

# CRSP/COMPUSTAT MERGED DATABASE RELEASE NOTES

JUNE 2008 ANNUAL UPDATE

## INTRODUCTION

These release notes accompany the June 2008 annual cut of the CRSP/Compustat Merged Database (CCM). The CCM files are updated monthly based on Compustat's final research cuts and shipped monthly, quarterly or annually. Files are delivered on a single compressed CD-ROM and installed using an InstallShield wizard. Please refer to our website for CRSPAccess Release Notes for full product information and for Documents and Guides for all CRSP products at:

<http://www.crsp.chicagobooth.edu/documentation>

**PLEASE NOTE:** Data edits for the 12 month period 200707-200806, sorted by month, can be found on the website at: [http://www.crsp.chicagobooth.edu/documentation/release\\_notes.html](http://www.crsp.chicagobooth.edu/documentation/release_notes.html)

## XPRESSFEED UPDATE

Xpressfeed represents a significant change in the Compustat data organization. Conversion to the new CCM database format that will contain these new items has been an extensive process. The new CCM format of the database will be shipped to subscribers in November 2008, along with new versions of CRSPSift for Windows and CRSPAccess that will have utilities to support both old and new database versions.

Annual subscribers will receive a re-cut of this 200806 annual data cut that is in the new format. Quarterly and

monthly subscribers will receive both old and new formats of the CRSP/Compustat Merged Database for the 200810 cut and through the calendar year-end cuts.

## **CCM DATA CREATED FROM XPRESSFEED INPUTS**

CRSP is now creating the CRSP/Compustat Merged database using new Xpressfeed inputs from Compustat. In this first phase of the transition, the format of the CCM databases is unchanged (CST) and it will continue to work with your existing tools. However, there are data and universe differences due to the new inputs.

## **UPCOMING BANK FILE CHANGE**

Effective August 31, 2008, Compustat will no longer produce the FTP Bank files and will provide bank data only through its Xpressfeed delivery. CRSP will include the FTP Bank files in the current CST format database for as long as they are available. The new Xpressfeed bank data will only be available in the new CCM format database that we expect to ship in September.

## **CONTINUING FTP INPUTS**

FTP formats for the Operating Segment files and the PDE files are still being used in CRSP's update process. CRSP will continue to use these FTP inputs for the CST format of the CRSP Compustat Merged database as long as they are available from Compustat. These FTP inputs are provided for each database update. CRSP uses the last FTP available prior to the time that the Xpressfeed files are released for the same update.

## **CANADIAN COMPANIES**

Xpressfeed provides one record per company, only in reported currencies, with US-to-Canadian exchange rates for each period. In order to match the FTP files, CRSP converts the fundamental Canadian data to US dollars in the CST format databases to create US records. To match the universe and data presented in Canadian Annual and Quarterly files,



CENTER FOR RESEARCH IN SECURITY PRICES

A Research Center at ChicagoGSB

GVKEYs in the >800,000 range are created for companies with fundamental data and a primary Canadian security is identified. These are presented in Canadian dollars. This action adds over a thousand Canadian records for inactive companies.

### UNIVERSE CHANGES

Only companies with fundamental data and a non-blank primary security field (PRIUSA or PRICAN) are included. As a result, a small number of GVKEYs were dropped and new ones added.

The US records that were dropped were either Canadian incorporated with no marked primary US security or with incomplete records from the previously used FTP expanded files. The Canadian records that were dropped included companies with no fundamental data and incomplete records from the previously used FTP expanded files.

The GVKEYs that were added include Canadian incorporated companies marked with primary US securities, new companies with newly available fundamental data, and new indices.

None of these added or deleted GVKEYs are linked with CRSP PERMNOs.

### LINK CHANGES

Using the Xpressfeed inputs, CRSP now generates new links for each security. The link from the primary US security identified by Compustat is appended to the historical CRSP company link. Only the company link is available in CST databases. Because Canadian GVKEYs over 800,000 are not present in Xpressfeed data, there are no links generated for these records. However, since each Canadian GVKEY in the CST file is connected to a US record, the link history is copied, but all link types are set to LX, the secondary link code. Since they represent issues trading on Canadian exchanges they can never be used as a direct link to a US-traded security.

Index records no longer have any link records. Previously there were link records but always a link type NU, indicating no link.

