CHICAGO BOOTH

Center for Research in Security Prices

CRSP

CRSP/COMPUSTAT MERGED DATABASE GUIDE

CHICAGO BOOTH Center for Research in Security Prices

105 West Adams, Suite 1700
Chicago, IL 60603
Tel: 312.263.6400
Fax: 312.263.6430
Email: Support@crsp.ChicagoBooth.edu

200902

CCM DATA DEFINITIONS

This section describes the CRSP Link and Compustat variables and structures supported by the CRSP/Compustat Merged database. For complete definitions, codes, and formulas for Compustat items, see the documentation and resources provided to you by Compustat.

Data Organization

Compustat data are organized by company and security around Compustat's Permanent SPC Identifier, GVKEY, and issue identifier, IID. Secondary identifiers are available in the header and link history that can be used to cross-reference companies to GVKEYs.

A defined structure for Compustat data is used to store all available Compustat and CRSP Link data for a GVKEY. Each structure is broken down into items. All structures are built from three basic data categories: headers, event data arrays, and time series.

- Headers have no time component. They are a collection of data items with one instance for each gvkey. Examples of header data items are current identifiers and date ranges.
- Event data arrays are collections of records, each describing a change in status or a new event. All data items describing the event type are included in each record. These always include one or more data items that describe the effective date range or the effective date of the event. A count of the number of events being referenced is available for each event data array.
- Time-series is a collection of records tied to a specific calendar of time periods. Each time-series has a beginning and ending period and exactly one record of information for each period in that range. A time-series record can include one or more data items describing the period.

The data items defined within each data category are determined by the available Compustat or CRSP Link data for that data type.

Data Items

Data definitions include data items provided by Compustat as well as structures and supplementary data items provided by CRSP. All data items include a mnemonic and field name. With the exception of roughly a dozen data items, the mnemonics of Compustat data items used in the CCM database match the name provided by Compustat. No further definition is provided in this guide except clarification on mnemonics and usage for a data item that may be used differently by Compustat in different files. Supplementary CRSP items include complete definitions.

Implicit in every structure is CCMID, which may be PERMNO, GVKEY, or GVKEYX, depending upon what identifier key is needed for data access. GVKEY is a unique permanent number assigned by Compustat, that can be used to identify a Compustat record in different updates if name or other identifying information changes. GVKEY is the primary key in the CRSP/Compustat Merged Database. Data are sorted and organized by this field.

Master, Header, Header History, and Link History Data

Descriptive structures include the Master, Company and Security Header and Header History, and Link History Data.

Master Defined Structure

The master structure contains CCM Company identification and range data.

Mnemonic	Field Name	Format
BEGQTR	Quarterly date of earliest data (yyyy.q)	integer
BEGYR	Annual date of earlist data (yyyymmdd)	integer
CBEGDT	First date of Compustat data	integer
CCMID	Permanent record identifier for Compustat company or index data, represents GVKEY for company, GVKEYX for index	integer
CCMIDTYPE	Type of key for Compustat data. 1 = company data, 2 = index data	integer
CENDT	Last date of Compustat data	integer
ENDQTR	Quarterly date of last data (yyyy.q)	integer
ENDYR	Annual date of last data (yyyymmdd)	integer

Company Defined Structure

The company structure contains CCM Company Header information.

Mnemonic	Field Name	Format
ADD1-4	Address lines 1-4	character
ADDZIP	Postal code	character
BUSDESC	Business description	character
CIK	CIK number	character
CITY	City	character
CONM	Company name	character
CONML	Company legal name	character
COSTAT	Postal code	character
COUNTY	County code	character
DLDTE	Research company deletion date	integer
DLRSN	Research company reason for deletion	character
EIN	Employer identification number	character
FAX	Fax number	character

FIC	ISO Country code of incorporation	character
FYRC	Fiscal year end (current)	integer
GGROUP	GICS groups	character
GIND	GICS industries	character
GSECTOR	GICS sectors	character
GSUBIND	GICS sub-industries	character
IDBFLAG	International/Domestic/ Both indicator	character
INCORP	State/Province of incorporation code	character
IPODATE	Company initial public offering date	integer
LOC	ISOCountry code/ headquarters	character
NAICS	North American Industry Classification Code	character
PHONE	Phone number	character
PRICAN	Primary Issue Tag - Canada	character
PRIROW	Primary Issue Tag – rest of world	character
PRIUSA	Primary Issue Tag - USA	character
SIC	SIC code	integer
SPCINDCD	S&P industry sector code - reference	integer
SPCSECCD	S&P economic sector code - reference	integer
STATE	State/Province	character
STKO	Stock ownership code	integer
WEBURL	Website address	character

IDX_Index Defined Structure

IDX_Index structure contains index header information.

Mnemonic	**Field Name**	**Format**
IDX13KEY	13 character key	character
IDXCSTFLG	Index constituent flag	character
INDEXCAT	Index category code	character
INDEXGEO	Index geographical area	character
INDEXTYPE	Index type	character
INDEXVAL	Index value	character
SPII	S&P industry index code	integer
SPMI	S&P major index code	integer
TICI	Issue trading ticker	character

XCONM	Company Name (Index)	character
XINDEXID	Index ID	character
XTIC	Ticker/trading symbol (index)	character

SPIND Defined Structure

The SPIND structure contains pre-GICS S&P Index header information.

Mnemonic	Field Name	Format
SPIID	S&P Industry ID	integer
SPIMID	S&P Major Index ID	integer
SPITIC	S&P Index ticker	character
SPIDESC	S&P Index industry description / reference	character

COMPHIST Defined Structure

The COMPHIST structure contains Compustat Company Header history.

