

**OTIMIZANDO PROCESSOS: UM ESTUDO DE CASO SOBRE A
IMPLEMENTAÇÃO DE UMA SOLUÇÃO ADOTADA PARA A CONVERSÃO DE
DOCUMENTOS PDF PARA TIFF**

**OPTIMIZING PROCESSES: A CASE STUDY ON THE IMPLEMENTATION OF A
SOLUTION ADOPTED FOR PDF TO TIFF DOCUMENT CONVERSION**

Guilherme Furtado Ribeiro ¹
Janaina Aparecida Alves Scaliza²
Géssica Mina Kim Jesus³
Paulo Roberto Alves⁴

RESUMO: Todas as serventias cartorárias em território nacional recebem diariamente milhares de arquivos eletrônicos enviados por órgãos governamentais, contendo conteúdo que deve ser arquivado digitalmente em sistemas de Gerenciador Eletrônico de Documentos (GED). A maioria desses arquivos é fornecida no formato *Portable Document Format* (PDF). No entanto, alguns sistemas de GED reconhecem apenas determinadas extensões de arquivo, como o *Tagged Image File Format* (TIFF), portanto é necessário converter esses arquivos de um formato para outro, dependendo da arquitetura do sistema. Por isso, esse trabalho teve como objetivo apresentar uma solução adotada por uma empresa de desenvolvimento de sistemas computacionais do interior do estado de São Paulo – SP, para atender alguns cartórios da região. O método utilizado foi um estudo de caso. Como resultado das alterações no código fonte e dos componentes necessários para a conversão, obteve-se uma solução que minimiza drasticamente a chance de corrompimento dos arquivos, mantendo segurança jurídica do documento original e a composição de seus metadados.

Palavras-chave: Arquivamento Digital. Conversão de Arquivos. Sistemas de Gerenciamento de Documentos.

ABSTRACT: All notary offices across the national territory receive thousands of electronic files daily, sent by government agencies, containing content that must be digitally archived in Electronic Document Management (EDM) systems. Most of these files are provided in the Portable Document Format (PDF). However, some EDM systems recognize only specific file extensions, such as the Tagged Image File Format (TIFF). Therefore, it becomes necessary to convert these files from one format to another, depending on the system's architecture. This study aims to present a solution adopted by a software development company located in the countryside of São Paulo State, Brazil, to meet the needs of several local notary offices. The applied method was a case study. As a result of the source code modifications and the integration of the required components for file conversion, the proposed solution significantly reduced the risk of file corruption, ensuring the legal integrity of the original document and the preservation of its metadata structure.

Keywords: Digital Archiving. Document Management Systems. File Conversion.

Bacharel em Engenharia da Computação – Faculdade Gran Tietê - E-mail:
guilhermerht@hotmail.com ¹

Doutora em Engenharia de Produção - Faculdade Gran Tietê - E-mail:
janainaalvesscaliza@hotmail.com ²

Doutora em Engenharia de Produção - Instituto Federal do Norte de Minas Gerais - E-mail:
gessica.jesus@ifnmg.edu.br ³

Mestre em Engenharia de Produção – Faculdade de Engenharia de Bauru/FEB - E-mail:
pr.alves@gmail.com ⁴

1 INTRODUÇÃO

Com o avanço tecnológico das últimas décadas, houve uma modificação nos padrões de arquivamento de documentos, com a utilização de formatos digitais de arquivos em detrimento de meios físicos (Banza, 2023).

A transição para o formato de arquivamento digital de documentos demandou uma ampla adaptação por parte de várias empresas e entidades prestadoras de serviços, incluindo os cartórios. A maioria destes costumava seguir a prática de arquivamento físico de documentos, utilizando papel timbrado de segurança (geralmente controlado por entidades governamentais) ou folhas simples no tamanho A4 (Martins, 2023).

