Chapter 15. Record Formats

The choice of proper record format for a data file is influenced by a number of factors. In general, the PDS strongly recommends a record format of fixed-length or stream be used whenever possible to ensure transportability across operating systems and computer platforms and to avoid potential difficulties with interpretation of the underlying data. Records of type FIXED_LENGTH are required for ASCII files described by TABLE Objects. Records of type VARIABLE_LENGTH may be used in cases where storage efficiency is a major consideration, as, for example, in storing compressed images. Records of type STREAM should be used for text files for ease of transportation to various computer systems. Input/output operations with stream files will generally use string-oriented access, retrieving one delimited record from the file each time.

The RECORD_TYPE element in the PDS label indicates the format of the records in the associated data file (attached or detached).

Table 15.1: Recommended Record Formats

	RECORD_TYPE= FIXED_LENGTH	RECORD_TYPE=STREAM	RECORD_TYPE=VARIABLE
Data format	BINARY, ASCII	ASCII	BINARY
Environment	STRUCTURED	AD HOC	STRUCTURED (VAX/VMS)
Data volume	LARGE	SMALL, MEDIUM	VERY LARGE
Input / Output	READ / WRITE	STRING I/O	CUSTOM, SPICE

15.1 FIXED_LENGTH Records

Records of type FIXED_LENGTH normally use a physical record length (RECORD_BYTES) that corresponds directly to the logical record length of the data objects (that is, one physical record for each image line, or one physical record for each row of a table). In some cases, logical records are blocked into larger physical records to provide more efficient storage and access to the data. This blocking is still an important consideration when storing data on magnetic tape, (which requires a gap on the tape between records), but is not generally a consideration in data sets stored on magnetic or CD-ROM disks. In other cases, the physical record length is determined by compatibility with external systems or standards, as in FITS-formatted files.

The PDS strongly recommends using a physical record length that matches the logical record length of the primary data object in the file for greatest compatibility with application software. In the data label, RECORD_BYTES defines the physical record length.

Figure 15.1 illustrates the physical and logical structure used to build a standard PDS FIXED_LENGTH file.

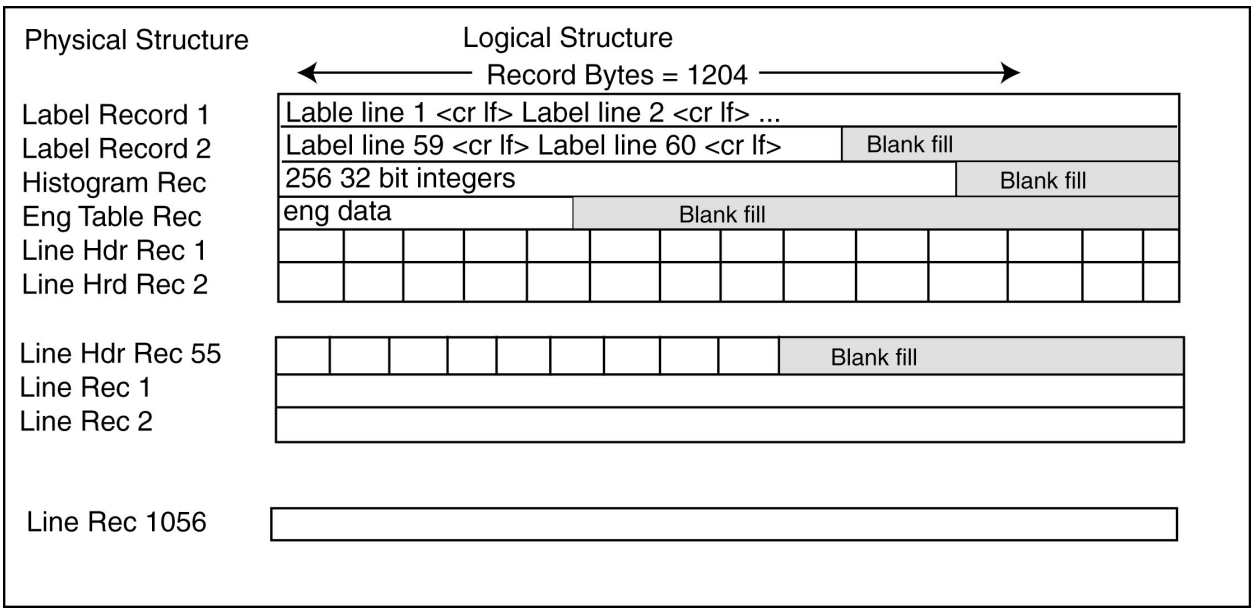

Figure 15.1 Physical and Logical Structure for Fixed Length Files

15.2 STREAM Records

The STREAM record type is reserved for ASCII text files. The records must be delimited by the two-character (carriage return, linefeed) sequence ("<CR><LF>" or "CR/LF"). This is the same record delimiter used for all PDS label and catalog files.

All major operating systems recognize one of either the carriage return, the line feed, or the CR/LF sequence as an ASCII record delimiter; thus, <CR><LF> will work in all cases. There are utilities available for Macintosh (*Apple File Exchange*) and Unix (*tr* translation utility) systems to remove the unneeded extra control character.

Note that the STREAM record type should only be used in those cases where the data contain delimited ASCII records that are *not* of fixed length. The FIXED_LENGTH specification should be used wherever possible.

15.3 VARIABLE_LENGTH Records

PDS data files using the VARIABLE_LENGTH record type must use the VAX/VMS counted byte string format. That is, each record string is preceded by a two-byte LSB integer containing the length of the record. The records may not contain carriage control characters.

The use of the VARIABLE_LENGTH record type is discouraged because of its inherent dependence on *a priori* knowledge of the record structure for proper reading and writing. Notwithstanding, VARIABLE_LENGTH records may be used in the following circumstances:

- When supporting software, which can be executed on a variety of hosts, is provided along with the data. For example, the Voyager CD-ROM disks contain variable-length compressed images along with a decompression program that can be compiled and

executed on VAX, PC, Macintosh and UNIX platforms. The decompression program reformats the data into a variety of forms.

- When the files are intended for use only in a specific environment that supports the selected record structure. For example, the Viking Infrared Thermal Mapper (IRTM) CDROM uses a VAX/VMS variable-length record format for software and command files. Note, however, that such proprietary formats are generally inappropriate for PDS deep archiving purposes and should be vigorously avoided in archive volumes.

15.4 UNDEFINED Records

Records with an undefined record type have no specific record structure. For files with attached labels, the label portion should be written using the STREAM conventions described above. When the record type is designated UNDEFINED, no record terminators are recognized and no record length is implied; the data are taken to be a continuous stream of bytes.

The use of the UNDEFINED record type when referring to a single data file is strongly discouraged. “RECORD_TYPE = UNDEFINED” is properly used in cases where a single label points to two or more *different* data files with different record types (i.e., one file with STREAM records and another with VARIABLE_LENGTH records).

(This page intentionally left blank.)