Mnemonic	Field Name	Format
HCHGDT	Comphist description effective date	integer
HCHGENDDT	Comphist description last effective date	integer
HDLDTE	Historical research company – deletion date	integer
HFYRC	Historical fiscal year end month / current	integer
HIPODATE	Historical company official public offering date	integer
HSIC	Historical SIC Code	integer
HSPCINDCD	Historical S&P Industry code	integer
HSPCSECCD	Historical S&P Economic sector code	integer
HSTKO	Historical stock ownership code	integer
HADD1…4	Historical address lines 1-4	character
HADDZIP	Historical postal code	character
HBUSDESC	Historical business description	character
HCIK	Historical CIK number	character
HCITY	Historical city	character
HCONM	Historical company name	character
HCONML	Historical legal company name	character

HCOSTAT	Historical active/inactive status marker	character
HCOUNTY	Historical county code	character
HDLRSN	Historical research company reason for deletion	character
HEIN	Historical employer identification number	character
HFAX	Historical fax number	character
HFIC	Historical ISO country code / incorporation	character
HGGROUP	Historical GICS group	character
HGIND	Historical GICS industries	character
HGSECTOR	Historical GICS sector	character
HGSUBIND	Historical GICS sub-industries	character
HIDBFLAG	Historical international, domestic, both indicator	character
HINCORP	Historical state/province of incorporation code	character
HLOC	Historic ISO country code/ headquarters	character
HNAICS	Historical NAICS codes	character
HPHONE	Historical phone number	character
HPRICAN	Historical primary issue tag - Cananda	character
HPRIROW	Historical primary issue tag – rest of world	character
HPRIUSA	Historical primary issue tag - US	character
HSTATE	Historical state/province	character
HWEBURL	Historical website url	character

CSTHIST Defined Structure

The CSTHIST structure contains the header history from the legacy CRSP/Compustat Merged database that was created from Compustat FTP files.

Mnemonic	**Field Name**	**Format**
CST_CHGDT	CST History effective date	integer
CST_CHGENDDT	CST History last effective date	integer
CST_DNUM	CST History industry code	integer
CST_FILE	CST History file identification code	integer
CST_ZLIST	CST History exchange listing and S&P Index code	integer
CST_STATE	CST History state identification code	integer
CST_COUNTY	CST History county identification code	integer
CST_STINC	CST History state incorporation code	integer
CST_FINC	CST History foreign incorporation code	integer
CST_XREL	CST History industry index relative code	integer

CST_STK	CST History stock ownership code	integer
CST_DUP	CST History duplicate file code	integer
CST_CCNDX	CST History current Canadian index code	integer
CST_GICS	CST History Global Industry Classification Standard Code	integer
CST_IPODT	CST History IPO date	integer
CST_FUNDF1	CST History fundamental file identification code 1	integer
CST_FUNDF2	CST History fundamental file identification code 2	integer
CST_FUNDF3	CST History fundamental file identification code 3	integer
CST_NAICS	CST History North American Industry Classification	character
CST_CPSPIN	CST History primary S&P Index marker	character
CST_CSSPIN	CST History subset S&P Index marker	character
CST_CSSPII	CST History secondary S&P Index marker	character
CST_SUBDBT	CST History current S&P subordinated debt rating	character
CST_CPAPER	CST History current S&P commercial paper rating	character
CST_SDBT	CST History current S&P senior debt rating	character
CST_SDBTIM	CST History current S&P senior debt rating - footnote	character
CST_CNUM	CST History CUSIP issuer code	character
CST_CIC	CST History issuer number	character
CST_CONAME	CST History company name	character
CST_INAME	CST History industry name	character
CST_SMBL	CST History stock ticker symbol	character
CST_EIN	CST History employer identification number	character
CST_INCORP	CST History incorporation ISO country code	character

LINK Defined Structure

Native Link usage provides access to Compustat records, regardless of whether or not securities are in the CRSP universe. All Compustat data including index data, Canadian records, and off-exchange ranges that cannot be directly linked to CRSP Data are accessed using GVKEY, GVKEY.IID, and GVKEYX. The native link reads Compustat data as organized and identified by Compustat identifiers and can choose CRSP data appropriate to those records. Decisions on handling overlaps or soft links are left to the user.

Mnemonic	Field Name	Format
LINKDT	linkdt is a calendar date in YYYYMMDD format marking the first effective date of the current link. It is derived from the first or last date of a CRSP exchange listing, the date of a CRSP name change corresponding to the beginning or end of the link the rows of available Compustat data, or the date of a Compustat description change corresponding to the beginning or end of the link.	integer
LINKENDDT	Last effective date of the link record. If the name represents current link information, the LINKENDDT is set to 99999999	integer
LPERMNO	CRSP PERMNO link during link period. It is set to zero if there is no CRSP link during the range.	integer
LPERMCO	CRSP PERMCO link during link period. It is set to zero if there is no CRSP link during the range.	integer
LIID	Security identifier	character
LNKTYPE	Link type code. Each link is given a code describing the connection between the CRSP and Compustat data. Values are: • LC – Link research complete. Standard connection between databases. • LU – Unresearched link to issue by CUSIP • LX – Link to a security that trades on another exchange system not included in CRSP data. • LD – Duplicate Link to a security. Another GVKEY/IID is a better link to that CRSP record. • LS – Link valid for this security only. Other CRSP PERMNOs with the same PERMCO will link to other GVKEYs. • NR – No link available, confirmed by research • NU – No link available, not yet confirmed	character
LINKPRIM	Primary issue marker for the link. Based on Compustat Primary/Joiner flag (PRIMISS), indicating whether this link is to Compustat's marked primary security during this range. P = Primary, identified by Compustat in monthly security data.	character

	J = Joiner secondary issue of a company, identified by Compustat in monthly security data. C = Primary, assigned by CRSP to resolve ranges of overlapping or missing primary markers from Compustat in order to produce one primary security throughout the company history. N = Secondary, assigned by CRSP to override Compustat. Compustat allows a US and Canadian security to both be marked as Primary at the same time. For Purposes of the link, CRSP allows only one primary at a time and marks the others as N.	

LINKUSED Defined Structure

The LINKUSED structure is a superset of the Link structure and is loaded each time data are accessed in CRSP-Centric mode, such as for ts_print, TsQuery, or when C and FORTRAN functions are run. The function builds a composite Compustat record from one or more Compustat GVKEYs and IIDs linked to a CRSP PERMNO. LINKUSED data are accessed using the composite PERMNO, APERMNO, or the Primary PERMNO, PPERMNO.

