

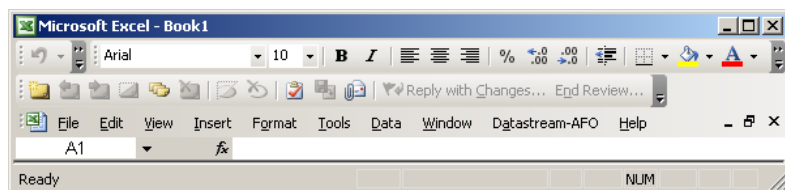
# Datastream

## INTRODUCTION ADVANCE FOR OFFICE

### 1. Introduction

Datastream is one of the most important financial databases available. It is provided by Thomson Financials and contains time series information on more than a million financial securities (such as equities, bonds, options, futures, swaps, mutual funds, indices, ...), with data going back as far as 50 years. To support this financial market data, a selected set of Worldscope company fundamental data and financial ratios for more than 30,000 companies is available. In addition, Datastream provides exchange rates and interest rates as well as some 400,000 economic data series sourced from central banks, national statistics offices, OECD, and IMF. Forecast data for many developed economies are also available.

Downloading data from this database is quite straightforward once the user is familiar with the query screens. The purpose of these notes is to show by example how most standard queries can be performed. We focus on the user interface integrated in Microsoft Excel, the **Advance for Office** tool. When the package is installed, you will see the item "Datastream-AFO" in the Excel menu:



In these notes we will first discuss how to set up a basic request to download data. Next, we introduce lists which allow downloading large amounts of similar data easily. In the last section we present a number of additional tools that can facilitate setting up requests.

### 2. Setting up a request

Data are downloaded by launching a "request". There are essentially three kinds of requests:

- A **time series request**: when you need data of one or more series over a certain time period;
- a **static request**: when you need data of one or more series on a specific date;
- a **company accounts request**: when you need company account data for one or more companies.

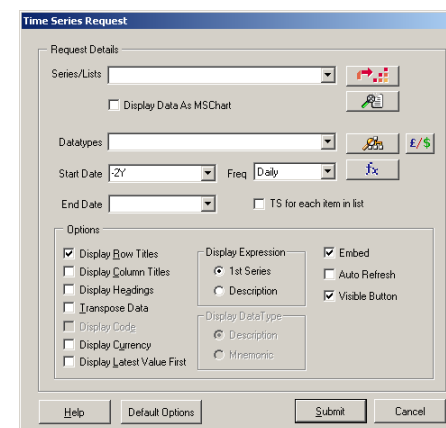
Each of these request types consists out of four steps:


1. specify the **series** for which information is needed;
2. specify the **type** of data needed (the 'datatypes');
3. specify the **date** or **time period** over which information is needed;
4. set **options**.

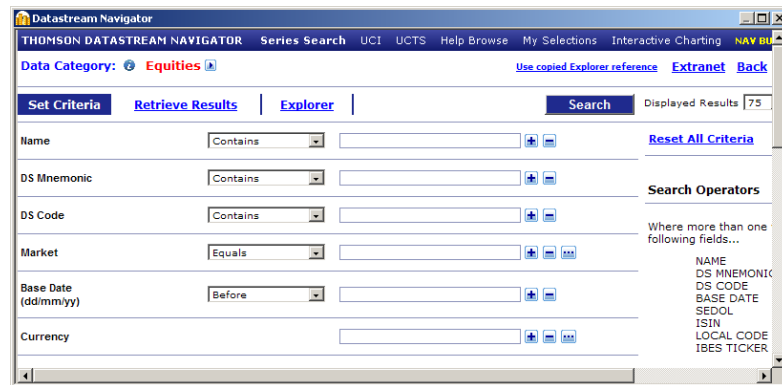
We will discuss all four steps using a time series request in which we need information about Delhaize.

#### First step: specify the series

From the Datastream-AFO menu, pick "Time Series Request". The following interface is opened:



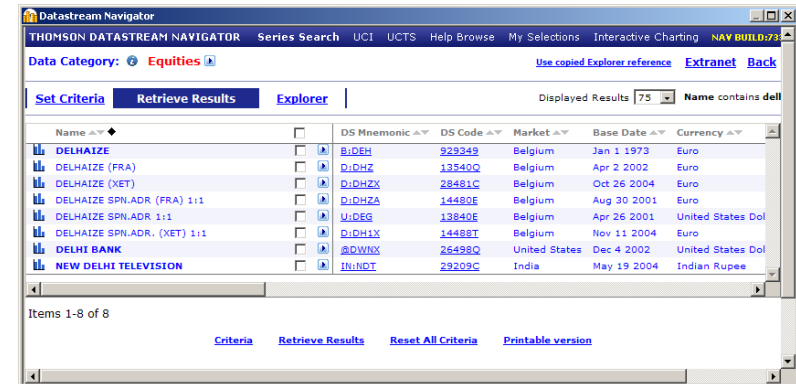
The requested series are identified by filling in their Datastream Mnemonic into the “Series/List” input box. If you do not know the exact mnemonic – this will certainly be the case at first – you can open the *Navigator window* by clicking the  button. The following screen opens:



This screen allows you to search series using a broad range of criteria. The screen shot above only shows the most important criteria. You will discover more criteria by scrolling down. The first “super” criterion is the Data Category. Here “Equities” is chosen, but in addition Datastream distinguishes the following criteria:




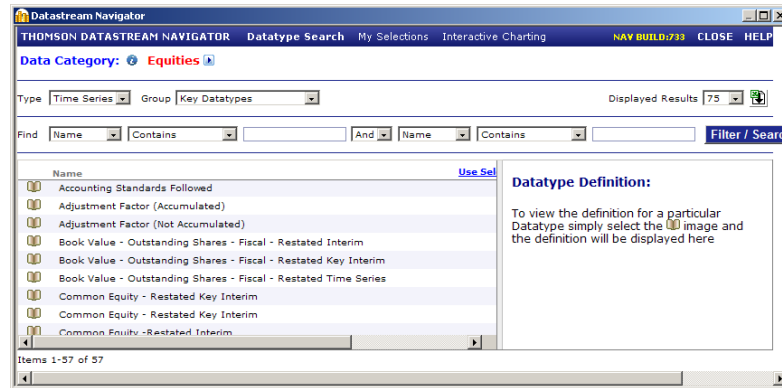
Be sure to use the category to which your requested series belongs. If uncertain, choose “Any Category”, but this may slow down the Search. As we need info about Delhaize, we leave the category choice on Equities. If we fill in “Delh” in the Name input box and click on the “Search” button, the following results are returned:



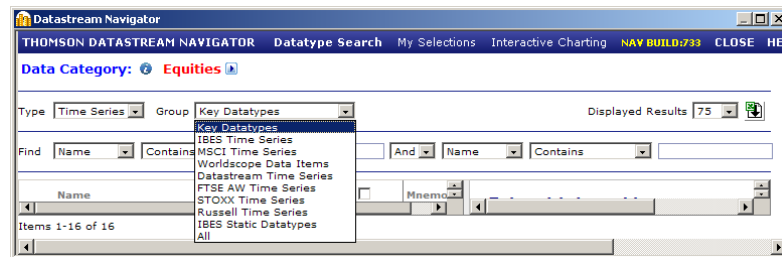
First of all, it is apparent that some results are represented in boldface. These are so-called primary series. For instance, the Delhaize stock is quoted on several markets, but the main market is Euronext (Brussels). This one is in boldface. Focusing on this line, you see the DS Mnemonic (B:DEH) as well as the DS Code, which is an alternative way to specify the series. Scrolling to the right more (straightforward) info is given, such as the base date (starting date) of the series and the currency. Clicking on the DS mnemonic returns this into the request screen. This concludes the first step.

## Second step: specify the datatype

Having identified the security for which information is required, the user has to specify *which kind* of information is needed. This is done in the “Datatypes” input box. If left blank, a default datatype is assumed. For equities this is simply the market price. To specify other datatypes, simply type their codes separated by commas in the input box. When you do not know the codes, you can open the *Datatype Mnemonic Picker* by clicking . The Datastream Navigator now opens on the following screen:

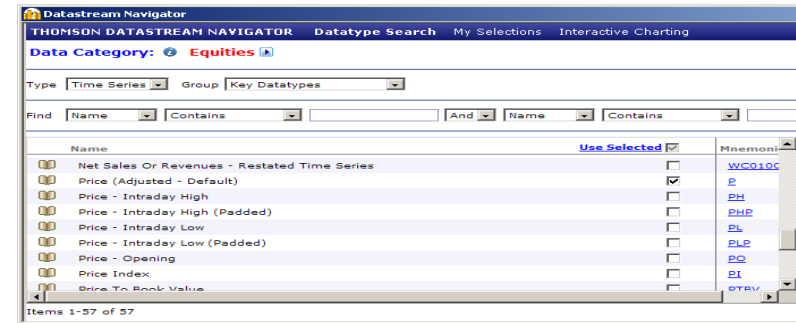


Again, note that the Data Category “Equities” is selected. Indeed, every data category has its specific datatypes. There are also two types of datatypes: one for time series requests and one for static requests. Finally, datatypes are assembled into groups. Above the most common datatypes – the so-called Key Datatypes – are selected, but you can easily switch to other groups:

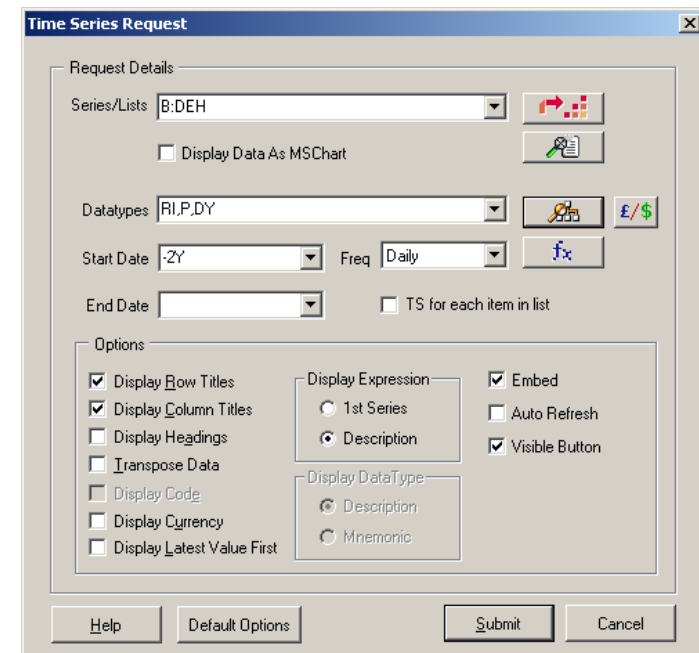


It should be noted that access to some none-key datatypes is restricted. Finally, the Navigator also allows filtering the datatypes to facilitate finding the datatypes you need.

Let's select the dividend yield (DY), price (P), and the total return index (RI) by checking the check boxes to the right of the datatypes:



If you are not sure about the exact content of a datatype, simply click on the open book icon to the left of the datatype to get its definition. To finalize the input, click on “Use Selected”. The datatypes are now entered in the input screen:



### Third step: specify the data period

To specify the data period, three input boxes are available:

1. The **Start Date** drop box: here you can enter the starting date of the requested period in two ways, a relative and an absolute way. The **relative way** is shown above: “-2Y” indicates that the last two years of data are to be downloaded. Likewise you can use the letters “D” for days, “W” for weeks, “M” for months, and “Q” for quarters. Another way to enter relative start dates is by using the “Start of Week/Month/Quarter/Year” entries. The advantage of the relative way is that you always have the most frequent data when you update the request.

When you need a fixed period, use the **absolute way** of entering starting (and end) date by simply typing the data in the following format dd/mm/yyyy. You can also enter “Base Date”, which means starting from when the data series becomes available.

2. The **Freq** drop box, where you indicate the frequency of the download: Daily, Weekly, Monthly, Quarterly, or Yearly.
3. The **End Date** drop box. If left blank, the period ends at the most recent date.

### Fourth step: specify options

Finally, you can set some options about how the output will be returned into Excel. For non-experienced users, it is advised to check “Display Row Titles”, “Display Column Titles”, “Display Headings” and “Display Currency” to have a clear overview about which data were exactly downloaded.

It is also useful to check the “Embed” and “Visible Button” boxes. Datastream will then embed the request into the Excel sheet, such that the user can easily update or edit the request. If the “Auto Refresh” box is checked the request is automatically updated when the Excel workbook is opened. Needless to say, when the workbook contains many requests and/or long requests this may slow down the opening of the file dramatically.

### Submitting the request

We are now ready to submit the request by clicking the “Submit” button. This is the result:

	A	B	C	D	E	F	G
1	Refresh	18/11/2005					
2	End	19/11/2007					
3	Frequency	D					
4	Name	DELHAIZE	DELHAIZE	DELHAIZE	DELHAIZE	DELHAIZE	DIVIDEND YIELD
5	Code	B:DEH(RI)	B:DEH(P)	B:DEH(DY)			
6	CURRENCY	E	E	E			
7	18/11/2005	7824.07	52.9	2.08			
8	21/11/2005	7868.44	53.2	2.07			
9	22/11/2005	7905.42	53.45	2.06			
10	23/11/2005	7971.97	53.9	2.04			
11	24/11/2005	7949.79	53.75	2.05			
12	25/11/2005	7986.77	54	2.04			

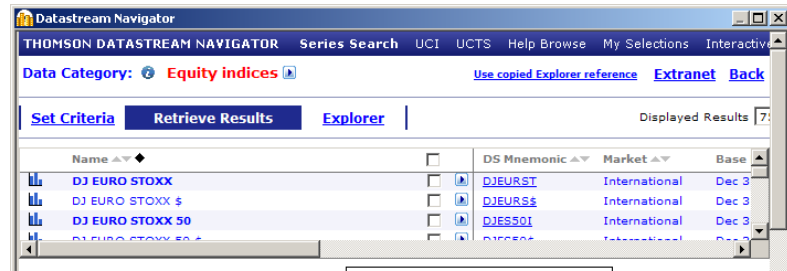
Datastream has returned a time series for each of the datatypes requested. Note that also all the essential information about the request is indicated in the Excel sheet. By having checked the “Embed” and “Visible” boxes, there is also a Refresh button. By clicking it, the request is updated. By right-clicking on it, you can cut, copy, or delete this request, but you can also open the Editor which brings you back to the input screen. You can then easily adapt the request if necessary. Try it out in order to download also information about Carrefour (F:CRFR).