Only PERMNOs in a distributed CRSP database

will appear in the link history. Therefore, by convention, all new LINKFLAGS are coded automatically to BBB, indicating the security resides in both CRSP Monthly and Daily Databases for annual, monthly, and quarterly releases.

### IPERM & ICOMP

Header IPERM and ICOMP link fields are no longer populated and are set to zero. Generally, users should use NPERMNO, the historical CRSP PERMNO link between CRSP data and Compustat records.

Four sample programs currently utilize IPERM and/or ICOMP. Cst\_samp1.c, cst\_samp2.c, and cst\_fsamp1.f90 do not do any processing on the items, but retrieve IPERM values, all zero. Cst\_fsamp5.f90 does a sample read on IPERM and ICOMP and will report "not found" where a record was previously found and will stop the program.

CSTSEARCH and the CRSPSift Find function utilize a header file in the CRSP/Compustat Merged database that contains IPERM. The search will return zeros in the PERMNO column. In CRSPSift, users can double click on the security to return and populate the entity field with the correct PERMNO. Users may retrieve GVKEYs from CSTSEARCH and key on them in cstprint to retrieve data. If a user needs to find PERMNOs, Link History, option /l, will return the full link history.

### INDICES

Deferred Taxes and Investment Tax Credit (Balance Sheet), annual item 35, corresponding to Xpressfeed item TXDITC, is missing for all indices. Common Shares Outstanding and Cumulative Adjustment Factor by Ex-Date, annual items 25 and 27, previously set to one for some indexes, are now missing.

### DATA RANGES

Ranges of items are now set based on the actual available data. In previous versions, ranges were set consistently across annual and quarterly items based on the period descriptor records available. This change shrinks the installed version of the CST

database slightly. Programs should be reviewed if a consistent range of data is expected across all data items. Each data item should be reviewed on its own merits.

## DATA ITEM DIFFERENCES

CRSP found a high degree of agreement between CST data loaded directly from FTP inputs and CST data translated from Xpressfeed inputs. However, there are differences that must be noted. Converting Xpressfeed inputs into FTP items relies on many non one-to-one mappings, including mappings to new calendars, mappings for currency exchange rates, mappings from security data, adjustment factors, and several specialized item-by-item mappings. In some cases, Compustat made corrections to data that are reflected in the Xpressfeed inputs. Following are instances of differences. This list may not be exhaustive, and with subsequent databases, some of these differences may be resolved.

- There are CUSIP differences between Xpressfeed and FTP. Xpressfeed no longer splits the issuer code and issue code of the CUSIP into separate CNUM and CIC fields and does not alter the check digit of the CUSIP. The CST database utilizes the Xpressfeed CUSIP mapping, but splits it into CNUM and CIC fields. In some cases a company is marked inactive by Compustat if filings are not available even though it has an active security, because the company CUSIP is frozen in FTP files when the GVKEY becomes inactive. In Xpressfeed, individual securities may continue to see CUSIP changes. The Xpressfeed security CUSIPs are provided.
- In the FTP data files, data item IA84 (DCLO) had two associated footnotes, #19 and #23. In Xpressfeed there is only one footnote assigned per footnote item (DCLO\_FN). Because there is now nothing to map into annual footnote slot #23, it is no longer available.
- Header fields DUP, and AVAILFLAG are not meaningful in Xpressfeed packages and are set to 0.
- Currency exchange rates are available with more precision in Xpressfeed than was used in presenting values in non-native currencies in the FTP data files. Slightly different values for data items that are converted from reported currencies

result from this increased precision.

- Occasional differences may be found in missing values between Compustat FTP and Xpressfeed data files. This can include the restriction of ranges where periods were previously filled with zeros but are now no longer provided.
- Annual and quarterly items based on security data, such as prices and dividend rates, are now derived directly from the actual security data provided in Xpressfeed. Xpressfeed includes all securities for a company, marks one as the current primary security, and includes a monthly history of primary security markets. Security-based annual and quarterly fundamental data items use the historical monthly primary issue market to identify the data source.

