



TIFF to JPEG2000: report on migration and characterisation experiments

Author : Johan van der Knijff
Date : 15-06-2010

Koninklijke Bibliotheek, National Library of the Netherlands
Afdeling Research & Development

TIFF to JPEG2000: report on migration and characterisation experiments

TIFF to JPEG2000: report on migration and characterisation experiments

I. Revision history

Revision	Date	Author	Comments
	15-6-2010	Johan van der Knijff	

II. Related documents

Document name	Date	Author
JPEG2000: adoptiegraad van JP2 en JPX bestandsformaten (in Dutch)		Johan van der Knijff

TIFF to JPEG2000: report on migration and characterisation experiments

III. Contents

I.	Revision history	3
II.	Related documents	3
III.	Contents	5
1	Introduction	7
1.1	Background and context	7
1.2	Outline of this report	7
2	Characterisation and identification tools.....	9
2.1	Exiftool.....	9
2.2	ImageMagick.....	9
2.3	JHOVE	9
2.4	DROID	9
3	Characterisation of a simple test image: source image.....	11
3.1	Introduction	11
3.1.1	<i>General object characteristics</i>	<i>11</i>
3.1.2	<i>Characteristics related to image size and resolution</i>	<i>11</i>
3.1.3	<i>Characteristics related to colour representation</i>	<i>11</i>
3.2	Source image	11
3.2.1	<i>Types of metadata in file</i>	<i>12</i>
3.2.2	<i>Main characteristics of the source image</i>	<i>12</i>
4	Characterisation of derived JPEG2000 images.....	15
4.1	Creating JPEG2000 images from the source image	15
4.2	Characterisation of JPEG2000 files with ExifTool	15
4.2.1	<i>Types of metadata in generated JPEG2000 files</i>	<i>15</i>
4.2.2	<i>Identification of JP2 and JPX file formats</i>	<i>16</i>
4.2.3	<i>Characterisation of resolution information</i>	<i>16</i>
4.2.4	<i>Characterisation of colour information</i>	<i>16</i>
4.2.5	<i>Reference to ICC profile in XMP metadata in KDU output.....</i>	<i>17</i>
4.3	Characterisation of JPEG2000 files with ImageMagick	18
4.3.1	<i>File size: kilobytes versus kibibytes.....</i>	<i>19</i>
4.3.2	<i>Identification of JP2 and JPX file formats</i>	<i>19</i>
4.3.3	<i>Characterisation of resolution information</i>	<i>19</i>
4.3.4	<i>Characterisation of colour information</i>	<i>19</i>
4.4	Characterisation of JPEG2000 files with JHOVE	20
4.4.1	<i>File size in bytes.....</i>	<i>20</i>
4.4.2	<i>Identification of JP2 and JPX file formats</i>	<i>20</i>
4.4.3	<i>Characterisation of resolution information</i>	<i>20</i>
4.4.4	<i>Characterisation of colour information</i>	<i>20</i>
4.4.5	<i>Validation.....</i>	<i>21</i>
4.5	Identification of JPEG2000 files with DROID.....	22
4.6	Observations on conversion tools.....	23
4.7	Implications for ongoing digitisation projects at the KB	24
5	Summary and conclusions.....	27
5.1	Identification of JP2 and JPX file formats.....	27
5.2	Characterisation of resolution information.....	27
5.3	Characterisation of colour information	27
5.4	Validation	27
5.5	General observations on the evaluated characterisation tools	28

TIFF to JPEG2000: report on migration and characterisation experiments

5.6	General observations on the evaluated conversion tools.....	28
5.7	Implications for ongoing digitisation projects at the KB	29
References		31
Annex A: characterisation of source image using ImageMagick and JHOVE...		33

1 Introduction

1.1 Background and context

This report describes a number of experiments on the characterisation (i.e. extraction of the ‘significant characteristics’) of JPEG2000 files. The objective of these experiments is twofold: first, the experiments serve to assess the performance of existing characterisation tools with JPEG2000 files. Second, the characterisation results provide information that can be used to verify the behaviour of tools that are used for the creation and processing of JPEG2000 files.

The used methodology is very simple: starting out with a simple test image (in TIFF format), I converted this image to JPEG2000 using a number of different tools. The resulting images were then analysed using different characterisation tools, and the characteristics of the images were compared to those of the source image.

1.2 Outline of this report

Chapter 2 gives a brief introduction on the characterisation (and identification) tools that were used. In chapter 3 I present a description of the source image that was used as the starting point for all tests. Chapter 4 starts off by describing how the source image was converted to JPEG2000 using different conversion pathways. The remainder of chapter 4 presents and discusses the characterisation of each of the converted images using the different characterisation tools. These results are then used to illustrate the shortcomings of the conversions tools that were used. The chapter ends with an overview that shows how these shortcomings are affecting the materials that have already been delivered to the KB within the various ongoing digitisation projects. Chapter 5 wraps up the main conclusions that can be drawn from the experiments.

TIFF to JPEG2000: report on migration and characterisation experiments

2 Characterisation and identification tools

For the characterisation experiments presented in this report, I have used 4 different tools. In this chapter I provide a brief description of each of them. Three tools (ExifTool, ImageMagick and JHOVE) are capable of both file format identification ('what type of file is this?') and characterisation ('what are the properties of this file?'). DROID is only capable of identification.

2.1 Exiftool

ExifTool is specifically aimed at reading, writing and editing metadata for over 116 different file formats [ExifTool, 2010]. It is also capable of file identification and characterisation. The identification appears to be based on magic numbers, although the documentation does not give any specific details on this. A report by Lindqvist [2008] on characterisation and metadata extraction concluded that the ExifTool compared favourably to 3 other characterisation tools (including JHOVE). A comparison of different characterisation tools by Jensen *et al.* [2009] found ExifTool to be the best choice for TIFF images. ExifTool is free, open source software that is available as a Perl library, a Windows executable and a Mac OS X package.

2.2 ImageMagick

ImageMagick is a collection of software tools that can be used to perform various operations on raster (bitmap) images [ImageMagick, 2010]. It includes an *identify* tool, which identifies files based on signatures ('magic numbers') and characterises them [Identify, 2010]. Since ImageMagick's main concern is bitmap images, identification is limited (a small number of exceptions aside) to bitmap formats. According its documentation, it supports over 100 formats. ImageMagick is open-source software, and all tools are available as command-line programs. In addition, it includes a set of programming interfaces that allow one to use ImageMagick's functionality directly from a number of widely-used programming languages, including C, C++, Java, Python, and many other languages.