No entanto, assim como existe o desafio físico do espaço para armazenar todos esses documentos, há também um contratempo equivalente no âmbito digital. Vários documentos podem abranger numerosas páginas, as quais precisam ser arquivadas digitalmente com uma qualidade de resolução adequada (Brasília, 2020). Isso impacta diretamente no tamanho do documento, principalmente se o formato dos arquivos for *Tagged Image File Format* (TIFF), formato esse que prioriza a qualidade e realiza uma compressão sem perdas de informação visual (Trigo, 2015).

Todas as serventias cartorárias presentes em território nacional recebem diariamente milhares de arquivos eletrônicos que são enviados por órgãos governamentais contendo conteúdo que deverá ser arquivado de forma digital em sistemas de Gerenciador Eletrônico de Documentos (GED), segundo Martins (2023). Em sua grande maioria esses arquivos são disponibilizados em formato *Portable Document Format* (PDF), porém existem sistemas de GED que, dependendo de sua arquitetura, apenas reconhecem algumas extensões de arquivos, como por exemplo o TIFF, dessa forma sendo necessária a conversão de um formato para outro (Raabe; Pohlmann Filho, 1998).

Visando destacar a preservação meticulosa da qualidade das imagens, aliada à mitigação dos potenciais riscos de corrompimento durante todo o processo de conversão de arquivos digitais do formato PDF para o formato TIFF, este artigo tem o objetivo de apresentar uma solução adotada por uma empresa de desenvolvimento de sistemas computacionais do interior do Estado de São Paulo – SP, para atender alguns cartórios da região.

2 MATERIAL E MÉTODOS

No contexto deste estudo de caso, foi feita uma análise detalhada da adaptação realizada por uma empresa desenvolvedora de *softwares* para serventias cartorárias ao enfrentar dificuldades relacionadas à conversão de documentos do formato PDF para TIFF entre seus sistemas.

A princípio, constatou-se que a ferramenta de conversão de documentos já integrada ao sistema encontrava-se sujeita a perturbações, resultando em alterações e, por vezes, na corrupção dos arquivos submetidos ao processo. Esta constatação foi feita com base nos chamados registrados no *HelpDesk* da empresa, provenientes dos clientes que fazem uso regular dessa ferramenta. Na maioria dos casos, os clientes reportavam que, ao escolher um documento em formato PDF para a conversão no sistema, ocorria um bloqueio completo do processo sistêmico. Após uma espera incomum de alguns minutos, o sistema apresentava uma mensagem indicando que a conversão do arquivo fora realizada com êxito. Entretanto, ao tentar acessá-lo, verificava-se que o documento estava apenas parcialmente convertido ou, em alguns casos, completamente corrompido.

Efetuando uma análise minuciosa sobre a origem dessa discrepância, foi confirmado que tal incoerência se manifestava em dois cenários preponderantes. O primeiro cenário se delineava quando o documento PDF escolhido para a conversão possuía uma quantidade significativa de páginas, ultrapassando a marca notável de 400. O segundo cenário, por sua vez, surgia quando o arquivo contava com páginas cujas dimensões discrepavam de forma substancial das demais, rompendo com o padrão convencional estabelecido pelo formato A4, que possui 297 milímetros de altura e 210 milímetros de largura (Penna *et al.*, 2014).

3 REFERÊNCIAL TEÓRICO

Atualmente existem diversos formatos de arquivos disponíveis no escopo computacional, cada um com suas especificidades e melhor cenário de aplicabilidade, alguns são melhores para fotografias, outros para arquivamento de mídia impressa e assim por diante (Da Silva; Cardia Neto; Hoffman, 2021). Imagens digitais podem ser preservadas de várias maneiras, cada um com seu código exclusivo, resultando em arquivos de tamanhos variados. Os formatos e suas características específicas têm

impacto direto no tamanho final tanto da imagem quanto do arquivo, seja para fins de armazenamento ou impressão.

O TIFF é um tipo de formato de arquivo que foi desenvolvido pela *Aldus* (atualmente integrada a *Adobe Corporation*) e *Microsoft*, fundamentado em marcações projetado para a preservação e a partilha de imagens. Ele funciona como um conversor versátil para uma variedade de codificações utilizadas em imagens em formato de bitmap. Estas codificações são responsáveis por aplicar diversos métodos de compressão para representar as cores de forma eficiente (Almeida; Cendón; Souza, 2012).

