocruz.br/index.php/reciis/article/view/2155>.
- ⁹³ OWENS T. *The theory and craft of digital preservation*. Baltimore, MD: Johns Hopkins University Press; 2017. Disponível em: <https://osf.io/preprints/lissa/5cpjt/download>.
- ⁹⁴ Obra citada na nota 18.
- ⁹⁵ Para consultar os estudos sobre preservação e durabilidade de materiais coloridos contemporâneos que estão disponíveis no site Wilhelm Imaging Research, ver: <http://www.wilhelm-research.com/>.
- ⁹⁶ WILHELM, Henry G.; BROWER, Carol. *The Permanence and Care of Color Photographs: Traditional and Digital Color Prints, Color Negatives, Slides, and Motion Pictures*. [S. l.], Preservation Publishing Company, 1993.
- ⁹⁷ WILHELM, Henry. A Survey of Print Permanence in the 4x6-Inch Consumer Digital Print Market in 2004-2007 *In: IS&T INTERNATIONAL SYMPOSIUM ON TECHNOLOGIES FOR DIGITAL FULFILLMENT*, 5 mar. 2007, Nevada. *Technical Program, Abstracts, and Proceeding*. Nevada: Society for Imaging Science and Technology, 2007. Disponível em: http://www.wilhelm-research.com/ist/ist_2007_03.html.
- ⁹⁸ EPSON SureColor P10000 and P20000 – Print Permanence Ratings. *Wilhelm Imaging Research*, [s. l.], 15 fev. 2019. Disponível em http://www.wilhelm-research.com/epson/WIR_Epson_SureColor_P10000_and_P20000_Printers_2019-02-15.pdf.
- ⁹⁹ Para mais informações sobre as tintas pigmentadas monocromáticas da Cone Editions, ver: <https://www.cone-editions.com/piezography>.
- ¹⁰⁰ ILFOCHROME. *Wikipédia, a enciclopédia livre*. Flórida: Wikimedia Foundation, 2022. Disponível em: <https://en.wikipedia.org/wiki/Ilfochrome>.
- ¹⁰¹ HEWLETT-PACKARD DEVELOPMENT COMPANY. *HP Indigo Print Permanence: Sustainable photo memories*. [S. l.]: Hewlett-Packard Development Company, 2012. Disponível em: http://www.wilhelm-research.com/hp/HP-WIR_Indigo_Print_Permanence_April_2012.pdf.
- ¹⁰² XEROX. *Xerox iGen Family with Matte Dry Ink: Making Memories More Permanent*. [S. l.]: Xerox, 2012. Disponível em: http://www.wilhelm-research.com/Xerox/Xerox-WIR_iGen_Photoalbum_Page_Permanence_October_2012.pdf.
- ¹⁰³ Para mais informações sobre DNA, ver: <https://www.yourgenome.org/facts/what-is-dna/>.
- ¹⁰⁴ WYSS INSTITUTE. Writing the Book in DNA. 16 ago. 2012. Disponível em: <https://wyss.harvard.edu/news/writing-the-book-in-dna/>.
- ¹⁰⁵ VITAK, Sarah. Technology alliance boosts efforts to store data in DNA: Trade group aims to advance technologies for storing big data. *Nature*, [s. l.], 3 mar. 2021. Disponível em: <https://www.nature.com/articles/d41586-021-00534-w>.