2.3 JHOVE

JHOVE is a tool for "format-specific identification, validation, and characterization of digital objects" [JHOVE, 2010]. JHOVE uses file signatures for identification, although the identification may optionally be based on a more sophisticated parsing of the whole file. Moreover, JHOVE can be used to extract and display metadata. JHOVE's characterisation output provides a high level of detail. Both identification and characterisation in JHOVE are limited to those formats for which dedicated modules are available. A JHOVE release comes with 12 standard modules, which includes one for JPEG2000. JHOVE is an open-source, platform-independent Java application.

2.4 DROID

DROID is an acronym for 'Digital Record Object Identification'. It is a file identification tool that is developed and maintained by The National Archives. According to its developers, "DROID is designed to meet the fundamental requirement of any digital repository to be able to identify the precise format of all stored digital objects, and to link that identification to a central registry of technical information about that format and its dependencies" [DROID, 2010]. DROID tries to identify a file using file signatures ('magic numbers'). If this does not yield any results, it makes a tentative attempt at identifying the file based on its extension.

TIFF to JPEG2000: report on migration and characterisation experiments

DROID is an open-source, platform-independent Java application. It can be used directly from the command line, or, alternatively, using a graphical user interface. The version of DROID at the time of writing of this report is 4.0.

3 Characterisation of a simple test image: source image

3.1 Introduction

The following chapters describe a number of characterisation experiments that are all based on one simple test image. First, I used a number of different software applications to convert the source image (which is in TIFF format) to JPEG2000. Next, I analysed all converted images using the four tools that were described in the previous chapter. Some of these tools provide extremely detailed information on the analysed files. It is beyond the scope of this study to describe every single bit of information that these tools may provide. Instead, I have limited the analysis to a basic set of characteristics. These characteristics can be seen as the ‘bare minimum’ amount of information that one would need to interpret a raster image file. They can be roughly subdivided into three groups.

3.1.1 *General object characteristics*

This category comprises the following file characteristics:

1. File size information (number of bytes)
2. File format information (e.g. format name, MIME type)
3. Information on the presence of features in a JPEG2000 file that are not included in Part 1 of the JPEG2000 standard (i.e. features that are only allowed in the JPX format, and not in JP2)¹.

3.1.2 *Characteristics related to image size and resolution*

These properties define the resolution of an image, and its (original) size:

1. Image width and length
2. Image resolution
3. Resolution units

3.1.3 *Characteristics related to colour representation*

These properties all describe how colour is defined:

1. Samples per pixel (i.e. the number of colour components or colour ‘channels’)
2. Bits per sample. This defines the ‘colour depth’, i.e. the number of possible values or levels of each colour component (e.g. 8 bits per sample corresponds to $2^8=256$ possible values for each colour component)
3. Colour space definition. This describes how the colours in a file are ‘mapped’ in an existing colour space
4. Information on ICC profiles. ICC profiles can also be used to link colours in a file to a specific colour space (or to colours on a specific output device, such as a printer or monitor).

The current chapter describes the characteristics of the source (TIFF) image. The next chapter addresses the characterisation of 4 JPEG2000 images that are derived from this source image.

3.2 Source image

¹ Unlike the other characteristics mentioned here, this one is specific to JPEG2000. In theory this should be entirely covered by file format information, but, as we will see later on, things aren’t quite as simple as that.

TIFF to JPEG2000: report on migration and characterisation experiments

Most of the tests that I present in this report are based on the simple test image that is shown in Figure 1. The original file is a small (895x650 pixels) TIFF image that contains an embedded ICC profile which defines the image's colour space (eciRGB v2). The image also contains some additional embedded (metadata) profiles, such as XMP and IPTC metadata.

3.2.1 *Types of metadata in file*

I first tried to extract all metadata from the file by using ExifTool with the following command line:

```
exiftool -a -u -g1 SGScott.tif > SGSCottExif.txt
```

This command "print[s] all meta information in an image, including duplicate and unknown tags, sorted by group"². The ExifTool output shows that the file contains a number of different metadata groups³:

1. IFD0. This metadata group contains the technical Exif metadata of the main image, such as image height, width, bits per sample, compression, samples per pixel and resolution. This is basically the file's header information.
2. ExifIFD. This is another group with Exif metadata with mainly photographic information (exposure time, ISO number, focal length, and so on).
3. ICC-header. This group contains information from the header of the embedded ICC profile.
4. ICC_Profile. This group contains the actual data in the embedded ICC profile, including a profile description (e.g. 'eciRGB v2'), media white point, and red, green and blue matrix columns. Optionally, the red, green and blue reproduction curves can be extracted as well (by default they are not).
5. XMP. This group is subdivided into subgroups that each contain XMP metadata. It is important to note here that most of the properties that are defined in the 'IFD0' group are duplicated in the 'XMP-tiff' group. Also, the 'XMP-photoshop' group contains an item that is called 'ICC Profile Name', which duplicates the 'Profile Description' out of the 'ICC_Profile' group!
6. IPTC 1 / IPTC2. These are two groups with IPTC metadata. For this image both groups are empty.
7. Photoshop. This is a group with metadata that are specific to Photoshop. Again there is some overlap here with the IFD0 metadata (X- and Y- resolution).
8. Composite. This group contains the following entries: Aperture, Image Size, Shutter Speed, Focal Length and Light Value.

The important thing here is that a couple of mechanisms are used to defined metadata within the file, and that there is a partial overlap in the information that is provided by these mechanisms. I will come back to this later.

3.2.2 *Main characteristics of the source image*

The main characteristics of the source image are summarised in Table 1. The table is based on the characteristics as defined by the 'System', 'File', 'IFD0' and 'ICC_Profile' groups in ExifTool's output. I verified these results by doing two additional characterisations of the file using ImageMagick and JHOVE. The results are included in Annex A of this report.

² See: http://www.sno.phy.queensu.ca/~phil/exiftool/exiftool_pod.html

³ In addition to these groups, the ExifTool output also includes a 'System' group, which contains information about the file system (file name, size, modification date/time and file permissions), and a 'File' group with information on file type and MIME type.

