

Standards Change Request

JPEG 2000 as a PDS Archive Format

SCR 3-1003

Steve Hughes (submitted)

Aug. 20, 2003

Joel Wilf (modified)

Apr. 19, 2004

Elizabeth D. Rye (modified)

Sept. 27, 2005

Elizabeth D. Rye (modified)

Sept. 28, 2005

Dick Simpson & Elizabeth Rye (editorial improvements + Sec. X.1)

Sept. 30, 2005

Elizabeth D. Rye (modified by request of MC)

Oct. 10, 2005

Problem:

The MRO HiRISE (High Resolution Imaging Science Experiment) team is currently predicting the production of 11,600 GB of RDR products. The maximum sizes of individual RDR products are expected to be about 5 GB. This large volume of data and the large size of the individual products raises several data management issues involving data storage and the distribution of products over the internet.

The MRO HiRISE team is suggesting that the JPEG 2000 compression scheme be used to compress the HiRISE products for use in both online data repositories and for the PDS archive. The JPEG 2000 compression scheme provides a good compression ratio (~3:1 to 4:1 lossless) and uses a wavelet encoding scheme to allow easy viewing of extremely large images. The use of the JPEG 2000 compression scheme will (in the words of the team) result in lower data volume requirements, smaller storage needs, lower media costs, more easily managed data sets, lower per-unit count of hard media, and faster transfer times. It will also preclude the need for separate product tiling.

Note that the MRO HiRISE team is proposing JPEG 2000 for HiRISE RDRs only.

Current Urgency:

The MRO HiRISE team and the PDS are both currently developing plans that include hardware and software procurement, operations center design, product distribution mechanisms, and archive media production. The volume of data and the size of the individual products are the design variables with the most impact at this time. The possible use of JPEG 2000 directly affects these two variables.

Since JPEG 2000 will most likely be used in the online repository, the decision on whether or not the PDS will allow JPEG 2000 in the archive impacts the proposed interface and must be made as soon as possible.

Proposed Solution:

Permit the JPEG 2000 compression algorithm to be used in archiving derived image products when the source data (EDRs) are archived in an uncompressed format. These compressed archives must be consistent with PDS policy on compression.

Permit PDS derived image products to be archived in the JPEG 2000 (or "JP2") format, consistent with PDS policy on the use of external data formats.

The PDS Management Council may wish to review and update policies on "compression" and "external data formats" prior to or in conjunction with consideration of this SCR.

Requested Changes:

1. Amend the Standards Reference to contain a new section describing JPEG 2000 compression. The proposed section is attached.
2. Modify Section 10.2.3 of the Standards Reference, "Reserved Extensions", to include "JP2 – JPEG 2000 (JP2) formatted image". Update "IMQ" to say "Image data that have been compressed. (Not for use with JPEG 2000 compressed data.)"
3. Add the new keyword, `ENCODING_TYPE_VERSION_NAME`, to the Planetary Science Data Dictionary. The element object definition for this keyword is attached.
4. Add a new standard value, "JP2", to the `ENCODING_TYPE` keyword. The updated element object definition for this keyword is attached.

Impact Assessment:

1. The PDS Standards Reference will need to be updated as described above.
2. Software libraries for accessing PDS data objects will need to be modified to include freely available algorithms for the decompression of JPEG 2000 formatted data.
3. NASAView (or other PDS-supported display tools) will need to be able to find the `UNCOMPRESSED_FILE` object in the combined detached label once the file in the `COMPRESSED_FILE` object has been decompressed. (It has been discovered that a similar upgrade is needed to support ZIP compressed files; the change can be implemented for both compressed formats at the same time.)

Additional Information:

The JPEG 2000 format is defined in the specification ISO/IEC 15444. Part one of this specification is the part most relevant to consideration of the format for inclusion in PDS archives. This specification may be ordered directly from the ISO by going to the following URL:

<http://www.iso.org/iso/en/StandardsQueryFormHandler.StandardsQueryFormHandler?scope=CATALOGUE&sortOrder=ISO&committee=ALL&isoDocType=ALL&title=true&keyword=15444>

X. JPEG2000

JPEG 2000 is defined as an “image coding system”. The ISO/IEC specification describing it includes not only the syntax for a compressed image codestream (mime type "J2C"), but also a description of the binary “JP2” file format that may be used to enhance the utility of the codestream.