## FOR MORE INFORMATION

The following subscriptions from Compustat are required in order to receive CCM:

- Compustat North America- Active
- Compustat North America-Inactive (Annual Data Only)
- Compustat Inactive Quarterly, Prices, Dividends and Earnings
- Compustat Bank-Annual/Quarterly
- Compustat Historical-Annual/Quarterly
- Compustat Historical-Segment

## FOR MORE INFORMATION

Compustat data definitions and additional information about Xpressfeed may be accessed directly from Compustat's website at [www.compustatresources.com](http://www.compustatresources.com).

## ABOUT THE JUNE 2008 CRSP/COMPUSTAT MERGED DATABASE:

The Compustat cut date for this release is June 14, 2008. The database is shipped in its entirety with historical edits for restatements and corrections for each release.

The CD label for the CCM database is CPZ.

CPZ: CRSP Stock and Indices with Compustat data. PDE market data is for listed and research companies.

The database is delivered and supported on three platforms: Windows, Linux and Solaris. It has been tested on CRSPAccess 2.97 and is supported on CRSPAccess versions 2.95 and higher.

### PARAMETERS:

- The number of CRSP Securities in the CPZ database is 23,856.
- The number of GVKEYs in the CPZ database is 33,341.
- The number of links in the CPZ database is 55,319.
- Database size is approximately 2.36 Gb.

### CALENDARS

The CCM database utilizes the CRSP stock calendars to align properly with CRSP data. Data in the CRSP daily Stock (and Indices) file use daily and annual calendars. Data in the monthly files use monthly, quarterly and annual calendars.

Calendar	Date Range	Date Index Range
Daily trading calendar	19251231-20080613	1-21886
Month-end trading calendar	19251231-20091231	1-1045
Year-end trading calendar	19251231-20091231	1-85
Quarter-end trading calendar	19251231-20091231	1-337
Weekly trading calendar	19260102-20080613	1-4302

### US BANK FILES

File Name	Description	Date Range
bankann	Bank Annual Current – roughly 650 US Banking companies	1988 – 2007
bankqtr	Bank Quarterly Current – roughly 650 US Banking companies	1997 – 2008

### PDE FILES

File Name	Description	Date Range
uspde	Price, Dividends & Earnings Active file – monthly history including Shares Outstanding for roughly 10,000 US companies and 540 US industry indices.	1962 – 2008
uspde.res	Price Dividends & Earnings Research file – Monthly history for inactive companies.	1962 – 2008

### SEGMENT FILES

For use reviewing and analyzing company activities that generate revenue and incur expenses.

File Name	Description	Date Range
seg.cmpy	Segment Company Current	2001 – 2007
seg.cmpy.back	Segment Company Backdata	1986 – 2001
seg.cmpy.r	Segment Company Research Current	2001 – 2007
seg.cmpy.r.back	Segment Company Research Backdata	1986 – 2001
seg.curr	Segment Currency Current	2001 – 2007
seg.cust	Segment Customer Current	2001 – 2007
seg.cust.back	Segment Customer Backdata	1986 – 2007
seg.cust.r	Segment Customer Research Current	2001 – 2007
seg.cust.r.back	Segment Customer Research Backdata	1986 – 2001
seg.dtail	Segment Detail Current	2001 – 2007
seg.dtail.back	Segment Detail Backdata	1986 – 2001
seg.dtail.r	Segment Detail Research Current	2001 – 2007
seg.dtail.r.back	Segment Detail Research Backdata	1986 – 2001
seg.geocd	Segment Geographical Area Codes Current	2001 – 2007
seg.geocd.back	Segment Geographical Area Codes Backdata	1986 – 2001
seg.geocd.r	Segment Geographical Area Codes Research Current	2001 – 2007
seg.geocd.r.back	Segment Geographical Area Codes Research Backdata	1986 – 2001
seg.item	Segment Item Value Current	2001 – 2007
seg.item.back	Segment Item Value Backdata	1986 – 2001
seg.item.r	Segment Item Value Research Current	2001 – 2007
seg.item.r.back	Segment Item Value Research Backdata	1986 – 2001
seg.naics	Segment NAICS Current	2001 – 2007
seg.naics.back	Segment NAICS Backdata	1986 – 2000
seg.naics.r	Segment NAICS Research Current	2001 – 2007
seg.naics.r.back	Segment NAICS Research Backdata	1986 – 2001
seg.prod	Segment Products current	2001 – 2007
seg.srce	Segment Source Current	2001 – 2007
seg.srce.back	Segment Source Backdata	1986 – 2001
seg.srce.r	Segment Source Research Current	2001 – 2007
seg.srce.r.back	Segment Source Research Backdata	1986 – 2001

## DATA INSTALLATION

Instructions and screen shots for installation are written for Windows. Differences related to Linux and Sun are noted.