TIFF to JPEG2000: report on migration and characterisation experiments

Table 1 Main properties of source image. Extracted using ExifTool 8.12.	
File Name	SGScott.tif
File Size	1736 kB
File Type	TIFF
MIME Type	image/tiff
Image Height	650
Image Width	895
Samples Per Pixel	3
Bits Per Sample	8 8 8
Compression	Uncompressed
X Resolution	300
Y Resolution	300
Resolution Unit	inches
Photometric Interpretation	RGB
Profile Description	eciRGB v2
Media White Point	0.9642 1 0.82491
Red Matrix Column	0.65027 0.32028 0
Green Matrix Column	0.17804 0.60205 0.06783
Blue Matrix Column	0.13588 0.07767 0.75708

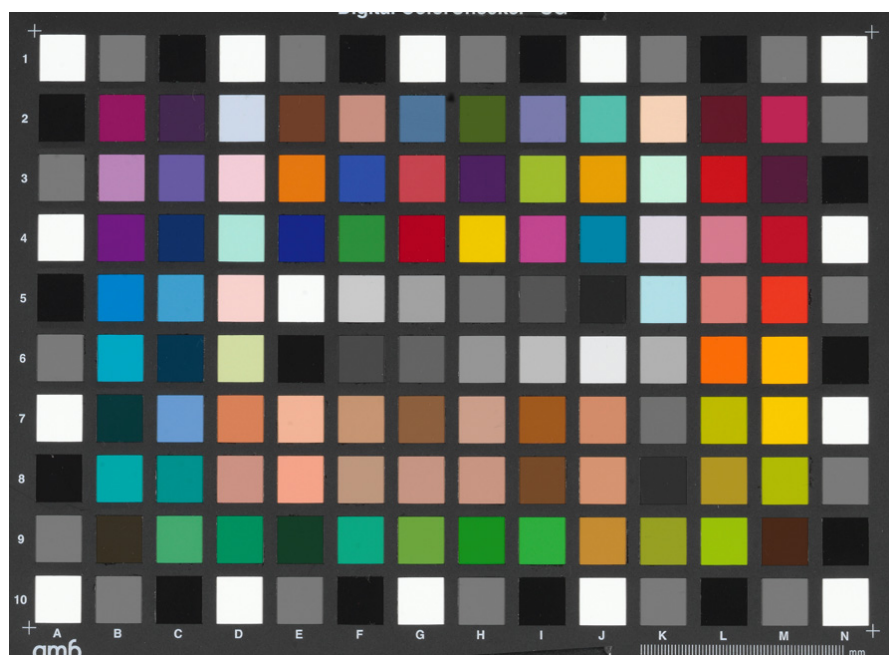


Figure 1 Source image.

TIFF to JPEG2000: report on migration and characterisation experiments

4 Characterisation of derived JPEG2000 images

4.1 Creating JPEG2000 images from the source image

Based on the source image that was described in the previous chapter, I created a number of JPEG2000 image files. The aim was to create images in the JP2 (JPEG2000 Part 1) format. There are different ways to do this conversion. For this study I investigated four conversion pathways, which are listed in Table 2. For the sake of brevity, I will refer to each of these by the identifiers as they are given in the table. Paths APS_1 and APS_2 are based on (two different) plugins in Adobe Photoshop CS4. The J2K plugin of APS_1 is based on the Kakadu library. The alternative plugin of APS_2 always produces files that are JPX (JPEG2000 Part 2) files, rather than JP2 files. However, this plugin has an option that produces file that are JP2 compatible, and this option was used for this test. The KDU path is based on the Kakadu library, and the IM path on the ImageMagick software suite.

ID	Description	Options / command line
APS_1	Adobe Photoshop CS4 using J2K (version: 2.01, Jun 7 2007) plugin	Checked boxes: lossless, JP2
APS_2	Adobe Photoshop CS4 using Adobe JPEG2000 (version: 2.0, 2007) plugin	Checked: lossless, quick encoding, JP2 compatible, include metadata
KDU	Kakadu 6.3 using kdu_compress.exe	kdu_compress -i SGScott.tif -o kdu.jp2 Creversible=yes
IM	ImageMagick 6.6.1-2 using convert.exe	convert SGScott.tif im.jp2

4.2 Characterisation of JPEG2000 files with ExifTool

The images were characterised with ExifTool using the following command line:

```
exiftool -a -u -g1 *.jpg* > jpeg2000Exif.txt
```

4.2.1 Types of metadata in generated JPEG2000 files

Just like the source (TIF) image, the generated JPEG2000 image files contain different types of metadata. Table 3 lists the metadata types that were identified by ExifTool. Since JPEG2000 does not support Exif metadata, none of the files contain an 'IFD0' group. Instead, most technical characteristics are defined in a 'Jpeg2000' group, which represents the technical header information in a JPEG2000 file. What is immediately apparent from Table 3 is that the four conversion pathways from TIFF to JPEG2000 show considerable variation in terms of the extent to which metadata and ICC profiles are preserved. With the exception of KDU (Kakadu), IPTC metadata are always lost. XMP metadata are lost for both APS_1 and IM. ICC profile data are lost for both KDU and IM. Also, note that the output of APS_2 contains *two* ICC profiles. We will have a closer look at this in the next section.

TIFF to JPEG2000: report on migration and characterisation experiments

Table 3 Metadata groups in generated JPEG2000 files. Extracted using ExifTool 8.12.

APS_1	APS_2	KDU	IM
Jpeg2000	Jpeg2000	Jpeg2000	Jpeg2000
ICC-header	ICC-header	XMP	Composite
ICC_Profile	ICC_Profile	IPTC	
Composite	ICC-header2	Composite	
	ICC_Profile2		
	Xmp		
	Composite		

Table 4 summarises the main characteristics of the generated images. The table is based on the characteristics as defined by the ‘System’, ‘File’, ‘Jpeg2000’ and ‘ICC_Profile’ groups in ExifTool’s output. The table shows a number of interesting things

4.2.2 Identification of JP2 and JPX file formats

All images are identified as JP2 (or image/jp2) files, even though the output of APS_2 is really a JPX file. Only the value of the ‘Compatible Brands’ field of this file would suggest that this may not be a ‘normal’ JP2 file. I initially thought that this was a shortcoming of ExifTool; however, further investigations revealed that Adobe’s JPEG2000 plugin actually writes an erroneous value to the file’s ‘brand’ header field. This value should be ‘jpx’ for a JPX file; however, the plugin always writes a value of ‘jp2’ (so the files written by the plugin are not what they purport to be). In practice this means that JPX files that were created by Adobe’s plugin cannot be easily distinguished from JP2 files.

