

Wharton Research Data Services

I. About WRDS

Wharton Research Data Services is an Internet-based business data service from The Wharton School, which allows access to most comprehensive sources of financial, accounting, economic, management, marketing, banking, and insurance data.

Restrictions on accessing the data through web-based query format:

1. Individual users are limited to submit only one query at a given time. That is, only after the submitted query is fully processed and the output retrieved, can another query be placed by a user.
2. There is a 2 Gigabyte limit to the web queries. If the limit is exceeded, the output will be automatically deleted. A possible solution is to break your query down into smaller pieces by date range, selected companies or IDs, or the numbers of variables selected, or use one of the compress options in the output step.

II. Extracting Full Financial Statements:

Balance sheets, Income Statements, Cash Flow Statements and Financial Ratios for each Corporation can be obtained by filling the query form on the following link,

<http://wrds.wharton.upenn.edu/aux/finstat/index.shtml> (Click on “Financial Statements” on the home page of WRDS after logging in.)

It would allow the user to extract financial statements in Long as well as Tabular form.

III. COMPUSTAT:

COMPUSTAT is a database of financial, statistical and market information covering publicly traded companies in the U.S. and Canada. A large portion of the COMPUSTAT data in WRDS is contained in the COMPANN and COMPQTR files that hold annual and quarterly data, respectively. To get a brief and separate description of the variables in the above two annual and quarterly data files, go through following paths,

Support Overview of COMPUSTAT on WRDS COMPANN variable table (for annual data variables)

Support Overview of COMPUSTAT on WRDS COMPQTR variable table (for quarterly data variables).

To get detailed information about COMPUSTAT go through following path,

COMPUSTAT COMPUSTAT data manuals (located at the bottom of the page).

Brief Description of different segments in COMPUSTAT:

Wharton Research Data Services

The segments most relevant for obtaining information about the American corporations are described as under,

1. *Bank*: The compustat Annual file contains data of the leading United States banking institutions.
2. *Canadian*: The compustat Canadian Industrial Annual file contains data for major Canadian industrial companies. All data is stated in U.S. dollars.
3. *Industrial*: The compustat Industrial Annual file contains data for major American industrial companies.
4. *Price, Dividends and Earnings*: The Compustat Price, Dividends, and Earnings file is a companion to the industrial Annual files and contains market information on all COMPUSTAT industrial companies and approximately 120 industry indexes and composites.
5. *Segment*: Provides segmental information as under,

Business Segment: Analyze the unique business of a specific company or similar business of a group of companies.

Geographic Segment: Review and analyze a company's exposure in both domestic and foreign markets and evaluate the company's dependence on foreign sales.

Operating Segment: Analyze a company's activities in comparison to other companies in a segment area(s) that is a hybrid of business and geographic types.

State Segment: Research the domestic activity of a North American company particular to the United States and Canada.

Procedure for placing query to retrieve information:

Retrieval of any information would depend upon the type of research intended. Information can be retrieved for individual companies or for group of companies meeting some common criteria or group of companies provided by feed them through a file. Based on the type of company(s) and information sought, select one of the segment of COMPUSTAT as listed on the left-hand side of COMPUSTAT index page (some of these segment has been discussed above.). Filling out the query form in any of these segments is a four-step procedure.

Step-1: To specify the Date range of the starting year and the ending year. Also specify, where necessary, whether the data sought is annual or quarterly.

