

PRODUÇÃO DE MATRIZES BND

ANÁLISE DE SOLUÇÕES ALTERNATIVAS AO FORMATO TIFF

SUMÁRIO:

1. INTRODUÇÃO

2. COMPRESSÃO TIFF LZW

3. FORMATO JPEG 2000

3.1 PROCESSO DE CONVERSÃO JP2

3.2 REQUISITOS TÉCNICOS JP2

3.3 PRINCIPAIS RISCOS JP2

3.4 AVALIAÇÃO DO FORMATO JP2

4. CONCLUSÕES

4.1 COMPARAÇÃO TIFF LZW/JP2

4.2 APLICAÇÃO PRÁTICA BNP

5. TRABALHO FUTURO

6. BIBLIOGRAFIA

ANEXO A - SOFTWARE DISPONÍVEL PARA JPEG 2000

ABREVIATURAS:

BL - British Library; NAUK - National Archives UK; DPC - Digital Preservation Coalition; KCL - Kings College London; OPF - Open Planets Foundation; LAC - Library and Archives Canada; NDNP LC - National Digital Newspaper Program Library of Congress

1. INTRODUÇÃO

A necessidade de diminuir o espaço de armazenamento destinado às matrizes da Biblioteca Nacional Digital (BND) é a principal motivação para a análise de soluções alternativas à atual utilização do formato TIFF sem compressão, ponderando as seguintes opções: compressão LZW sem perdas dos ficheiros TIFF; conversão para o formato JPEG 2000.

O quadro abaixo resume os ganhos em espaço de armazenamento conseguidos quer pela conversão dos ficheiros no formato JPEG 2000, quer pela compressão LZW dos TIFF (Gillesse, 2008).

TIFF SEM COMPRESSÃO	TIFF LZW	JP2 Lossless	Jp2 Lossy
100%	30%	52%	91-98%

Mas este não é o único parâmetro de análise, neste relatório serão também ponderadas as consequências da utilização destes formatos ao nível da qualidade de imagem e da sustentabilidade a longo prazo (Gillesse, 2008).

2. Compressão TIFF LZW

A compressão LZW (Lempel-Ziv-Welch) é um algoritmo criado em 1988 e que pode ser usado dentro dos ficheiros TIFF. Apesar do TIFF não ser um standard ISO, a descrição do formato está disponível livremente e de forma aberta no site da Adobe, sendo esta uma das principais razões para a adopção deste formato para as matrizes de imagens criadas por a grande maioria das organizações do sector cultural. O mesmo não poderá, contudo, afirmar-se relativamente à compressão LZW, pois a Unisys reclama ter ainda a patente sobre esse algoritmo, pese embora a mesma ter expirado em 2003, nos EUA, e em 2004, na Europa e Japão (Gillesse, 2008).