4.2.3 Characterisation of resolution information

All conversion pathways result in some modification of the resolution information (which is expressed in pixels per inch in the original TIFF image). The resolution information for the result of both APS_1 is expressed in pixels per 0.01 mm, whereas it is expressed in pixels per meter for APS_2. The KDU output does not contain any ‘Capture’ X- or Y resolution information. Upon further inspection it turned out that the resolution information is not completely lost after the conversion, but that it is written to the header fields that are reserved for the display resolution (‘Display Y Resolution’, ‘Display X Resolution’ and the corresponding resolution units). However, these fields are only intended to provide image viewer applications a default resolution for displaying an image, whereas ‘Capture Resolution’ specifies the resolution at which the source was digitised [Boliek *et al.*, 2010]. The IM output does not contain any resolution information whatsoever!

4.2.4 Characterisation of colour information

The four images show some major differences in the way colour information is preserved. The embedded ICC profile is completely lost in the results of both KDU and IM. In both cases, the header information suggests that the files use a simple ‘sRGB’ enumerated colour space. The result of APS_2 contains *two* ICC profiles, where the first one is labelled ‘Modified eciRGB v2’ and the second one ‘eciRGB v2’. This is a result of the ‘JP2 compatible’ switch that was used upon the creation of this file in Photoshop. The Photoshop JPEG2000 plugin only *writes* files in JPX format. Within JPX, it is possible to embed any possible ICC profile. However, the JP2 format only supports a restricted set of ICC features, and as a result the ICC profile in a JPX file cannot usually be understood by a JP2 reader. When the ‘JP2 compatible’ switch is used, the Photoshop plugin writes a JPX file that includes both the original (full) ICC profile⁴, as well as a ‘restricted’ version of the profile hat

⁴ The JPX file specification refers to this as the ‘Any ICC’ method, whereas the term ‘Restricted ICC’ method is used for the restricted type.

TIFF to JPEG2000: report on migration and characterisation experiments

can be understood by any JP2 reader. Internally, Photoshop adds a 'Modified' prefix to the name of the latter (e.g.: 'Modified eciRGB v2'). However, the Photoshop documentation is not specific as to the extent to which the 'modified' version actually differs from the full ICC profile. Although ExifTool identifies both ICC profiles, it does not tell one whether a profile is of the 'full' or the 'restricted' type.

4.2.5 Reference to ICC profile in XMP metadata in KDU output

As I explained in section 3.2.1, the XMP metadata in the original source image contain a reference to the ICC profile. Table 3 shows that the ICC profile is lost in the result of KDU. However, since the XMP metadata in that image are preserved, this means that the reference to the ICC profile is preserved as well! In fact, the full ExifTool output contains an 'ICC Profile Name' entry whose value is 'eciRGB v2'. When Exifool is invoked without any special command-line switches, it will simply report these values without referring to the fact that they originate from the XMP metadata. This could easily lead to erroneous conclusions about the presence of ICC profiles (or other features). Because of this, it is important to always use the '-a' ('allow duplicate tag names in output') and '-g' ('organize output by tag group') switches. This way, the origins of each metadata element are always known.

TIFF to JPEG2000: report on migration and characterisation experiments

Table 4 Main characteristics of generated JPEG2000 images. Extracted using ExifTool 8.12.				
Conversion path	APS 1	APS 2	KDU	IM
File Name	aps_1.jp2	aps_2.jp2	kdu.jp2	im.jp2
File Size	504 kB	548 kB	518 kB	499 kB
File Type	JP2	JP2	JP2	JP2
MIME Type	image/jp2	image/jp2	image/jp2	image/jp2
Major Brand	JPEG 2000 Image (.JP2)	JPEG 2000 Image (.JP2)	JPEG 2000 Image (.JP2)	JPEG 2000 Image (.JP2)
Minor Version	0.0.0	0.0.0	0.0.0	0.0.0
Compatible Brands	jp2	jp2 , jpxb, jpx	jp2	jp2
Image Height	650	650	650	650
Image Width	895	895	895	895
Number of Components *	3	3	3	3
Bits per Component **	8 Bits, Unsigned	8 Bits, Unsigned	8 Bits, Unsigned	8 Bits, Unsigned
Compression	JPEG 2000	JPEG 2000	JPEG 2000	JPEG 2000
Capture X Resolution #	0.118103	11811	-	-
Capture Y Resolution #	0.118103	11811	-	-
Capture X-/Y-resolution unit ###	0.01 mm	m	-	-
Colorspace	-	-	sRGB	sRGB
Profile Description	eciRGB v2	Modified eciRGB v2	-	-
Media White Point	0.9642 1 0.82491	0.9642 1 0.82491	-	-
Red Matrix Column	0.65027 0.32028 0	0.65028 0.32028 0	-	-
Green Matrix Column	0.17804 0.60205 0.06783	0.17804 0.60205 0.06783	-	-
Blue Matrix Column	0.13588 0.07767 0.75708	0.13588 0.07767 0.75708	-	-
Profile Description (2)	-	eciRGB v2	-	-
Media White Point (2)	-	0.9642 1 0.82491	-	-
Red Matrix Column (2)	-	0.65027 0.32028 0	-	-
Green Matrix Column (2)	-	0.17804 0.60205 0.06783	-	-
Blue Matrix Column (2)	-	0.13588 0.07767 0.75708	-	-
* : Corresponds to 'Samples per Pixel' in Table 1				
** : Corresponds to 'Bits per Sample' in Table 1				
# : Corresponds to 'X-/Y Resolution' in Table 1				
### : Corresponds to 'Resolution Unit' in Table 1				

4.3 Characterisation of JPEG2000 files with ImageMagick

In addition to the ExifTool characterisation, I also tried to characterise the files with ImageMagick's 'identify' tool using the following command line:

```
identify -verbose *.jp* > jpeg2000IM.txt
```

TIFF to JPEG2000: report on migration and characterisation experiments

The results are summarised in Table 5. Again, some interesting observations can be made.

4.3.1 *File size: kilobytes versus kibibytes*

The file size values that are reported by ImageMagick are slightly different from those given by ExifTool. This is caused by the fact that both tools follow different definitions of what a kilobyte is. ExifTool follows the convention that 1 kilobyte equals 1024 bytes (which is actually a *kibibyte*), whereas ImageMagick uses the formally correct definition of a kilobyte as 1000 bytes⁵.