## 3. Working with lists

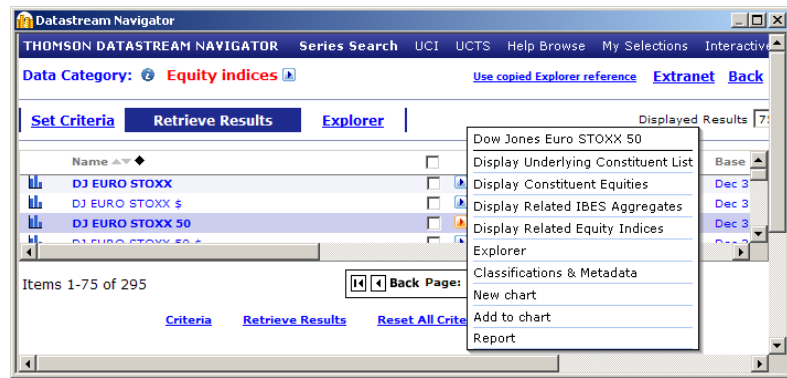
When you need data for a large number of securities, you may consider working with lists. As the name indicates, lists are a collection of Datastream codes or Mnemonics. There are lists created and maintained by Datastream, but you can also create your own lists. We will first download a so-called constituents list, which contains the codes of the constituent securities of a stock market index. We will then use this list to create our own list and use it to download data.

### Using a constituents list

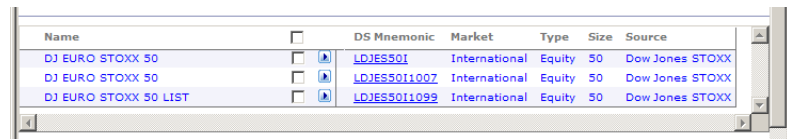
When we want to know the stocks that constitute the Dow-Jones Euro Stoxx 50 index, we open a static request, where we look for Euro Stoxx (be sure to look under Data Category “Equity Indices”):



Next, position the mouse pointer on the right-arrow button, which is located to the right of DJ Euro Stoxx 50. The following screen should appear:

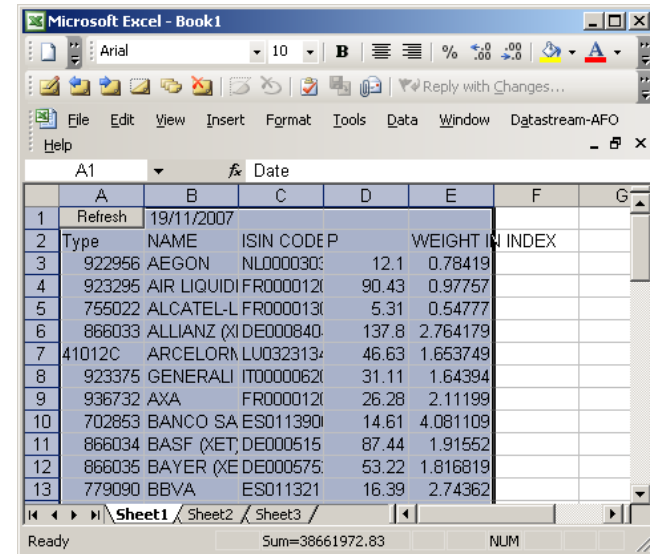


You should click on "Display Underlying Constituent List", which results in



The first one is the present constituent list. Their Datastream mnemonic is always an L followed by the underlying index's mnemonic. You also see two other lists, where 1007 and 1099 are added to the mnemonic. These are the lists on October 2007 and October 1999. The lists for other dates are not available. Now click on the first mnemonic to enter it in the input screen.

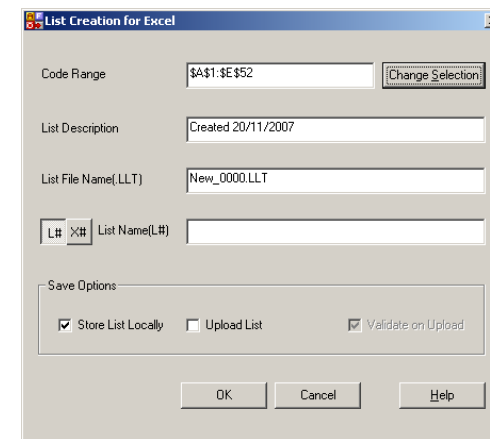
Then, the datatypes have to be specified. With the Datatype Picker we choose the following datatypes: NAME (the name of the company), ISIN (ISIN-code, a unique security identifier), P (stock price), and WTIDX (weight in the index). This is the result:



The first column reports the Datastream code, the remaining columns show the requested information.