- 106 JACOBS, Angelika. Data-storage for eternity. *ETH Zürich*, [s. l.], 13 fev. 2015. Disponível em: <https://ethz.ch/en/news-and-events/eth-news/news/2015/02/data-storage-for-eternity.html>.
- 107 GRASS, Robert N. *et al.* Robust Chemical Preservation of Digital Information on DNA in Silica with Error-Correcting Codes. *Angewandte Chemie International Edition*, v. 54, n. 8, p. 2552-2555, 2015. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1002/anie.201411378>.
- 108 ROTHENBERG, Jeff. Ensuring the Longevity of Digital Documents. *Scientific American*, v. 272, n. 1, p. 42-47, 1995. Disponível em: https://iconnectdots.com/wp-content/uploads/2010/03/1995_Jan_SciAm.pdf.
- 109 HAND, Eric. Hologramas desafiam limite de armazenamento de dados. *Terra*, [s. l.], 8 fev. 2009. Byte. Disponível em: <https://www.terra.com.br/noticias/tecnologia/eletronicos/hologramas-desafiam-limite-de-armazenamento-de-dados,421853ba037ea310VgnCLD200000bbcceb0aRCRD.html>.
- 110 Para mais informações sobre a empresa e seus serviços, ver: <https://www.piql.com/>.
- 111 PENNINGTON, Adrian. Film preservation for 500 years on film. *The Broadcast Bridge*, [s. l.], 10 nov. 2014. Disponível em: <https://www.thebroadcastbridge.com/content/entry/940/film-preservation-for-500-years-on-film>.
- 112 LANGSTON, Jennifer. Project Silica proof of concept stores Warner Bros. 'Superman' movie on quartz glass. *Microsoft*, [s. l.], 4 nov. 2019. News. Disponível em: <https://news.microsoft.com/source/features/innovation/ignite-project-silica-superman>.
- 113 DENT, Steve. '5D' storage could fit 500TB on CD-sized glass disc. *Engadget*, [s. l.], 1 nov. 2021. Disponível em: <https://www.engadget.com/5-d-storage-could-fit-500-tb-on-cd-sized-disc-095039455.html#:~:text=Using%20high%2Dspeed%20lasers%2C%20researchers,for%20archival%20and%20other%20purposes>.
- 114 Figura disponível em: <https://commons.wikimedia.org/wiki/File:NBSFirstScanImage.jpg>.
- 115 KIRSCH, Russel A. SEAC and the Start of Image Processing at the Nacional Bureau of Standards. *IEEE Anals of the History of Computing*, v. 20, n. 2, p. 7-13, 1998. Disponível em: https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=821701.
- 116 NOURSE, Lee. Why Russel Kirsch Apologized for his Pixel Technology. *Medium*, [s. l.], 22 ago. 2020. Disponível em: <https://medium.com/illumination/why-russell-kirsch-apologized-for-his-pixel-technology-946b6433e647>.
- 117 BIOLOGICAL computing. *Wikipédia, a enciclopédia livre*. Flórida: Wikimedia Foundation, 2022. Disponível em: https://en.wikipedia.org/wiki/Biological_computing
- 118 THE INTERGOVERNMENTAL PANEL ON CLIMATE CHANGE. *Climate Change 2021: The Physical Science Basis*. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. [S. l.]: IPCC, 2021. Disponível em: <https://www.ipcc.ch/report/ar6/wg1/>.
- 119 WASKOW, David; GERHOLDT, Rhys. Mudanças Climáticas: veja grandes resultados do relatório do IPCC. *WRI Brasil*, 9 ago. 2021. Disponível em: https://wribrasil.org.br/pt/blog/clima/ipcc-relatorio-mudancas-climaticas-2021?gclid=Cj0KCQjw8p2MBhCiARIsADDUFVGGQFD5ue-Cpi1tpKs6cyV3t8eYD0VDCi1tyfzZ079Ve0fDgIV09J7e4aAsNpEALw_wcB.
- 120 HENKEL, Linda A. Point-and-Shoot Memories: The Influence of Taking Photos on Memory for a Museum Tour. *Association for Psychological Science*, v. 25, n. 2, p. 396 -402, 2014. Disponível

em: https://www.researchgate.net/publication/259207719_Point-and-Shoot_Memories_The_Influence_of_Taking_Photos_on_Memory_for_a_Museum_Tour.

- 121 HENKEL, Linda A. No Pictures, Please: Taking Photos May Impede Memory for a Museum Tour. *Association for Psychological Science*, [s. l.], 9 dez. 2013. Disponível em: <https://www.psychologicalscience.org/news/releases/no-pictures-please-taking-photos-may-impede-memory-of-museum-tour.html>.
- 122 CGI SCI-FI Short Film: "Lost Memories" by – Francois Ferracci. Paris: [s. n.], 2012. Direção de François Ferracci. 1 vídeo (2 min 51 s). Publicado pelo canal TheCGBros. Trailer do filme *Lost Memories*. Disponível em: <https://www.youtube.com/watch?v=PMgoiJ48H-k>
- 123 SLOW food. *Wikipédia, a enciclopédia livre*. Flórida: Wikimedia Foundation, 2022. Disponível em: https://en.wikipedia.org/wiki/Slow_Food.
- SLOW photography. *Wikipédia, a enciclopédia livre*. Flórida: Wikimedia Foundation, 2022. Disponível em: https://en.wikipedia.org/wiki/Slow_photography.
- 124 BUDLIGER, Kurt. Slow Photography Movement. *Kurt Budliger Photography*, [s. l.], [2012]. Disponível em: <https://kurtbudliger.com/slow-photography-movement/>.
- 125 WU, Tim. The Slow-Photography Movement: What is the point of taking pictures? *Slate*, [s. l.], 18 jan. 2011. Disponível em: <https://slate.com/human-interest/2011/01/the-slow-photography-movement-asks-what-is-the-point-of-taking-pictures.html>.