Unlike many older compression algorithms, JPEG 2000 provides a great deal of flexibility in the way in which data may be stored in the codestream and retrieved from it. This flexibility allows for the progressive decompression of “layers” of the image with increasing resolution or precision. It also permits the extraction and decompression of only a portion or “tile” of the image. Specific portions of the image of particular interest to the intended audience may also be stored at the beginning of the codestream so that they may be accessed and decompressed first. (This would be of potential interest for approach images where the target of the observation fills only a small portion of the field of view.)

All of the information necessary to successfully decompress a JPEG 2000 image is contained in the J2C codestream. However, the information necessary to take advantage of the additional capabilities is only available in the JP2 format.

A JP2 file essentially consists of a set of “boxes” that encapsulate both the J2C codestream and the meta data that describe it (Figure X). The first two of these boxes provide information that identifies the file as a JP2 formatted file. The following “superbox” is the JP2 header box which contains information about the image size, resolution, colorspace, etc. Following this, in no particular order, are contiguous codestream boxes containing the compressed image data and (optionally) intellectual property rights boxes, XML boxes containing vendor-specific meta data, and UUID boxes containing reference URLs. In this document, all of these non-image boxes will be collectively referred to as the "JP2 binary wrapper."

PDS requires the presence of the JP2 binary wrapper so that external software may take full advantage of the JPEG 2000 capabilities. PDS software will have the capability to fully decompress the entire data file, but will not necessarily have the capability to decompress subsets of the codestream such as individual resolution layers or tiles.

The ISO/IEC specification defining JPEG 2000 is entitled “Information technology – JPEG 2000 image coding system” and may be ordered from the ISO by going to their web site (<http://www.iso.ch/>) and searching on “JPEG 2000”.

