Chapter 9. Documents

Supplementary or ancillary reference materials are usually included with archive products to improve their short- and long-term utility. These documents augment the internal documentation of the product labels and provide further assistance in understanding the data products and accompanying materials. Typical archive documents include:

- Flight project documents
- Instrument papers
- Science articles
- Volume information
- Software Interface Specifications (SISs)
- Software user manuals

The PDS criteria for inclusion of a document in the archive are:

1. Would this information be helpful to a data user?
2. Is the material necessary?
3. Is the documentation complete?

In general, the PDS seeks to err on the side of completeness.

Each document to be archived must be prepared and saved in a PDS-compliant format, including a PDS label. Documents are delivered in the DOCUMENT directory of an archive volume (see the *Volume Organization and Naming* chapter of this document).

A flat, human-readable ASCII text version of each document must be included on the volume, although additional versions may be included in other supported formats at the option of the data producer. “Flat ASCII text” means the file may contain only the standard, 7-bit printable ASCII character set, plus the blank character and the carriage-return and linefeed characters as record delimiters. A file is “human-readable” if it is not encoded and if any special markup tags which may be included do not significantly interfere with an average user’s ability to read the file. So, for example, simple HTML files and TeX/LaTeX files with relatively little markup embedded in the text are generally considered human-readable and may, therefore, be used to satisfy the above ASCII text version requirement.

Note that the PDS takes the requirement for complete documentation very seriously. Documents that are essential to the understanding of an archive are considered as important as the data files themselves. Furthermore, including a document in a PDS archive constitutes publication (or re-publication) of that document. Consequently, documents prepared for inclusion in an archive are expected to meet not only the PDS label and format requirements, but also the structural, grammatical and lexical requirements of a refereed journal submission. Documents submitted for archiving which contain spelling errors, poor grammar or illogical organization will be rejected and may ultimately lead to the rejection of the submitted data for lack of adequate documentation.

9.1 PDS Objects for Documents

PDS labels of documentation files use either the TEXT or DOCUMENT object, as appropriate. The DOCUMENT object is usually used with documentation files found in the DOCUMENT directory of an archive volume. Files described by a DOCUMENT object may be in any of the formats described in Section 9.2.

The TEXT object may only be used with ASCII text files containing no markup. TEXT objects are most often used for small text files occurring anywhere in the archive volume (for example, the AAREADME.TXT file in the root directory or the DOCINFO.TXT file in the DOCUMENT directory).

9.1.1 TEXT Objects

TEXT objects are preferred for stand-alone documents with a narrow focus. For example, the AAREADME.TXT or DOCINFO.TXT files on the archive volume are usually labeled using a TEXT object. Files described by a TEXT object must:

a) Be plain, flat ASCII files without markup tags (i.e., no HTML or TeX files), encoded graphics (as in PostScript files), or programmatic structures (i.e., no source code files or scripting commands); and

b) Have a file extension of ".TXT"

9.1.2 DOCUMENT Objects

DOCUMENT objects are preferred when several versions of the same file are provided or when there are several component files constituting a single version of the document - for example, when graphics are included in separate files from the text. Any file labeled using a DOCUMENT object must:

a) Be in one of the PDS-approved formats listed below; and

b) Use the appropriate object characteristics (listed below) for the DOCUMENT object parameters and the file extension.

DOCUMENT labels are most often combined detached labels, since attaching them to most of the formats listed below would make the combined file unusable in its customary environment (Microsoft *Word*, for example, cannot recognize ".DOC" files with attached PDS labels).

Format	Object	Interchange Format	Document Format	File Extension
Plain ASCII Text	ASCII_DOCUMENT	ASCII	TEXT	.ASC
HTML	HTML_DOCUMENT	ASCII	HTML	.HTM or .HTML*
TeX	TEX_DOCUMENT	ASCII	TEX	.TEX
LaTeX	LATEX_DOCUMENT	ASCII	LATEX	.TEX
Adobe PDF	PDF_DOCUMENT	BINARY	ADOBE PDF	.PDF
MS Word	WORD_DOCUMENT	BINARY	MICROSOFT WORD	.DOC
Rich Text	RTF_DOCUMENT	BINARY	RICH TEXT	.RTF
GIF	GIF_DOCUMENT	BINARY	GIF	.GIF
JPG	JPG_DOCUMENT	BINARY	JPG	.JPG
Encapsulated Postscript	EPS_DOCUMENT	BINARY	ENCAPSULATED POSTSCRIPT	.EPS
PNG	PNG_DOCUMENT	BINARY	PNG	.PNG
Postscript	PS_DOCUMENT	BINARY	POSTSCRIPT	.PS
Tagged Image File Format	TIFF_DOCUMENT	BINARY	TIFF	.TIF or .TIFF*

* See chapter *File Specification and Naming* regarding extensions with more than three characters.