O PDF é um formato de arquivo muito utilizado para o armazenamento de arquivos digitais, oferecendo alta compatibilidade com diversos *hardwares*, *softwares* e sistemas operacionais. Segundo Trigo (2015), este formato, centrado na estrutura de páginas, destaca-se por sua capacidade de manter a integridade visual do conteúdo, garantindo uma experiência consistente para os usuários, desde a criação até o arquivamento e o uso final.

Como definido por Falcão (2009), a conversão de imagens entre os formatos citados passa por diversas etapas de compressão e alteração de metadados, o que pode muitas vezes corromper o arquivo final gerado, situação que pode perdurar por decorrência do sistema operacional utilizado no momento da conversão.

Os GED necessitam da preservação total da estrutura original dos arquivos mesmo após sua conversão ser realizada, para que a validade jurídica antes concedida ao documento permaneça imutável. Como concluído por Da Silva *et al.* (2003), a segurança jurídica de um acervo digital garante a boa fé na prestação de informações aos indivíduos solicitantes de tais documentos.

4 RESULTADOS E DISCUSSÕES

Com base em todas as análises realizadas sobre o problema de conversão, a equipe de desenvolvimento tomou uma decisão. Para evitar esses cenários problemáticos, optou-se por adaptar o sistema para adotar o método *Single-TIFF*. Esse método não transforma o PDF de entrada diretamente em um único arquivo TIFF, mas cria, em vez disso, um arquivo TIFF único separado para cada página do PDF. Todas essas páginas são armazenadas em uma pasta dedicada, proporcionando uma solução robusta e segura para o desafio enfrentado.

Tendo como base o ponto de partida, foi realizada a alteração no código fonte do sistema e no principal componente utilizado para realizar o processo, componente esse que possui o nome *GhostScript*. Na versão antiga do sistema, onde o problema ainda era presente, quando o usuário requisitava a conversão de um documento PDF para TIFF, o componente realizava o comando apresentado na Figura 1.

Figura 1: Comando antigo executado

```
gs -sDEVICE=tiffg4 -r300x300 -o output.tiff input.pdf
```

Fonte: Elaborado pelos Autores (2023).

As opções, argumentos do comando apresentado na Figura 1 representam Os dados apresentados na Figura 1 representam:

- -sDEVICE=tiffg4 define o dispositivo de saída como TIFF com compressão CCITT G4, que é comum para imagens em preto e branco.
- -r300x300 define a resolução de saída para 300 DPI (*dots per inch* ou pontos por polegada).
- -o output.tiff especifica o nome do arquivo de saída TIFF, ou seja, nesse momento o sistema define a formatação do seu nome. O padrão utilizado pelo GED para realizar a visualização da imagem oito caracteres, como por exemplo: 00001598.TIFF.
- input.pdf é o nome do arquivo PDF de entrada que será convertido.

Após o comando ser executado, conforme a Figura 1, o resultado obtido na conversão do arquivo está representado na Figura 2.

Figura 2: Conversão inicial



Fonte: Elaborado pelos Autores (2023).

Após a definição de como seria a nova tratativa de conversão, o código apresentado sofreu algumas alterações para que o arquivo fosse PDF fosse convertido em páginas TIFF separadas e dentro de uma pasta específica. Ao final, o

código fonte ficou conforme apresentado na Figura 3.

Figura 3: Comando novo executado

```
gs -sDEVICE=tiffg4 -r300x300 -o output-%03d.tif -sOutputFile=/caminho/para/pasta/%03d.tif input.pdf
```

Fonte: Elaborado pelos Autores (2023).

Sobre a figura 3, destaca-se:

- o output-%03d.tif especifica o padrão de nomenclatura dos arquivos de saída. %03d é um espaço reservado que será substituído pelo número da página. Isso garantirá que os arquivos TIFF sejam nomeados sequencialmente com três dígitos (por exemplo, output-001.tif, output-002.tif, etc.).
- -sOutputFile=/caminho/para/pasta/%03d.tif especifica o diretório onde os arquivos TIFF serão salvos.