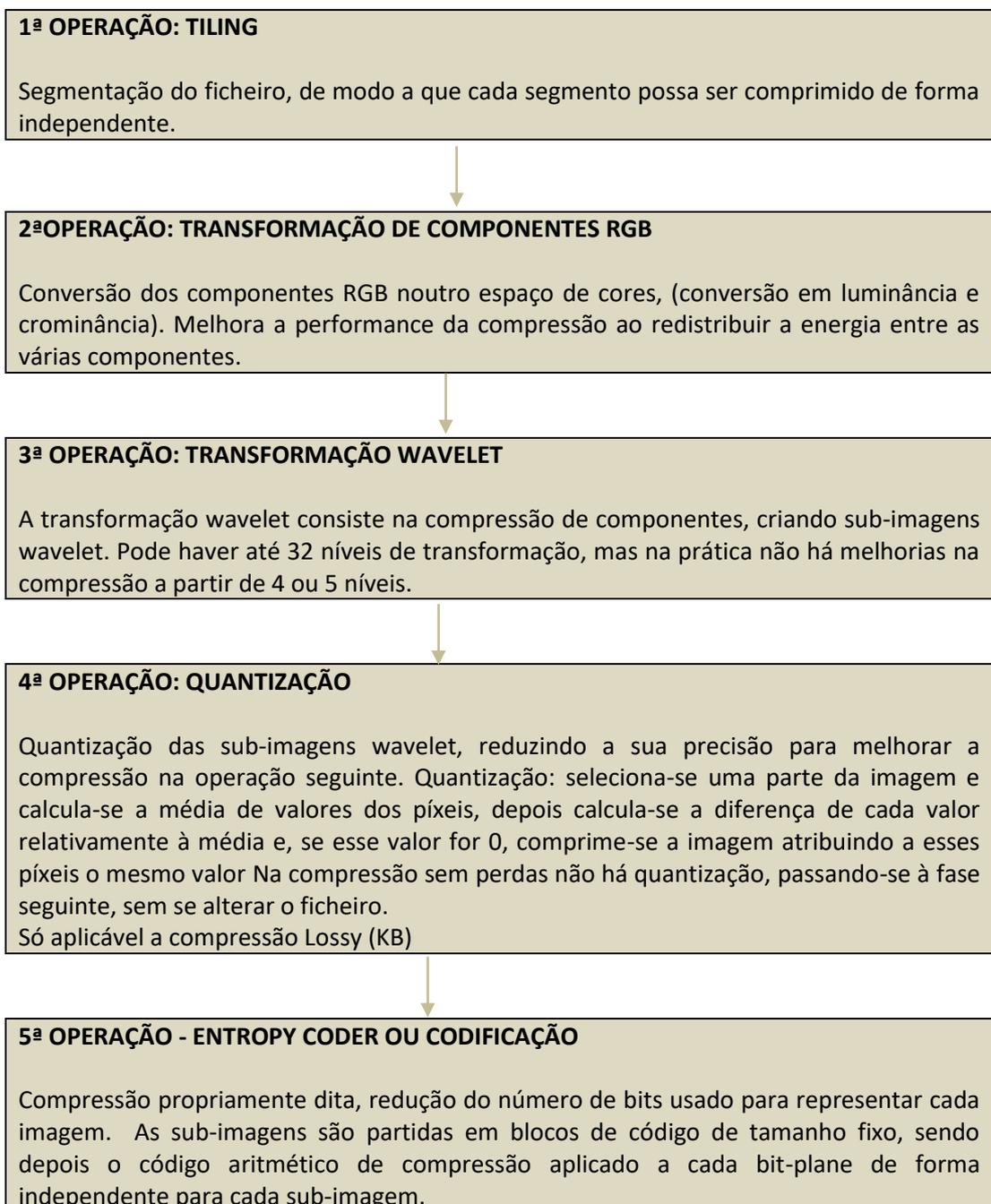
Relativamente à avaliação da compressão do formato TIFF, replicamos abaixo a análise da KB (Gillesse, 2008), pois permite quantificar os resultados em cada parâmetro e, assim, facilita a comparação entre os diferentes formatos.

Avaliação KB - TIFF LZW		
Parâmetro	Avaliação	Pontuação (1-5)
Espaço de armazenamento	Ganhos de 30% face ao TIFF sem compressão	1
Qualidade de imagem	Lossless, sem perdas, não tem impacto nenhum na qualidade da imagem	5
Sustentabilidade a longo prazo	Patente LZW reclamada pela Unisys e consequente baixa taxa de adopção no sector cultural	1

3. Formato JPEG 2000

O formato JPEG 2000 é um standard (ISO/IEC 1544-1) para a compressão de imagens digitais fixas, desenvolvido com o duplo objetivo de melhorar a performance do formato JPEG e de lhe acrescentar características, que permitiram o desenvolvimento de novos atributos do formato e de que se destaca a denominada “descodificação inteligente”, que possibilita o acesso/descodificação apenas da quantidade de código necessária para a tarefa que se pretende realizar (Buckley, 2008).

3.1 Processo de conversão jp2



Isto permite o acesso e a descodificação apenas de algumas regiões da imagem e que os bit-planes de cada sub-imagem sejam decompostos em camadas de qualidade.

3.2 Requisitos técnicos jp2

Na análise das especificações técnicas do formato JPEG2000, foram considerados os requisitos constantes nos perfis JPEG2000 das bibliotecas que já adotaram este formato e cujo resumo se pode consultar em: http://www.dpconline.org/component/docman/doc_download/529.

Campo	Descrição	Valores mais utilizados	Exemplos
Formato do ficheiro	<p>Formatos JPEG 2000:</p> <p>JP2 (.jp2), descrito na Parte 1 da ISO</p> <p>JPX (.jpx), pode incluir mais do que uma imagem, características avançadas, descrito na Parte 2 ISO</p> <p>MJ2, imagens em movimento, cada frame é um ficheiro, descrito na Parte 3 da ISO</p> <p>JPM (.jpm), imagens compostas por imagem, texto e gráficos, na Parte 6 da ISO)</p>	JP2 (.jp2)	Maioria das BN
Capture resolution	Resolução da imagem resultante da digitalização. Não usar o campo "Display resolution"	Resolução do TIFF de origem.	
Compressão	Lossy, compressão com perdas de bits, que portanto é irreversível KCL - A compressão lossy deve ser "minimally lossy", com este coder nenhuns dados são deitados fora (apesar de continuar a haver perdas por causa das transformações irreversíveis).	Lossy	Maioria das BN; KCL
	Lossless, compressão que preserva todos os bits e que, portanto, é reversível	Lossless	BN Checa; BN Noruega; KB; LAC (matrizes cartografia/iconografia)