Example: "MYDOC" is a documentation file to be included in the DOCUMENT directory of an archive volume. Two versions will be supplied: a flat ASCII version with the graphics in separate TIFF files; and a Microsoft Word version with in-line graphics in a single file. In the PDS label, "MYDOC" will be described using a DOCUMENT object for each different file format provided. The files included in the directory will be:

1. MYDOC.ASC required ASCII version
2. MYDOC.DOC optional Microsoft Word version to retain all graphics
3. MYDOC001.TIF optional scanned TIFF version of selected pages
4. MYDOC002.TIF optional scanned TIFF version of other selected pages
5. MYDOC.LBL PDS label defining DOCUMENT object(s) for these files

Optional versions of the document should have the same file name as the required ASCII version but with different extensions. Optional versions should be defined as additional DOCUMENT objects in the single PDS label; the name of the required ASCII file should be indicated in the text of the DESCRIPTION keyword.

9.2 Document Format Details

9.2.1 Flat ASCII Text

Line Length and Delimiters - PDS recommends plain text files have line length restricted to 78 characters or fewer, to accommodate printing and display on standard devices. Each line must be terminated by the two-character carriage-return/linefeed sequence (ASCII decimal character codes 13 and 10, respectively).

Page Length and Breaks - Block paragraph style is preferred, with paragraphs being separated by at least one blank line. The form feed character (ASCII decimal code 12) may be used to indicate page breaks, in which case pages should contain no more than 60 lines of text. A formfeed character should be inserted immediately after the END statement line of an attached PDS label in these files.

9.2.2 ASCII Text Containing Markup Language

Line Length and Delimiters - The 78-character line length recommendation is dropped for these files. Notwithstanding, the lines must be delimited by the carriage return/linefeed character combination.

Page Length and Breaks - Page breaks are controlled by the markup in these files. Consequently, there are no specific page length recommendations.

Note: ASCII files containing extensive markup may not pass the "human-readable" test. Also, some automatic converters producing, for example, HTML files that might be expected to be human-readable in fact add so many additional marks and notations that those files also fail the "human-readable" test. Consult a PDS data engineer for help in determining whether a particular file can be considered "human-readable" for archive purposes.

9.2.2.1 Hyper-Text Markup Language (HTML) Files

PDS archive products must adhere to Version 3.2 of the HTML language, a standard generalized markup language (SGML) conforming to the ISO 8879 standard. All files are subject to validation against the HTML 3.2 SGML Declaration and the HTML Document Type Definition.

Note: Constructs not defined in the HTML 3.2 standard (e.g., FRAME, STYLE, SCRIPT, and FONT FACE tags) are not allowed in PDS documentation files.

9.2.2.2 Location of Files

PDS strongly recommends that targets of all HTML links be present on the archive volume. In cases where external links are provided, the link should lead to supplementary information that is not essential to understanding or use of the archival data.

PDS recommends that all files comprising an HTML document or series of documents be located in a single directory. However, locating ancillary files (e.g., images, common files) in subdirectories may be required under certain circumstances (e.g., to avoid conflicts in file names or to minimize replication of common files).

9.2.2.3 Discouraged HTML 3.2 Capabilities

Although the APPLET tag is advertised to be supported by all Java enabled browsers, not all

applets execute on all browsers on all platforms. Further, some browsers require that the user explicitly enable use of Java applets before the applet will execute. Consequently, applets are permitted in PDS document files only when the information they convey is not essential to understanding or use of the archival data.

Use of the TAB character is permitted but strongly discouraged because of variations in implementation among browsers and resulting misalignments within documents.

Use of animated GIF image files is discouraged.

9.2.3 Non-ASCII Formats

Wherever possible the specific encoding and version level information should be included in the label for all non-ASCII documents. The ENCODING_TYPE keyword is used to indicate the base encoding type (e.g., PostScript, GIF, etc.), while the specific version information should be included in the text of the DESCRIPTION keyword. See the PSDD for a list of standard encoding types. Additional types may be added at the discretion of the PDS data engineer.

9.2.4 Validation

Documentation files prepared to accompany a data set or data set collection must be validated. Validation consists of checking to ensure that the files can be copied or transmitted electronically, and can be read or printed by their target text-processing program. Documentation files should be spell-checked prior to being submitted to PDS for validation.