Wharton Research Data Services

Step-2: Select the variable to be used for making the search, such as SMBL (Stock Ticker Symbol), DNUM (Industry Classification code), NAICS (North American Classification System) etc. To find value of any of these variables for any corporation click on 'code lookup' link provided there. Fill in the name of the corporation for which the code is sought and it would provide the value of each of the above said variables for that corporation. It is important that the value of only that variable, which has been specified for making search in the database, is used for making the search. There are three ways of retrieving information about select group of companies enumerated below,

1. Individual Company Search: One way of retrieving company information is to specify value of the variables, which is selected for making the search, for one or more companies. This can be done by selecting first option given under the 'Step Two: Search' section of the query form and specifying in its text area the value of above mentioned variables.
2. Group Search for companies through file: The second option requires a text file containing the value, for each company, of the variable, which is selected for making the search. The file should be so formatted that each line has only one company code. The path to file in the local computer has to be specified by using browse feature provided in the query form.
3. Searching Entire Database: The third option allows retrieval of information for all the companies in the database. Some segments of the compustat would provide access to active (currently traded and existing companies) companies only, whereas other gives the option to access either active or inactive or both types of companies. For this option of searching entire database, the search can be restricted by specifying one or two conditional statements. The conditions are specified in terms of variables, comparison operators, the desired values, and the logical operators (and/or). A brief description of what each variable in the drop down list given in the section names 'Conditional Statements' can be accessed by clicking the link 'documentation' provided on the left-hand side of the query form.

Step-3: This step requires selecting variables for which the information is sought for each of the companies specified in step-2. The information about the meaning of each variable can be obtained by clicking on the link 'documentation' on left-hand side of the query form.

Step-4: Specify the format for the output file. Preferable formats can be *.txt (for MS Word), *.CSV (for Excel), *.htm, *.sas7bdat (for SAS users), *.dbf (for MS Access). Since, only 2 gigabytes of output is allowed per query, for very large outputs, it is advisable to get the file in zipped format.

Wharton Research Data Services

Data Manuals: Important sections,

Chapter 2:

1. Restated Annual and Quarterly Data (Page 8/9) – Data restated for mergers, acquisitions, accounting changes or discontinued operations.
2. Database Comparability between industrial segment and non-industrial segment (Bank, Utility and Insurance) and among non-industrial companies. (Page – 26 to 74).

Chapter 3: Refer this chapter to know detailed financial formulas used to define certain output variables, such as ratios etc., in compustat.

Chapter 4: This chapter classifies each variable, separately for annual and quarterly, as either a balance sheet or Income statement item or cash flow statement item or restated item etc. and also provides the corresponding data item number used to fill the query form.

Chapter 5: Part 1 through 3 contains detailed definition and description for each data items.

Chapter 6: Briefly describes the footnotes and relates to corresponding data item number(s). It ordered by footnote code.

Chapter 7: Lists the combined data items, that is, data items derived from two or more other data items.

Chapter 8: Lists the data items and corresponding data item numbers under different categories. In each category the data items are listed in ascending order of data item numbers. Quick way of looking what a particular data item number refers to.

Wharton Research Data Services

III. CRSP:

The Center for Research in Security Prices (CRSP) maintains the most comprehensive collection of standard and derived security data available for the NYSE, AMEX and NASDAQ Stock Market.

Out of different segments, only 'Stocks' Monthly Treasuries' and 'CRSP/COMPUSTAT merged database' are only available for UIC users. CRSP/COMPUSTAT merged database is very similar to compustat.

The CRSP NYSE/AMEX/NASDAQ Monthly Stock file provides five kinds of information on individual securities: (1) identity information, which includes complete name histories, all historical CUSIPs, share classes, tickers, SIC codes, and other identifiers; (2) price histories and trading volumes for the period a security is traded on the NYSE, the AMEX, or The NASDAQ markets; (3) information on why a security stopped trading; (4) complete distribution histories including all dividends, stock splits, and special distributions; and (5) shares outstanding values. Filling the query form for the retrieval of the data entails similar 4-step procedure as was enumerated for Compustat above. A brief description of variable can be obtained by clicking on them. Some of the frequently used data items are described as under,

Ask or High Price:

daily: Ask or High Price is the highest trading price during the day, or the closing ask price on days when the closing price is not available. The field is set to zero if no Ask or High Price is available.

monthly: Monthly files contain the highest daily Price or Bid/Ask Average during the month. Closing price values are positive, Bid/Ask Averages negative. The negative sign is a symbol used to differentiate between price and bid/ask average. The bid/ask average does not have a negative value. The field is set to zero when no Ask or bid/ask average was available during the month.

