EXECUTIVE SUMMARY OF DIFFERENCES

CRSP US STOCK AND INDEXES FLAT FILE 1.0 (SIZ) TO FLAT FILE 2.0 (CIZ)

EASIER AND FASTER TABULAR ACCESS

CRSP's Stock and Indexes Flat File Format 1.0 (SIZ) replicated the internal structures of CRSPAccess's hierarchal design which loaded all data for a security at once. In designing CRSP's Stock and Indexes Flat File Format 2.0 (CIZ) files, additional, and sometimes redundant, items, were added in order to simplify query writing and speed the execution of the data in a tabular format.

Three major categories of data were added to improve access: previous daily period data, intermediate calculation amounts, and calculation summary flags.

- **Previous Daily Period Data**

  The previous price (DlyPrevPrc - used for the calculation of the daily return) and the previous capitalization (DlyPrevCap - used as a weight for market-value-weighted indexes) and their associated date and flags were added to the main daily file, StkDlySecurityData.

  These new fields provide efficient access to these values without the need to use a lag function or a self-join to obtain the values from a previous row.

- **Intermediate Calculation Amounts**

  While CRSP maintains the detailed distribution event history in the StkDistribution file, which includes multiple rows for the same security and ex-Distribution Date and detailed code information about every event, widespread usage of distribution information is to calculate how distributions impact the daily return. The new main daily file, StkDlySecurityData, includes the dividends amounts (DlyOrdDivAmt and DlyNonOrdDivAmt), and the price factor (DlyFacPrc) used to calculate the daily return. These denormalized items allow many queries that previously would have had to aggregate the distribution file before joining to the daily file to be executed without that complexity.

- **Calculation Summary Flags**

  Many new calculation summary flags allow efficient filtering and grouping of data. Three examples of these new flags are DlyDistRetFlg, DlyRetDurFlg, and AnnCompFlg:

  - DlyDistRetFlg (Daily Distribution Return Flag) provides information about the summary distribution values tfrom above so that returns with specific categories of distributions (e.g., No distribution, ordinary cash only, stock split only, etc.) can be easily identified for further analysis.
  - DlyRetDurFlg (Daily Return Duration Flag) provides information about the number of periods and days between the prices used to calculate the daily return, whether there were any missing prices, and if there were no missing prices, whether they were adjacent days or intervening weekends or holidays.

AnnCompFlg (Annual Completeness Flag) provides information about the completeness of the underlying daily date used to create the row in the annual aggregate file (StkAnnSecurityData). The expectation is that this flag, perhaps in conjunction with other flags in the annual file, will allow a user to subset the annual data to what best fits their use case while ideally eliminating (or at least greatly reducing) the queries that have to access the very large (about 100,000,000 rows) daily stock file.

CRSP® | **CENTER FOR RESEARCH IN SECURITY PRICES**
An Affiliate of the University of Chicago

FILE RESTRUCTURING

INTRODUCTION OF ISSUER LEVEL — StkIssuerInfoHdr AND StkIssuerInfoHist

Issuer level was introduced to allow an easy way to have non-duplicated counts of issuers (versus securities), enforce consistency for Issuer Level fields: (ICB, UES, SICCD, USIncFlg, NAICS, IssuerNm), and allow clean non-duplicated information about CRSP (Issuer) Cap-Based Statistics.

For convenience, most issuer fields are also included in the security level files StkSecurityInfoHdr and StkSecurityInfoHist.

- Vast majority of issuers (about 98 %) have only one security, and of those with more than one security, the vast majority have exactly two securities (e.g., Class A and Class B). However, there are a couple dozen of issuers that have dozens or even hundreds of securities.
- As a result of implementing the issuer level, a few differences exist between COMPNO and HCOMNAM, and the new fields, NASDCompno and IssuerNm. In addition, during the past year some historical edits were made to SICCD to ensure consistency for all securities for the same issuer.

MONTHLY HOLDING PERIOD RETURNS

STOCK MONTHLY

There is no direct equivalent to SFZ_MTH's holding period returns, but SFZ_MTH's Mret and Mretx data items map most closely to CIZ MthRet and MthRetx respectively. The reason for this is that the calculation of legacy Mret and Mretx uses different assumptions than the calculation of MthRet and MthRetx. For example, MRet is a month to month holding period return with dividends reinvested at month-end. MthRet is a compound daily return with dividends reinvested on their ex-dates. These two methods will yield different results, in most cases small but they can be significant. This change to using daily compounded returns rather than monthly also applies to delisting returns (dlret) where the chance of larger differences is also present. See Cross Reference Guide for more information.