4.3.2 *Identification of JP2 and JPX file formats*

ImageMagick identifies all images as JP2 files. As the output of APS_2 (which is actually a JPX file) has erroneous header information, the identification of this file as JP2 is not unexpected.

4.3.3 *Characterisation of resolution information*

ImageMagick appears to be unable to derive the correct resolution information from any of the files. In all cases it report an incorrect 'Resolution' value of '72x72'; in addition, it incorrectly reports that the resolution units are undefined.

4.3.4 *Characterisation of colour information*

For all four files, ImageMagicks reports the use of an RGB colour space. It correctly identifies the embedded ICC profile in the output of APS_1. For APS_2, the presence of the 'restricted' ICC profile (i.e. 'Modified eciRGB v2') is reported, but ImageMagick fails to identify the second (full) ICC profile that is also embedded in the file. Also, the profile reported in APS_2 is about half the size of the profile in APS_1, even though both are (or should be) 'restricted' ICC profiles. Interestingly, ImageMagick reports the presence of a generic sRGB ICC profile for both the KDU and the IM output. According to ExifTool, neither of these files contains an embedded ICC profile at all!

⁵ More about the kilobyte-kibibyte ambiguity here: <http://en.wikipedia.org/wiki/Kilobyte> (accessed 4 May 2010)

TIFF to JPEG2000: report on migration and characterisation experiments

Conversion path	APS_1	APS_2	KDU	IM
Image	aps_1.jp2	aps_2.jp2	kdu.jp2	im.jp2
Filesize	516 KB	561 KB	530 KB	511 KB
Format	JP2	JP2	JP2	JP2
Geometry	895x650+0+0	895x650+0+0	895x650+0+0	895x650+0+0
Depth**	8-Bit	8-Bit	8-Bit	8-Bit
Compression	JPEG 2000	JPEG 2000	JPEG 2000	JPEG 2000
Resolution [#]	72x72	72x72	72x72	72x72
Units ^{###}	Undefined	Undefined	Undefined	Undefined
Colourspace	RGB	RGB	RGB	RGB
Profiles	Profile-icc: 1992 bytes eciRGB v2	Profile-icc: 1112 bytes Modified eciRGB v2	Profile-icc: 3024 bytes IEC 61966-2.1 Default RGB colour space - sRGB	Profile-icc: 3024 bytes IEC 61966-2.1 Default RGB colour space - sRGB
* : Corresponds to 'Samples per Pixel' in Table 1 ** : Corresponds to 'Bits per Sample' in Table 1 (also specified separately for each channel by ImageMagick) # : Corresponds to 'X-/Y Resolution' in Table 1 ### : Corresponds to 'Resolution Unit' in Table 1				

4.4 Characterisation of JPEG2000 files with JHOVE

In addition to the ExifTool and ImageMagick characterisation, I also tried to characterise the files with the JHOVE software, using the following command line:

```
jhove -m jpeg2000-hul *.jp* > jpeg2000Jhove.txt
```

The results are summarised in Table 6.

4.4.1 *File size in bytes*

JHOVE reports the file size in bytes.

4.4.2 *Identification of JP2 and JPX file formats*

Again, all images are identified as JP2 files, even though the output of APS_2 is a JPX file. As the brand header field in this file is incorrect, this result is hardly surprising.

4.4.3 *Characterisation of resolution information*

For APS_1 and APS_2, JHOVE correctly reports the image resolution; unlike ExifTool, JHOVE reports the resolution in samples (pixels) per centimetre in both cases. For the output of IM, no resolution information is reported at all. This is consistent with the ExifTool results. Like ExifTool, JHOVE does not report the resolution information for KDU either. However, the JHOVE output for KDU contains a group of properties that are called 'DefaultDisplayResolution', even though it contains no information about the resolution units. These properties appear to correspond to the 'Display Y Resolution' and 'Display X Resolution' values reported by ExifTool.

4.4.4 *Characterisation of colour information*

JHOVE correctly identifies the embedded ICC profiles in the output of APS_1 and APS_2. With respect to ExifTool, JHOVE provides some additional information: it explicitly mentions the method(s) used for defining colours (e.g. whether colours are defined using an enumerated colour space, a restricted ICC profile or a full ICC profile). The presence of multiple colour definitions, such as in APS_2, is reported correctly. However, JHOVE does

TIFF to JPEG2000: report on migration and characterisation experiments

not report the names of the ICC profiles at all (e.g. 'Modified eciRGB v2'). For KDU and IM JHOVE reports that an enumerated (sRGB) colour space is used. This is consistent with the ExifTool results.

4.4.5 *Validation*

A difference between JHOVE and the other tools that are evaluated in this report is that JHOVE is also capable of checking files for validity and well-formedness. JHOVE also provides this functionality for JPEG2000. According to JHOVE, all analysed files are 'well-formed and valid'. For the result of APS_2 this is remarkable: this file contains features from the JPX specification, but at the same time the 'Brand' field in the header defines it as a JP2 file. Despite this, according to JHOVE the file is a well-formed and valid JP2 file. This leads to the conclusion that it would be unwise to rely upon JHOVE for checking the validity and well-formedness of any JP2 files.