Campo	Descrição	Valores mais utilizados	Exemplos
)
Níveis de compressão	Os níveis de compressão expressam-se por x:y, em que x é o nº de bits original e y é o nº bits comprimido. A DPC (2008) aconselha 24:1 como sendo o valor máximo, a partir é o valor máximo, a partir do qual começa a haver artefactos na imagem e perda de qualidade visível	2:1 ou 3:1	BN Checa Lossless
		2:1	DPC (lossless)
		6:1 ou 10:1	Wellcome Lossy
		4:1	KCL Lossy
		Não especificado (Têm de ser determinados em função dos conteúdos a digitalizar e dos recursos financeiros disponíveis)	BL Lossy
Não especificado (O resultado deve ser: 1 oitavo do tiff)	NDNP LC		
Regiões de interesse?	Seleção de zonas da imagem que são codificadas e transmitidas com melhor qualidade do que o resto da imagem. Estas regiões são descodificadas antes do resto da imagem	Não	Maioria das BN
Espaços de cor	Há dois métodos alternativos para especificar o sistema de cor: o método enumerativo (valor numérico que representa determinado espaço de cor) e método de embeber um perfil ICC no ficheiro.	Enumeração de valores: sRGB; greyscale ou SYCC	
		Perfil ICC embebido - únicos valores suportados, sem riscos, pelo jp2: - Restricted ICC Profile - Three componente matrix - Input device	BN Checa
Checksum	Deteção de erros de corrupção de ficheiro	Sim	OPF

Campo	Descrição	Valores mais utilizados	Exemplos
Profundidade de bits		24	LAC
		Determinada pelo scanner	
Dimensão da imagem	Altura e largura do ficheiro	Ficheiro com a mesma altura e largura do TIFF	NDNP LC, KCL
1ª OPERAÇÃO: TILING			
Tiling?	Com esta segmentação consegue-se reduzir a quantidade de memória necessária para implementar o formato, garantindo acesso independente a cada uma das regiões da imagem	Não: Apenas 1 tile que corresponde à imagem inteira. A BL comparou a utilização de tiles ou precints em termos de tempo de descodificação , tendo concluído que é melhor utilizar precints pois são mais rápidos	BL e NAUK
		Sim: Para evitar corrupção do ficheiro, recomenda-se que não haja só um tile (OPF) O tiling, permite a utilização de menos memória na descodificação da imagem (KB)	KB OPF

Campo	Descrição	Valores mais utilizados	Exemplos
Tamanho de cada tile ¹	O tamanho de cada tile não tem impacto no tamanho do ficheiro. Se os tiles forem muito pequenos e se a compressão for simultaneamente lossy e muito elevada, pode criar artefactos de blocos (blocking). A utilização de um só tile impede o blocking (KB)	1 tile para a imagem inteira	BL
		1024x1024	BN Noruega, Wellcome, NDNP LC LAC
		4096X4096	BN Checa
Marcadores TLM (Tile Length Markers)	Podem melhorar a performance e o acesso, pois ajudam a encontrar as “fronteiras” de cada tile num datastream (KCL)	Yes	Welcome
		No	BL
Tile header	Não contém coding style default, coding style component, quantization default ou quantization componente marker segments		
2ª OPERAÇÃO: TRANSFORMAÇÃO DE COMPONENTES RGB			
Transformação de componentes?	A transformação de componentes RGB é opcional, mas altamente recomendável para espaços de cor RGB (KB)	Sim	BL
Nº de componentes	Especificação do nº de componente: 1=cinzas 3=cores	3	BL; LAC
		1	NDNP LC
3ª OPERAÇÃO: TRANSFORMAÇÃO WAVELET			
Filtros wavelet filter	Os valores mais usados para especificar o tamanho dos filtros usados para implementar a transformação wavelet são: 9-7 (aplicável a lossy) ou 5-3 (aplicável a lossless) (DPC+KCL)	9-7	BL , NDNP LC
		5-3	BN Checa, BN Noruega LAC: Matrizes cartografia/ iconografia
Tipo transformação	ICT (Irreversible Component Transformation): gera maior compressão, mas implica perda de informação, pelo que pode haver erros de quantificação (KB). É irreversível e só se aplica a compressões lossy	Irreversível ICT	BL , NDNP LC
	RCT (Reversible Component transformation): gera menor	Reversível RCT	BN Checa, BN

