An Introduction to **Workspace Datastream**

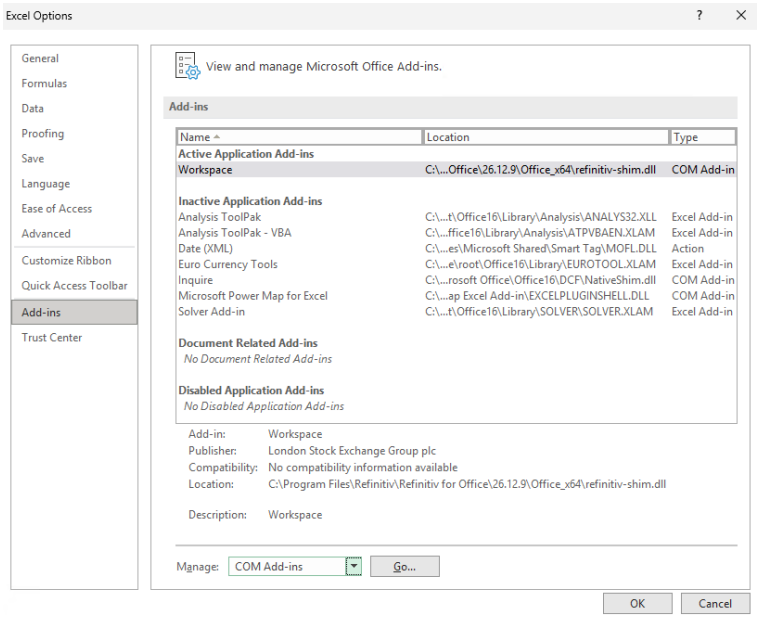
Ulm University

Institute of Finance

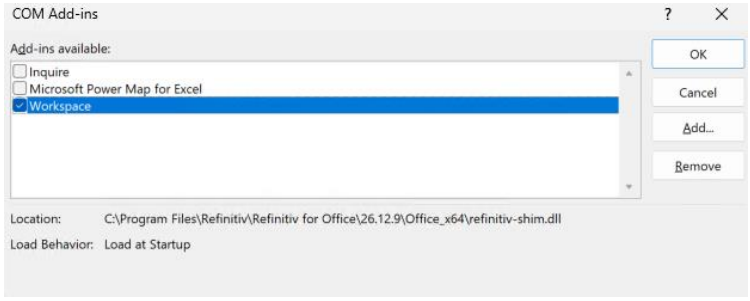
Content

1. How to plug into Datastream

You can use Datastream through Excel via the Workspace add-in. If you do not see the “Workspace” add-in in Excel, go to File -> Options -> Add-Ins and select the Manage “COM Add-ins” from the bottom dropdown.



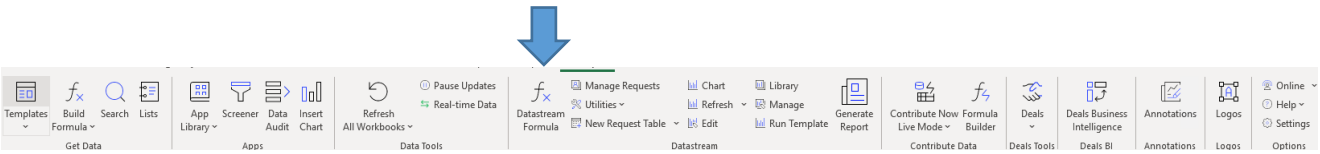
Check the “Workspace” checkbox and select OK



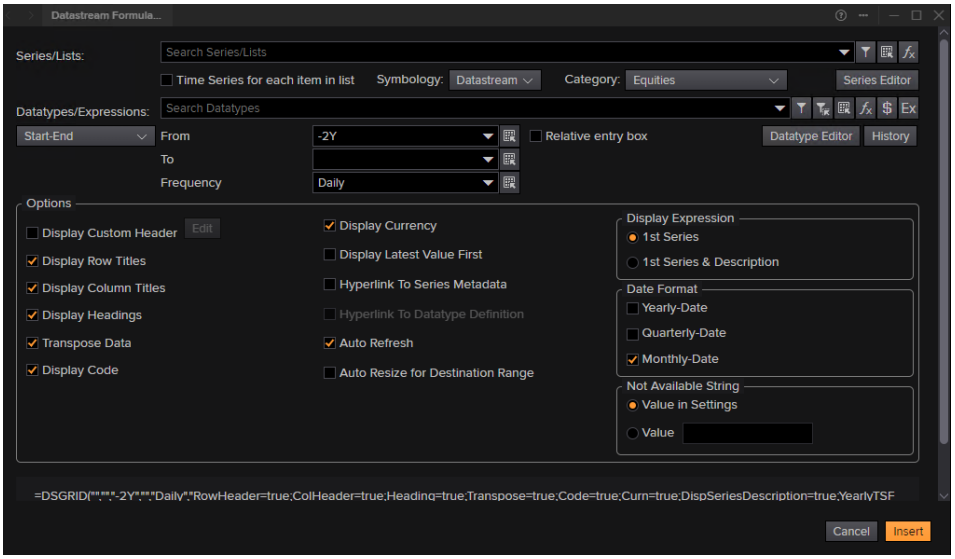
In the following sections, we give examples on how to work with Datastream in Excel.

2. Finding what you are looking for

From the *Workspace* ribbon in Excel, select *Datastream Formula*.



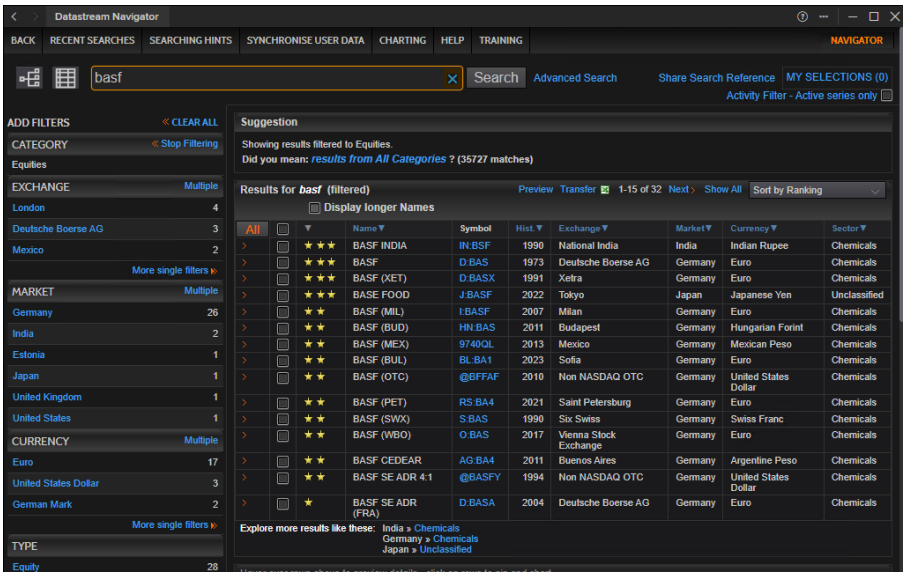
A new window opens:



You need to determine the series (e.g. the DAX, BASF, German unemployment rate) and the type of data (e.g. price, depreciation, net income). For your request, you also need to specify the interval and the frequency. Series can be identified through their specific Datastream symbol, also called the mnemonic. However, there are often alternative ways of identifying a series. For example, you can identify most German stocks by putting a “D” before the WKN (security identification number).

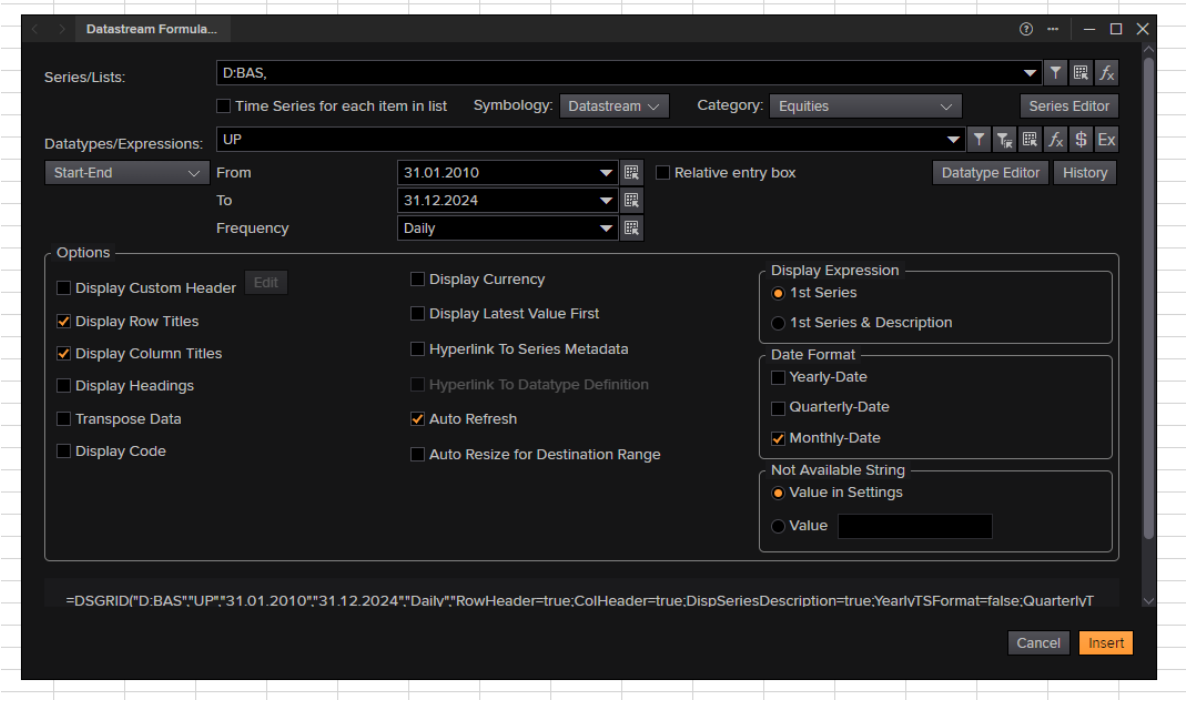
2.1. First example

You need the daily stock price of BASF SE over the years 2010 to 2024. In the *Datastream Formula* window, either start typing (e.g., basf) in the Series/Lists field to get suggestions of Datastream symbols, or click on *Find Series* button  (right to the Series/Lists field). The *Datastream Navigator* opens: Insert your request (basf) into the upper input field. Results are shown immediately.



The result list contains many entries for BASF because BASF shares are traded on several exchanges and Datastream keeps individual records of different data sources. Just “BASF” leads you to the data from Deutsche Börse, the major German exchange. Click on the *Symbol* “D:BAS”, and the symbol will be pasted into the request window of Excel.

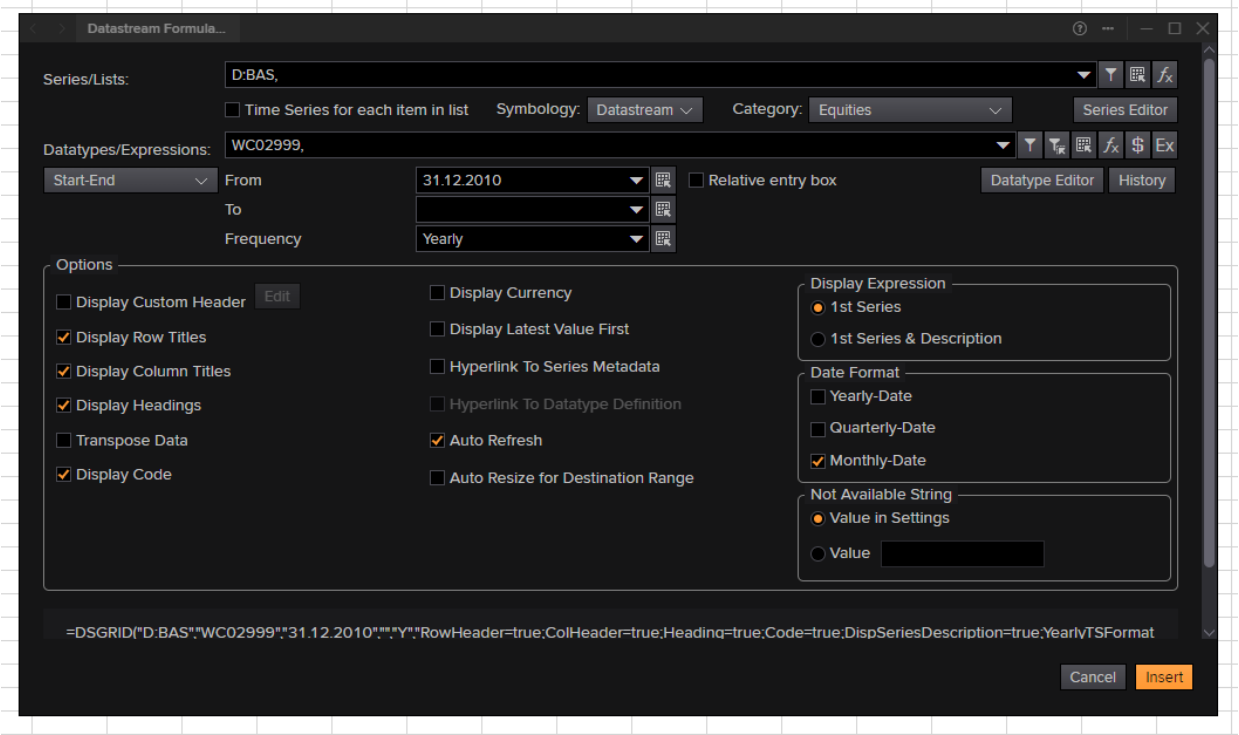
Search for the appropriate type of data by typing something into the *Datatypes/Expressions* field or by clicking on the *Datatypes* button  (right to the *Datatypes/Expressions* field). For example “UP” gives you the unadjusted price, the price that was actually quoted at the respective day in the past. Select a *Start Date* and an *End Date* as well as your *Frequency*. Have a look at the screenshot below:



Press *Insert* in the Datastream window (or *Enter* on your keyboard) and the data will be written into the sheet. The upper left corner of the output range is the cell that was selected in the sheet before you opened the request window.

2.2 Second example

You need annual data on total assets of BASF starting in 2010. By typing in the *Datatypes/Expressions* field or by clicking on the button next to it on the right, you can search for the appropriate type of data, here WC02999 for “Total Assets”.



Select time period and frequency, press *Insert* and you obtain the following results:

	A	B	C	D
1	Start	2010		
2	End	2024		
3	Frequency	Y		
4	Name	BASF SE - TOTAL ASSETS		
5	Code	D:BAS(WC02999)		
6	2010	58788000		
7	2011	60626000		
8	2012	63369000		
9	2013	63390000		
10	2014	69166000		
11	2015	69045000		
12	2016	73983000		
13	2017	76650000		
14	2018	84214000		
15	2019	84063000		
16	2020	76906000		
17	2021	84783000		
18	2022	83592000		
19	2023	76778000		
20	2024	79841000		
21				

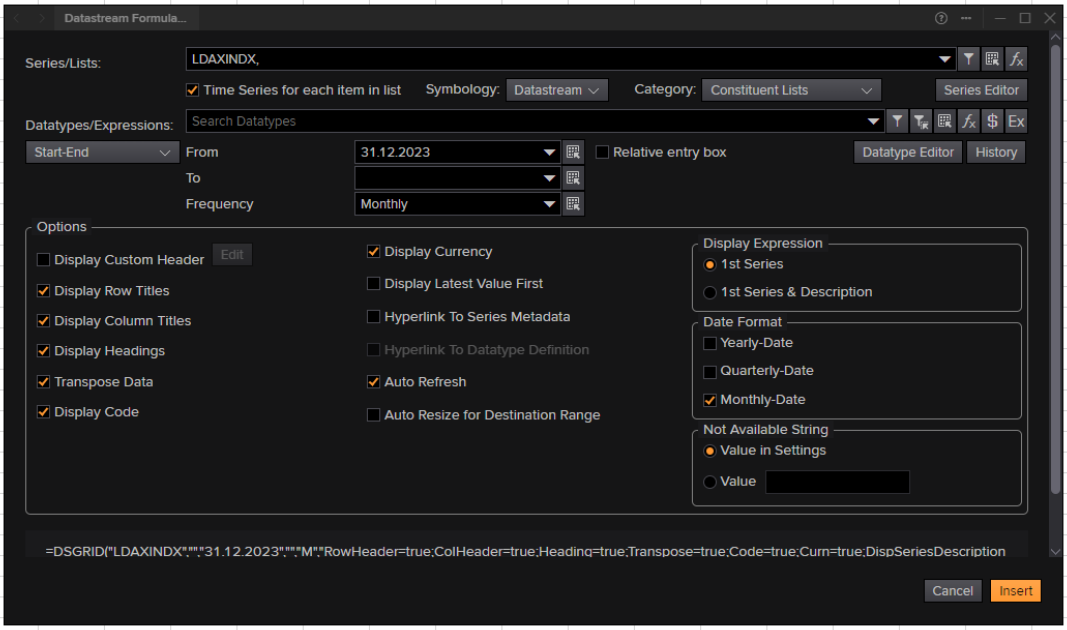
2.3 Third example

You are looking for data on unemployment in Germany. In the *Datastream Navigator*, select the data category *Economics* and use the *Advanced Search* button. You may want to type “unemployment” in the *Name* field, and select “Germany” as the market.

Press *Ok* and you will get a long list of different data in unemployment (e.g. differing in the data provider, or unemployment in various regions). Try to identify what you need. If the list is too long to go through, try to narrow down your search in the search criteria window of the *Advanced Search* in the *Datastream Navigator*.

3. Download data for the stocks contained in a stock market index

Select the *Datastream Formula Builder* window. Enter the Code for the index in *Series/Lists* field, e.g. “LDAXINDX” for the DAX index. Select the *Datatype* that you want, e.g. “RI” for total return index and select dates and frequencies as before. Tick “Time Series for each item in list”. Press *Insert*.