### WINDOWS:

To install data, insert your CD into the drive and from the Start Menu, select Run `d:\setupwin32.exe`. Click OK.

### SUN SOLARIS:

After inserting the CD, a terminal will open with the file, `setupsolaris.bin`. Double click on this setup file to begin the installation process.

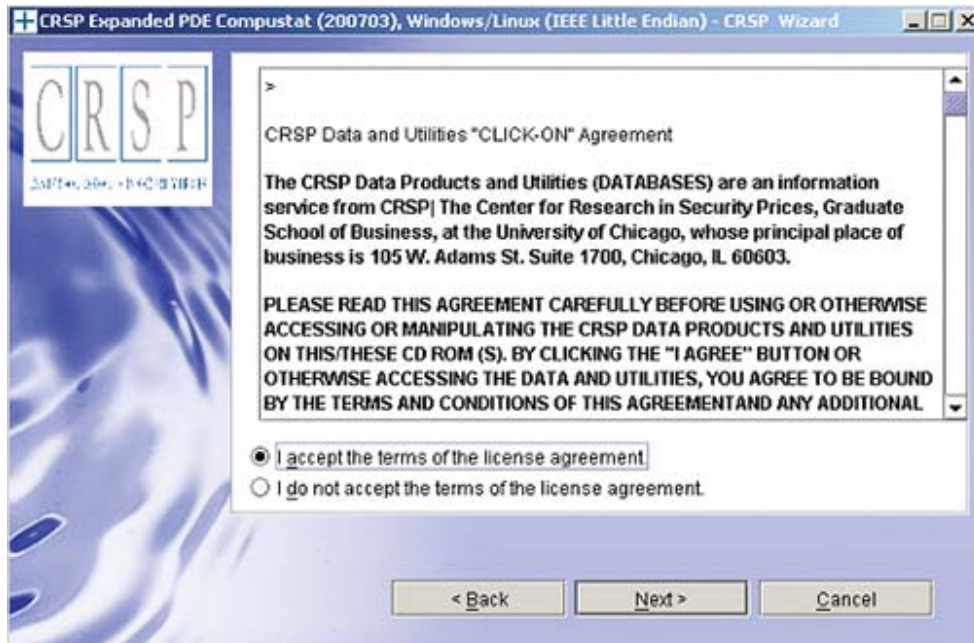
### LINUX:

Upon insertion of the CD, a terminal will open with the file, `setuplinux.bin`. Double click on this setup file and open a terminal to begin the installation process.





After you have clicked **Next** on the Welcome screen, scroll through and read the CRSP Data and Utilities Agreement. Click to accept the terms of the license agreement, and assuming you do, click **Next**.