¹ Parâmetros relacionados com o acesso e a performance (KCL)

Campo	Descrição	Valores mais utilizados	Exemplos
	compressão e só é aplicável a lossless.		Noruega LAC: Matrizes cartografia/ iconografia
4ª OPERAÇÃO: QUANTIZAÇÃO			
5ª OPERAÇÃO - ENTROPY CODER OU CODIFICAÇÃO			
Code block size		6, 64x64	Maioria das BN
Code block style	O "Bypass" impede a codificação de bit-planes mais baixos. Isto permite poupar tempo e tem pouco impacto na compressão (DPC)	Bypass	BL, BN Noruega, NDNP LC
Nº de níveis de decomposição	Cada nível de decomposição cria componentes com diferentes resoluções	6	Maioria das BN NDNP LC LAC
		5	KCL
Nº de layers de qualidade ²	<p>Cada camada de qualidade adiciona mais dados codificados e melhora a qualidade da imagem descomprimida. Adicionar camadas de qualidade, diminui a compressão. Muitas aplicações usam apenas 1 camada de qualidade (KCL)</p> <p>Para ficheiros master, deve haver múltiplas camadas de qualidade, sempre que a compressão for lossy (DPC).</p> <p>Cada layer tem um determinado nº de bits por pixel (1, 0.84, 0.7, etc).</p>	Múltiplos	BL Depende do material. Não há vantagem em extrair camadas de qualidade mais elevada, se a resolução for baixa, pois isso não tem impacto ao nível da visualização e só vai tornar o acesso mais lento
		1	BN Checa No formato jp2 só pode haver 1

² Parâmetros relacionados com o acesso e a performance (KCL)

Campo	Descrição	Valores mais utilizados	Exemplos
			camada
		25	NDNP LC
		8	Wellcome
		10-25	LAC
N.º de níveis de resolução ³	Adicionar níveis de resolução aumenta a compressão (DPC). Dois ou três níveis de resolução facilitam a compressão, mas mais do que isso não adianta. Esses níveis são adicionados apenas até que a sub-imagem de resolução mais baixa seja um thumbnail no tamanho desejado (normalmente: 5 níveis). Para ficheiros master, deve haver níveis múltiplos níveis de resolução, sempre que a compressão for lossy (KCL)	Depende do material	BL
Progression order	Cada layer de qualidade é uma colecção de pacotes. Os pacotes são os elementos fundamentais do codestream de um ficheiro JPEG2000. Cada layer incrementa a qualidade da imagem no seu todo. Cada pacote incrementa a qualidade numa posição específica de um componente de um tile. A ordem de progressão dos pacotes determina a ordem por que os dados comprimidos são recebidos e descomprimidos. Esta ordem de progressão resulta da forma como os pacotes são intercalados no codestream. Se a resolução (R) vier em primeiro lugar, os pacotes de menor resolução vêm em primeiro lugar, seguidos das resoluções mais elevadas (DPC). Os valores possíveis são RPCL, RLCP, LRCP e CPRL. O RPCL é o que permite melhor velocidade de decoding (o PCRL e CPRL são os piores).	RPCL (Resolution; Position; Component; Layer)	BN Checa, BL, BNNoruega
		RLCP ou RPCL	NDNP LC
Code Stream	A utilização destes marcadores para o	Packet-length	BL

³ Parâmetros relacionados com o acesso e a performance (KCL)

Campo	Descrição	Valores mais utilizados	Exemplos
markers	encoding torna mais rápido o decoding das imagens.	markers	
Start-Of-Packet marker?	Propicia pontos de referência para a re-sincronização periódica do decoder, de forma a limitar os efeitos dos erros e a sua propagação no codestream (DPC). Limita os erros de corrupção de ficheiro e permite que o sw de decodificação recupere dos mesmos (OPF)	Sim	DPC; OPF
End-Of-Packet marker?		Sim	OPF
Localização dos marcadores		Main header	LAC
Segmentation Symbols?	Limita os erros de corrupção de ficheiro e permite que o sw de decodificação recupere dos mesmos (OPF)	Sim	OPF
Precints size	Um precinct é um conjunto de blocos de código. A BL comparou a utilização de tiles ou precincts em termos de decoding time, tendo concluído que é melhor utilizar precincts pois são mais rápidos.	256x256 para as primeiras duas componentes	BL BN Checa
		128x128 para os componentes de níveis mais baixos	
		Não especificado de forma explícita	Wellcome; KCL

3.3 Principais riscos do formato jp2

Dos problemas descritos nesta secção, salienta-se como mais importante a falta de robustez dos ficheiros JPEG2000 (3.3.4 - Corrupção de ficheiros), quer pelo impacto no acesso ao objeto digital, quer pela gravidade desta ocorrência relativamente a documentos originais de natureza frágil (por exemplo, jornais).