Mnemonic	Field Name	Internal Storage
ULINKDT	ulinkdt is a calendar date in YYYYMMDD format marking the first effective date of the current link. It is derived from the first or last date of a CRSP exchange listing, the date of a CRSP name change corresponding to the beginning or end of the link the rows of available Compustat data, or the date of a Compustat description change corresponding to the beginning or end of the link.	integer
ULINKENDDT	Last effective date of the link record. If the name represents current link information, the ULINKENDDT is set to 99999999	integer
ULINKID	Unique ID per link associated with PERMNO. This is used to join with range data in the LINKRANGE table that describes the data ranges applied from used GVKEYs.	integer
UGVKEY	GVKEY used in the link	integer
UPERMNO	CRSP PERMNO link during link period. It is set to zero if there is no CRSP link during the range.	integer
UPERMCO	CRSP PERMCO link during link period. It is set to zero if there is no CRSP link during the range.	integer
UIID	Used Security ID	character
USEDFLAG	Flag marking whether link is used in building composite record	character
ULINKPRIM	Primary issue marker for the link. Based on Compustat Primary/Joiner flag (PRIMISS), indicating whether this link is to Compustat's marked primary security during this range. P = Primary, identified by Compustat in monthly security data.	character

	J = Joiner secondary issue of a company, identified by Compustat in monthly security data. C = Primary, assigned by CRSP to resolve ranges of overlapping or missing primary markers from Compustat in order to produce one primary security throughout the company history. N = Secondary, assigned by CRSP to override Compustat. Compustat allows a US and Canadian security to both be marked as Primary at the same time. For Purposes of the link, CRSP allows only one primary at a time and marks the others as N.	
ULINKTYPE	Link type code. Each link is given a code describing the connection between the CRSP and Compustat data. Values are: • LC – Link research complete. Standard connection between databases. • LU – Unresearched link to issue by CUSIP • LX – Link to a security that trades on another exchange system not included in CRSP data. • LD – Duplicate Link to a security. Another GVKEY/IID is a better link to that CRSP record. • LS – Link valid for this security only. Other CRSP PERMNOs with the same PERMCO will link to other GVKEYs. • NR – No link available, confirmed by research • NU – No link available, not yet confirmed	character

LINKRNG Defined Structure

CRSP generates a range table with information on the fiscal periods associated with each used link for each time series calendar frequency and keyset. This shows ranges in each of the fiscal and calendar calendars available in the CCM. This range table shows the ranges from the GVKEY for each type of time series data used to build the composite record for the PERMNO selected.

Mnemonic	Field Name	Format
RLINKID	Linkused row identifier	integer
RKEYSET	Keyset applicable to range	integer
RCALID	Calendar applicable to range	integer
RBEGIND	Beginning time series range of link	integer
RENDIND	Ending time series range of link	integer
RPREVIND	Time series range immediately prededing the link	integer

RBEGDT	Beginning calendar range of link	integer
RENDDT	Ending calendar range of link	integer
RPREVDT	Ending calendar range preceding the link	int(4)
RFISCAL_DATA_FLG	Type of time series, C-calendar or F-fiscal.	char(8)

CRSP-Centric Link Usage

Accessing Compustat data through ts-print and TsQuery is done through the CRSP-centric mode, meaning that the primary access key in this mode is CRSP PERMNO or PERMCO rather than GVKEY, as used in the Native Access mode. The CRSP identifiers are the access keys while the Compustat identifiers become attributes.

In CRSP-Centric mode a composite record is built using the CRSP Link reading one or more GVKEYs. All GVKEYS with some presence of the PERMNO in the link are accessed. A used-link history is built from these link records by identifying those that cover the ranges of Compustat data needed to link to the CRSP identifier. For time series items that are stored on a fiscal period basis, the link ranges are translated to a fiscal range. This translation simplifies the selection of fundamental data that are applicable to the range and allows for the creation of a composite Compustat record from the applicable ranges that correspond to a CRSP security.

Records in CRSP-Centric form are identical in layout to the native records, but use CRSP PERMNO as the effective key. The Compustat component identifiers – GVKEY, IID, and PRIMISS are available in a Link Used table in the CRSP records.

Using the CRSP-Centric view simplifies access when viewing Compustat data through CRSP. One drawback, however, is that only data considered a direct link to CRSP, applied using CRSP link rules, are available.

Accessing two separate GVKEYs from the Link table, see that both share a single PERMNO.

GVKEY = 011947

Link History

LINKDT	LINKENDDT	LPERMNO	LPERMCO	LIID	LINKTYPE	LINKPRIM
19820701	19860304	0	0	00X	NR	C
19860305	19890228	10083	8026	01	LU	P

GVKEY = 015495

Link History

LINKDT	LINKENDDT	LPERMNO	LPERMCO	LIID	LINKTYPE	LINKPRIM
19880101	19890227	0	0	00X	NU	C
19890228	19930909	10083	8026	01	LC	C
19930910	19990304	0	0	01	NR	C

Using CRSP-Centric access in the USEDLINK table, access the composite history using the Primary PERMNO (LINKPRIM=P)

PERMNO = 10083

Link Used

LINKDT	LINKENDDT	GVKEY	IID	LINKID	PERMNO	PERMCO	USEDFLAG	LINKPRIM	LINKTYPE
19820701	19860304	11947	00X	5	0	0	-1	C	NR
19860305	19890228	11947	01	6	10083	8026	1	P	LU
19880101	19890227	15495	00X	0	0	0	-1	C	NU
19890228	19930909	15495	01	1	10083	8026	1	C	LC
19930910	19990304	15495	01	2	0	0	-1	C	NR
19990305	20051019	15495	01	3	86787	16430	-1	C	LC
20051020	99999999	15495	01	4	0	0	-1	C	NR

Company Data

ADJFACT Defined Structure

The ADJFACT structure contains company adjustment factor history.

Mnemonic	Field Name	Internal Storage
EFFDATE	Effective date- company cumulative factor	integer
THRUDATE	Thu date – company cumulative factor	integer
ADJEX	Cumulative adjustment factor by Ex-date	floating point double precision
ADJPAY	Cumulative adjustment factor by Pay-date	floating point double precision

HGIC Defined Structure

The HGIC structure contains company level GICS history.