9.3 Examples

9.3.1 Simple Example of Attached label (Plain ASCII Text)

The following label could be attached to a plain ASCII text file describing the content and format of Mars Pathfinder Imager Experiment Data Records.

```
PDS_VERSION_ID    = PDS3
RECORD_TYPE       = STREAM
OBJECT            = TEXT
NOTE              = "Mars Pathfinder Imager Experiment Data Record SIS"
PUBLICATION_DATE  = 1998-06-30
END_OBJECT        = TEXT
END
```

9.3.2 Complex Example of Detached Label (Two Document Versions)

If the data producer chose to provide the same document in both plain ASCII text and as a Microsoft Word document, the detached label would have the name EDRSIS.LBL and would be as follows:

```
PDS_VERSION_ID          = PDS3
RECORD_TYPE             = UNDEFINED
^ASCII_DOCUMENT         = "EDRSIS.ASC"
^WORD_DOCUMENT          = "EDRSIS.DOC"

OBJECT                  = ASCII_DOCUMENT
 DOCUMENT_NAME          = "Mars Pathfinder Imager Experiment Data Record"
 PUBLICATION_DATE       = 1998-06-30
 DOCUMENT_TOPIC_TYPE    = "DATA PRODUCT SIS"
 INTERCHANGE_FORMAT     = ASCII
 DOCUMENT_FORMAT        = TEXT
 DESCRIPTION            = "This document contains a textual description of
                          the VICAR and PDS formatted Mars Pathfinder IMP
                          Experiment Data Records. This is the ASCII text
                          version of the document required by PDS."
END_OBJECT              = ASCII_DOCUMENT

OBJECT                  = WORD_DOCUMENT
  DOCUMENT_NAME         = "Mars Pathfinder Imager Experiment Data Record"
  PUBLICATION_DATE      = 1998-06-30
  DOCUMENT_TOPIC_TYPE   = "DATA PRODUCT SIS"
  INTERCHANGE_FORMAT    = BINARY
  DOCUMENT_FORMAT       = "MICROSOFT WORD"
  DESCRIPTION           = "This document contains a textual description of
                          the VICAR and PDS formatted Mars Pathfinder IMP
                          Experiment Data Records. The document was
                          created using Microsoft Word 6.0.1 for the
                          Macintosh."
END_OBJECT              = WORD_DOCUMENT
END
```

9.3.3 Complex Example of Detached Label (Documents Plus Graphics)

The following label (EDRSIS.LBL) illustrates the use of an HTML document as the required ASCII document. The same document is also included as a PDF file, and four PNG images are included separately.

```
  PDS_VERSION_ID          = PDS3
  RECORD_TYPE             = UNDEFINED
  ^HTML_DOCUMENT          = "EDRSIS.HTM"
  ^PDF_DOCUMENT           = "EDRSIS.PDF"
  ^PNG_DOCUMENT           = ("FIG1.PNG","FIG2.PNG","TAB1.PNG","TAB2.PNG")

  OBJECT                  = HTML_DOCUMENT
    DOCUMENT_NAME         = "Mars Pathfinder Imager Experiment Data
                            Record"
    PUBLICATION_DATE      = 1998-06-30
    DOCUMENT_TOPIC_TYPE   = "DATA PRODUCT SIS"
    INTERCHANGE_FORMAT    = ASCII
    DOCUMENT_FORMAT       = HTML
    DESCRIPTION           = "This document contains a description
                            of the VICAR and PDS formatted Mars
                            Pathfinder IMP Experiment Data Records.  This
                            is an HTML version of the document."
```

```
END_OBJECT                = HTML_DOCUMENT

OBJECT                    = PDF_DOCUMENT
  DOCUMENT_NAME           = "Mars Pathfinder Imager Experiment Data
                            Record"
  PUBLICATION_DATE        = 1998-06-30
  DOCUMENT_TOPIC_TYPE     = "DATA PRODUCT SIS"
  ENCODING_TYPE           = "PDS-ADOBE-1.1"
  INTERCHANGE_FORMAT      = BINARY
  DOCUMENT_FORMAT         = "ADOBE PDF"
  DESCRIPTION             = "This document contains a description
                            of the VICAR and PDS formatted Mars
                            Pathfinder IMP Experiment Data Records. This
                            is a PDF version of the document."
END_OBJECT                = PDF_DOCUMENT

OBJECT                    = PNG_DOCUMENT
  DOCUMENT_NAME           = "Mars Pathfinder Imager Experiment Data
                            Record"
  PUBLICATION_DATE        = 1998-06-30
  DOCUMENT_TOPIC_TYPE     = "DATA PRODUCT SIS"
  FILES                   = 4
  ENCODING_TYPE           = "PNG1.0"
  INTERCHANGE_FORMAT      = BINARY
  DOCUMENT_FORMAT         = PNG
  DESCRIPTION             =  "This document is a PNG representation of two
                            figures and two tables from the Mars
                            Pathfinder IMP Experiment Data Record SIS."
END_OBJECT                = PNG_DOCUMENT
END
```