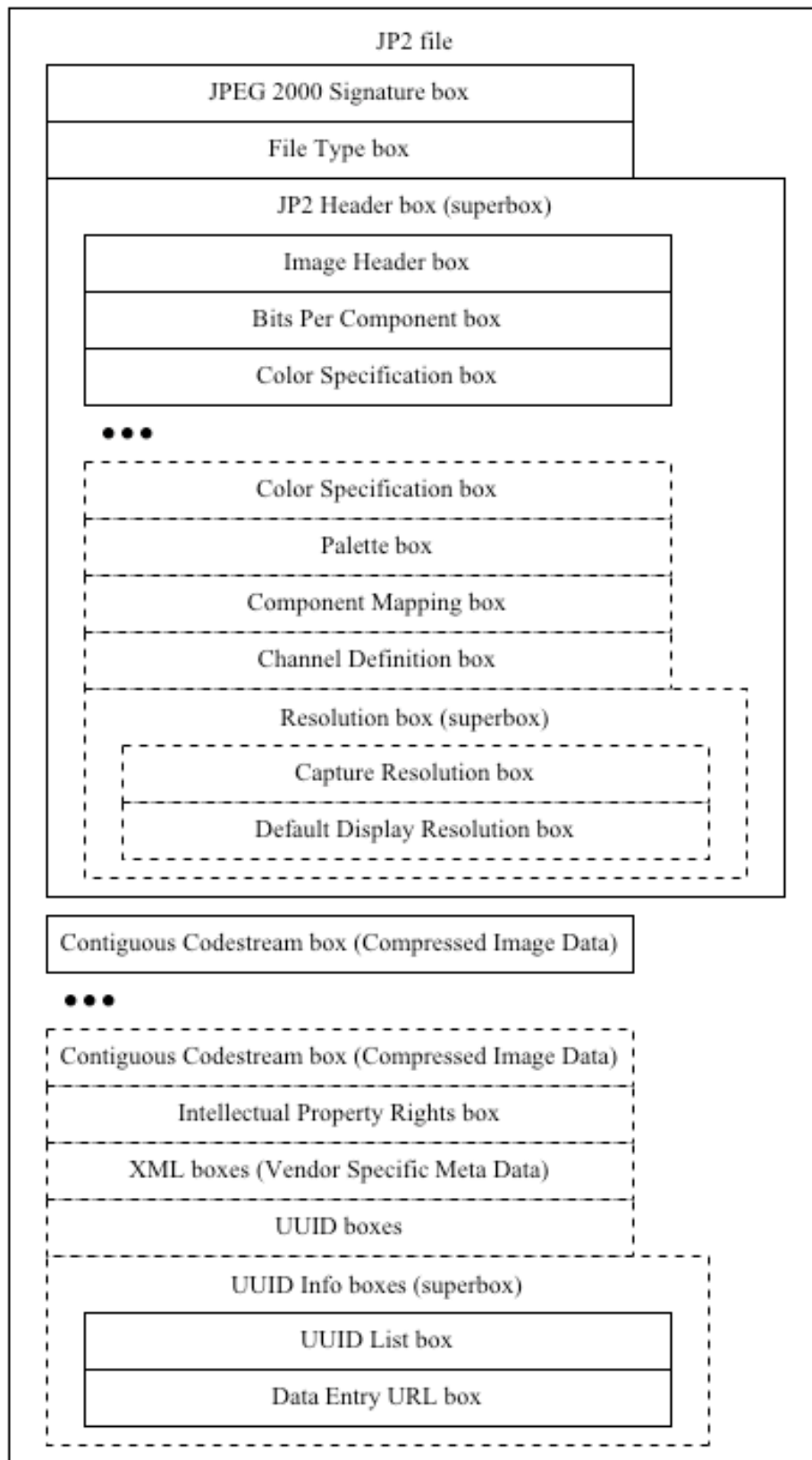


Figure x – Graphical representation of a JP2 file. Dashed lines indicate optional boxes. (Modified from ISO/IEC 15444-1:2004, “Information technology – JPEG 2000 image coding system: Core coding system”, figure T.800_FI.1)

X.1 Table of Compression Ratios

The JPEG 2000 compression algorithm was tested on two Mars Express HRSC images. In both cases, the binary headers and line prefix information on these images were retained in order to provide some additional stress testing of the compression algorithm. With the inclusion of this artificially included binary noise, both of the following images cover the full 16-bit data range.

The first image, “h0068_0000_s22.img”, has 2,618 samples and 119,757 lines, with 16 bit pixels. The image data alone have a dynamic range of 67 to 308 DN, with a standard deviation of 52.5.

The second image, “h0068_0009_s22.img”, has 2,618 samples and 11,013 lines, with 16 bit pixels. The image data alone have a dynamic range of 62 to 188 DN, with a standard deviation of 9.2.

Both images were also converted to 8 bit data for comparison purposes. Selecting for various tile sizes in the JPEG 2000 compression software, the following lossless compression ratios were obtained:

tile size	h0068_0000_s22		h0068_0009_s22	
	16-bit	8-bit	16-bit	8-bit
1024	5.83	2.89	6.00	2.25
512	5.82	2.88	5.99	2.25
256	5.78	2.87	5.94	2.24
128	5.64	2.81	5.80	2.20

These results can be compared with a compression ratio of 3.8 for both of the 16-bit versions of the h0068_0000_s22 and h0068_0009_s22 images, when using Zip compression.

Re-projected images containing large areas of no data and half word image data which do not utilize a full 16-bit data range should result in even higher compression ratios.

X.2 PDS Implementation Rules

The JPEG 2000 compression algorithm must be implemented on PDS archive volumes as a codestream encapsulated by the JP2 binary wrapper (ie., not as a bare codestream).

Only lossless compression may be used.

Furthermore, the syntax and features of the compressed codestream must conform to part 1, the “Core coding system”, of the ISO/IEC specification defining JPEG 2000, namely 15444-1.

Use of the JPEG 2000 compression algorithm and format is restricted to derived image data where the source, or EDR, products have been archived in an uncompressed format.

X.3 Labeling

For each image archived in JPEG 2000 format, two files need to be considered: (1) the compressed file physically included in the archive, and (2) the dynamically generated data file produced by decompressing the JP2 file. These two files have the same name but different extensions: “.JP2” for a JP2 formatted file and “.IMG” for the decompressed file. (The “.JP2” file extension is reserved exclusively, and must be used, for JPEG 2000 compressed files within the PDS).