Mnemonic	Field Name	Format
INDFROM	Effective from (start) date	integer
INDTHRU	Effective through (last) date	integer
GGROUPH	Industry group name	character
GINDH	Group industry	character
GSECTORH	Group industry sector	character
GSUBINDH	Group sub-industries	character

OFFTITL Defined Structure

The OFFTITL structure contains company officer title data.

Mnemonic	Field Name	Format
OFID	Officer ID	integer
OFCD	Officer title	character
OFNM	Officer Name(s)	character

CCM_FILEDATE Defined Structure

The CCM_FILEDATE structure contains company filing date data.

Mnemonic	Field Name	Format
FDATADATE	Company filing data date	integer
FCONSOL	Company consolidation level filedate	character
FPOPSRC	Population source filedate	character
SRCTYPE	Document source type filedate	character
FILEDATE	Company filing date	integer

CCM_IPCD Defined Structure

The CCM_IPCD structure contains company industry presentation code data.

Mnemonic	Field Name	Format
IPDATADATE	Industry presentation code data date	integer
IPCONSOL	Level of consolidation (Industry presentation code)	character
IPPOPSRC	Population source (Industry presentation code)	character
IPCD	Industry presentation code	character

Security Data

SECURITY Defined Structure

The SECURITY structure contains security level header data.

Mnemonic	Field Name	Format
EXCHG	Stock exchange	integer
DLDTEI	Security inactivation date	integer
IID_SEQ_NUM	IID sequence number	integer
SBEGDT	First date of Compustat data for issue	integer
SENDDT	Last date of Compustat data for issue	integer
IID	Issue ID	character
SCUSIP	CUSIP	character
DLRSNI	Security inactivation code	character
DSCI	Security description	character
EPF	Earnings participation flag	character
EXCNTRY	Stock exchange country code	character
ISIN	International security identification number	character
SSECSTAT	Security status marker	character
SEDOL	SEDOL	character
TIC	Ticker/trading symbol	character
TPCI	Issue type	character

SECHIST Defined Structure

The SECHIST structure contains security header history data.

Mnemonic	Field Name	Format
HSCHGDT	Historical security change date	integer
HSCHGENDDT	Historical security change end date	integer
HEXCHG	Historical stock exchange	integer
HDLDTEI	Historical security inactivation date	integer
HIID_SEQ_NUM	Historical issue ID sequence number	integer

HIID	Historical issue ID	character
HSCUSIP	Historical CUSIP	character
HDLRSNI	Historical security inactivation code	character
HDSCI	Historical security description	character
HEPF	Historical earnings participation flag	character
HEXCNTRY	Historical stock exchange country code	character
HISIN	Historical international security identification number	character
HSSECSTAT	Historical security status marker	character
HSEDOL	Historical SEDOL	character
HTIC	Historical ticker/trading symbol	character
HTPCI	Historical issue type	character

SEC_MTHSPT Defined Structure

The SEC_MTHSPT structure contains security monthly split event data.

Mnemonic	Field Name	Format
DATADATEM	Monthly adjustment factor data date	integer
RAWPM	Raw adjustment factor – pay date - monthly	floating point double precision
RAWXM	Raw adjustment factor – ex date - monthly	floating point double precision

SEC_MSPT_FN Defined Structure

The SEC_MSPT_FN structure contains security monthly split event footnotes.

Mnemonic	Field Name	Format
DATADATEMF	Monthly adjustment factor footnote data date	integer
DATAITEMMF	Monthly split footnote dataitem	character
RAWPM_FN1..FN5	Raw adjustment factor – pay date – monthly – footnotes 1-5	character
RAWXM_FN1..FN5	Raw adjustment factor – ex date – monthly – footnotes 1-5	character

SEC_MDIV_FN Defined Structure

The SEC_MDIV_FN structure contains security monthly dividend event footnotes.

Mnemonic	Field Name	Format
DIVDATADATEMF	Monthly dividend footnote data date	integer
DIVDATAITEMMF	Monthly dividend footnote data item	character
DVPSPM_FN1..FN5	Dividend per share – pay date – monthly – footnotes 1-5	character
DVPSXM_FN1..FN5	Dividend per share – ex date – monthly – footnotes 1-5	character

SEC_SPIND Defined Structure

The SEC_SPIND structure contains data associated with security S&P Industry events.

Mnemonic	Field Name	Format
SPBEGDATE	S&P Index event beginning date	integer
SPENDDATE	S&P Index event ending date	integer
SPHIID	S&P holdings industry index ID	integer
SPHMID	S&P holdings major index ID	integer
SPHSEC	S&P holdings sector code	integer
SPH100	S&P holdings S&P 100 marker	integer
SPHCUSIP	S&P holdings CUSIP	character
SPHNAME	S&P holdings name	character
SPHTIC	S&P holdings ticker	character
SPHVG	S&P holdings value/growth indicator	character

IDXCST_HIS Defined Structure

The IDXCST_HIS structure contains security historical industry constituent data.

Mnemonic	Field Name	Format
XFROM	S&P constituent from event date	integer
IDX13KEY	S&P 13 character key	character
XGVKETX	S&P constituent event index GVKEY	integer

SPIDX_CST Defined Structure

The SPIDX_CST structure contains security S&P index constituent events.

Mnemonic	Field Name	Format
SXBEGDATE	S&P constituent event beginning date	integer
SXENDDATE	S&P constituent event ending date	integer
SPFLOAT	S&P float shares	integer
INDEXID	S&P major index ID	character
EXCHGX	S&P constituent exchange	character
TICX	S&P holdings ticker	character
CUSIPX	S&P constituent CUSIP	character
CONMX	S&P constituent name	character
CONTYPE	S&P constituent type	character
CONVAL	S&P constituent value	character

Segment Data

CCM_SEGCUR Defined Structure

The CCM_SEGCUR structure contains operating segment currency rate data.

Mnemonic	Field Name	Format
SC_DATAYR	Data year	integer
SC_DATAFYR	Data fiscal year end month	integer
SC_CALYR	Data calendar year	integer
SC_SRCYFYR	Source fiscal year end month	integer
SC_XRATE	Period end exchange rate	floating point double precision
SC_XRATE12	12-month moving exchange rate	floating point double precision
SC_SRCCUR	Source currency code	character
SC_CURCD	ISO currency code (USD)	character

CCM_SEGSRC Defined Structure

The CCM_SEGSRC structure contains operating segment source data.