Referências complementares

INSTITUTO MOREIRA SALLES. A imagem digital no contexto pessoal: sobre memória pessoal e coletiva na atualidade. Evento *on-line*. 19 maio - 14 jul. 2020. Disponível em: <https://ims.com.br/eventos/a-imagem-digital-no-contexto-pessoal/>.

KROGH, Peter. *The DAM Book: Digital Asset Management for Photographers*. 2. ed. Sebastopol: O'Reilly Media, 2009.

RUSSOTTI, Patricia; ANDERSON, Richard. *Digital Photography Best Practices and Workflow Handbook: A Guide to Staying Ahead of the Workflow Curve*. Abingdon: Routledge, 2009.

Créditos de figuras e tabelas

Figura 1 [p. 12]

Imagem utilizada com permissão de seu autor, Ross Hemera.

Figura 2 [p. 15]

Reprodução de Millard Schisler.

Figura 3 [p. 18]

Fotografia de Millard Schisler.

Figura 5 [p. 19]

Imagem do autor.

Figura 6 [p. 20]

Fotografia de um dos pioneiros ambientalistas, Howard Cleaves.

Imagem do Freshkills Park.

Figura 7 [p. 21]

Reprodução de Millard Schisler, *The Dictionary of Photography*, 10ª edição de 1920.

Figura 8 [p. 24]

Permissão de uso da imagem cedida por Andy Maltz do Science and Technology Council, AMPAS.

Figura 10 [p. 29]

Imagem CC-BY-SA-2.0 atribuído a John Vetterl.

Figura 11 [p. 31]

Imagem utilizada com permissão da autora da foto Joanna Americano.

Figura 13 [p. 33]

Imagem disponível em <https://www.apple.com/legal/internet-services/icloud/br/terms.html>.

Figura 22 [p. 54]

Imagem utilizada com permissão da Debora Schisler.

Figura 23 [p. 55]

Imagem utilizada com permissão de Mark McKinney.

Figura 24 [p. 56]

Imagem CC BY-SA 3.0 atribuída a SharkD.

Figura 25 [p. 56]

Imagem CC BY-SA 4.0 atribuída a Hannes Grobe.

Figura 26 [p. 57]

Imagem utilizada com permissão do Ciro Albano.

Figura 36 [p. 66]

Imagem CC BY-SA 3.0 atribuída a MD.

Figura 37 [p. 67]

Imagem do autor.

Figura 38 [p. 68]

Imagem do autor.

Figura 40 [p. 70]

Imagem utilizada com permissão de Beatriz Franco.

Figura 41 [p. 71]

Imagem utilizada com permissão de Beatriz Franco.

Figura 59 [p. 89]

Imagem do autor.

Figura 60 [p. 90]

Imagem do autor.

Figura 65 [p. 94]

Desenho do autor.

Figura 66 [p. 95]

Imagem do autor.

Figura 67 [p. 96]

Imagem utilizada com permissão do autor Peter Krogh. Disponível em: <https://dpbestflow.org/>.

Tabela 1 [p. 100]

Elaborada pelo autor.

Figura 71 [p. 101]

Imagem disponível em: <https://serverfault.com/questions/232986/overcoming-maximum-file-path-length-restrictions-in-windows/654756#654756>.

Tabela 2 [p. 104]

Tradução livre desta tabela autorizada pelo Getty Research Institute - GRI.

Figura 72 [p. 106]

Fotografia original utilizada com autorização de Silvano de Gennaro.
<https://www.theatlantic.com/technology/archive/2012/07/the-story-behind-the-first-photograph-ever-posted-on-the-web/259655/>.

Figura 73 [p. 107]

Acervo do autor.

Figura 77 [p. 109]

Acervo do autor.

Figura 79 [p. 111]

Imagem utilizada com permissão de Marcos Issa / Banco de imagens Argosfoto.

Figura 82 [p. 114]

Imagem de domínio público da NASA.