TIFF to JPEG2000: report on migration and characterisation experiments

Table 6 Main characteristics of generated JPEG2000 images. Extracted using JHOVE 1.4.				
Conversion path	APS 1	APS 2	KDU	IM
File Name	aps_1.jp2	aps_2.jp2	kdu.jp2	im.jp2
Size (bytes)	516131	560763	530438	511144
Format	JPEG 2000	JPEG 2000	JPEG 2000	JPEG 2000
Status	Well-Formed and valid	Well-Formed and valid	Well-Formed and valid	Well-Formed and valid
MIMEtype	image/jp2	image/jp2	image/jp2	image/jp2
Profile	JP2	-	JP2	JP2
Brand	jp2	jp2	jp2	jp2
MinorVersion	0	0	0	0
Compatibility	jp2	jp2 , jpxb, jpx	jp2	jp2
ImageWidth	895	895	895	895
ImageLength	650	650	650	650
SamplesPerPixel *	3	3	3	3
BitsPerSample **	8, 8, 8	8, 8, 8	8, 8, 8	8, 8, 8
CompressionScheme	JPEG 2000	JPEG 2000	JPEG 2000	JPEG 2000
XSamplingFrequency #	118.103	118.11	-	-
YSamplingFrequency #	118.103	118.11	-	-
SamplingFrequencyUnit ###	centimeter	centimeter	- †	-
ColorspaceUnknown	true	true	true	false
ColorSpec:Method (1)	Restricted ICC Profile	Restricted ICC Profile	Enumerated Colorspace	Enumerated Colorspace
ColorSpec:Precedence (1)	0	1	0	0
ColorSpec:Approx (1)	0	Approximation with reasonable quality	0	0
ColorSpec:RestrictedICCProfile (1)	Three-Component Matrix-Based Input Profile	Three-Component Matrix-Based Input Profile	-	-
ColorSpec:EnumCS (1)	-	-	sRGB	sRGB
ColorSpec:Method (2)	-	Any ICC Method	-	-
ColorSpec:Precedence (2)	-	2	-	-
ColorSpec:Approx (2)	-	Accurate representation	-	-
* : Corresponds to 'Samples per Pixel' in Table 1 ** : Corresponds to 'Bits per Sample' in Table 1 # : Corresponds to 'X-/Y Resolution' in Table 1 ## : Corresponds to 'Resolution Unit' in Table 1 † : Jhove does report a number of properties that define this (and only this) image's 'default display resolution'. This is expressed in terms of a numerator (3870) and a denominator (32768), which, divided, result in a value of 0.118103. However, no units are given.				

4.5 Identification of JPEG2000 files with DROID

In addition to the characterisation experiments, I also tried to identify all JPEG2000 files with DROID 4.0 (using signature file version 29). Table 7 shows the results. Note that also DROID identifies all images as JP2 files.

TIFF to JPEG2000: report on migration and characterisation experiments

Table 7 Identification of generated JPEG2000 images with DROID 4.0 (signature file version 29).

Conversion path	APS 1	APS 2	KDU	IM
File Name	aps_1.jp2	aps_2.jp2	kdu.jp2	im.jp2
Format	JPEG2000	JPEG2000	JPEG2000	JPEG2000
MIME	image/jp2	image/jp2	image/jp2	image/jp2

4.6 Observations on conversion tools

The characterisation experiments revealed a number of limitations and shortcomings of the tools that were used for the TIFF to JPEG2000 conversion. Tables 8 and 9 summarise these findings.

Table 8 Summary of issues with TIFF to JPEG2000 conversion using 4 conversion pathways (see Table 2 for detailed description of conversion pathways).

Conversion path	APS 1	APS 2	KDU	IM
Results in JP2 file?	Yes	No *	Yes	Yes
IPTC metadata preserved?	No	No	Yes	No
XMP metadata preserved?	No	Yes	Yes	No
ICC profile preserved?	Yes	Yes, but 2 versions	No	No
Resolution info preserved?	Yes	Yes	Capture resolution erroneously written to display resolution header fields	No

* File format is JPX, but 'brand' field has erroneous value of 'JP2'; as a result these files will always be (wrongly) identified as 'JP2'.

Table 9 Summary problems with tools that were used for the TIFF to JPEG2000 conversion

Tool	Main problems
Photoshop CS4 (using J2K plugin)	Poor preservation of metadata (both XMP and IPTC)
Photoshop CS4 (using Adobe JPEG2000 plugin)	Plugin only writes JPX files. However, it adds an erroneous 'JP2' entry to the file header's 'brand' field, which has the result that these JPX files will always be identified as JP2 files. IPTC metadata are not preserved.
Kakadu kdu_compress	Capture resolution is erroneously written to display resolution header fields; ICC profiles are not preserved
ImageMagick convert	Metadata are not preserved; ICC profiles are not preserved; resolution information is not preserved.

4.7 Implications for ongoing digitisation projects at the KB

At the time of writing of this report, the KB is involved in a number of digitisation projects that use JPEG2000 (more specifically its JP2 and JPX formats) for storing access and preservation masters. The current digitisation workflows are based on some of the tools that are listed in Table 9. Consequently, some of the problems that were outlined in the previous section also affect these materials. For each digitisation project I analysed a set of sample files using JHOVE and ExifTool. Table 10 lists the general characteristics of these sample batches.

Dataset	Format	Creation tool	ICC profile(s)
Ddd-access	JP2	Accusoft ImageGear (uses Kakadu library)	'Restricted'
Ddd-master	JP2	Accusoft ImageGear (uses Kakadu library)	None (grayscale)
Dpo-access	JPX	Photoshop using Adobe JPEG2000 plugin	'Restricted' + 'Any'
Dpo-master	JPX	Photoshop using Adobe JPEG2000 plugin	'Restricted' + 'Any'
Dts-access	JPX	Photoshop using Adobe JPEG2000 plugin	'Any'
Dts-master	JPX	Photoshop using Adobe JPEG2000 plugin	'Any'

Table 11 below lists the specific problems that I found in each of the sample batches, and for each problem the associated long-term sustainability risks. These results are representative of all files in each specific category. Obviously some of these risks are not really relevant to the access masters, but for the sake of completeness they are included in the table. The most important observation though is that *all* files that have been delivered to the KB so far have some issues. Although all files can be displayed normally, this results in a variety of preservation risks. In their current form, *none* of the currently delivered files (JP2 and JPX) meet the basic requirements for sustainable long-term storage.