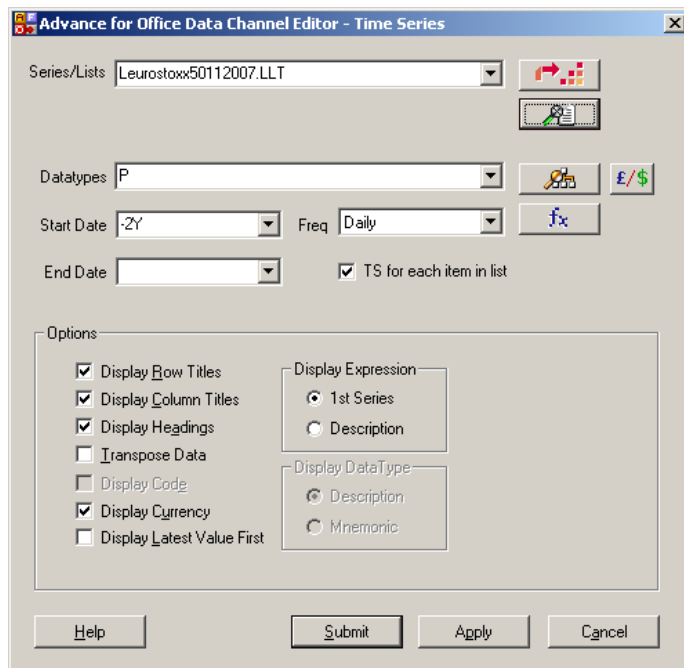
### Making a list

Making your own list is easily done using the "Tools" item under the Datastream-AFO menu. There you can choose the "Create List from Range" option. The following interface is shown:




In the Code Range input box you enter the Excel range containing the Datastream codes to be assembled in a list. As an illustration, let's recreate the Euro Stoxx 50 list using the codes we obtained from the previous step. These codes are contained in the range A3:A52. We then enter a list description, which should be sufficiently clear such that we still understand it later, e.g. Eurostoxx 50 November 2007. Then we enter a List File Name, e.g. Leurostoxx50112007. This is the name of the file where the list is stored. Finally, check the "Store List Locally" box. This means that the list information is stored on your computer. Alternatively, you can choose for "Upload List", in which case the list is stored on the Datastream mainframe. This implies that you can use the list from whichever computer you log on. In this case you may also provide a name in the List Name box, starting with L.

We can now use our own list in any request. Let's download the recent price history of the constituent stocks:



Note: do not forget to add the .LLT extension behind the list name: it indicates that the name refers to a user list. If you forgot the exact name of the list, you can al-

ways open the List Picker, by clicking on the  button. You will then get an overview of all the lists you have created.

Submitting the request returns the following data:

	A	B	C	D	E	F	G
1	Refresh	18/11/2005					
2	End	19/11/2007					
3	Frequency	D					
4	Name	AEGON	AIR LIQUI	ALCATEL	ALLIANZ	ARCELOR GEN	
5	Code	922956(P)	923295(P)	755022(P)	866033(P)	41012C(P)	9233
6	CURRENCY	E	E	E	E	E	E
7	18/11/2005	13.19	69.64	10.6	121.9	NA	23
8	21/11/2005	13.22	70.27	10.38	122.02	NA	24
9	22/11/2005	13.23	69.91	10.3	122.1	NA	23
10	23/11/2005	13.3	69.73	10.4	121.86	NA	23
11	24/11/2005	13.22	69.68	10.32	122	NA	23
12	25/11/2005	13.17	69.55	10.37	122.63	NA	23
13	26/11/2005	13.27	69.23	10.45	123.15	NA	23

Note the "NA" entries in column F, which indicates that the prices for Arcelor are not available over this period. You can change this with the "Options..." item in the "Datastream-AFO" menu. For instance, you can change this into "=NA()", such that Excel recognizes this as not available.

In this illustration we have used the Datastream codes to build the list. We can also use the ISIN codes (or SEDOL codes). This is especially convenient as this information is often provided by a third party (e.g. an index provider).