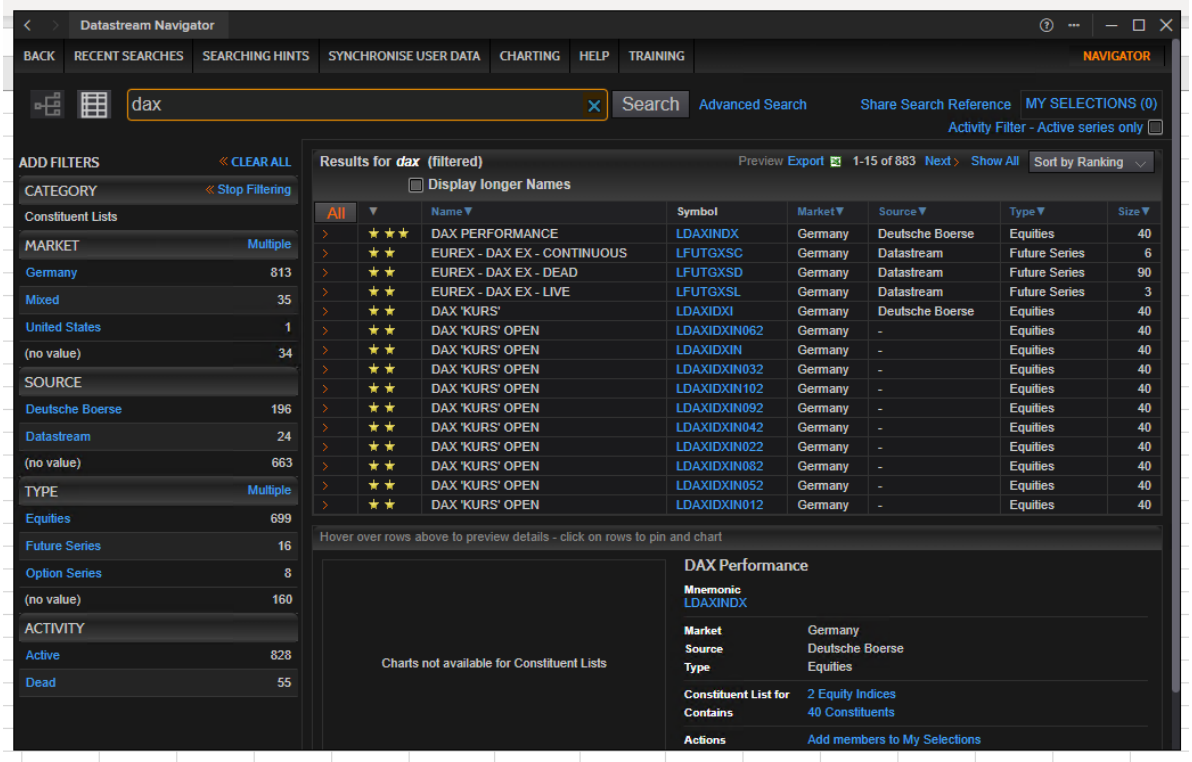


The result should look similar to this:

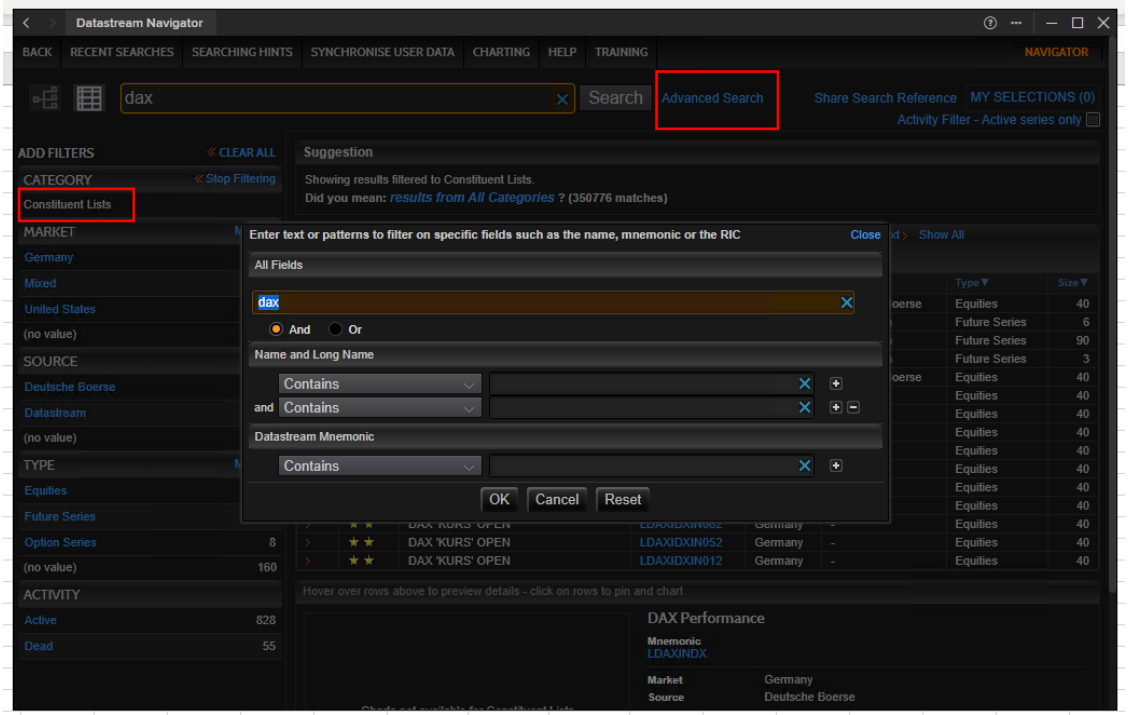
	A	B	C	D	E	F	G	H	I	J	K	L
1	Start	31.12.2023										
2	End	30.04.2025										
3	Frequency	M										
4	Name	Code	CURRENCY	31.12.2023	31.01.2024	29.02.2024	29.03.2024	30.04.2024	31.05.2024	28.06.2024	31.07.2024	30.08.20
5	ADIDAS (X	870874	E	184.16	176.12	187.12	207	226.4	231.5	223	231.3	2
6	AIRBUS (X	289746	E	140.08	147.8	153.02	170.76	154.64	155.9	128.26	140.02	138.
7	ALLIANZ (	866033	E	241.95	247.8	253.85	277.8	266.6	268.3	259.5	260.8	280
8	BASF (XET	866034	E	48.78	44.46	47.075	52.93	49.155	48.41	45.185	43.1	45.
9	BAYER (XE	866035	E	33.63	28.9	28.09	28.43	27.35	28.24	26.38	27.495	27.8
10	BEIERSDO	870528	E	135.7	135.85	132.55	134.95	140.55	144.35	136.55	134.2	130.
11	BMW (XET	866038	E	100.78	96.78	109.22	106.96	102.45	93.18	88.38	85.82	83.
12	BRENNTA	69026P	E	83.22	82.26	84.38	78.08	74.86	65.98	62.96	65.82	67.
13	COMMER	866039	E	10.76	10.68	10.695	12.73	13.96	15.54	14.185	15.09	13.
14	CONTINE	866040	E	76.92	76.2	74.08	66.9	60.84	62.24	52.9	56.7	61.
15	DAIMLER	2628M2	E	34.02	33.26	37.78	46.96	42.31	39.17	37.17	35.74	34.
16	DEUTSCHE	866043	E	12.364	12.034	12.362	14.582	15.01	15.226	14.908	14.436	14.7
17	DEUTSCHE	13454V	E	186.5	184.75	193.6	189.65	181.15	182.9	191.1	189.4	202

How do I find out the index list code to make such a download?