TIFF to JPEG2000: report on migration and characterisation experiments

Table 11 Problems found in analysed sample batches and resulting risks			
Dataset	Format	Problems	Risks
Ddd-access	JP2	<ul style="list-style-type: none"> Capture resolution stored in display resolution header fields 	<ul style="list-style-type: none"> Loss of resolution and size information after re-saving or future migration to other format
Ddd-master	JP2	<ul style="list-style-type: none"> Capture resolution stored in display resolution header fields 	<ul style="list-style-type: none"> Loss of resolution and size information after re-saving or future migration to other format
Dpo-access	JPX	<ul style="list-style-type: none"> Format is JPX, but 'brand' header field has value 'jp2' Two versions of ICC profile (using 'Restricted' and 'Any ICC' method, respectively) 	<ul style="list-style-type: none"> Files will be identified as JP2, even though actual format is JPX. This could lead to various (largely unpredictable) problems when these files are subjected to further processing or when migrating these files When re-saving or migrating these files, it may not be completely predictable which version of the ICC profile is used
Dpo-master	JPX	<ul style="list-style-type: none"> Format is JPX, but 'brand' header field has value 'jp2' Two versions of ICC profile (using 'Restricted' and 'Any ICC' method, respectively) 	<ul style="list-style-type: none"> Files will be identified as JP2, even though actual format is JPX. This could lead to various (largely unpredictable) problems when these files are subjected to further processing or when migrating these files When re-saving or migrating these files, it may not be completely predictable which version of the ICC profile is used
Dts-access	JPX	<ul style="list-style-type: none"> Format is JPX, but 'brand' header field has value 'jp2' ICC profile uses 'Any ICC' method, only supported by JPX reader software 	<ul style="list-style-type: none"> Files will be identified as JP2, even though actual format is JPX. This could lead to various (largely unpredictable) problems when these files are subjected to further processing or when migrating these files When re-saving or migrating these files, ICC profiles will be lost if used tool doesn't support JPX
Dts-master	JPX	<ul style="list-style-type: none"> Format is JPX, but 'brand' header field has value 'jp2' ICC profile uses 'Any ICC' method, only supported by JPX reader software 	<ul style="list-style-type: none"> Files will be identified as JP2, even though actual format is JPX. This could lead to various (largely unpredictable) problems when these files are subjected to further processing or when migrating these files When re-saving or migrating these files, ICC profiles will be lost if used tool doesn't support JPX

TIFF to JPEG2000: report on migration and characterisation experiments

5 Summary and conclusions

5.1 Identification of JP2 and JPX file formats

- All files in the experiments were identified as JP2 by all tested characterisation and identification tools, even though 1 file is actually a JPX file. As it turns out, this is not caused by limitations of the characterisation tools, but by the fact that the application that was used to create the JPX file (Adobe's JPEG2000 plugin) erroneously writes a 'JP2' 'Brand' field to the header of the files it creates

5.2 Characterisation of resolution information

- Different tools often report resolution information in different units. This is not necessarily a problem, but it should be taken into account when such tools are used within an automated workflow.
- ImageMagick's 'Identify' tool provides resolution information for JPEG2000 images that is unreliable: resolution is *always* reported as '72x72', irrespective of the actual image resolution (or the presence of any resolution in an image at all); moreover, it does not report the resolution units.
- Both ExifTool and Jhove were unable to extract any resolution information for an image that was created using ImageMagick. However, this appears to be caused by the fact that ImageMagick simply doesn't include this information in any JP2 files it creates.
- Both ExifTool and JHOVE were unable to extract any (capture) resolution information for an image that was created using Kakadu. However, this appears to be caused by the fact that Kakadu writes the 'capture resolution' information to the 'display resolution' fields.
- Overall, ExifTool appears to be the most reliable tool for extracting resolution information from JPEG2000 images.

5.3 Characterisation of colour information

- ExifTool is capable of identifying multiple ICC colour profiles in a single JPEG2000 image. For each ICC profile, the profile description (name), media white point, and red, green and blue matrix column data are reported. However, ExifTool does not report whether an ICC profile is of the 'restricted' or the 'full' ('any ICC') type.
- JHOVE is also capable of identifying multiple ICC colour profiles in a single image. For each ICC profile, it reports which method is used ('restricted' or 'any ICC'). However, the actual profile description (name) is not included in JHOVE's output, and neither are media white point, and red, green and blue matrix column data.
- ImageMagick's 'Identify' tool only detects ICC profiles that are of the 'restricted' type; only the description (name) and the size (in bytes) of the profile are reported. For images that do not contain any ICC profile at all, 'Identify' nevertheless reports the presence of a generic sRGB profile.

5.4 Validation

Only one of the investigated tools (JHOVE) has the capability to check files for validity and well-formedness. The results of the tests suggest that at this stage JHOVE's validation of

TIFF to JPEG2000: report on migration and characterisation experiments

JPEG2000 files is not reliable, and existing workflows should not rely upon JHOVE's validation results.

5.5 General observations on the evaluated characterisation tools

Overall, ExifTool and JHOVE appear to be the best tools for characterising JPEG2000 images. The main advantage of ExifTool over JHOVE is that its output is more concise, and that the output format is highly configurable. However, one should keep in mind that ExifTool is primarily a tool for metadata extraction, and not for characterisation *per se*. A potential pitfall is that the values of some properties (e.g. ICC profiles) may be derived from metadata elements that do not reflect the true characteristics of the image. An example of this situation was shown in section 4.2.5. Here, a JP2 image contained a reference to an ICC profile that was not present in the actual file. There are two ways to deal with this:

1. Use ExifTool with the '-a' ('allow duplicate tag names in output') and '-g' ('organize output by tag group') switches. This way, it is possible to see the origins of all extracted properties. In the case of JPEG2000 images, only the properties that are part of the 'Jpeg2000', 'ICC-header' and 'ICC-Profile' groups should be taken into account.
2. Use ExifTool's -X switch. When this switch is activated, ExifTool reports its output in XML format. Namespaces are used to indicate the groups to which the extracted elements belong⁶. For example, the following command will extract all metadata in file 'aps_1.jp2' to an XML file:

```
exiftool -X aps_1.jp2 > aps_1.xml
```

Within an automated workflow, the second method is preferable.

JHOVE provides some useful additional information on ICC profiles, so, ideally, it should be used in conjunction with ExifTool. Unfortunately, JHOVE does not report the description (name) of any of the ICC profiles, which means that this information cannot be easily linked to the information provided by ExifTool.

Of all tools, ImageMagick's 'Identify' tool shows the poorest performance: the information it provides on resolution is erroneous and incomplete. It only detects ICC profiles that are of the 'restricted' type⁷. Moreover, it reports non-existent ICC profiles when colour is defined using an enumerated (e.g. sRGB) colour space. Because of this, I would advise against the use of ImageMagick for the characterisation of JPEG2000 files.

I should add here that the poor performance of ImageMagick appears to be caused by limitations of the underlying JasPer JPEG2000 library⁸. Therefore, the results are most likely not representative of the way ImageMagick handles other file formats.

