About the COMPUSTAT Character ASCII File Formats …

5

File Description

COMPUSTAT files are available in the Character ASCII file format.

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 used as figures.

Not Available Figure — If data is not available, that element of the data array contains a left-adjusted minus sign, a "1" in the third position to the right of the implied decimal point, and zeros in the remainder of the field. For example, items with a field precision of 10.3 contain "-000000Δ001" codes.

Not Meaningful — If data is not meaningful, that element of the data array contains a left-adjusted minus sign, a "7" in the third position to the right of the implied decimal point, and zeros in the remainder of the field. For example, items with a field precision of 10.3 contain "-000000Δ007" codes.

Combined Figure — If data has been combined into another data item, that element of the array contains a left-adjusted minus sign, a "4" in the third position to the right of the implied decimal point, and zeros in the remainder of the field. For example, items with a field precision of 10.3 contain "-000000Δ004" codes. *(This code is not used on the Prices, Dividends, and Earnings (PDE) files.)*

Insignificant Figure — If data has been reported by the company as insignificant, that element of the data array contains a left-adjusted minus sign, an "8" in the third position to the right of the implied decimal point, and zeros in the remainder of the field. For example, items with a field precision of 10.3 contain "-000000Δ008" codes. (*This code is not used on the PDE files.*)

Semi-Annual Figure — If data is available only on a semi-annual basis, the first and third quarters of the year contain a left-adjusted minus sign, a "2" in the third position to the right of the implied decimal point, and zeros in the remainder of the field. For example, items with a field precision of 10.3 contain "-000000Δ002" 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 Business Information or PDE files.)*

Annual Figure If data is available only on an annual basis, the first, second, and third quarters of the year contain a left-adjusted minus sign, a "3" in the third position to the right of the implied decimal point, and zeros in the remainder of the field. For example, items with a field precision of 10.3 contain "-000000Δ003" codes. The code indicates that data in the fourth quarter represents an annual figure rather than a quarterly figure. *(This code is used only on the Industrial Quarterly File, and not on the Business Information or PDE files.)*

Data Fields

All fields designated to contain numeric data are zero-filled and contain no blanks. There are no decimals physically present in the data fields. All decimals are implied by the designated field specification. For example, a 7.3 field containing Earnings per Share data might look like "0001Δ250". In this case, earnings would be interpreted as $1.25.

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.

Negative Numbers

Negative values are preceded by a left-adjusted minus sign in the data field.

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. The sorting order for the Reference File of SIC Codes is ascending by SIC Code.

Trailer Records

One or more trailer records may appear at the end of the file. The annual files will have one trailer record that appears at the end of each file. The Quarterly files will have three (3) trailer records which appear at the end of each file. 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 values. In some cases, it may be necessary to check for these records to avoid using them in the program.

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.