Like all PDS data files, both the compressed and the decompressed data files require labels. Both files must be described by a single, detached PDS label file using the combined-detached label approach (see Section 5.2.2). Attached labels are not permitted for JPEG 2000 compressed data, because an attached PDS header would violate the JP2 format. In a combined-detached label, each individual file is described using an explicit FILE object. The general framework is:

```
PDS_VERSION_ID          = PDS3
DATA_SET_ID             = ...
PRODUCT_ID              = ...
    (other parameters relevant to both compressed and decompressed
     files)

OBJECT                  = COMPRESSED_FILE
    (parameters describing the compressed file)
END_OBJECT              = COMPRESSED_FILE

OBJECT                  = UNCOMPRESSED_FILE
    (parameters describing the first uncompressed file)
END_OBJECT              = UNCOMPRESSED_FILE
END
```

The compressed file is described by a “minimal label” (see Section 5.2.3), and the following keywords are required:

```
FILE_NAME               = name of the compressed file
RECORD_TYPE             = UNDEFINED
ENCODING_TYPE           = "JP2"
ENCODING_TYPE_VERSION_NAME = version of the JPEG 2000 specification
                        consistent with the data product
INTERCHANGE_FORMAT      = BINARY
UNCOMPRESSED_FILE_NAME  = name of the decompressed file
REQUIRED_STORAGE_BYTES  = approximate total number of bytes in the
                        decompressed data file
DESCRIPTION              = brief description of the JPEG 2000
                        format, including a reference to the
                        full specification
```

Typically, the DESCRIPTION is given as a pointer to a file called “JP2INFO.TXT” found in the

DOCUMENT directory on the same volume.

The subsequent UNCOMPRESSED_FILE object contains a complete description of the data file obtained by decompressing the JPEG 2000 file.

X.4 Label Examples

The following combined detached label describes a hypothetical JP2 formatted image and the decompressed PDS formatted image derived from it:

```
PDS_VERSION_ID          = PDS3

/* IDENTIFICATION DATA ELEMENTS */

MISSION_NAME            = "MARS RECONNAISSANCE ORBITER"
INSTRUMENT_HOST_NAME    = "MARS RECONNAISSANCE ORBITER"
INSTRUMENT_NAME         = "HIGH RESOLUTION IMAGING SCIENCE
                          EXPERIMENT"
TARGET_NAME             = "MOON"
DATA_SET_ID             = "MRO-L-HIRISE-5-DIM-V1.0"
PRODUCT_ID              = "CRU_000004_1200_RED2_2"
START_TIME              = 2005-09-08T23:16:44.863
STOP_TIME               = 2005-09-08T23:16:51.569
SPACECRAFT_CLOCK_START_COUNT = "810688604:56542"
SPACECRAFT_CLOCK_STOP_COUNT  = "810688611:37300"
PRODUCT_CREATION_TIME    = 2005-09-09T15:35:45
(etc.)

/* DESCRIPTIVE DATA ELEMENTS */

(etc.)

OBJECT                  = COMPRESSED_FILE
  FILE_NAME              = "FILENAME.JP2"
  RECORD_TYPE            = UNDEFINED
  ENCODING_TYPE          = "JP2"
  ENCODING_TYPE_VERSION_NAME = "ISO/IEC15444-1:2004"
  INTERCHANGE_FORMAT     = BINARY
  UNCOMPRESSED_FILE_NAME = "FILENAME.IMG"
  REQUIRED_STORAGE_BYTES  = 240000000
  ^DESCRIPTION           = "JP2INFO.TXT"
END_OBJECT              = COMPRESSED_FILE

OBJECT                  = UNCOMPRESSED_FILE
  FILE_NAME              = "FILENAME.IMG"
  RECORD_TYPE            = FIXED_LENGTH
  RECORD_BYTES           = 40000
  FILE_RECORDS           = 60000

/* POINTERS TO DATA OBJECTS */
```



```
^IMAGE = "FILENAME.IMG"

/* DATA OBJECT DEFINITIONS */

OBJECT = IMAGE
  LINES = 60000
  LINE_SAMPLES = 20000
  SAMPLE_TYPE = UNSIGNED_INTEGER
  SAMPLE_BITS = 16
  SAMPLE_BIT_MASK = 2#0011111111111111#
  (etc.)
END_OBJECT = IMAGE
END_OBJECT = UNCOMPRESSED_FILE
END
```

PDS_VERSION_ID = PDS3
LABEL_REVISION_NOTE = "Submitted as part of SCR3-1003
(JPEG 2000 as PDS Archive Format) 2005-09-28 EN:E. Rye"

OBJECT = ELEMENT_DEFINITION
ELEMENT_NAME = ENCODING_TYPE_VERSION_NAME
BL_NAME = encypevernm
DESCRIPTION = "

The ENCODING_TYPE_VERSION_ID element indicates the version of a standard or specification with which a particular ENCODING_TYPE complies."