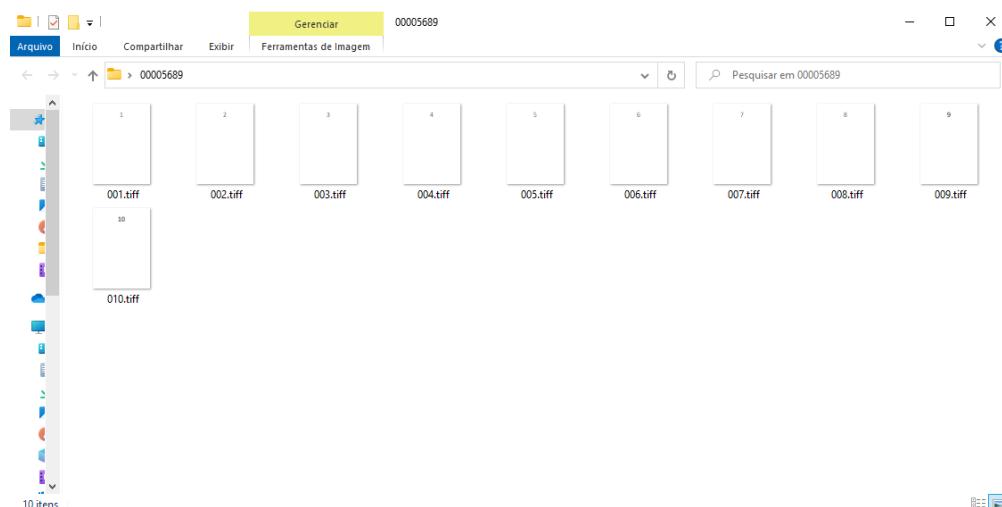
O comando representado pela Figura 3 converterá cada página do arquivo PDF em um arquivo TIFF separado e os salvará na pasta especificada. O *Ghostscript* criará automaticamente os arquivos TIFF e a pasta de saída se eles não existirem (Figuras 4 e 5).

Figura 4: Conversão Final



Fonte: Elaborado pelos Autores (2023).

Figura 5: Pasta com imagens



Fonte: Elaborado pelos Autores (2023).

Seguindo a mesma linha de pensamento de Da Silva *et al.* (2003) em relação à segurança jurídica dos documentos digitais, esta nova forma de conversão incorporada no sistema é crucial para preservar a integridade do arquivo, especialmente quando se trata de serventias cartorárias.

Neste contexto, é fundamental destacar que o PDF original permanece inalterado em sua composição original de metadados, ponto esse que é destacado por Trigo (2015). Além disso, ao final do processo, o documento TIFF gerado passa por uma separação das páginas, o que não apenas simplifica a organização, mas também minimiza significativamente o risco de corrupção ou perda de resolução do arquivo, fato esse que era comum de ocorrer antes das alterações realizadas, principalmente se o sistema operacional utilizado no terminal em que o sistema estava sendo executado se tratada do *Windows*, como apontado por Falcão (2009).

Essa abordagem não apenas assegura a segurança jurídica dos documentos digitais, mas também garante a preservação integral de seu conteúdo no formato original e depois de convertido pelo sistema. Dessa forma, torna-se possível manter a confiabilidade e autenticidade dos registros cartorários, proporcionando uma solução eficaz e confiável para o armazenamento e preservação de documentos importantes.

5 CONSIDERAÇÕES FINAIS

A demanda de arquivos eletrônicos crescerá cada vez mais no âmbito cartorário, e as serventias deverão estar preparadas no âmbito tecnológico para realizar o devido armazenamento e desmaterialização dos arquivos recebidos, além de manter a segurança jurídica dos documentos. Por conta desse cenário foi necessário realizar um estudo detalhado a respeito de como tais documentos devem ser manipulados digitalmente, em especial suas conversões.