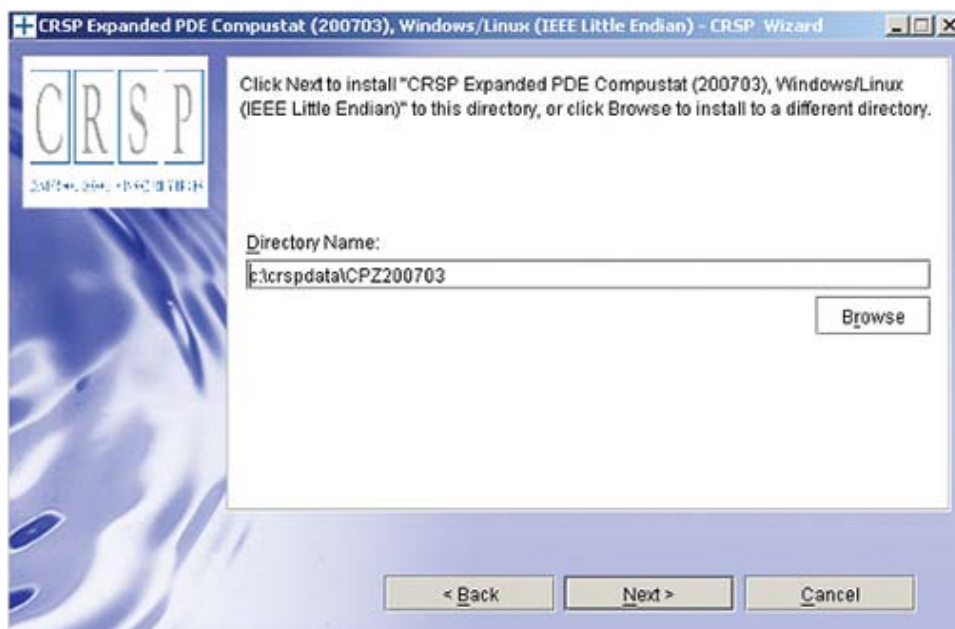


#### WINDOWS:

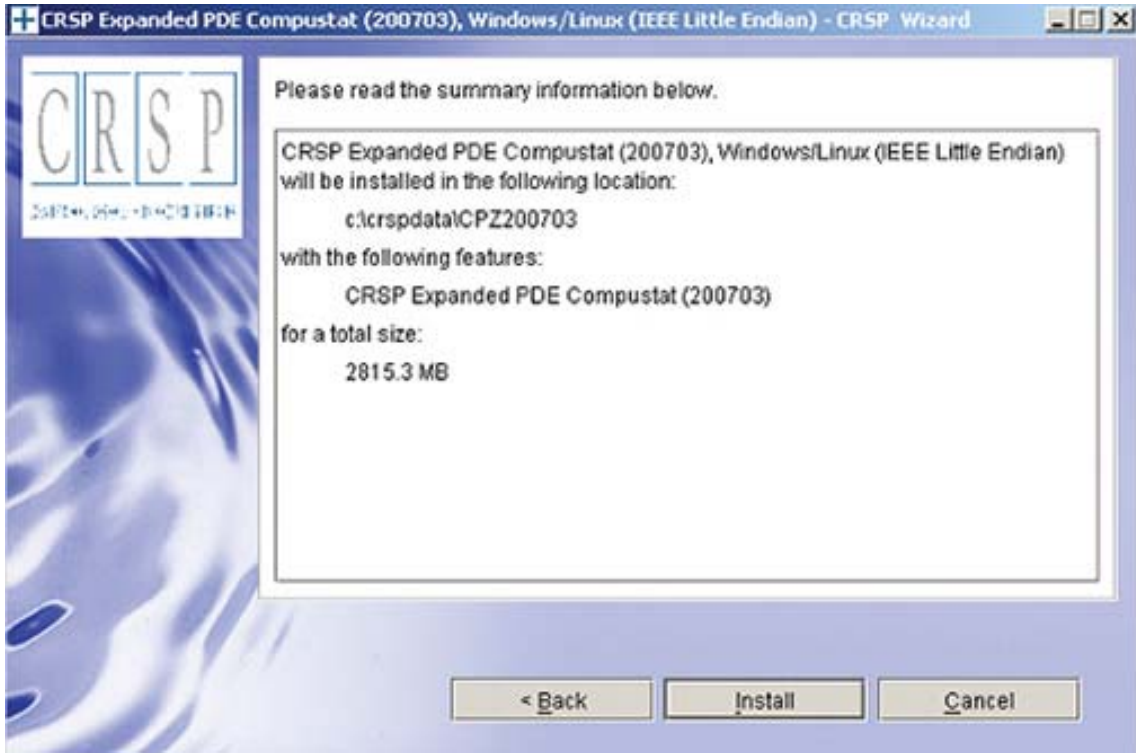
The default directory for the data installation is `c:\crspdata\cczyyyymm`. You may accept or modify it. Click **Next**. **Note:** InstallShield does not verify available space. Please see the database parameters on page 1 for space required for installation.