This can be done with the *Datastream Navigator*. In the *Datastream Formular Builder* window, press the *Find Series* button (first button from the left in the upper right corner). Datastream Navigator opens.



Select the data category *Constituent lists* on the left side of Datastream Navigator and just type *dax* in the *Search* field (screenshot above) or click on *Advanced Search* and search for *dax* in *All Fields* or in *Name and Long Name* field (screenshot below).



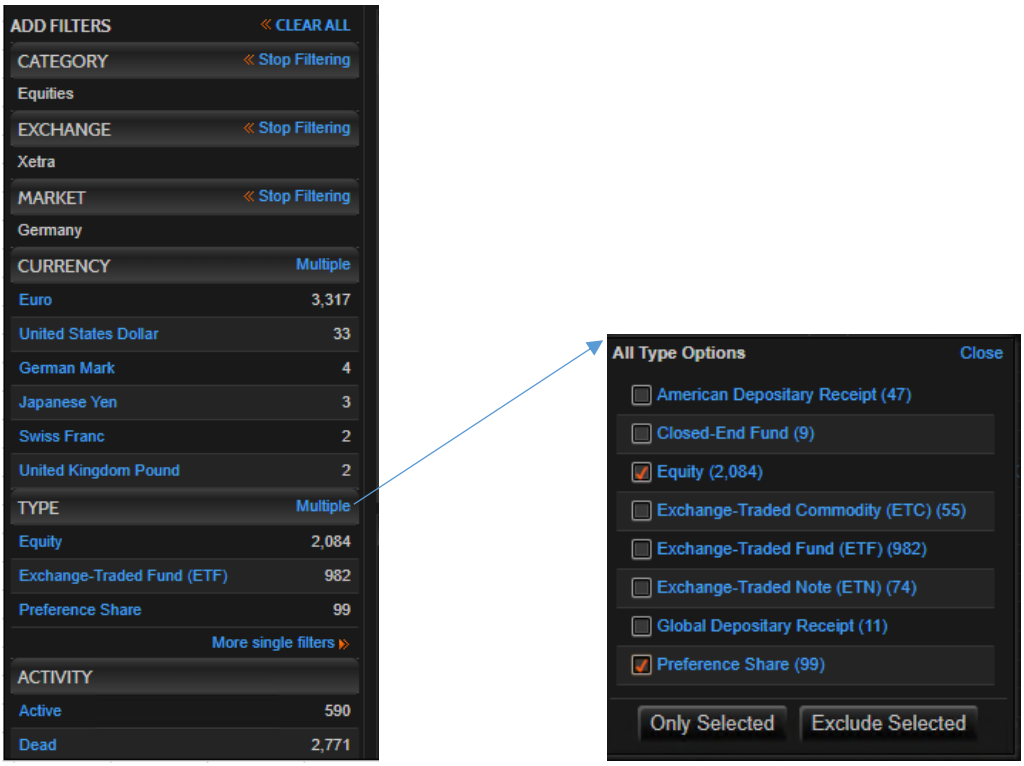
The example above focuses on downloading data for current index constituents. However, for many empirical tasks it is important to obtain lists of historical index constituents. For example, you might want to backtest a trading strategy to see how it would have performed in the past. In this case, it is important to run the backtest on the index as it was in the past to avoid survivorship bias. For some indices, Datastream also provides historical constituent lists. The code for these lists is the code for the current list followed by the month and year (MMYY) for which you want to retrieve the constituents, i.e. if you want to retrieve the constituents for the DAX 40 in December 2021, the code is as follows LDAXINDX1221.
If you want to retrieve data for historical constituents over multiple periods, it is a good idea to structure your download in request tables. A brief description of request tables is given in the next chapter. An example of how to download data for historical constituents over multiple periods can also be found in chapter 8 of the Introduction to Datastream from the University of Bern.

4. Downloading data for a large number of stocks

Sometimes you need to compile a dataset with all stocks that are listed at a specific stock exchange as well as those that were listed at the exchange at some point in the past.

In this example, it is shown how to download all German stocks from the Xetra Exchange.

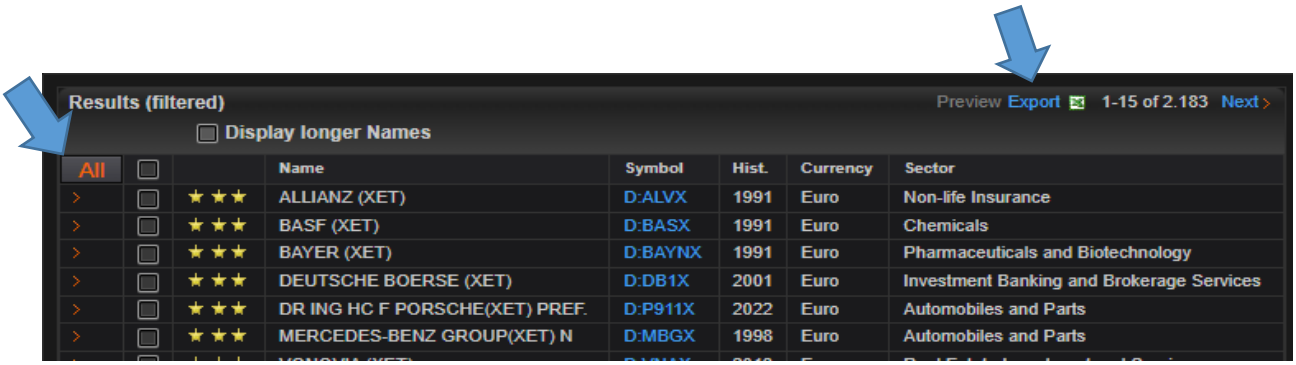
- Go to the *Datastream Navigator*.
- Firstly, choose a single category: *Equities*.
- Next search option is exchange – select *Xetra*.
- Then you can select a market – select *Germany*.
- Under *Type* choose *Equity* to limit your search to common stocks. If you also want to include *preferred stocks*, simply click on *Multiple* and the new window opens. Then you see all possible options and you can select as many types as you want.



- If you do not choose between active and dead stocks under *Activity*, both of them will be selected.