Merecem também especial atenção os riscos associados à especificação de espaços de cor (3.3.3 - Espaços de cor) e os problemas de implementação do formato pelo software (3.3.5 - Problemas de software), pois as soluções apresentadas exigem o domínio de ferramentas de validação complexas, bem como a alteração da própria norma ISO ou a alteração de software de conversão para que a norma seja implementada.

Por último, os problemas de identificação correta do formato (3.3.1 Formato JPX, em vez de JP2) e de especificação do campo relativo à resolução (3.3.2 Resolução especificada no campo errado do header) são facilmente ultrapassáveis, bastando para o efeito escolher o software de conversão adequado.

3.3.1 Formato JPX, em vez de JP2

Ficheiros identificados como sendo JP2, quando na realidade são JPX (exemplo: Adobe JPEG 2000 plugin do Photoshop). O domínio e implementação do formato pelo software atual é ainda fraco, pelo que estes ficheiros poderão perder informação e não ser rendidos no futuro (OPF).

Soluções:

- Validação por Jpylyser, recusa de ingestão
- Não usar Photoshop

3.3.2 Resolução especificada no campo errado do header

Ambiguidade da norma na descrição dos campos “capture resolution” e “display resolution”. Há sw que apenas implementam valores no campo “display”. Se a resolução não estiver especificada de forma rigorosa, pode haver problemas de impressão de originais ou até na utilização de sw de ocr (OPF)

Soluções:

- Revisão da definição ISO relativamente a estes campos
- Validação por Jpylyser, recusa de ingestão
- Não usar Kakadu (até à v. 6), imageMagik ou OpenJPEG, pois não implementam valores no campo “capture resolution”.

3.3.3 Espaços de cor

A opção pelo método enumerativo, permite a especificação de espaços de cor RGB através de um código numérico. Esta forma de especificação não é standard.

Solução deste problema:

- ❖ Revisão da norma ISO JPEG 2000 (proposta em Outubro 2012)

Em alternativa ao método enumerativo acima especificado, é possível optar pelo método de embeber determinado perfil ICC. O formato jp2 não permite embeber informação relativa a espaços de cor RGB que sejam do tipo “Display” (ex: Adobe), pois só admite perfis do tipo “Input”. Se for especificado um perfil “Display”, há sw que altera o formato do ficheiro, convertendo-o em jpx.

Soluções (Knijff, 2011):

- ❖ Forçar o valor “Input”, mesmo que não corresponda à realidade, pois é preferível manter o formato jp2, do que a conversão automática do mesmo em jpx.
- ❖ Validação dos seguintes valores pelo Jpylyser:
 - Perfil ICC restrito
 - “Three componente matrix” (valor não permitido: “N-component Lut based method”)
 - “Input device” (valores não permitidos: “Display Device” e “Output Device”)

3.3.4 Corrupção de ficheiros

Corrupção de bits ou de bytes que pode tornar os ficheiros total ou parcialmente ilegíveis. As causas podem consistir em falha de hw ou sw no processo de criação do ficheiro, erro na transferência de ficheiros ou deterioração do suporte físico de armazenamento (OPF).

“Variable length coders, such as the arithmetic coder in JPEG 2000 and the Huffman coder in JPEG, can be vulnerable to errors. For example, errors that appear in the middle of the code stream can cause the code to lose synchronization and make all the data from there on in is not decodable” (DPC).

A complexidade da codificação do jp2 leva a que a corrupção do ficheiro possa afetar na totalidade a decodificação da imagem, ao contrário do tiff que, apesar de corrompido, pode permitir ver grande parte do ficheiro. Por isso é muito importante a aplicação de métodos de controlo de qualidade de imagem em todos os passos do workflow, especialmente utilizando as ferramentas Scape (BL).

O facto de a codificação não ser progressiva, mas sim sequencial, leva a que a corrupção dos dados interrompa o processo de codificação/decodificação, perdendo-se o resto da imagem. A imagem pode ficar totalmente corrompida apenas com 0,001% de erro, pois um simples erro pode afetar todo o bitstream e impedir a decodificação correta da imagem (Buonora, 2008).