#### SUN SOLARIS & LINUX:

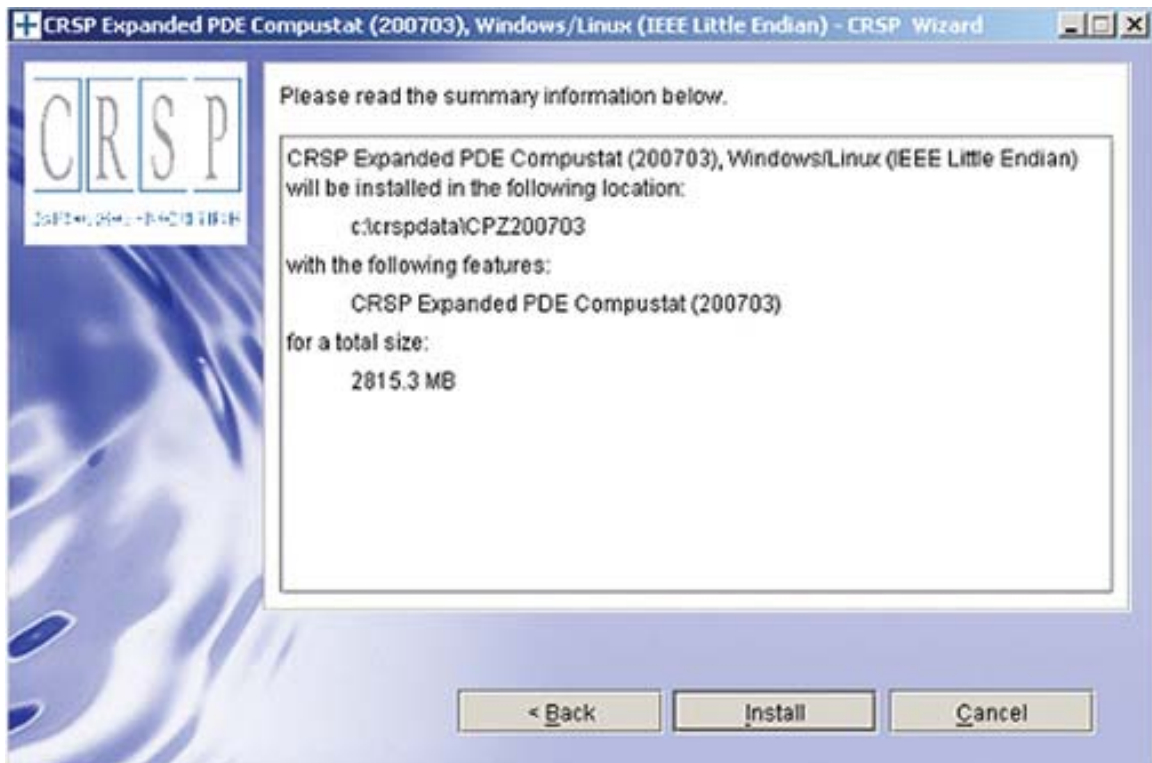
The default directory for the data installation is `/crspdata`. You must have write permission to the directory you specify, for example, `/home/username/crspdata`. **Note:** InstallShield does not verify available space. Please see the database parameters on page 1 for space required for installation.



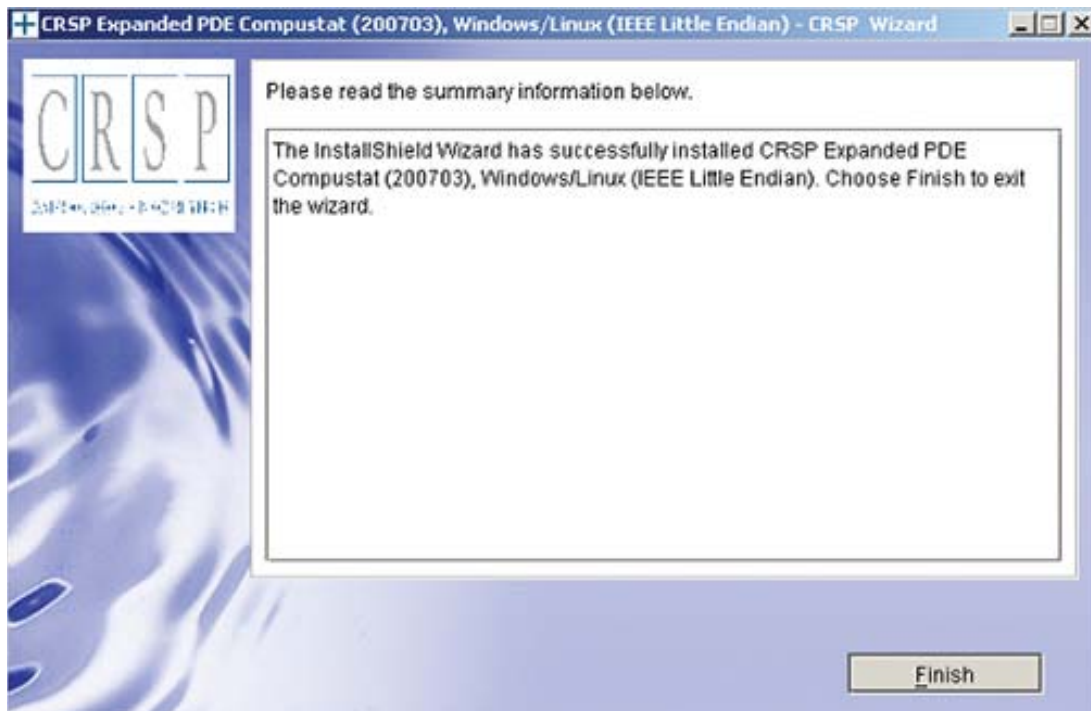
Summary information is displayed: Location, data components and the amount of space that is required for the data. Click on **Install** to proceed.