Refining search on the left side of the Datastream Navigator automatically updates the results on the right side of the Datastream Navigator. You can see a long list with names, symbols and other information.

Next, you can see how to export your data into Excel.



You have now two possibilities to proceed. You press *All* (left blue arrow) and Datastream pastes all results to the *Time Series Request* window. Or you proceed with the Excel symbol (right blue arrow). The former method is straight forward but you are fixed to the results of this single search. Moreover, the number of characters that are allowed to be put in the *Series/List field* of the *Datastream Formula Builder* window may be limited (so check whether all the symbols you want to select have actually been pasted into the *Series/List field*) . The second method (export to Excel) allows the combination of

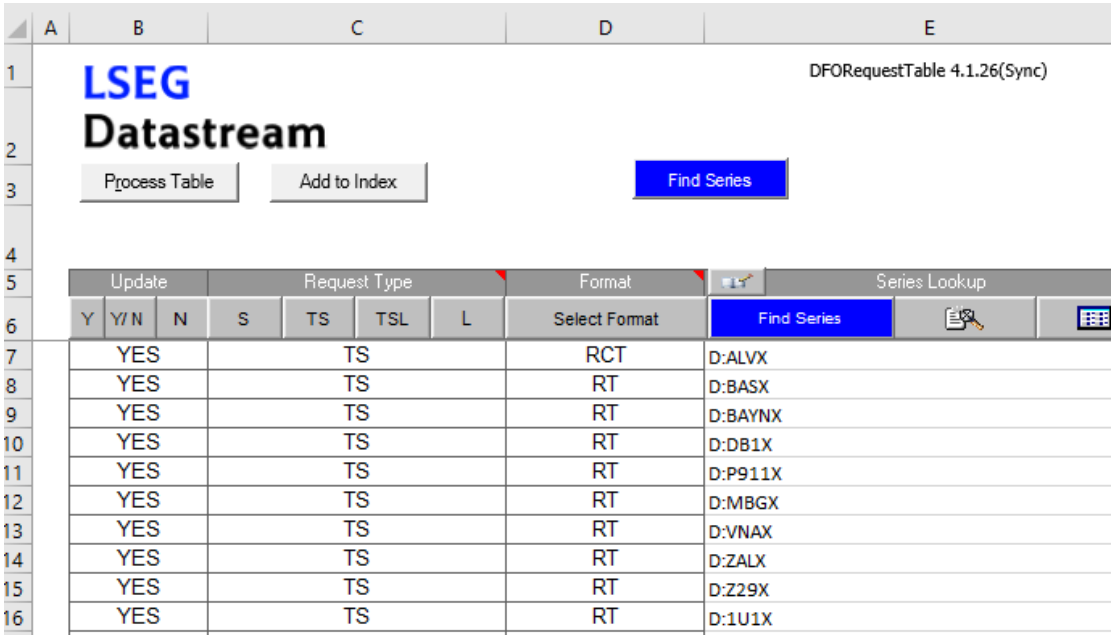
different search results. Again there may be restrictions on the number of lines that are exported but since this has been changing over time we do not provide you with a maximum number but simply point out that you should check whether your export is complete.

Click on the Excel symbol and you can download the data. Close the navigator and request window and open the downloaded file.

	A	B	C	D
1	Name	Symbol	RIC	Start Date
2	ALLIANZ (XET)	D:ALVX	ALVG.DE	01.07.1991
3	BASF (XET)	D:BASX	BASFn.DE	01.07.1991
4	BAYER (XET)	D:BAYNX	BAYGn.DE	01.07.1991
5	DEUTSCHE BOERSE (XET)	D:DB1X	DB1Gn.DE	02.02.2001
6	DR ING HC F PORSCHE(XET) PREF.	D:P911X	P911_p.DE	29.09.2022
7	MERCEDES-BENZ GROUP(XET) N	D:MBGX	MBGn.DE	27.10.1998
8	VONOVIA (XET)	D:VNAX	VNAn.DE	10.07.2013
9	ZALANDO (XET)	D:ZALX	ZALG.DE	30.09.2014
10	029 GROUP (XET)	D:Z29X	Z29.DE	10.10.2022
11	1&1 (XET)	D:1U1X	1U1.DE	12.10.1998
12	11 88 0 SOLUTIONS (XET)	D:TGTX	TGTG.DE	23.04.1999
13	123FAHRSCHULE (XET)	D:123FX	123F.DE	02.11.2021
14	2G ENERGY (XET)	D:2GBX	2GBG.DE	30.07.2007

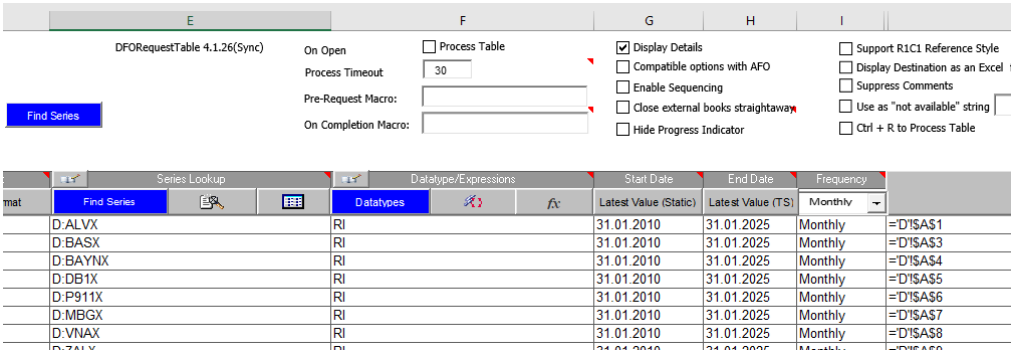
Select the column with Symbols and copy it (the highlighted column in the above example). Now open a new Excel file. From the *Workspace* ribbon choose *New Request Table*.

In the *Series Lookup* column, paste your Symbols or refer to Symbols when they are stored outside the request table.



Now we discuss the filling of the remaining fields. Let us start with the left part. “Yes” in the *Update* column is obvious. “TS” is for the time series request because we want to download a time series for

each code. The *Select Format* tells Datastream whether to display series names, dates, etc. and whether the data belonging to one code should be written in one row or one column. Say we would like to arrange the data such that the data for one stock are in one row. Then choose a "T" for transpose because by default a series is displayed in a column. The row title "R" in this case is the name of the series – we would like to display it for each series, which is why there is an "R" everywhere. The column titles "C" are the dates. We need these dates only once because they will be the same for each series if we choose the same start and end date for each series. So we add the "C" only for the first series. Also note that you get a selection of possible choices by clicking on the grey buttons.



