

Top 10 Things to Know about WRDS

1. Do I need special software to use WRDS?

WRDS was built to allow users to use standard and popular software. There is no WRDS-specific software to install.

For example, our web queries are designed to work with any browser. We recommend Internet Explorer, however, which has been found to be most reliable with different java scripts that we use on the WRDS web site.

Almost any telnet program or SSH client will provide access to user accounts on the WRDS Unix system. You may also use an FTP program to download files from your personal directories and use Exceed to obtain an X-windows display of the UNIX system.

For work on a personal computer, PC SAS will provide the greatest compatibility with WRDS data, but other programs such as SPSS and EXCEL can easily import our text and comma delimited output files.

2. How can I change my password and what do I do if I forget my password?

To change your password, use the following web page:

<http://wrds.wharton.upenn.edu/changepassword>

A link to this page as well as a 'lost password' form is found on the WRDS login page.

Passwords also can be changed using the Unix system by following these steps.

1. Telnet to wrds.wharton.upenn.edu and after logging in, enter: **passwd**
2. Follow the prompt and type in your current or old password (then press the Enter key).
3. Type in a new password. Your password should be at least 6 characters long.
4. You will be prompted asked to re-enter the same password (to ensure that you typed your new password correctly). If you type the same password both times, the new password will be in effect for your next login (both UNIX and web) and the system prompt will return.

Note that on our system, passwords are case sensitive: lower-case and upper-case letters are treated as different characters.

3. What databases are accessible through WRDS?

WRDS data includes stock exchange (market-wide and company specific), balance sheet (company specific), income statement (company specific), bank, and macroeconomic information. Data on bonds, bills, inflation, stock and currency options (PHLX) is also included.

The following chart is a quick reference for the most commonly used data. Reference manuals for several distributors may be downloaded from the Online Manuals section under the Support.

DATA	DISTRIBUTOR	DATA SETS
Stock Quotes	CRSP	Monthly (Daily) Stock File
Stock Indices	CRSP	Monthly (Daily) Stock Index
Stock Quotes	TAQ	Intraday Quotes and Trades
Balance Sheet/ Income Statement	COMPUSTAT	Industrial Annual (Quarterly) Full Coverage Annual (Quarterly) Business Industry Segment Price, Dividends, Earnings
Earnings per Share: Estimates and Actual Reported Values	IBES	Daily Detailed History Summary History
Banks	COMPUSTAT	Bank Annual (Quarterly)
Banks	FDIC	all FDIC files
Macro Economic Data	DRI	Yearly (Quarterly, Monthly or Weekly) Data
Bonds	CRSP	Bonds, Bills, and Inflation files all Government Bond files
Options	PHLX	Stock and Currency Option Prices (on the Philadelphia Stock Exchange)

For details, see the Data File Index (under Support) for a list of the databases – and all datasets - that are available in the WRDS system.

4. How do I perform a web query to extract data from WRDS?

The web query has the same general format for all databases in WRDS. In almost all cases, you need to complete four steps before submitting your request.

- Select data range and frequency of the data.
- Choose one of the 3 modes to Search the database:
 - (1) Input identifying codes into a box.
 - (2) Use file with a list of identifiers.
 - (3) Use entire database.
- Select the variables.
- Choose the format for the output file.

Note: In a few cases, there is a preliminary selection step and in other databases, not all steps are needed. For the largest databases, choosing the ‘entire database’ is not an option because web-based downloads are limited to 2 Gigs in size.

You may want to review our “Web Query Demo for Compustat” (under Support) to visualize these steps.

5. Are there any restrictions on web queries?

There are two important restrictions:

No simultaneous requests for individual users.

You must submit data request one at a time. The error message “**ERROR: User Exceed Simultaneous Request**” will be produced when attempting to run a second job while the first is still running. Please note that the error message will remain even if you log off and log back in, unless the output is completed. The query can be ‘killed’ before it finishes using a button on the output page if you decide you do not need the results or if you want to modify the selection criteria.

Output may not exceed a limit of 2 Gigabytes.

If the output from your web query exceeds 2 Gigabytes, you will see the following error message “**ERROR: No Output Available**” and the results will be automatically deleted from the WRDS temporary storage space. In these cases, we suggest that you break your query into smaller pieces by date range, selected companies or IDs, and/or the number of variables selected. In addition, use one of the compress options in the output step. Alternatively, you can use our Unix system and write a SAS program to access the parts of any database that you need.

6. How do I link stock prices in CRSP with company data from Compustat and other databases?

Ticker symbols are available in most of databases on WRDS, but are not reliable because firms can change their symbols and these symbols can be reused by other firms.

The first six digits of a company’s CUSIP codes, which are unique in the sense that they are never used or reused by other firms, are the traditional method for matching companies in different databases. This method is not full proof, however, because companies often change their entire CUSIP identification code when they reorganize or create a replacement issue of their stock. We carry the CRSP ‘events’ datafile and create a ‘names’ dataset that includes a large

array of historical CUSIPs and the associated PERMNO that CRSP created to follow equity issues over time, to assist in this type of matching work.

The most accurately method to link the two most popular database on WRDS is to use the CRSP/COMPUSTAT Merged Database (CCM for short). This database merges CRSP and COMPUSTAT data using a PERMNO-PERMCO-GVKEY link.