Mnemonic	Field Name	Format
SS_SRCYR	Source year	integer
SS_SRCFYR	Source fiscal year end month	integer
SS_CALYR	Data calendar year	integer
SS_RCST1	Reserved 1	integer
SS_SSRCE	Source document code	character
SS_SUCODE	Source update code	character
SS_CURCD	ISO currency code	character
SS_SRCCUR	Source ISO currency code	character
SS_HNAICS	Segment primary historical NAICS	character

CCM_SEGPROD Defined Structure

The CCM_SEGPROD structure contains operating segment product data.

Mnemonic	Field Name	Format
SP_SRCYR	Source year	integer
SP_SRCFYR	Source fiscal year end month	integer
SP_PDID	Product identifier	integer
SP_PSID	Segment link – segment identifier	integer
SP_PSALE	External revenues	floating point single precision
SP_RCST1	Reserved 1	floating point single precision
SP_PNAICS	Product NAICS code	character
SP_PSTYPE	Segment link- segment type	character
SP_PNAME	Product name	character

CCM_SEGCUST Defined Structure

The CCM_SEGCUST structure contains operating segment customer data.

Mnemonic	Field Name	Format
SC_SRCYR	Source year	integer
SC_SRCFYR	Source fiscal year end month	integer
SC_CDID	customer identifier	integer
SC_CSID	Segment link – segment identifier	integer
SC_CSALE	customer revenues	floating point single precision
SC_RCST1	Reserved 1	integer
SC_CTYPE	Customer type	character
SC_CGEOCD	Geographic area code	character
SC_CGEOAR	Geographic area type	character
SC_CSTYPE	Segment link – segment type	character
SC_CNAME	Customer name data	character

CCM_SEGDTL Defined Structure

The CCM_SEGDTL structure contains operating segment detail data.

Mnemonic	Field Name	Format
SD_SRCYR	Source year	integer
SD_SRCFYR	Source fiscal year end month	integer
SD_SID	Segment identifier	integer
SD_RCST1	Reserved 1	integer
SD_STYPE	Segment type	character
SD_SOPTP1	Operating segment type 1	character
SD_SOPTP2	Operating segment type	character
SD_SGEOTP	Geographic segment type	character
SD_SNAME	Segment name	character

CCM_SEGITM Defined Structure

The CCM_SEGITM structure contains operating segment item data.

Mnemonic	Field Name	Format
SI_DATYR	Data year	integer
SI_FISCYR	Data fiscal year end month	integer
SI_SRCYR	Source year	integer
SI_SRCFYR	Source fiscal year end month	integer
SI_CALYR	Data calendar year	integer
SI_SID	Segment identifier	integer
SI_EMP	Employees	integer
SI_SALE	Net sales	floating point single precision
SI_OIBD	Operating income before depreciations	floating point single precision
SI_DP	Depreciation & amortization	floating point single precision
SI_OIAD	Operating income after depreciation	floating point single precision
SI_CAPX	Capital expenditures	floating point single precision
SI_IAT	Identifiable assets	floating point single precision
SI_EQEARN	Equity in earnings	floating point single precision
SI_INVEQ	Investments at equity	floating point single precision
SI_RD	Research & development	floating point single precision
SI_OBKLG	Order backlog	floating point single precision
SI_EXPORTS	Export sales	floating point single precision

SI_INTSEG	Inter-segment eliminations	integer
SI_OPINC	Operating profit	floating point single precision
SI_PI	Pretax income	floating point single precision
SI_IB	Income before extraordinary earnings	floating point single precision
SI_NI	Net income (loss)	floating point single precision
SI_RCST1	Reserved 1	floating point single precision
SI_RCST2	Reserved 2	floating point single precision
SI_RCST3	Reserved 3	floating point single precision
SI_SALEF	Footnote 1 - sales	character
SI_OPINCF	Footnote 2 – operating profit	character
SI_CAPXF	Footnote 3 – capital expenditures	character
SI_EQEARNF	Footnote 4 – equity in earnings	character
SI_EMPF	Footnote 5 - employees	character
SI_RDF	Footnote 6 – research & development	character
SI_STYPE	Segment type	character

CCM_SEGNAICS Defined Structure

The CCM_SEGNAICS structure contains operating segment NAICS data.

Mnemonic	**Field Name**	**Format**
SN_SRCYR	Source year	integer
SN_SRCFYR	Source fiscal year end month	integer
SN_SID	Segment identifier	integer
SN_RCST1	Reserved 1	integer
SN_STYPE	Segment type	character
SN_SNAICS	NAICS code	character
SN_RANK	Ranking	integer
SN_SIC	Segment SIC code	integer

CCM_SEGGEO Defined Structure

The CCM_SEGGEO structure contains operating segment geographic data.

Mnemonic	**Field Name**	**Format**
SG_SRCYR	Source year	integer
SG_SRCFYR	Source fiscal year end month	integer

SG_SID	Segment identifier	integer
SG_RCST1	Reserved 1	integer
SG_STYPE	Segment type	character
SG_SGEOCD	Geographic area code	character
SG_SGEOTP	Geographic area type	character

Missing Data

Notes on Missing Values

Compustat provides specific codes for data items which are:

- not available
- combined with other data items
- considered insignificant by the company
- available only on a semi-annual or annual basis

The data provided in the CRSPAccess format have constants representing each case. Missing value codes conform with Compustat's Strategic Insight and binary conventions for missing values.

Compustat Missing Value Codes		
Missing Value Code	**C Constant**	**Numeric Value**
No data for data item	CST_MISS_NA	.0001
Data has been combined into another item	CST_MISS_COMB	.0004
Data has been reported by the company as insignificant	CST_MISS_INSIG	.0008
Data is only reported on a semi-annual basis	CST_MISS_SEMI	.0002
Data is only reported on an annual basis	CST_MISS_ANN	.0003

CRSP Link®

Overview

CRSP and Compustat data are commonly linked to match CRSP event and market data history with Compustat fundamental and supplemental data. Because of different identification conventions, universes, available historical information, and conventions unique to each organization, linking is not a straightforward process. Through using the CRSP Link, a data array which contains a history of links using CRSP and Compustat identifiers, subscribers may accurately combine CRSP and Compustat data into a single source of clean, reliable data.