Now to the right part. Choose your type of data, e.g. "RI" for total return index. Choose *Start Date* and *End Date*. Choose *Freq* (Frequency). Now the trickier part: Choose the *Data Destination*. I have inserted a new worksheet in the meantime, which I named "D".

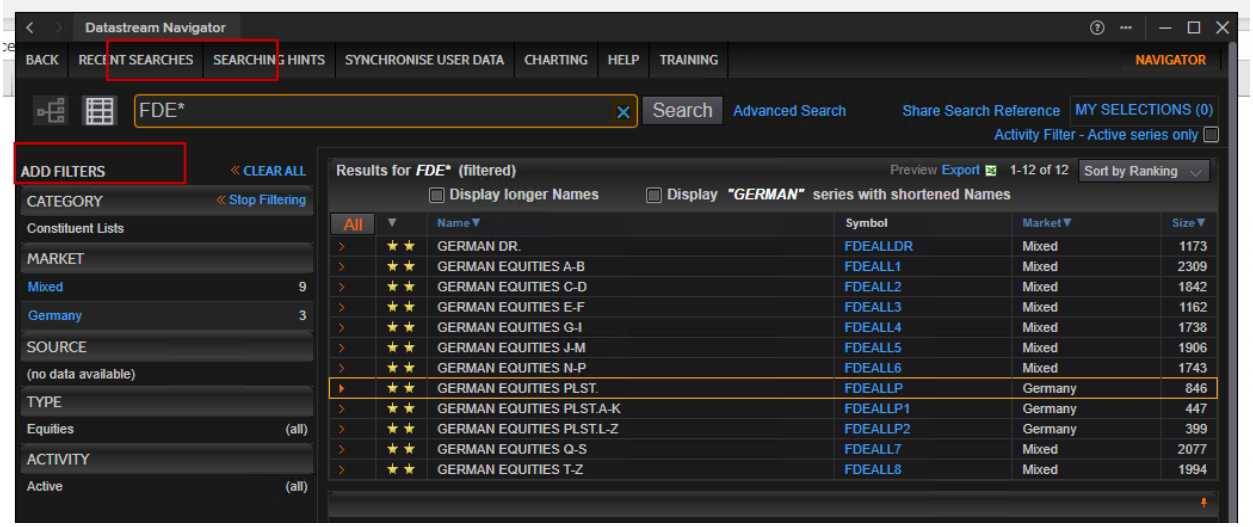
The data for the first series starts in D!A1; it spreads over two rows because it includes the dates. For the other series, we need just one row, which is why we then have D!A3, D!A4 and so on. Note that it is very easy to fill this down. Assume that you have already filled down the frequency cells. Now select two data destination cells which differ by one row, e.g. the ones with D!A3 and D!A4. Double click on the lower right corner and the cells will be filled down in the way you want it to be.

Then press *Process Table* and the data will be downloaded.

Research and Worldscope lists

An alternative to constructing a list of stocks as described above is to use Refintiv's pre-defined constituent lists. These lists have been used in previous financial research (e.g. Ince and Porter (2006)) and can be found by selecting "Constituent Lists" as Category in the navigator. In general, these lists aim to cover all stocks in a given market. However, certain stocks may not be included while other non-equity instruments may be included in a list. For example, Worldscope lists have very sparse coverage for years prior to 1985. To ensure the quality of your data, it is therefore important to follow the screens outlined in previous research (e.g. Landis and Skouras (2021)).

Research lists only include stocks that are currently active in a given market. In order to include stocks that have gone bankrupt or have been delisted, it is also necessary to download data for these companies using so-called "dead lists". Dead lists include stocks that were traded in a particular market but are no longer active. Most research lists begin with an "F" followed by the country code for the market. Dead lists begin with "DEAD" followed by the country code. As the number of stocks covered for some markets would be too large for a single list, research and dead lists are often divided into several sub-lists. A number is added to the name of the list to identify each sub-list. For example, to see all the research lists available for active stocks Germany, select 'Constituent lists' as the category and enter 'FDE' in the search field.



Worldscope lists cover currently active and dead stocks of a market and for most markets only one list exists which covers all stocks. To search for Worldscope lists, select “Constituent Lists” as *Category* and enter “WS” or “WSCOPE” in the search field.

Important hints for working with such data:

You will see a lot of cases where you do not get any data. This is not surprising with such a download. Some series will not have data in the period chosen by you; some may not have data for the type of data chosen by you, etc.

If a stock stops trading on an exchange, there are no more prices and you may expect that Datastream has no values for the return series “RI”. This is not the case. Datastream rather keeps the value from the last trading day in its database and displays it as the current one. So you need to find a way of dealing with such cases. In other words: if the values for one series do not change from some time on until the end, you should discard those repeated values or set them to missing (e.g. to “NA”). Do not set these data points (or missing data in general) to zero. In some cases, this may be equivalent to discarding them, but in other cases it may introduce big errors. Here’s what you can do:

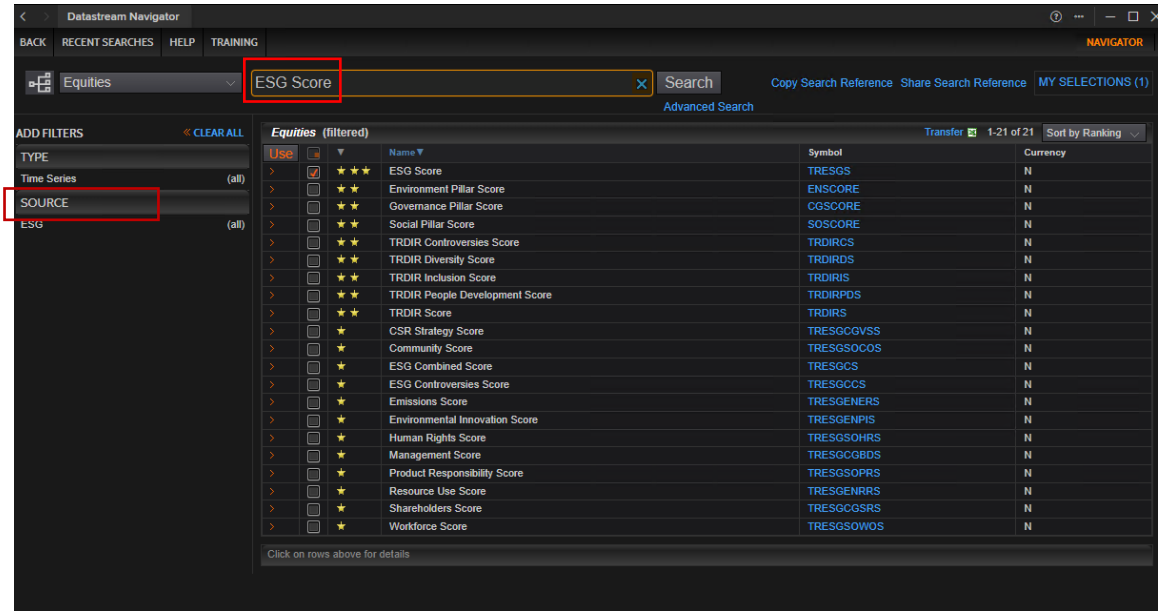
You can write a macro to eliminate the repetitive values at the end of a series. To figure out when a stock stopped trading, you can download (with a static request) the datatype “TIME”, which gives you the date when a stock was last traded.