GENERAL_DATA_TYPE = CHARACTER
MAXIMUM = "N/A"
MINIMUM = "N/A"
MAXIMUM_LENGTH = 60
MINIMUM_LENGTH = 1
STANDARD_VALUE_TYPE = SUGGESTED
STANDARD_VALUE_SET = {"ISO/IEC15444-1:2004"}
STANDARD_VALUE_SET_DESC = "N/A"
KEYWORD_DEFAULT_VALUE = "N/A"
UNIT_ID = "none"
SOURCE_NAME = "PDS EN/E. Rye"
FORMATION_RULE_DESC = "N/A"
SYSTEM_CLASSIFICATION_ID = "COMMON"
GENERAL_CLASSIFICATION_TYPE = "SOFTWARE"
CHANGE_DATE = "2005-09-28"
STATUS_TYPE = "PROPOSED"
STANDARD_VALUE_OUTPUT_FLAG = "Y"
TEXT_FLAG = "N"
TERSE_NAME = "encypevernm"
SQL_FORMAT = "CHAR(60)"
BL_SQL_FORMAT = "char(60)"
DISPLAY_FORMAT = "JUSTLEFT"
AVAILABLE_VALUE_TYPE = "N/A"
END_OBJECT = ELEMENT_DEFINITION
END

PDS_VERSION_ID = PDS3
 LABEL_REVISION_NOTE = "Updated as part of SCR3-1003
 (JPEG 2000 as PDS Archive Format) 2005-09-28 EN:E. Rye"

OBJECT = ELEMENT_DEFINITION
 ELEMENT_NAME = ENCODING_TYPE
 BL_NAME = encodingtype
 DESCRIPTION = "

The ENCODING_TYPE element indicates the type of compression or encryption used for data storage.
 cf. inst_cmprs_name."

GENERAL_DATA_TYPE = CHARACTER
 MAXIMUM = "N/A"
 MINIMUM = "N/A"
 MAXIMUM_LENGTH = 30
 MINIMUM_LENGTH = "N/A"
 STANDARD_VALUE_TYPE = DYNAMIC
 STANDARD_VALUE_SET = {"CLEM-JPEG-0",
 "CLEM-JPEG-0 DECOMPRESSED",
 "CLEM-JPEG-1",
 "CLEM-JPEG-1 DECOMPRESSED",
 "CLEM-JPEG-2",
 "CLEM-JPEG-2 DECOMPRESSED",
 "CLEM-JPEG-3",
 "CLEM-JPEG-3 DECOMPRESSED",
 "DECOMPRESSED",
 "GIF87A", "GIF89A",
 "HUFFMAN FIRST DIFFERENCE",
 "JP2", "N/A",
 "PDF-ADOBE-1.1", "PNG",
 "PREVIOUS PIXEL", "PS-ADOBE-1.0",
 "PS-ADOBE-2.0", "PS-ADOBE-3.0",
 "RUN LENGTH", "ZIP"}
 STANDARD_VALUE_SET_DESC = "The standard values for
 ENCODING_TYPE are described in
 the PDS Standards document."
 KEYWORD_DEFAULT_VALUE = "N/A"
 UNIT_ID = "none"
 SOURCE_NAME = ""
 FORMATION_RULE_DESC = "N/A"
 SYSTEM_CLASSIFICATION_ID = "COMMON"
 GENERAL_CLASSIFICATION_TYPE = {"IMAGING", "STRUCTURE"}
 CHANGE_DATE = "2005-09-28"
 STATUS_TYPE = "APPROVED"
 STANDARD_VALUE_OUTPUT_FLAG = "Y"
 TEXT_FLAG = "N"
 TERSE_NAME = "encodingtype"
 SQL_FORMAT = "CHAR(30)"
 BL_SQL_FORMAT = "char(30)"
 DISPLAY_FORMAT = "JUSTLEFT"
 AVAILABLE_VALUE_TYPE = "N/A"
 END_OBJECT = ELEMENT_DEFINITION
 END