#### 4. More tools


##### Exchange rate conversions

Most series are denoted in their original currency. But sometimes it is useful to convert them into a common currency, e.g. when you want to compare the investment performance of international investments. This can be specified in the request interface. As an illustration, we will download the price of gold in euro. Normally,

this is denoted in USD per troy ounce. Its Datastream mnemonic is GOLDBLN. To convert this into another currency, we enter the following expression in the “Series/List” input box: (GOLDBLN)~E0. So note (1) the brackets around the mnemonic, (2) the tilde, and (3) the currency code (here: E0 for euro<sup>1</sup>). The result is:

	A	B	C	D	E	F
1	Refresh	18/11/2005				
2	End	19/11/2007				
3	Frequency	D				
4	Name	Gold Bullion U\$/Troy Ounce (~E0)				
5	Code	(GOLDBLN)~E0				
6	CURRENCY	E				
7	18/11/2005	416.28				

This only works when the datatype field is not filled in, i.e. for the default datatype. If you need other datatypes or several datatypes, the exchange rate conversion is done in the datatype input box. Let's download the price index (PI) and return index (RI) for Microsoft (@MSF) in euro. We simply enter the requested information as follows:

Then we click the currency converter button  where we select the euro:

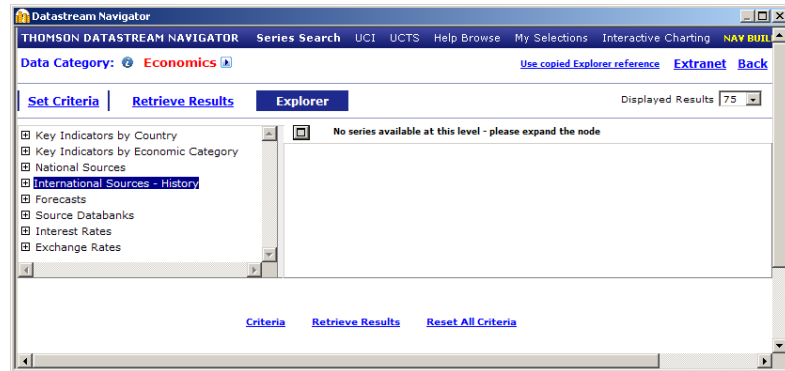
<sup>1</sup> Alternatively, also “E” can be used. In the next example we will show how you can find the currency mnemonics.

Then we click the “Apply All” button to convert *all* selected datatypes – use “Apply Last/Selected” for converting selected datatypes. The input screen now shows

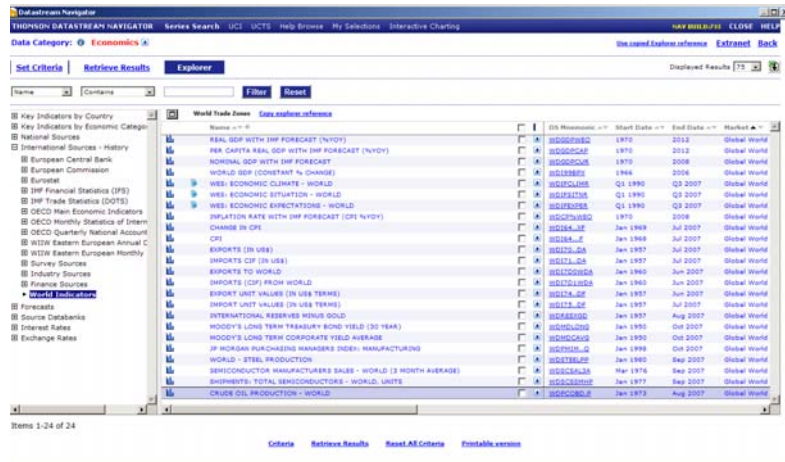
If you want to enter these conversions manually, remember to add an “X”, which is a dummy for the series code, to enter the datatype between brackets, and then the tilde (~) followed by the currency mnemonic.

### Using the Datastream Explorer

In the Datastream Navigator series can easily be found by using one or more criteria. However, it offers an alternative way to look for series, the *Explorer*. After clicking on “Explorer” the following interface shows (we selected the Economics data category):