An alternative and usually simpler procedure is to exploit the fact that Datastream has NAs for other datatypes once a stock stopped trading, e.g. for the datatype P. Using the expression X(RI)*(X(P#T)/X(P#T)) in the *Datatype/Expression* field should give the desired result: a series with the return index RI that is set to NA if P is NA.

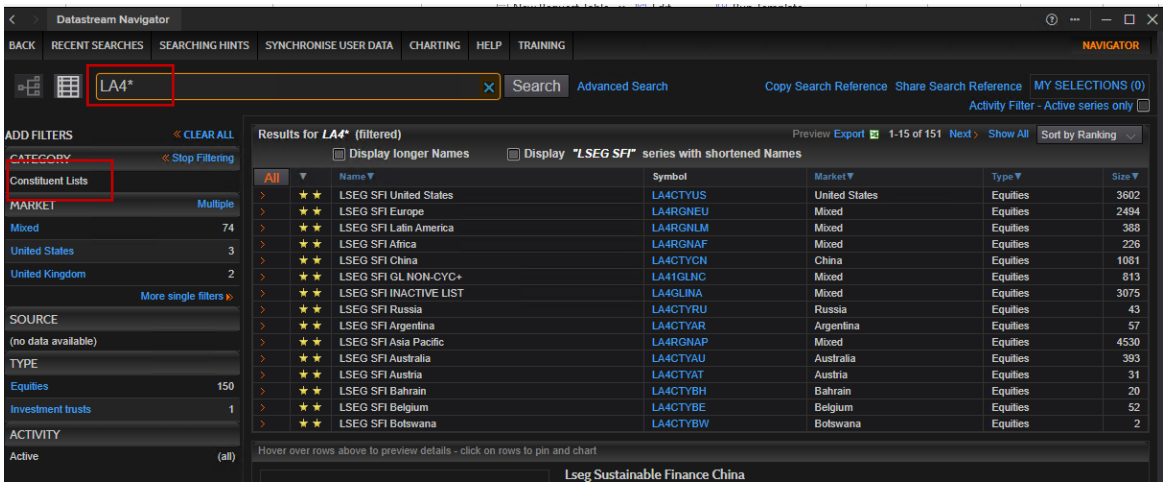
5. ESG data in Datastream

Datastream provides access to environmental, social and governance (ESG) data for over 9000 companies globally. More than 600 individual ESG-related datatypes are available, from which aggregated ESG scores are calculated to reflect a company’s performance in this area. The data availability spans back until 2002.

For an overview of all available aggregated ESG scores click on *Find Datatypes* in the Datastream Formula Builder. Search for “ESG Score” and select *ESG* as Source on the left side.



Datastream also provides constituent lists for all companies with available ESG data. To get an overview of all available lists in the ESG universe, select “Constituent Lists” in the navigator and type “LA4*” in the search field.



If you want to retrieve lists for a specific country, the code for the list starts with "LA4CTY", followed by the country code, e.g. for Germany the code would be "LA4CTYDE". Regional lists start with "LA4RGN" and end with the region code and industry lists start with "LA43" followed by the industry code.

6. Further important hints for working with Datastream

a) Date conventions

In finance and many other fields, the convention is to work with end-of-period values for monthly and yearly intervals. When working with monthly data, for example, you should therefore use month-end values. This can be accomplished by entering month-end dates into the date fields, as was done in the examples above.

With weekly data, it is more customary to select a date from the middle of the week. When combining different data sources, make sure that they are not based on different conventions.

b) Identifying non-trading days

Datastream usually exports data for exchange holidays even though there is no trade on such days. For some empirical applications, you need to highlight and purge the data of these days. There are several possible solutions.

- When you want to get for example total return index data for BASF, type "D:BAS" in the *Series/List* field and "X(RI)*IF#(X(P#S),NNA,ONE)" in the *Datatype/Expression* field of the *Time Series Request* window. You will get the total return index of BASF, but it will be negative when there is a non-trading day.
- Find a reliable source of trading days, e.g. https://www.xetra.com/xetra-en/newsroom/trading-calendar/trading-calendar-archive, or take the dates from research data files (e.g. https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html)

c) Which price/ index type to use

For most purposes such as determining volatilities or investment performance, use the total return index (RI). UP (unadjusted price) is the price as it was historically obtained on the exchange. Use it to implement stock price restrictions (e.g. exclude all stocks with a stock price below one dollar). P (price) is the adjusted price, which takes stock splits and similar corporate actions into account, but does not adjust for dividends. It is Datastream's default datatype but it should not be yours. If you use it instead of the total return index, make sure that you provide good reasons for doing so.

d) Data quality

A discussion of Datastream data quality is given in:

Ince, O. S./Porter, R. B. (2006): "Individual Equity Return Data From Thomson Datastream: Handle with Care!", Journal of Financial Research, 29(4), 463-479

Landis, C./ Skouras, S. (2021). Guidelines for asset pricing research using international equity data from Thomson Reuters Datastream. Journal of Banking & Finance, 130, 106128. (working version available at: https://ssrn.com/abstract=3225371)

Brückner, R. (2013): "Important Characteristics, Weaknesses and Errors in German Equity Data from Refinitiv Eikon Datastream and their Implications for the Size Effect", Working Paper.

This literature suggests a number of screens to deal with quality problems, which you should also consider for your analysis.