Soluções (OPF):

- ❖ Na criação do ficheiro:
 - Utilização de marcadores “start-of-packet”, “end-of-packet” e de símbolos de segmentação (limita o impacto dos erros de e permite ao decodificador recuperar de erros)
 - Não usar um só tile
 - Validação com Jpysyer
 - Comparar pixéis das imagens de fonte com as do destino. (Comparação TIFF/JP2 utilizando a ferramenta Scape)
 - Gerar checksum
 - Usar precints mais pequenos (Buonora, 2008)

- ❖ Na pré-ingestão:
 - Verificar checksum
 - Validação com Jpysyer

- ❖ Manutenção:
 - Verificação regular de checksum
 - Manter cópias múltiplas

- ❖ Utilização de sw de remediação FixIt (Buonora, 2008)

3.3.5 Problemas sw

Os programas open source para a codificação/decodificação de ficheiros jp2 são lentos, instáveis e não suportam todas as características do formato. Há poucas ferramentas de controlo de qualidade, o que aumenta o risco de erros no workflow da digitalização. A reutilização destes ficheiros pelo grande público torna-se mais difícil pela ausência de programas eficientes e gratuitos de leitura dos ficheiros. A utilização de programas pagos pode ser um risco, porque pode limitar a acessibilidade aos ficheiros a longo prazo, caso se perca o interesse comercial no formato (OPF).

Mesmo os programas pagos têm problemas na correta implementação do formato. O validador Jpylyser é executado por linhas de comando.

Soluções:

- Apoiar o desenvolvimento do OpenJPEG (OPF)

Consultar o Anexo A deste relatório, que contém listagem dos programas de criação de ficheiros JPEG2000.

3.4 Avaliação do formato jp2

Adaptamos o método de avaliação da KB (Gillesse, 2008), pois permite quantificar os resultados de análise em cada parâmetro, assim, facilita a comparação entre os diferentes formatos.

Paralelamente e em alguns parâmetros, apresentamos avaliação alternativa à da proposta em 2008 pela KB, pela natural evolução do contexto de utilização do formato JPEG2000.

Avaliação - JP2 Lossy			
Parâmetro	Avaliação	Pontuação (1-5)	
		KB	BNP
Espaço de armazenamento	Ganhos de 91 a 98%, face ao TIFF sem compressão	5	5
Qualidade de imagem	Não se atribui a pontuação máxima neste parâmetro, pois a compressão lossy pode degradar a imagem ou introduzir artefactos na mesma. Os artefactos “woolly” podem ser muito visíveis em texto ou documentos com linhas finas. Por outro lado, se os tiles forem muito pequenos e a compressão for lossy, podem surgir artefactos de blocos na imagem (blocking effect) (KB). Com efeito, esses artefactos, ainda que imperceptíveis, influenciam o processamento da imagem, por exemplo para efeitos de OCR (Scape, 2012), tendo de nesses casos de se efetuar o OCR a partir dos TIFF, com prejuízo da performance (maior lentidão) e de posteriores reconhecimentos de texto, uma vez que os TIFF são eliminados (BNP). ⁴	5	3
Sustentabilidade a longo prazo	Norma ISO, standard aberto. Principal problema: baixa taxa de aceitação pelo sector cultural (KB). Atribuímos uma pontuação mais elevada do que a KB, pois nos últimos anos o formato JPEG2000 tem tido uma implementação crescente por parte de bibliotecas e outros organismos do sector cultural (BNP).	2	3

⁴ Ao contrário do KCL, que afirma que os ganhos de performance nas operações de codificação e decodificação (20 a 30% mais rápidos no lossy do que no lossless) compensam os erros que podem surgir depois da descompressão de uma imagem lossy, consideramos que os erros resultantes da compressão têm um impacto no OCR e, eventualmente, noutras operações futuras que não justificam a opção por este método de compressão para ficheiros matrizes. Por estes mesmos motivos, também discordamos do argumento de que não se percebe a diferença da imagem com perdas, nem no papel, nem em ecrã (KCL).

Avaliação - JP2 LossLess			
Parâmetro	Avaliação	Pontuação (1-5)	
		KB	BNP
Espaço de armazenamento	Ganhos de 50%, face ao TIFF sem compressão	3	3
Qualidade de imagem	Lossless, sem perdas, não tem impacto nenhum na qualidade da imagem	5	5
Sustentabilidade a longo prazo	Norma ISO, standard aberto. Principal problema: baixa taxa de aceitação pelo sector cultural (KB). Os riscos inerentes à corrupção de ficheiros e os problemas na especificação de espaços de cor e de implementação do formato pelo software existente impedem a atribuição da pontuação máxima neste parâmetro (BNP).	5	3

4. Conclusões

4.1 Comparação TIFF LZW/JP2

O quadro abaixo resume a avaliação dos formatos JPEG2000 e da compressão LZW do TIFF efetuada ao longo deste relatório.