Compustat Xpressfeed provides new security level data requiring adjustments to the linking process between CRSP and Compustat databases. Previously, Compustat included one security per record. Now all securities are available with a new identifier, IID, which can be used along with GVKEY to permanently identify all securities tracked by Compustat, and marker items that identify the security that Compustat considers Primary.

CRSP provides two views of the data through the CRSP Link. While the standard form is the native data and linking information that is organized by Compustat GVKEY, CRSP provides tools to use the link to build CRSP-centric records linked by PERMNO, as needed.

Identifiers used by the link:

- GVKEY – Compustat's permanent company identifier.
- IID – Compustat's permanent issue identifier. An identifying relationship exists between IID and GVKEY. Both must be accessed as a pair to properly identify a Compustat security. One GVKEY can have multiple IIDs. Because Compustat company data ranges can extend earlier than security ranges, there may be some time periods with no identified IID for a GVKEY. In these cases, CRSP assigns a dummy IID ending in "X" as a placeholder in the link. This range may or may not be associated with a CRSP PERMNO, but there is no Compustat security data found during the range.
- PRIMISS – Compustat provides a primary marker indicating which security is considered primary for a company at a given time.
- PERMCO – CRSP's permanent company identifier.
- PERMNO – CRSP's permanent issue identifier. There is a non-identifying relationship between PERMNO and PERMCO. One PERMNO belongs to only one PERMCO. One PERMCO can have one or more PERMNOs.

The Linking Process

Prior to the introduction of Xpressfeed, Compustat included only one security per record. The links between CRSP and Compustat were between CRSP PERMNO and Compustat GVKEY. Because PERMNO is a security identifier and GVKEY is a company identifier, the linking was a many to one relationship. More than one PERMNO may be linked to a single GVKEY.

CRSP addressed the security links in phases. The initial phase addressed security links for issues after mid-April in 2007, for that was when the first Compustat security-level information was available. In this phase, links prior to this time were maintained by using the old CST link information as a foundation onto which updates and refinements were applied.

The primary goal of the second phase of building the security links was to remove the April 2007 starting limitation to the security-based links and move to a full security link history. Once the full security history was built, it would be used to generate company –based historical linking broken down into primary issue ranges and indicators.

This process is laborious and demanding of CRSP researchers and programmers. The new links are reflected beginning with the release of the 200806 annual (CMX200806) and the 200810 monthly and quarterly (CMX200810) release.

Native Link Access

The native link, accessing data using GVKEY, GVKEY.IID, and GVKEYX is used to access all Compustat data including index data, Canadian records, and off-exchange ranges that cannot be directly linked to CRSP. The native link reads Compustat data as organized and identified by Compustat identifiers and can choose CRSP data appropriate to those records. Decisions on handling overlaps or soft links are left to the user.

CRSP provides security level link data with a flag indicating whether or not each link is to Compustat's identified primary issue. The primary issue flag can be used to restrict the link to one security per company for each range as it was done with the original CRSP link.

Accessing two separate GVKEYs in Native Mode from the Link table, see that both share a single PERMNO.

GVKEY = 011947

Link History

LINKDT	LINKENDDT	LPERMNO	LPERMCO	LIID	LINKTYPE	LINKPRIM
19820701	19860304	0	0	00X	NR	C
19860305	19890228	10083	8026	01	LU	P

GVKEY = 015495

Link History

LINKDT	LINKENDDT	LPERMNO	LPERMCO	LIID	LINKTYPE	LINKPRIM
19880101	19890227	0	0	00X	NU	C
19890228	19930909	10083	8026	01	LC	C
19930910	19990304	0	0	01	NR	C

CRSP_CCM_LINK – Security Link History

Itm_name	Type	Description
GVKEY*	integer, primary key (1)	Compustat GVKEY
LIID	char(3), primary key (2)	Compustat IID. Dummy IID assigned with an "X" suffix during a range when company data exists but no Compustat security is identified.
LINKDT	integer (date), primary key (3)	First effective calendar date of link record range
LINKENDDT	integer (date)	Last effective calendar date of link record range

LPERMNO	integer	Linked CRSP PERMNO, 0 if no CRSP security link exists
LPERMCO	integer	Linked CRSP PERMCO, 0 if no CRSP company link exists
LINKPRIM	char(3)	Primary issue marker for the link. Based on Compustat Primary/Joiner flag (PRIMISS), indicating whether this link is to Compustat's marked primary security during this range. P = Primary, identified by Compustat in monthly security data. J = Joiner secondary issue of a company, identified by Compustat in monthly security data. C = Primary, assigned by CRSP to resolve ranges of overlapping or missing primary markers from Compustat in order to produce one primary security throughout the company history. N = Secondary, assigned by CRSP to override Compustat. Compustat allows a US and Canadian security to both be marked as Primary at the same time. For Purposes of the link, CRSP allows only one primary at a time and marks the others as N.
LINKTYPE	char(3)	Link type code. Each link is given a code describing the connection between the CRSP and Compustat data. Values are: • LC – Link research complete. Standard connection between databases. • LU – Unresearched link to issue by CUSIP • LX – Link to a security that trades on another exchange system not included in CRSP data. • LD – Duplicate Link to a security. Another GVKEY/IID is a better link to that CRSP record. • LS – Link valid for this security only. Other CRSP PERMNOs with the same PERMCO will link to other GVKEYs. • NR – No link available, confirmed by research • NU – No link available, not yet confirmed

** - The GVKEY is the primary key of all Compustat company records when using the native link. In CRSPAccess programming this field is not present in the structure but inherited from the CCMID item in the master structure for the company. In standalone usage the GVKEY field is included.*

Only one set of link information is presented for each calendar range in the Compustat GVKEY and IID history. Soft LX and LD links are included if there is a match that indicates an alternate record or a security on a non-US exchange. CRSP provides no automated methods to use these soft links to connect to CRSP data, but the information is available for the user.

Native Link usage provides access to all Compustat records, regardless of whether or not securities are in the CRSP universe.