The screen will display the status of your installation.



A message indicating the success of the installation appears when the process is complete. Click **Finish**.



#### SUN SOLARIS & LINUX:

After clicking Finish, a file named `crsp.kshrc` must be run to set the local environment variable for the database. At the prompt, type:

```
> . ./crsp.kshrc <enter>
```

**Note:** If you are not using `ksh`, modifications will be necessary. Check with your Unix administrator.



## WINDOWS:

Location of the software and data may be checked by opening **CRSP Client Environment** from the CRSPAccess menu. This new utility provides users a quick and simple way of both checking and switching environment variables if necessary. Where, in the past, users had to go through the Control Panel to manipulate the environment variables, this can now be done by quick editing of the Client Environment screen.

The screenshot shows a window titled "CRSPAccess Client Environment" with a standard Windows title bar. The window contains five sections, each with a label, a description, a text input field, and a "Browse" button:

- CRSP Root**  
CRSP Root Directory  
This folder/directory contains the CRSP root subfolders that will allow the CRSP utilities to function properly.  
Input field: N:\crsprot294\  
Browse button
- CRSP Monthly Stock**  
CRSP Monthly Stock Database Directory  
Input field: c:\crspdata\MIZ200611\  
Browse button
- CRSP Daily Stock**  
CRSP Daily Stock Database Directory  
Input field: c:\crspdata\DIZ200611\  
Browse button
- CRSP Work**  
CRSP Work Directory  
This folder/directory is used for storage of temporary files, log files and contains subfolders that may hold user created files  
Input field: C:\CRSP\work\  
Browse button
- CRSP Compustat CCM**  
CRSP Compustat Merged Database Directory  
Input field: c:\crspdata\CPZ200703\  
Browse button

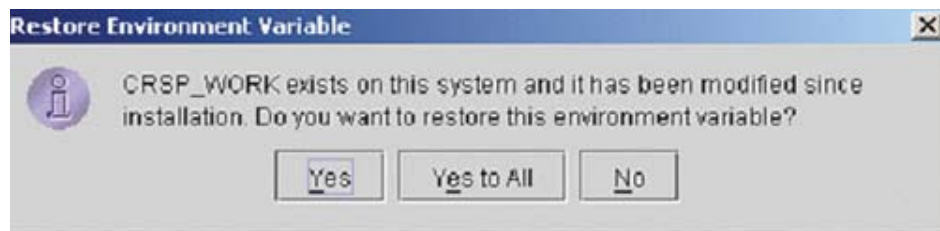
At the bottom of the window, there are three buttons: "Do", "Remove", and "Exit".

## UNINSTALL

For each data install, an uninstall is created on the CRSPAccess menu. Data uninstalls do not touch software. Likewise, software uninstalls do not touch data. Each uninstall is done separately. As with the install, the uninstall is wizard-driven. When installation of either software or data is complete, users may be prompted with an option to either reboot the computer or wait until later.

### Uninstall & CRSP Client Environment

Each time a path is modified in client environment, the action is recorded. Upon running `uninstall`, a *Restore Environment Variable* prompt will appear:



This prompt appears for each variable for each modification. Answer each prompt and continue cycling through. **Yes** is the default, which will restore environment variables back to what they were prior to the install.