Avaliação (1 a 5)			
Parâmetro	TIFF LZW	JP2 Lossy	JP2 Lossless
Espaço de armazenamento	1	5	3
Qualidade de imagem	5	3	5
Sustentabilidade a longo prazo	1	3	3

Da análise comparativa destes formatos conclui-se o seguinte:

- ❖ Não adoção do TIFF LZW como alternativa ao atual formato de matrizes na BND, pois apresenta a pior pontuação em termos de ganhos em espaço de armazenamento e de sustentabilidade a longo prazo;
- ❖ Teste de adoção do JP2 Lossless como novo formato para as matrizes da BND, por se conseguir ganhos de 50% em espaço de armazenamento; maior facilidade na manipulação/transmissão de matrizes e maior flexibilidade/rapidez na derivação de cópias de consulta, sem prejuízo da qualidade das imagens e com garantias de sustentabilidade a longo prazo;
- ❖ Teste de implementação de fluxo de produção de matrizes com validação sistemática de ficheiros JP2, por forma a minimizar os riscos de corrupção de ficheiros nos momentos de conversão e ingestão no sistema de armazenamento;
- ❖ Ponderação de adoção futura do formato JP2 Lossy para as cópias de acesso, pois para além dos ganhos em espaço de armazenamento, poder-se-ia melhorar a gestão da coleção (menos instâncias de cópias de acesso a armazenar, manter e sincronizar) e obter funcionalidades mais ricas.

4.2 Aplicação prática BNP

Nesta primeira fase, visa-se testar a criação de matrizes de imagens de jornais no formato JP2 LossLess. Para o efeito, apresenta-se em seguida os requisitos técnicos desses ficheiros, o sw a testar e o fluxo de trabalho a implementar.

4.2.1 Requisitos para matrizes BND

Campo	Valor
Formato do ficheiro	JP2 (.jp2)
Capture resolution	Resolução do TIFF de origem.
Display resolution	Não utilizar
Compressão	Lossless
Níveis de compressão	2:1
Dimensão da imagem	Altura e largura do TIFF
Tiling?	Sim
Tamanho de cada tile	1024x1024
Tile Length Markers?	Sim
Tile header	Não contém coding style default, coding style component, quantization default ou quantization componente marker segments
Transformação de componentes?	Sim
Nº de componentes	3
Filtros wavelet filter	5-3
Tipo transformação	RCT - Reversível
Code block size	6, 64x64
Code block style	Bypass
Nº de níveis de decomposição	6
Nº de layers de qualidade	1
N.º de níveis de resolução	3
Progression order	RPCL (Resolution; Position; Component; Layer)
Code Stream markers	Packet-length markers
Start-Of-Packet marker?	Sim
End-Of-Packet marker?	Sim
Localização dos marcadores	Main header
Segmentation Symbols?	Sim
Precints size	256x256 para as primeiras duas componentes; 128x128 para os componentes de níveis mais baixos
Regiões de interesse?	Não
Espaços de cor	Enumeração de valores: sRGB
Checksum	Sim
Profundidade de bits	24

4.2.2 Software a testar

Ponderados os riscos inerentes à adoção de sw para a criação de ficheiros no formato JPEG2000, serão testados apenas os seguintes programas: Leadtools; Kakadu v.7.1; LuraWave (v. listagem de programas de criação de ficheiros JPEG2000 no Anexo A).

Por outro lado, também terá de se testar a utilização de programas de validação do formato, como o Jpylyser (ferramenta de validação melhor do que o jhove, pois permite validar integridade e cumprimento de requisitos de perfil) e a ferramenta Scape SIFT matching para a deteção de omissão ou duplicação de ficheiros.

4.2.3 Teste conversão, integração com aplicações BND e fluxos trabalho

Os primeiros testes de conversão no formato JP2 terão por objeto as matrizes de imagens de jornais, por corresponderem aos ficheiros que mais espaço de armazenamento exigem.

Visa-se testar os softwares de conversão para o JP2 e de validação do formato, a integração do novo formato com as aplicações atualmente existentes na BND e, por último, o fluxo de trabalho mais adequado para este formato.