CRSP-Centric Link Usage

Accessing Compustat data through ts-print and TsQuery is done through the CRSP-centric mode, meaning that the primary access key in this mode is CRSP PERMNO rather than GVKEY, as used in the Native Access mode. The CRSP identifier is the access key while the Compustat identifiers become attributes. There are two options: Primary only, which mirrors the company-level link by ignoring links not to the primary security, and All, which allows use of any link to the PERMNO.

In the CRSP-Centric mode a composite record is built using the CRSP Link reading one or more GVKEYs. All GVKEYS with some presence of the PERMNO in the link are accessed. A used-link history is built from these link records by identifying those that cover the ranges of Compustat data needed to link to the CRSP identifier. For time series items that are stored on a fiscal period basis, the link ranges are translated to a fiscal range. This translation simplifies the selection of fundamental data that are applicable to the range and allows for the creation of a composite Compustat record from the applicable ranges that correspond to a CRSP security.

Records in CRSP-Centric form are identical in layout to the native records, but use CRSP PERMNO as the effective key. The Compustat component identifiers – GVKEY, IID, and PRIMISS are available in the Link Used table in the CRSP records.

Using the CRSP-Centric view simplifies access when viewing Compustat data through CRSP. One drawback, however, is that only data considered a direct link to CRSP, applied using CRSP link rules, are available.

The two securities extracted using Native access via GVKEY returns separate series.

GVKEY = 011947

Link History

LINKDT	LINKENDDT	LPERMNO	LPERMCO	LIID	LINKTYPE	LINKPRIM
19820701	19860304	0	0	00X	NR	C
19860305	19890228	10083	8026	01	LU	P

GVKEY = 015495

Link History

LINKDT	LINKENDDT	LPERMNO	LPERMCO	LIID	LINKTYPE	LINKPRIM
19880101	19890227	0	0	00X	NU	C
19890228	19930909	10083	8026	01	LC	C
19930910	19990304	0	0	01	NR	C

Using CRSP-Centric access, the LINKUSED data show which GVKEYs and IIDs are used to build a composite record by PERMNO. Only the rows with USEDFLAG=1 show the GVKEYs and calendar ranges used tobuild the composite record for PERMNO 10083. table, access the composite history using the Primary PERMNO (LINKPRIM=P)

PERMNO = 10083

Link Used

LINKDT LINKENDDT GVKEY IID LINKID PERMNO PERMCO USEDFLAG LINKPRIM LINKTYPE

19820701 19860304 11947 00X 5 0 0 -1 C NR

19860305 19890228 11947 01 6 10083 8026 1 P LU

19880101 19890227 15495 00X 0 0 0 -1 C NU

19890228 19930909 15495 01 1 10083 8026 1 C LC

19930910 19990304 15495 01 2 0 0 -1 C NR

19990305 20051019 15495 01 3 86787 16430 -1 C LC 20051020 99999999 15495 01 4 0 0 -1 C NR

CRSP_CCM_LINKUSED – CRSP-Centric Link Used History

Itm_name	Type	Description
PERMNO*	integer, primary key (1)	CRSP PERMNO used as basis for this history
ULINKID	integer	Unique ID per link associated with PERMNO. This is used to join with range data in the LINKRANGE table that describes the data ranges applied from used GVKEYs.
UGVKEY	integer	Compustat GVKEY
UIID	char(3)	Compustat IID
ULINKDT	integer (date), primary key (2)	First effective calendar date of link record range
ULINKENDDT	integer (date)	Last effective calendar date of link record range
UPERMNO	integer	Linked CRSP PERMNO, 0 if no CRSP security link exists
UPERMCO	integer	Linked CRSP PERMCO, 0 if no CRSP company link exists
ULINKPRIM	char(3)	Primary issue marker for the link. Based on Compustat Primary/Joiner flag (PRIMISS), indicating whether this link is to Compustat's marked primary security during this range. P = Primary, identified by Compustat in monthly security data. J = Joiner secondary issue of a company, identified by Compustat in monthly security data. C = Primary, assigned by CRSP to resolve ranges of overlapping or missing primary markers from Compustat in order to produce one primary security throughout the company history.

		N = Secondary, assigned by CRSP to override Compustat. Compustat allows a US and Canadian security to both be marked as Primary at the same time. For Purposes of the link, CRSP allows only one primary at a time and marks the others as N.
ULINKTYPE	char(3)	Link type code. Each link is given a code describing the connection between the CRSP and Compustat data. Values are: • LC – Link research complete. Standard connection between databases. • LU – Unresearched link to issue by CUSIP • LX – Link to a security that trades on another exchange system not included in CRSP data. • LD – Duplicate Link to a security. Another GVKEY/IID is a better link to that CRSP record. • LS – Link valid for this security only. Other CRSP PERMNOs with the same PERMCO will link to other GVKEYs. • NR – No link available, confirmed by research • NU – No link available, not yet confirmed
USEDFLAG	integer	1 = this link is applicable to the selected PERMNO and used to identify ranges of Compustat data from a GVKEY used to build a composite GVKEY record corresponding to the PERMNO. -1 = this link is informational, indirectly related to the PERMNO, but not used.

** - The PERMNO is the CRSP security identifier used as the basis for a composite Compustat record and serves as the primary identifier for the composite record. In CRSPAccess programming this field is not present in the structure but inherited from the master structure for the company. The APERMNO or PPERMNO key types store the PERMNO in the CCM structure CCMID field and marks the CCMIDTYPE as 3. In standalone usage the PERMNO field is included.*

CRSP_CCM_LINKRNG – CRSP-Centric Link History Range

The link history is presented by calendar range. If data are presented on a fiscal basis the calendar dates must be interpreted as the proper fiscal period. In this case there can be overlaps generated when links change across GVKEYS or fiscal year end month changes.

CRSP generates a range table with information on the fiscal periods associated with each used link for each time series calendar frequency and keyset. This shows ranges in each of the fiscal and calendar calendars available in the CCM. When there is an overlap and used links provide data for the same fiscal period, the link with the latest filing data date is chosen for the fiscal period. This range table shows the ranges from the GVKEY for each type of time series data used to build the composite record for the PERMNO selected.