Figura 84 [p. 118]

Imagem utilizada com permissão da autora Jenn Riley.

Figura 85 [p. 120]

Imagem do site DPBestFlow, utilizada com permissão de Peter Krogh.

Figura 86 [p. 120]

Imagem do site DPBestFlow, utilizada com permissão de Peter Krogh.

Figura 91 [p. 130]

Imagem utilizada com permissão da empresa Backblaze.

Figura 94 [p. 134]

Imagem utilizada com permissão de Ralph Clevenger, com texto adicional do autor desta obra.

Figura 97 [p. 139]

Imagem CC BY 2.0 atribuída a Andrew Magill.

Figura 98 [p. 140]

Imagem CC BY-SA 3.0 de Evan-Amos, texto adicionado pelo autor.

Figura 99 [p. 141]

Imagem disponível em: <https://www.backblaze.com/blog/drive-failure-over-time-the-bathtub-curve-is-leaking/>.

Figura 100 [p. 143]

Imagem CC BY-SA 3.0 atribuída a Vladsinger.

Figura 101 [p. 144]

Imagem CC BY-SA 4.0 atribuído a Dmitry Nosachev.

Figura 102 [p. 144]

Imagem de divulgação da Samsung.

Tabela 5 [p. 145]

Tabela utilizada com permissão da Memoriav, Association for the preservation of the audiovisual heritage of Switzerland.

Figura 103 [p. 149]

Imagem disponível em:

<https://www.drobo.com/storage-products/capacity-calculator/>.

Figura 104 [p. 150]

Imagem disponível em: https://www.google.com/search?q=nas+network+attached+storage&tbm=isch&ved=2ahUKEwjI2tOFg-z4AhUMZ_EDHR1gDxoQ2-cCegQIABAA&oq=NAS+-+network+att&gs_lcp=CgNpbWcQARgAMgQIABAeMgYIABAeEAgyBggAEB4QCD0ECCMQJzoFCAAQgAQ6BAgAEEM6BggAEB4QBVCkB1jvLGDkQGgDcAB4AIAB8gGIAaUYkgEGMC4xNS4zmAEAoAEBqgELZ3dzLXdpei1pbWfAAQE&scient=img&ei=hZHJYqWEBYzOxc8PncC90AE&rlz=1C5CHFA_enBR905BR905. Acesso em: 15 set. 2021.

Figura 105 [p. 150]

Obra de Scottkipp, CC BY-SA 4.0.

Imagem disponível em:

<https://commons.wikimedia.org/w/index.php?curid=49675363>.

Figura 106 [p. 152]

Imagem CC BY-SA 3.0 Hugovanmeijere.

Figuras 107 e 108 [p. 155]

Imagens utilizadas com permissão da Spectra Logic.

Disponíveis em: <https://spectralogic.com/features/lto-8/>. Acesso em: 7 jul. 2022.

Figura 109 [p. 155]

Imagem utilizada com permissão da Spectra Logic.

Figura 110 [p. 157]

Imagem CC BY-SA 4.0 Omnisource5.

Figura 116 [p. 168]

Imagem utilizada com permissão de Jonas Palm.

Figura 118 [p. 181]

Imagem disponível em:

https://imaging.kodakalaris.com/sites/default/files/wysiwyg/pro/Make_Memories_Last_with_True_Photo_Papers_2019.pdf. Acesso em: 7 jul. 2022.

Imagem disponível em: <https://truephotopaper.com/products>. Acesso em: 8 jul. 2022.

Figura 119 [p. 181]

Imagem disponível em: <https://www.fujifilm.com/it/en/business/photofinishing/photographic-paper>. Acesso em: 8 jul. 2022.

Figura 122 [p. 183]

Imagem disponível em:

<https://produto.mercadolivre.com.br/MLB-983229717-revelaco-90-fotos-10x-15-profissional-no-e-impreso-JM>.

Acesso em: 14 set. 2021.

Figura 123 [p. 184]

Gráfico elaborado pelo autor.

Figura 124 [p. 185]

Imagem de divulgação.

Figura 125 [p. 185]

Imagem de divulgação.

Figura 126 [p. 187]

Fonte: U.S. National Library of Medicine.

Figura 127 [p. 189]

Imagem utilizada com permissão de Robert Grass. ©Robert Grass, ETH Zurich.

Figura 128 [p. 191]

Imagem da Wikimedia.

Figura 129 [p. 192]

Imagem de domínio público.