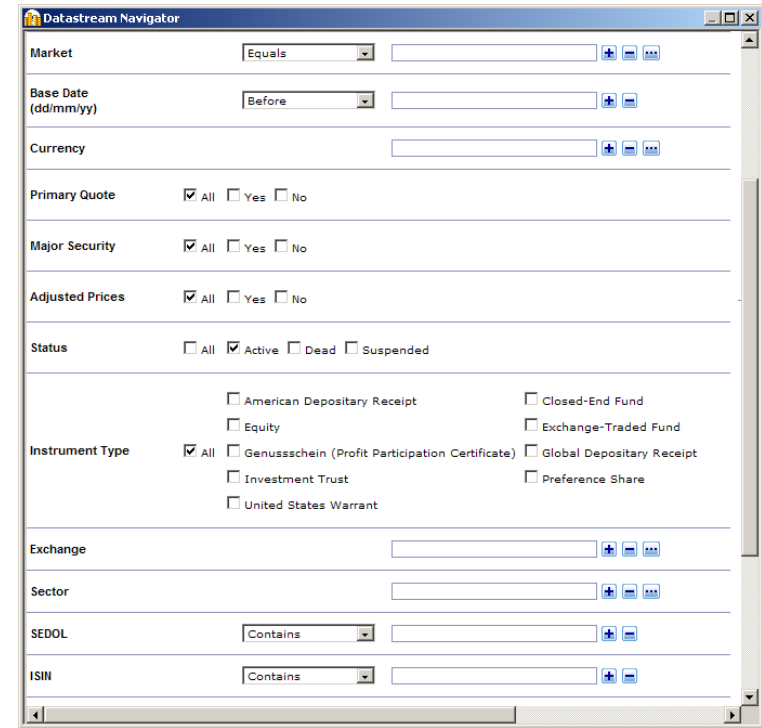
It presents different ways to navigate through the database. The categories can be expanded by clicking on the plus signs on the left of the screen, e.g.:



As always, in the right-hand side panel you will find more information about the series (Datastream code, source, start date, currency, ...). You can sort the series on each of these fields to retrieve information more efficiently using the arrow signs next to the column headings. In addition, you can restrict the selection of series by using the filter on top of the screen.

### More search criteria

Going back to the Navigator screen, you see more search criteria than the few we have used so far:

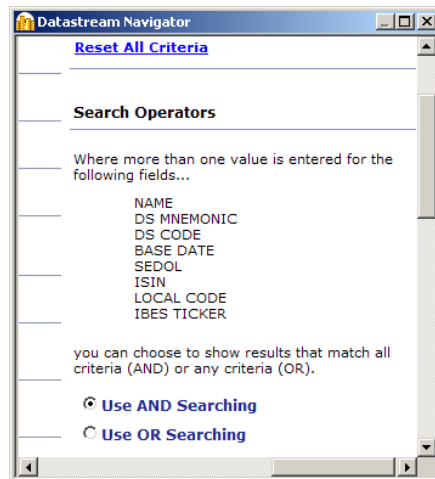


- You see that you can also select on Market or Exchange. Here you have to enter the Datastream code for market of exchange. To get an overview of these codes, simply click the "..." button, which opens the following screen for "Market":





- Another selection criterion is “Primary quote”, which allows you to restrict your selection to e.g. only the primary quotes (and not the cross-listings).
- Very useful is also the criterion “Status”. By default this is set to “Active”, meaning that only securities that are still listed will be returned. Sometimes it is necessary to have also access to securities (or series in general) that have delisted or are suspended.
- To end this short discussion, notice that you can combine criteria using both “AND” and “OR” Boolean operators. This is set using the buttons at the right of the Navigator screen:



It is often useful that the Navigator remembers your last search criteria. Of course, when you want to start with an entirely new request, it is recommendable to click the “Reset All Criteria” option to be sure that no previous search criterion is still active.

### More help...

This note only provides the most basic skills needed to work with Datastream. Many more facilities are available, such as making charts, making reports, set up batch requests (which allows you to download overnight, for instance), etc. Additional help and manuals are available on the Thomson Datastream Extranet, <http://extranet.datastream.com/Index.htm>, available through the Web browser item on the Datastream-AFO menu. In any case, in the end there is only one way to learn the necessary Datastream tricks: practice a lot! Good luck!

## 5. More on equities datatypes

Datastream provides several datatypes for equities. Many of them are very similar and can cause confusion. Here, we want to illustrate how the user can best capture the return evolution of a particular stock. We downloaded the following information for Deutsche Börse (Datastream mnemonic D:DB1):

Datatype:	UP	P	PI	RI	AF	AX	NOSH	DT	XDDE	DD	UDD
11/05/2007	168.67	84.33	503.5	540.61	0.5	#N/A	100000	#N/A	#N/A	#N/A	#N/A
14/05/2007	167.21	83.6	499.1	546.83	0.5	#N/A	100000	YR	14/05/2007	1.7	3.4
06/06/2007	167	83.5	498.5	546.14	0.5	#N/A	100000	#N/A	#N/A	#N/A	#N/A
07/06/2007	164.9	82.45	492.2	539.28	0.5	#N/A	100000	#N/A	#N/A	#N/A	#N/A
08/06/2007	166.75	83.37	497.8	545.33	0.5	#N/A	100000	#N/A	#N/A	#N/A	#N/A
11/06/2007	84.4	84.4	503.9	552.03	1	0.5	200000	#N/A	#N/A	#N/A	#N/A
12/06/2007	83.04	83.04	495.8	543.13	1	#N/A	200000	#N/A	#N/A	#N/A	#N/A