DAILY PRIMARY DATA (PRIMARY VS. FULL)

STOCK DAILY DATA— StkDlySecurityData AND StkDlySecurityPrimaryData

- StkDlySecurityData is a complete file with 32 columns and closely matches data items in SIZ's SFZ_DP_DLY and SFZ_DS_DLY. The size of a complete file is approximately 20 GB.
- Contains Prices, Returns, Volume, Capitalization, Previous Values (e.g., Previous Price, Previous Capitalization), Bid/Ask, High/Low, and Open.
- Prices, Returns, Volume, and Capitalization are also included in the StkDlySecurityPrimaryData file.
- Subset of delisting columns has been added, reducing the need to include the StkDelists file in some common queries.
- StkDlySecurityPrimaryData file has the same number of rows but fewer columns; all 12 columns in this file are also included in the StkDlySecurityData. StkDlySecurityPrimaryData is analogous to SIZ's SFZ_DP_DLY and contains the most commonly used columns. The StkDlySecurityPrimaryData file size is approximately 7 GB, less than a quarter of the size of the StkDlySecurityData. Since this file is smaller, obtaining results might be easier and faster.

NEW DATA ITEMS

- Addition of ICB Industry field (high level only — Energy, Telecom, Tech, etc.) in the following files: StkIssuerInfoHdr, StkIssuerInfoHist, StkSecurityInfoHdr, StkSecurityInfoHist
- Addition of UES Industry field (high level only — Energy, Telecom, Tech, etc.) in the following files: StkIssuerInfoHdr, StkIssuerInfoHist, StkSecurityInfoHdr, StkSecurityInfoHist
- Addition of the index returns, index levels, and index membership for the investable CRSP Indexes, including the new

CRSP ISS ESG Indexes.

- Addition of a new file, StkMthFloatShares, that contains monthly float time series data for a subset of securities starting as early as December 1999.

Addition of Completeness Flag and Completeness SubFlag field in the monthly, quarterly, and annual aggregate files to allow for easier filtering of complete underlying daily data from infrequent circumstances where the data is not complete for a variety of reasons.

NEW FLAG ITEMS FOR DELISTS DISTRIBUTION AND SHARE CODES

Overloaded data items, such as numeric delists and distribution codes, are "unpacked" and replaced by alphanumeric flag items. As a result, information is more visible, and end-users can filter more efficiently on a more granular level.

Detailed information may be found in the CRSP US Stock & Indexes Flat File Layout Guide 2.0.

METADATA FILES

With the CRSP Stock and Indexes Flat File Format 2.0 (CIZ), CRSP has included 10 metadata files (see Metadata Guide - File Format 2.0). These files allow for improved ingestion into proprietary systems and provide descriptions and definitions that may be incorporated more easily into online access tools. Within the 10 metadata files, there are five groups of files:

- Three metadata schema files containing information about the files, items, and columns can help speed the importing of the ASCII files. In addition, the item names, descriptions, and definitions can be more easily searched within the access tool of choice than a stand-alone PDF document and potentially even incorporated into an online access tool.

- Numeric codes were unpacked into alphanumeric fields, and two metadata flag files provide more flexible and sophisticated searching of all the flag values, descriptions, and definitions used throughout the files.

- Two metadata coverage files contain the results of data profiling done by CRSP that are intended to be useful in providing a more three-dimensional description of the data than a text-only description. For example, the percent of securities for which an item has non-missing values or the number of instances that a particular flag value is used can help determine if a specific item or flag is appropriate for a study.

- Two metadata calendar files are intended to improve transparency related to exchange holidays and closures and ease of use by supplementing and complementing date arithmetic functions with pre-calculated information about the CRSP daily, monthly, quarterly, and weekly periods.

- The tenth metadata file is helpful for those familiar with the Flat File Format 1.0 (SIZ) files or the CRSPAccess files by providing a mapping between the previous item names and the names of the new items. Details on that file are found in the Cross Reference Guide.