Diante disso, o presente trabalho teve como objetivo principal demonstrar uma solução adotada para realizar a conversão de documentos no formato PDF para o formato TIFF, assegurando o mantimento de toda a composição de metadados e minimizando a chance de corrompimento dos arquivos nesse processo.

O principal ponto limitador deste trabalho foi o cenário e situação em que a ferramenta foi desenvolvida. Como tudo foi baseado e estudado dentro de uma arquitetura específica de sistema computacional, pode haver discrepâncias na adoção dessa solução para outro sistema informatizado. Cada sistema possui sua tratativa e

arquitetura, seja ela de pastas ou de código fonte, o que pode influenciar diretamente no uso da ferramenta desenvolvida, além de que o componente adotado para a conversão pode não satisfazer totalmente as necessidades em outro cenário.

Existe uma grande necessidade em estudar outros componentes de conversão de arquivos de imagem, não apenas nos formatos PDF e TIFF, mas também em toda a gama de documentos presentes hoje em dia no universo computacional. Cada componente de conversão pode possuir propriedades e tratativas específicas para cada tipo de extensão, o que pode gerar um maior dinamismo e otimização nos processos solicitados, gerando melhores resultados e melhorias.

REFERÊNCIAS BIBLIOGRÁFICAS

ALMEIDA, Maurício Barcellos; CENDÓN, Beatriz Valadares; SOUZA, Renato Rocha. Metodologia para implantação de programas de preservação de documentos digitais a longo prazo. Encontros Bibli: **Revista eletrônica de biblioteconomia e ciência da informação**, v. 17, n. 34, p. 103-130, 2012.

BANZA, Ana Paula. O texto desmaterializado: explorando os limites da Filologia Digital. **Revista LaborHistórico**, v.9, n.1, e51304, 2023.

BRASÍLIA. Presidência da República. Subchefia Para Assuntos Jurídicos. **DECRETO Nº 10.278, DE 18 DE MARÇO DE 2020**. 2020. Presidência da República Secretaria-Geral. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2019-2022/2020/decreto/D10278.htm. Acesso em: 30 ago. 2023.

DA SILVA, Danielle Pereira et al. GED–Gerenciamento Eletrônico de Documentos A tecnologia que está mudando o Mundo. **INICIA**, v. 37, p. 38, 2003.

DA SILVA, Guilherme Antonio Noma; NETO, Rubens Cardia; HOFFMANN, Maria Luisa. Tamanho e qualidade: ferramentas para a melhor utilização das fotografias em diferentes finalidades. **Revista Informação em Cultura, Qualis Capes B2**, v. 3, n. 2, p. 07-26, 2021.

FALCÃO, Livia Maria Rocha. **Evaluation the reproducibility of cephalometric landmarks on digital lateral headfilms at the file formats DICOM, TIFF and JPEG**. 2009. 89 f. Dissertação (Mestrado em Ortodontia) - Universidade Metodista de São Paulo, São Bernardo do Campo, 2009.

GIL, Antônio Carlos. Como classificar as pesquisas. **Como elaborar projetos de pesquisa**, v. 4, n. 1, p. 44-45, 2002.

GIL, Antônio Carlos. **Métodos e técnicas de pesquisa social**. 6. Ed. Editora Atlas SA, 2008.

MARTINS, Kyrianny Faria. Os impactos na prática registral após a adaptação

tecnológica dos cartórios durante o enfrentamento da pandemia da COVID-19. **Revista de Direito Notarial**, v. 5, n. 1, 2023.

PENNA, Luiz Fernando da Rocha *et al.* Diagnóstico do consumo de papel A4: o caso do Instituto Federal Minas Gerais - Campus Governador Valadares-MG. **V Congresso Brasileiro de Gestão Ambiental**. Belo Horizonte, p. 1-10. 27 nov. 2014.

RAABE, André; POHLMANN FILHO, Omer. Estudo comparativo entre sistemáticas de digitalização de documentos: formatos HTML e PDF. **Ciência da Informação**, v. 27, p. 300-310, 1998.

TRIGO, T. **Equipamento fotográfico: teoria e prática**. 6. ed. São Paulo: Senac, 2015.