Os testes serão realizados na seguinte sequência de tarefas:

- 1º) Criação do Tiff
- 2º) Conversão em Jp2 lossless (testar Leadtools, Kakadu v. 7.1 e LuraWave)
- 3º) Controlo de qualidade: comparação TIFF/JP2 (testar Jpylyser e Scape SIFI)
- 4º) Checksum
- 5º) Comparar performance e resultados do OCR a partir de jp2 lossless e de TIFF (usar LURA sem compressão)
- 6º) Destruição de tiff
- 7º) Derivação de cópias de consulta JPEG (usar Photoshop) e PDF feito diretamente a partir do JPEG ou do JP2 (usar LURA com e sem compressão)
- 8º) Nomeação de matrizes JP2
- 9º) Testar processamento pelo ContentEPeriodics e PeriodicsInfo
- 10º) Testar geração de calendário

5. Trabalho futuro

A adoção do formato JPEG2000 pela BNP exige que sejam ainda desenvolvidos os seguintes trabalhos de investigação, especificação de requisitos e teste:

- ❖ Metadados específicos ou relacionados com o formato JPEG2000
- ❖ Requisitos JP2 Lossy para cópias de consulta;
- ❖ Teste de criação de JP2 Lossy para cópias de consulta, processamento para publicação e derivação de PDF;
- ❖ Análise de sw de visualização/exibição de jpeg2000 em web browsers: Insight (LUNA), ContentDM (DIMEMA) e Digitool (Ex Libris).

6. Bibliografia

BUCKLEY, Robert; TANNER, Simon - **JPEG 2000 as a preservation and access format for the Wellcome Trust Digital Library**. S.l.: King's College London, 2009. Disponível em: <http://library.wellcome.ac.uk/content/documents/22082/JPEG2000-preservation-format>

BUCKLEY, Robert - **JPEG 2000: a practical digital preservation standard?** S.l.: Digital Preservation Coalition, 2008. Disponível em: www.dpconline.org/docs/reports/dpctw08-01.pdf

BUONORA, Paolo; LIBERATI, Franco - A format for digital preservation of images: a study on JPEG 2000 file robustness. "D-Lib Magazine" Vol. 14, nº 7/8 (Jul-Aug 2008) ISSN 1082-9873. Disponível em: <http://www.dlib.org/dlib/july08/buonora/07buonora.html>

CLARK, Richard - JPEG 2000 standards in digital preservation. "Journal of Physics: Conference Series" Nº 231. S.l.: IOP Publishing, 2010. Disponível em: http://iopscience.iop.org/1742-6596/231/1/012015/pdf/1742-6596_231_1_012015.pdf

GILLESSE, Robèrt; ROG, Judith; VERHEUSEN, Astrid - **Alternative file formats for storing master images of digitization projects**. S.l.: National Library of the Netherlands, 2008. Disponível em: http://www.kb.nl/sites/default/files/docs/Alternative_File_Formats_for_Storing_Masters_2_1.pdf

VAN DER KNIJFF, Johan - JPEG 2000 for long-term preservation: JP2 as a preservation format. "D-Lib Magazine" Vol. 17, nº 5/6 (May-Jun 2011) ISSN 1082-9873. Disponível em: <http://www.dlib.org/dlib/may11/vanderknijff/05vanderknijff.html>

WILSON, Carl [et al.] - **LSDR Executable workflows for experimental execution**. S.l.: SCAPE Project, 2012. Disponível em: http://www.scape-project.eu/wp-content/uploads/2012/05/SCAPE_D16.1_BL_V1.0.pdf

ANEXO A - Software disponível para JPEG 2000

1) Programas para a criação de ficheiros jp2

Nome	Preço	Vantagens	Desvantagens	Utilização
Leadtools	995 \$			British Library
Kakadu v. 7.1	500 \$ (trial)		Não usar versões anteriores à 6.3, pois não implementa valores no campo "capture resolution", nem respeita as especificações de perfil ICC para jp2	LOC, BN Checa, BN Noruega
LuraWave	3.300 € + 120 € por core	A linha de comandos da versão 2.1.22.0 é a única que respeita as especificações de perfil ICC para jp2	As versões mais actuais conseguem respeitar as especificações ICC Profile do formato, mas para isso alteram os dados originais do ficheiro (Knijff, 2011)	Wellcome, NA UK, KB
ImageMagik			Não usar pois não implementa valores no campo "capture resolution" (OPF), nem respeita as especificações de perfil ICC para jp2	
Photoshop			Na realidade implementa o formato JPX, denominando-o por JP2 (OPF) Versão CS4 não respeita as especificações de perfil ICC para jp2	
Aware		Implementa o restrict method ICC Profile		
OpenJPEG	GRATUITO	Open source	Não usar pois não implementa valores no campo "capture resolution" (OPF)	
			Lento, instável e não suporta todas as características do formato (OPF)	
JASPER, JJ2000	GRATUITO	Open source	Lentos, instáveis e não suportam todas as características do formato (OPF)	

2) Programas de validação de ficheiros jp2

Jhove; Jpylyser; SCape SIFI

3) Programas para visualização de ficheiros jp2 em web browsers

Programas para exibição jpeg2000 em web browsers: Insight (LUNA), ContentDM (DIMEMA) e Digitool (Ex Libris).