

Chapter 15

Record Formats

The choice of the proper record format is determined by the applications which the data will support. In general, fixed length records are well-suited to the storage of binary data files, such as images, binary tables or cubes. These files are expected to be transported and used in structured environments. They shall also be used for ASCII tables to promote transportability. Input/output operations with `FIXED_LENGTH` files will use read and write statements which read `RECORD_BYTES` number of bytes with each operation.

Variable length files are less transportable and require special software to read. Their use is discouraged except in instances where they may optimize storage efficiency or access. An example of such an application is the compressed image format being used for CD-ROM storage.

For CD-ROMs that are meant to be VAX/VMS-compatible (ie., for CDs with XARs), it is recommended that all records in fixed length or variable length files contain an *even* number of bytes. Thus records which contain an odd number of bytes would be padded by one byte to give them an even length.

Stream records should be used for text files for ease of transportation to different computer systems. Input/output operations with stream files will generally use string-oriented access, retrieving a record from the file each time.

Table 15.1: Recommended Record Formats

	RECORD_TYPE=FIXED	RECORD_TYPE=STREAM	RECORD_TYPE=VARIABLE
Data format	BINARY, ASCII	ASCII	BINARY
Environment	STRUCTURED	ADHOC	VERY STRUCTURED
Data volume	LARGE	SMALL, MEDIUM	VERY LARGE
Input / Output	READ / WRITE	STRING I/O	CUSTOM, SPICE

15.1 Fixed Length Record Formats

Fixed length record formats normally use a physical record length (`RECORD_BYTES`) which corresponds directly to the logical 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 arbitrary, and only specifies a unit of data for input/output operations, as in FITS format files or USGS PICS images.

The use of a record length which matches the size of the primary data object in a file is recommended, to provide fairly simple file access with a variety of applications. In this approach, objects within a file are all stored in physical records of RECORD_BYTES length. Figure 15.1 illustrates the physical and logical structure used to build a standard PDS FIXED_LENGTH file.

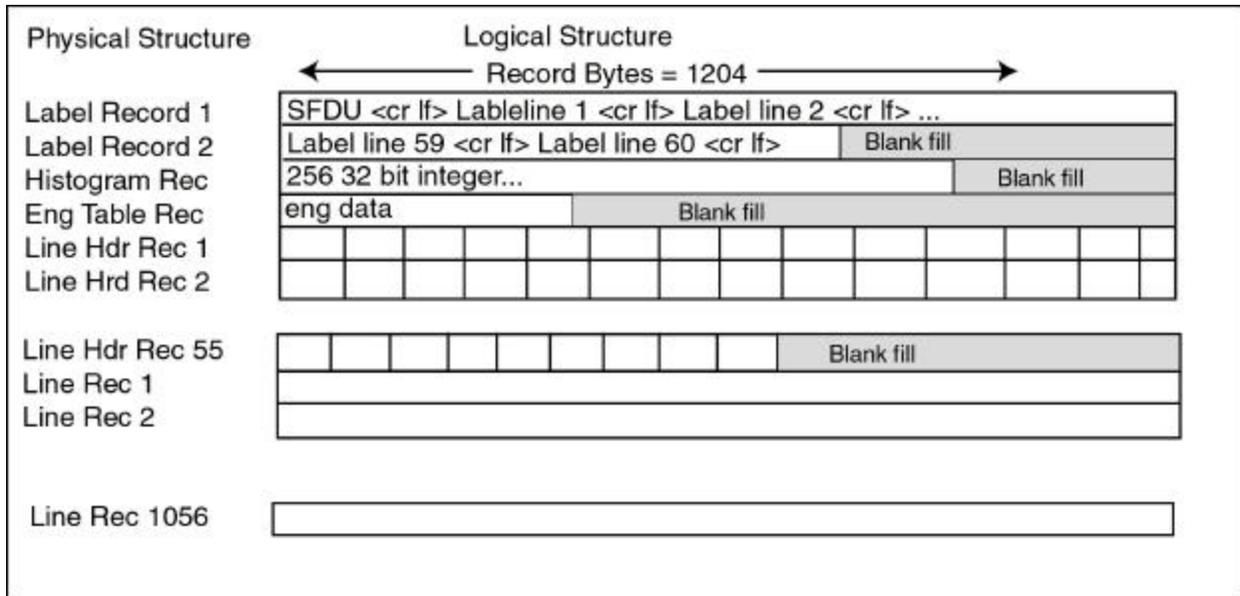


Figure 15.1 Physical and Logical Structure for Fixed Length Files

15.2 Stream Record Formats

Stream records consist of ASCII text delimited with a carriage return (CR) and line feed (LF) sequence. Different computers interpret these codes differently. For example, IBM PC's use the two-byte CR/LF sequence to terminate a line of text. UNIX systems use only a line feed. The Macintosh uses only a carriage return. VAX computers support these various formats as stream files, but prefer to store text files internally as variable length records.

Despite the confusion, stream files can easily be transmitted via text-oriented communications facilities like NASAMAIL, Eudora, or QuickMail. In addition, most file transfer protocols (KERMIT, FTP) will automatically make the needed conversions when stream files are transported between different computers.

PDS has adopted the CR/LF as the standard line delimiter for archival products. Note, in particular, that CR/LF is the required line terminator for all PDS labels and catalog files. This is the only end-of-line sequence that insures that text file will be viewable on all computer systems.

System utilities are available on the various computer types to convert this format to the internal format if necessary.

Macintosh - Apple File Exchange, MS-DOS to Mac option.

Unix - Translate utility (tr-d'\15' <input_file>output_file)

The stream format is recommended for the transfer and archive of text and for files containing detached labels. While stream format can be used for ASCII tables, it is recommended that the FIXED_LENGTH format be used when storing these tables on archival or distributable media (CD-ROM).

15.3 Variable Record Formats

A third category of record type is variable length. The use of variable length records is discouraged, since they are operating-system dependent. They should only be used in the following circumstances:

- Software that can operate 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 for VAX, PC, Macintosh and UNIX systems. These programs will reformat the data to a variety of user-selectable formats.
- The files are only intended for use on one computer system. For example, the Viking IRTM CD-ROM utilizes VAX/VMS variable length formats for software and command files because the software cannot be used unless it is in this format.

PDS data files using variable length records shall follow the VAX/VMS conventions where the records are preceded by a 2-byte (LSB first or swapped) integer which defines the length of the record with no carriage control. The reason for this choice is that VAX/VMS supports variable length records and numerous planetary science data files are stored in this format.

15.4 Undefined Record Formats

Undefined record formats are those which have no implied record structure. For files with attached labels, the label portion should be written using undefined record format and should use record terminators as in the stream case. When data are written using undefined format, no record terminators or specific record length is implied; it is assumed to be a stream of bytes. It is recommended that fixed length records rather than undefined record format be used whenever possible.

15.5 Detached Label Files

Detached label files should be in stream record format. The data elements in a detached label ALWAYS REFER TO THE DATA FILE, not to the detached label file. Thus a RECORD_TYPE = FIXED_LENGTH data element in a label file refers to the record type of the

data file, not the label file itself. Detached label files shall carry the file extension “.LBL” so that they can be easily identified by users.