

About the COMPUSTAT IBM 360/370 General File Format ...

2

About the COMPUSTAT IBM 360/370 General File Format ...	i
File Format Description	1
Arrays	1
Data Codes	1
Data Units	1
Default Data	1
Sorting Order	2
Trailer Records	2

File Format Description

COMPUSTAT files are available in the IBM General file format. Most data is carried in IBM single-precision floating point binary notation. These files are designed to be read by the higher-level programming languages supported by IBM. The files use IBM standard block and record descriptor words.

Arrays

Data arrays are organized with the oldest information in the first part of the array and the most recent information in the last part of the array. The logic for the annual and quarterly data arrays is item by year or quarter with item varying the most rapidly. For example, the 176th element in the annual data array of (175, 20) is Data Item #1 of the second oldest year.

Data Codes

Standard & Poor's provides specific codes for data items that are not available, combined with other data items, considered insignificant by the company, or available only on a semi-annual or annual basis. Programmers must make provisions to recognize these codes so that they are not interpreted as figures.

Not Available Figure	If data is not available, that element of the data array contains a .0001 code.
Combined Figure	If data has been combined into another data item, that element of the array contains a .0004 code. <i>(This code is not used on the Prices, Dividends, and Earnings (PDE) files.)</i>
Insignificant Figure	If data has been reported by the company as insignificant, that element of the data array contains a .0008 code. <i>(This code is not used on the PDE files.)</i>

Not Meaningful

If data is not meaningful, that element of the data array contains a .0007 code. For example, a .0007 appearing in Book Value per Share indicates data has not been calculated since divisor is 0 (zero).

Semi-Annual Figure

If data is available only on a semi-annual basis, the first and third quarters of the year contain .0002 codes. The code indicates that data in the second and fourth quarters represents semi-annual figures rather than quarterly figures. *(This code is used only on the Industrial Quarterly File and not on the PDE files.)*

Annual Figure

If data is available only on an annual basis, the first, second, and third quarters of the year contain .0003 codes. The code indicates that data in the fourth quarter represents an annual figure rather than quarterly figures. *(This code is used only on the Industrial Quarterly File and not on the PDE files.)*

Data Units

Most data items are represented in units of \$ millions. Data is rounded to the nearest thousand. Earnings per share items are presented in actual dollars and cents. See *Chapter 11, Reference*, of this document for more specific information.

Default Data

Header and Trailer Records may contain Default Data. For Floating Point fields, the default is 0 or NA for data slots. For Character fields, the default is Character Blanks or zeros. For Integer fields, the default is zero.

Sorting Order

Companies are sorted numerically in ascending order by Industry Classification Code, CUSIP Issuer Code, and CUSIP Issue Number and Check Digit. Within each Industry Classification Code, companies are arranged by CUSIP and Record Number.

Trailer Records

One or more trailer records may appear at the end of the file. The annual files will have one trailer record which appears at the end of each file (before the end-of-file mark). The Quarterly files will have three (3) trailer records which appear at the end of each file (before the end-of-file mark). These trailer records are identified by zeros in the 13-character key (DNUM + CNUM + CIC). The data array for these trailer records contains default data. In some cases, it may be necessary to check for these records to avoid using them in the program.