The datatype UP is the **Unadjusted Price**, i.e. the closing price as it was historically determined on the stock exchange. In contrast, P is the **Adjusted Price** and is the default datatype (if you don't specify the datatype, Datastream will return P). The most recent price will also reflect the price obtained on the stock exchange – see for instance the data on 12 June 2007: there is no difference between P and UP. However, the stored historical price data are occasionally recomputed by Datastream to take into account capital operations. As an example, compare P and UP on 8 June 2007. UP is exactly twice P. This is because Deutsche Börse underwent a 2 for 1 stock split: every shareholder received an additional share for each share she owned. This can be checked using the NOSH datatype, the number of shares in thousands. We notice an increase from 100 million to 200 million shares. Of course, such a stock split doesn't change anything for the company, so the share price would drop by 50%. This is approximately what we see in the UP column: a drop from € 166.75 to €84.40. In order to make historical data comparable, Datastream adjusts the previous prices, in this case by dividing UP by 2. This adjustment is reflected in the **Adjustment Factor** (AF). It indicates by how much you have to multiply UP to obtain P. Note that each time there is a capital operation, both AF and P will be recalculated. The datatype AX is the **Nonaccumulated Adjustment Factor**. It indicates at *which date* the adjustment has been made, and by how much. It is non available at dates without adjustment.

The stock return can now be computed in two equivalent ways (assuming no dividends have been paid out):

$$R_t = \frac{UP_{t+1} \cdot AF_{t+1}}{UP_t \cdot AF_t} - 1 = \frac{P_{t+1}}{P_t} - 1.$$

An alternative way to compute returns is by using the **Price Index** (PI) or the **Return Index** (RI). Both are set equal at 100 at the equity's base date, i.e. the first date for which price data are available in Datastream. The price index grows at the capital growth rate, i.e. the return net of dividends. It is also called the **capital appreciation index**. In contrast, the return index, or total return index, grows at the total return (inclusive of dividends) rate. As no dividends have been paid out in June 2007, both series grow by the same rate.

To see the effect of dividend payments, see the data on 14 May 2007. In the column DT (Dividend Type) you see the entry (YR). This datatype has an entry whenever dividends are paid out, in which case it indicates which kind of dividend is concerned. Here, YR stands for yearly dividend. Other possibilities are:

CMB	Combination
CPG	Capital Gains
CPL	Long Term Capital Gains
CPS	Short Term Capital Gains
CPU	Undefined Capital Gains
FIN	Final
HYR	Half yearly
INT	Interim
MTH	Monthly
QTR	Quarterly
RST	Restricted dividend, not payable to all shareholders
SPL	Extraordinary payment
UND	Undefined
YR	Yearly.

The next column, XDDE, is the **ex-dividend date of the dividend**, i.e. the day on which the shares are traded without the right on this dividend payment. This is the day on which the share price will drop by the dividend. The actual dividend paid is indicated by the UDD (**Unadjusted Dividend**) datatype. Here, €3.4 per share was paid out. The datatype DD adjusts the dividend payment, to make it comparable to the adjusted price (P). So, the same adjustment factor (AF) is used to convert UDD into DD. Note that DD will be recomputed any time the AF changes. The total return can be computed using any of the following expressions:

$$R_t = \frac{(UP_{t+1} + UDD_{t+1}) \cdot AF_{t+1}}{UP_t \cdot AF_t} - 1 = \frac{P_{t+1} + DD_{t+1}}{P_t} - 1 = \frac{RI_{t+1}}{RI_t} - 1.$$

Note that on 14 May 2007 the growth rate of RI is higher than for PI as the former takes into account the dividend payment.

It is important to realize that the dividend payment does not necessarily occur at the ex-dividend day. Indeed, often the dividends are paid out days or even weeks later. In that case, DD and UDD will be entered on the dividend payment date, whereas the dividends are included in the return on the ex-dividend date.

## 6. References

Datastream is increasingly used as a source for empirical research in finance, more specifically for countries that are not covered by high-quality databases such as CRSP. This raises the issue of the quality of Datastream data. Ince & Porter (2006) warn researchers for potential flaws in the Datastream coverage. They compare the US coverage to that of CRSP over the period 1975 and 2002 and find that certainly in the earlier years a large number of equities are missing in Datastream. For instance in 1975 only 20% of common equity issues found in the CRSP database is also included in Datastream. The problem is more important for small caps (bottom 20% of stocks measured by market cap). In addition, Datastream suffers to some extent from survivorship bias. The authors suggest a number of screens that alleviate the problems somewhat. Some of these screens are easy to implement (e.g. multiplying the return index when rounding errors become important, or deleting returns higher than 300% when they are reversed next month), but other are much more labour-intensive. See the paper for more details.

Ince, O.S. & R.B. Porter, 2006, "Individual equity return data from Thomson Datastream: handle with care!," *Journal of Financial Research* 29(4), 463-479.