Essentially, the CCM database adds the CRSP PERMNO and PERMCO keys as extra variables to various COMPUSTAT databases. Thus an extract of CCM data can be easily matched to CRSP stock prices, via either PERMCO or PERMNO keys. (PERMCOs are company specific, while PERMNOs identify a stock. Some companies have more than one PERMNO.) The WRDS web query page for a CCM extract is <http://wrds.wharton.upenn.edu/ds/crsp/ccm/>

For a list of the complete set to CRSP, COMPUSTAT and CCM databases, see the Data File Index under Support.

7.What are the benefits of using the Unix System?

The Unix system is suitable for customized and refined downloads. In many cases, the web interface may not be flexible enough to provide the specific data in the format that you need.

An example: An extract of data from Compustat based on 3 or more common features (searching the entire database) such as belonging to a specific industry, having a certain level of assets or market value, and belonging to the S&P 500. In this case, you would need to run several web queries, merge and then additionally process the results to create the desired dataset. Alternatively, you can use the WRDS UNIX system to write and run a relatively straightforward SAS or Fortran program to retrieve such data.

SAS programs that are run on the Unix system can also take advantage of the fact that our data libraries are organized to facilitate cross merging. For example, if you would like to obtain and merge data for a certain set companies from Compustat, CRSP and IBES, the most efficient way to proceed would be by using SQL language in a SAS program. In additional, users of the UNIX system can run custom programs to process and analyze a database that they create or to analyze a subset of an existing WRDS database.

For numerous applications, WRDS provides sample programs which are tried and true starting points for basic data exploration and published academic research.

8. How do I download data using the Unix System?

Start by connecting via telnet or another 'terminal' type program to **wrds.wharton.upenn.edu**.

After logging in, your UNIX prompt will be at your home directory. See the document: "How to Telnet to WRDS" and "Useful Unix Commands" to learn how to navigate through the Unix system.

Simple downloads can be accomplished using sample programs that have been created for almost all of the databases on WRDS. These programs are in sub-directories such as:

```
/wrds/crsp/samples  
/wrds/compustat/samples  
/wrds/ibes/samples
```

For example, to extract data from the CRSP Monthly Stock File (MSF), copy `/wrds/crsp/samples/msf.sas` to your home directory. After modifying the selection criteria in `msf.sas` to meet your needs, run the program in background using the command

```
sas msf &
```

The resulting text file in `msf.lst` can be downloaded using an FTP program.

You can write a SAS or Fortran program from scratch using a standard UNIX editor or upload a program you have created on your PC (via FTP, and be sure to use an ASCII transfer method for a text file). See "Using SAS in WRDS" and "Using FORTRAN in WRDS" to get started with SAS or FORTRAN programming in WRDS.

9. Do different download methods and formats affect the precision of the data?

Rounding and issues pertaining to the numerical precision of computer storage affect all numerical representations and calculations, but these problems are typically inconsequential.

For example, a data query request through the WRDS website may produce text output with numbers such as .330 and .220. In contrast, if you directly view the SAS dataset or create output that displays a larger number of decimal places, you may see .3299999991 and .2199999996. The differences in representation are due to the fact that the web output implicitly rounds and the values .33 and .22 can not be exactly represented in binary or computer form. In the case above, the web output format was set at 3 decimal places. The output you will see in many cases will be driven by a computer algorithm that attempts to find the 'best' format for display.

In a SAS program, using the round function may make sense, depending on your application. In the example above, you can round to the hundredth to control the display and numerical precision. Both `x6=round(x,10**(-6))` vs `x2=round(x,10**(-2))` will produce variables that equal .33 and .22 but the latter would handle other cases with more precision, such as 0.325555. Computers cannot store or compute all numbers to their actual value, however, and while more decimals is generally better than less decimals for computation, there always will be numerical imprecision when comparing data stored and used on different systems. Fortunately, the differences are usually in the 8th decimal place.

10. How are missing values reported in WRDS?

In our SAS datasets, we make wide use ‘.’ to represent missing values and translate special numeric value in various databases, such as -999999, to a true missing value code. Thus, there is a need to be careful when comparing the COMPUSTAT and CRSP manuals to the WRDS SAS format databases and when importing the data into another software package.

Examples of the translation of COMPUSTAT data:

FORTTRAN	SAS	Meaning
-0.001	= .	= Not Available Figure - data is not available
-0.002	= .S	= Semi-Annual Figure – appears in 1st and 3rd quarters of the year to indicate that data in the 2nd and 4th quarters represents semi-annual figures rather than quarterly figures.
-0.003	= .A	= Annual Figure - appears in the 1st, 2nd, and 3rd quarters of the year to indicate that data in the 4th quarter represents an annual figure rather than quarterly figures.
-0.004	= .C	= Combined Figure - appears if data has been combined into another data item.
-0.007	= .M	= Not Meaningful
-0.008	= .I	= Insignificant Figure - appears if data has been reported by the company as insignificant

You also can find the letter translations at the bottom of the documentation page for any COMPUSTAT or CRSP query.

Note that these pages may include a reference to a preceding ‘.’ such as .S, .C, or .I but data pulled from the web queries and put into text, tab delimited or other formats will generally only show the letter code. (References to .S or .C are consistent with the SAS documentation for special missing value codes, but default output formats exclude the . part. Also note that this only applies to numerics. Character variables will have blanks when missing.)