Itm_name	Type	Description
PERMNO*	integer, primary key 1	PERMNO key built
RLINKID	integer, primary key 2	unique ID set in the link used record, used for joining range data with the appropriate link.
RKEYSET	integer, primary key 3	Keyset of time series object
RCALID	integer, primary key 4	CRSP calendar of time series
RFISCAL_DATA_FLG	char(1)	Type of time series data, F = fiscal, C= calendar
RBEGIND	integer	first index in time series with valid data for this used link
RENDIND	integer	last index in time series with valid data for this used link
RPREVIND	integer	index of previous data
RBEGDT	integer	first calendar date in time series with valid data for this used link.
RENDDT	integer	last calendar date in time series with valid data for this used link
RPREVDT	integer	date of previous data

* - see note on CRSP_CCM_LINKUSED PERMNO.

Link Actions

This table shows the types of links that are supported by the CRSP CCM link and how they are achieved. A date range is associated with each link so all actions imply an event history.

#	Action	Input Identifier Type	Output Identifier Type	Link Table
1	Find all securities in CRSP for Compustat Company data	GVKEY	PERMNO (PERMCO)	crsp_ccm_link (all links used)
2	Find primary security in CRSP for Compustat Company data	GVKEY	PERMNO	crsp_ccm_link (only links where LINKPRIM is P or C)
3	Find data in CRSP for a specific Compustat Company and issue	GVKEY/IID	PERMNO	crsp_ccm_link (links with desired IID)
4	Find Compustat data for a given CRSP security.	PERMNO	GVKEY/IID	crsp_ccm_linkused (history used to build a composite GVKEY record in

				link used)
5	Find Compustat company and security data for a CRSP security, only if it is considered primary.	PERMNO	GVKEY/IID	crsp_ccm_linkused (only use links where LINKPRIM is P or C)
6	Find Compustat company and security data for a CRSP company.	PERMCO	GVKEY	crsp_ccm_linkused (for each PERMNO of that PERMCO)

Link Action Notes:

1. CRSP_CCM_LINK contains valid links for all securities provided by Compustat. Each record with a valid link to a PERMNO can be followed to the appropriate CRSP data. The user has the option of restricting links by LINKTYPE to ignore soft links, and using the CRSP PERMCO to identify other issues of the same company not addressed in the link. All PERMNOs found with this method share the company-level data from the GVKEY. The link record IID is needed to match the CRSP PERMNO data to the proper Compustat security level data.
2. Link records with the security not marked Primary are ignored. Otherwise this is the same as #1. The result is that even if multiple CRSP PERMNOs are found, there should be no overlap in the CRSP history used. All PERMNOs found will share the company-level data from the GVKEY, but will match only the Compustat IID indicated in the link record.
3. Given a GVKEY and IID from Compustat, use CRSP_CCM_LINK to get the history of CRSP PERMNOs linked to that company and security. The user has the option of restricting soft links using LINKTYPE. No consideration is given to whether the security is considered primary any time during its history. The link can produce multiple CRSP PERMNOs, but only one link should be found at any time.
4. Given a CRSP PERMNO, use CRSP_CCM_LINKUSED to find Compustat data. Access with APERMNO key type will build a composite GVKEY record from the used link records. CRSP_CCM_LINKRNG is used to find ranges of data for the composite record. Secondary links are ignored, and only the Compustat security data matching the permno are included. There will be one composite security record created with a pseudo IID of 01X.
5. Same as #4, but a link record is ignored if the security matched is not primary. This will result in a smaller range, and a not-found if the PERMNO is never primary for the company. Access with PPERMNO key type is used to select this method.
6. PERMCO is not directly supported with linkused, but attached PERMNOs can be found from the PERMCO and the user can select securities with PERMNO. To avoid double-counting company data, the primary flag can be used to ensure that only one security is represented during each time range.

4,5. A user can use secondary index on PERMNO or PERMCO to find GVKEYs with matching information and see the Compustat data in native form, then handle processing as desired. These reads are not necessarily unique, so it is left to the user to select information from the correct ranges corresponding to the desired CRSP identifier.

Table vs. CRSPAccess Usage Notes

The Link Actions table includes the primary identifiers for the databases: GVKEY for CCM and PERMNO for CRSP Stock. In a standalone setup where data are dumped and stored as a table these identifiers are included in each table and used to join data.

CRSPAccess programming access always organizes all data for one GVKEY (CCM) or PERMNO (CRSP Stock) in a single structure. The primary identifier is set at the full structure level and inherited by all substructures. Therefore the field is not explicitly included in the substructures. When a CCM composite record is built by the crsp_ccm_read_all function the primary identifier becomes the PERMNO used as the key, which is stored in the CCM_ID field of this structure. The LOADTYPE flag is set to 1 to signify that the structure is loaded with a composite record.

Security Level Link Data Considerations

Consider the following in order to access the new security level link data.

- Additional security links allow multiple PERMNOs of the same company to link to the same company level data. Users must be aware that the same company data can be retrieved in multiple ways.

- The PERMCO link is no longer needed since a secondary security can link directly between CRSP and Compustat. PERMCO can still be used to find other securities when no direct link is found.

- Security level links are available only during the range of Compustat security data. In some cases, Compustat security data are not available as far back as company data. In others, there may be gaps of security data within a company range. CRSP fills in the available Compustat company data range so at least one link record covers all time periods in the range. If no securities are available during a range, a dummy security is generated for purposes of the link. These dummy securities always have an IID ending with X.

- CRSP assigns a LINKPRIM marker to all link records, based on the Compustat PRIMISS marker. PRIMISS is used to identify the primary security for the company at any given time. LINKPRIM values are

    - P = Primary, identified by Compustat in monthly security data.
    - J = Joiner secondary issue of a company, identified by Compustat in monthly security data.
    - C = Primary, assigned by CRSP to resolve ranges of overlapping or missing primary markers from Compustat in order to produce one primary security throughout the company history.
    - N = Secondary, assigned by CRSP to override Compustat. Compustat allows a US and Canadian security to both be marked as Primary at the same time. For Purposes of the link, CRSP allows only one primary at a time and marks the others as N..

- CRSP supports an access option of primary PERMNO, or ppermno, which restricts links to only those marked primary.

- The legacy CST format databases remain based on the old company-based links, thus using only the rows marked as primary.