Bid or Low Price:

daily: Bid or Low Price is the lowest trading price during the day or the closing bid price on days when the closing price is not available. The field is set to zero if no Bid or Low Price is available.

monthly: Monthly files contain the lowest daily Price or Bid/Ask Average during the month. Closing price values are positive, Bid/Ask Averages negative. The negative sign is a symbol used to differentiate between price and bid/ask average. The bid/ask average does not have a negative value. The field is set to zero when no Ask or bid/ask average is available.

Wharton Research Data Services

Dividend Cash Amount:

Dividend Cash Amount is the US dollar value per share of distributions resulting from cash dividends, spin-offs, mergers, exchanges, reorganizations, liquidations, and rights issues. Dividend Cash Amount includes the cash value of ordinary and non-ordinary (return of capital) dividends. When the distribution is paid in shares of a trading security, the Dividend Cash Amount is set to the price of the security at the close of the *Ex-Distribution Date*.

Holding Period Total Return:

A return is the change in the total value of an investment in the security over some period of time per dollar of initial investment. Holding Period Total Return is the return for a sale on the given day. It is based on a purchase on the most recent time previous to this day when the security had a valid price. Usually, this time is the previous calendar period. See “Returns” on page 156.

daily: In daily databases, dividends are reinvested on the *Ex-Distribution Date*.

monthly: In monthly databases, returns are holding period returns from month-end to month-end, not compounded from daily returns, and ordinary dividends are reinvested at month-end.

Price or Bid/Ask Average:

Price or Bid/Ask Average is the closing price or the bid/ask average for a trading period. To differentiate them, prices have positive values and bid/ask averages have negative values. If the closing price is not available for any given period, the number in the price field is a bid/ask average, not an actual closing price. Please note that in this field the negative sign is a symbol and that the value of the bid/ask average is not negative. If neither price nor bid/ask average is available on a date, Price or Bid/Ask Average is set to zero.

daily: Price or Bid/Ask Average is the closing price or the bid/ask average (negative) for a trading day. If the closing price is not available on any given trading day, the number in the price field is a bid/ask average, not an actual closing price.

monthly: In a monthly database, Price or Bid/Ask Average is the price on the last trading date of the month. The price series begins the first month-end after the security begins trading and ends the last complete month of trading. If the closing price is not available on any given end of month trading day, the number in the price field is a bid/ask average, not an actual closing price.

Wharton Research Data Services

Return (Excluding Dividends) on Equal-Weighted Index:

Return (Excluding Dividends) on Equal-Weighted Index contains returns, excluding all dividends, on an equally-weighted market portfolio (including ADRs).

Return (Excluding Dividends) on Value-Weighted Index:

Return (Excluding Dividends) on Value-Weighted Index contains, returns, excluding all dividends, on a value-weighted market portfolio (excluding ADRs).

Return (Including all Distributions) on Equal-Weighted Index:

Return (Including all Distributions) on Equal-Weighted Index contains returns, including all distributions, on an equally-weighted market portfolio (including ADRs).

Return (Including all Distributions) on Value-Weighted Index:

Return (Including all Distributions) on Value-Weighted Index contains the returns, including all distributions, on a value-weighted market portfolio (excluding ADRs).

Return without Dividends:

Return without Dividends is the capital appreciation of a security. Ordinary dividends and certain other regularly taxable dividends are excluded from the returns calculation.

Spread between Bid and Ask:

monthly: Spread between Bid and Ask is the difference between the closing bid and ask quotes for a security. It is only available when *Ask or High Price* and *Bid or Low Price* are available and *Closing Price or Bid/Ask Average* is a bid/ask average. It is set to zero if unavailable.

For more information, refer the data manuals. Data manuals can be accessed by clicking on the 'data manual' link on the top of the query form for 'stock' segment of CRSP.