5.6 General observations on the evaluated conversion tools

Out of the four tools that were used for the TIFF to JPEG2000 conversion, all resulted in information loss after conversion. In some cases metadata are not kept after conversion (Photoshop using J2K plugin, ImageMagick), in some other cases embedded ICC profiles are

⁶ For more details, see: <http://www.sno.phy.queensu.ca/~phil/exiftool/metfiles.html> ('ExifTool XML Files')

⁷ Note that this is not a shortcoming of the software, since ImageMagick only supports JPEG2000 Part 1 (JP2), and 'restricted' ICC profiles are the only ones that are allowed within a JP2 file.

⁸ See e.g. : <http://studio.imagemagick.org/discourse-server/viewtopic.php?f=1&t=15807>

TIFF to JPEG2000: report on migration and characterisation experiments

lost (Kakadu, ImageMagick). Resolution information is lost with ImageMagick, whereas Kakadu stores this information in the wrong header fields. Using Adobe's JPEG2000 plugin only IPTC metadata are lost. However, the Adobe plugin always writes files in the (rarely used) JPX format. To make matters worse, it plugin incorrectly adds a 'JP2' value to the 'brand' header field, which means that other applications will identify these files as JP2 file (which they are not!). Also, the presence of two different versions of the ICC profile in each file is not really desirable from a long-term preservation point of view, as this may make the outcome of future migration actions somewhat unpredictable.

5.7 Implications for ongoing digitisation projects at the KB

As some of the conversion tools that were evaluated in this report are also being used within ongoing digitisation projects, this implies that the corresponding materials are also affected. A scan of a cross-section of these revealed that *all* files that have been delivered to the KB so far have some issues. Although all files can be displayed normally, this results in a variety of preservation risks (see Table 11 in Chapter 4). In their current form, *none* of the currently delivered files (JP2 and JPX) meet the basic requirements for sustainable long-term storage.

TIFF to JPEG2000: report on migration and characterisation experiments

References

- Boliek, M., Christopoulos, C. & Majani, E. Coding of Still Pictures. JPEG 2000 Part I Final Committee Draft Version 1.0, 2000. Link: <http://www.jpeg.org/public/fcd15444-1.pdf> (accessed 6 April 2010).
- DROID (Digital Record Object Identification). Link: http://sourceforge.net/apps/mediawiki/droid/index.php?title=Main_Page (accessed 3 March 2010).
- European Color Initiative. eciRGB_v2 - the update of eciRGB 1.0 - Background information. Link: <http://www.eci.org/doku.php?id=en:colourstandards:workingcolorspaces> (accessed 7 April 2010).
- ExifTool by Phil Harvey. Link: <http://www.sno.phy.queensu.ca/~phil/exiftool/> (accessed 8 March 2010).
- Identify (ImageMagick command). ImageMagick Studio LLC. Link: <http://www.imagemagick.org/script/identify.php> (accessed 4 March 2010).
- ImageMagick. ImageMagick Studio LLC. Link: <http://www.imagemagick.org/> (accessed 4 March 2010).
- ISO/IEC 15444-2. Information technology — JPEG 2000 image coding system: Extensions, Annex M: JPX extended file format syntax, 2004. Link: <http://www.jpeg.org/public/15444-2annexm.pdf> (accessed 6 April 2010).
- ISO/IEC 15444-5:2003. Information technology -- JPEG 2000 image coding system: Reference software. Link: http://www.iso.org/iso/catalogue_detail.htm?csnumber=33877 (accessed 15 April 2010).
- JasPer Project Home Page. Link: <http://www.ece.uvic.ca/~mdadams/jasper/> (accessed 15 April 2010).
- Jensen, C., Larsen, T., Jurik, B.O., Hansen, T.S., Blekinge, A.A., Frellesen, J.L. & Zierau, E. Evaluation report of additional tools and strategies. Planets subproject PC WP PC/4. Kongelige Bibliotek, Statsbiblioteket, 2009.
- JHOVE. Link: <http://hul.harvard.edu/jhove/> (accessed 3 March 2010).
- JHOVE JPEG2000-hul Module. Link: <http://hul.harvard.edu/jhove/jpeg2000-hul.html> (accessed 6 May 2010).
- Kulovits H., Rauber, A., Kugler, A., Brantl, M., Beinert, T. & Schoger, A. From TIFF to JPEG 2000? Preservation Planning at the Bavarian State Library Using a Collection of Digitized 16th Century Printings. D-Lib Magazine 15 (11/12), 2009. Link: <http://www.dlib.org/dlib/november09/kulovits/11kulovits.html> (accessed 15 April 2010).
- Taubman, D. Overview of Kakadu V6.2.1: A foundation for JPEG2000 applications. Link: <http://www.ee.unsw.edu.au/~taubman/kakadusoftware/Overview.txt> (accessed 27 April 2010).

TIFF to JPEG2000: report on migration and characterisation experiments

Annex A: characterisation of source image using ImageMagick and JHOVE

Image	SGScott.tif
Filesize	1.777MB
Format	TIFF (Tagged Image File Format)
Geometry	895x650+0+0
Depth **	8-Bit
Compression	None
Resolution #	300x300
Units ###	PixelsPerInch
Colorspace	RGB
Profiles	Profile-8bim: 10122 bytes Profile-icc: 1992 bytes eciRGB v2 Profile-iptc: 8 bytes unknown[2,0]: Profile-xmp: 19201 bytes
<p>* : Corresponds to 'Samples per Pixel' in Table 1 ** : Corresponds to 'Bits per Sample' in Table 1 (also specified separately for each channel by ImageMagick) # : Corresponds to 'X-/Y Resolution' in Table 1 ## : Corresponds to 'Resolution Unit' in Table 1</p>	

File Name	SGScott.tif
Size (bytes)	1777268
Format	TIFF
Version	6.0
Status	Well-Formed and valid
MIMEtype	image/tiff
Profile	Baseline RGB (Class R), DLF Benchmark for Faithful Digital Reproductions of Monographs and Serials: color, Exif 2.1 (JEIDA-49-1998)
ImageWidth	895
ImageLength	650
SamplesPerPixel *	3
BitsPerSample **	8, 8, 8
CompressionScheme	uncompressed
XSamplingFrequency #	300
YSamplingFrequency #	300
SamplingFrequencyUnit ###	inch
ColorSpace	RGB
photoshop:ICCProfile	eciRGB v2
<p>* : Corresponds to 'Samples per Pixel' in Table 1 ** : Corresponds to 'Bits per Sample' in Table 1 # : Corresponds to 'X-/Y Resolution' in Table 1 ## : Corresponds to 'Resolution Unit' in Table 1</p>	