

Data Patterns: Testing storage to your target environment

There are several levels or stratus of testing that should be considered. The first level is basic functionality i.e., “does the storage device work at a functional level”? Can it write and read without getting any errors?

The next level to consider is performance, i.e. “does the storage device maintain a minimum data transfer rate?

At this level test results can start to become dependent upon the data pattern used for testing. For example, testing a tape device that does data compression using a repeating highly compressible data pattern can lead you to false assumptions regarding transfer rates and amount of tape media used. Or testing using large records when the real-world target application transfers small data sets may not be optimal.

For this reason consideration should be given to the data pattern used during testing. Ideally you would want to test using a data pattern that closely approximates the type of data that will be stored on the data storage system in its “real-world” setting.

Most standard test tools will provide a number of different data patterns, usually either random data or various types of repeating data, such as alternating 1/0, incrementing byte, and so forth. How do these patterns relate to real-world data? The answer is they may not relate well at all, therefore another source of data patterns should be available. Being able to load a user-defined data set for testing is very important, as is having a selection of data files which closely represents real-world types of data.

To provide a number of data sets that closely approximate real-world data we will turn to a sample of data file sets which has been developed specifically for testing new data compression algorithms. The files were selected based on their ability to provide representative performance results.

These files were originally compiled to be used as a standard data set for testing and measuring data compression methods – they also work perfectly in our application as real-world data patterns that closely mimic real-world user data such as spread sheets, database, text, binary executables, and other types of data.

There are 11 files included in this set:

File	Abbrev	Category	Size
alice29.txt	text	English text	152089
asyoulik.txt	play	Shakespeare	125179
cp.html	html	HTML source	24603
fields.c	Csrc	C source	11150
grammar.lsp	list	LISP source	3721
kennedy.xls	Excl	Excel Spreadsheet	1029744
lct10.txt	tech	Technical writing	426754
plrabin12.txt	poem	Poetry	481861
ptt5	fax	CCITT test set	513216
sum	SPRC	SPARC Executable	38240
xargs.1	man	GNU manual page	4227

In addition, to represent random data we supply a text file consisting of the first one-million digits of the number π .

Now it is simply a matter of using these data sets as data patterns during testing. In the case of using either the Disk Manufacturing Module (DMM) or the Tape Manufacturing Module (TMM) go to the Test Setup tab and select “user defined (binary)” as the data type:



And browse to and select the data pattern that best fits your target data:



The test data buffer will be filled with repeating copies of the selected data pattern.

We have made these files available for download in a zip file – simply go to <http://www.scsitoolbox.com/downloads/StandardDataPatterns.zip> , download the file and unzip it into your STB folder.