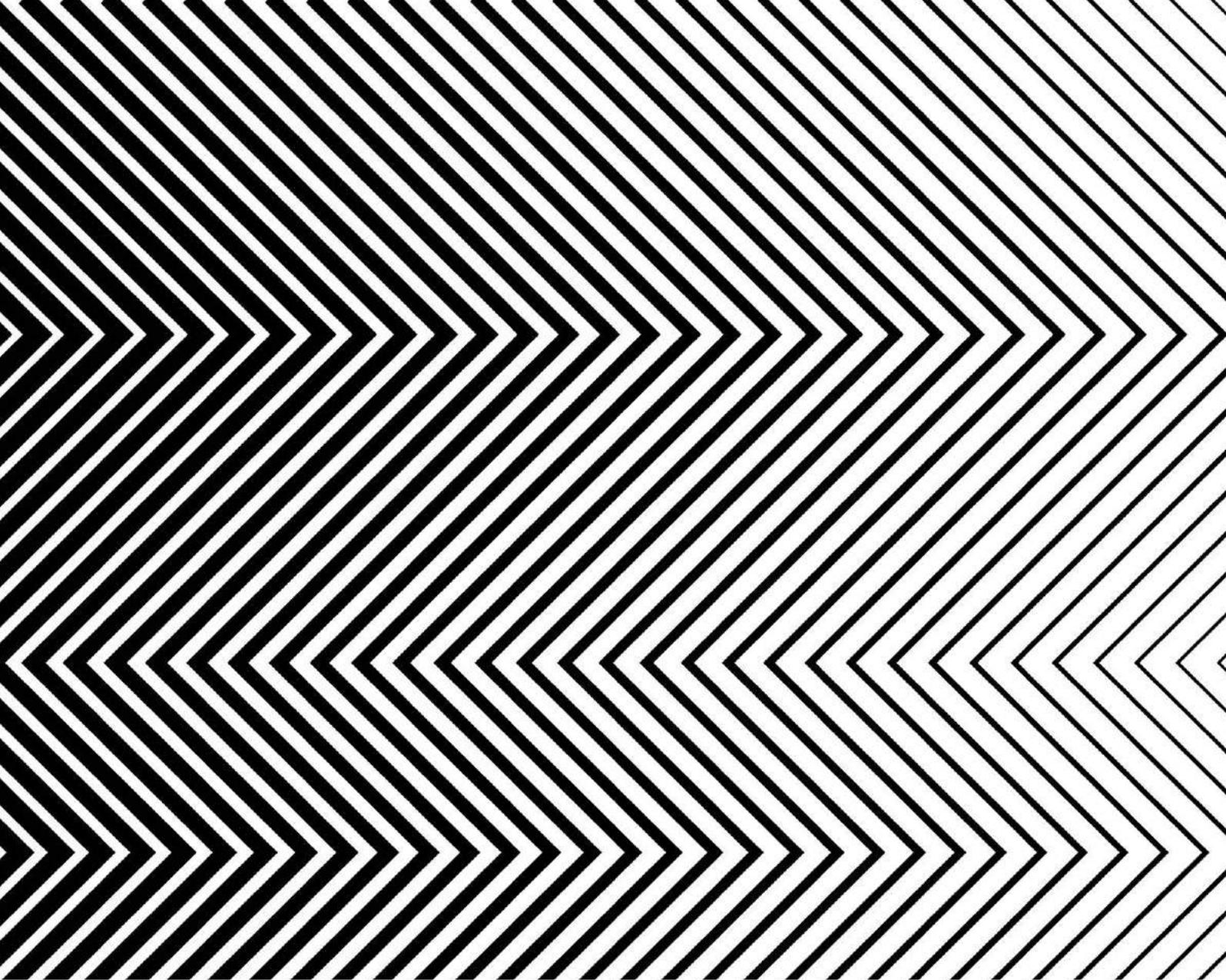
Figura 130 [p. 195]

Imagem de palestra do autor.

Figura 132 [p. 198]

Fotografia de Gabriele Chiesa. Imagem utilizada com permissão do autor.

SOBRE O LIVRO
Formato: 21 x 28 cm
Mancha: 15 x 21,5 cm
Tipologia: Adobe Garamond Pro | Lato | Acumin Variable Concept



FUNDAÇÃO NACIONAL DE ARTES
funarte

ISBN 978-65-5845-013-9



9 786558 450139