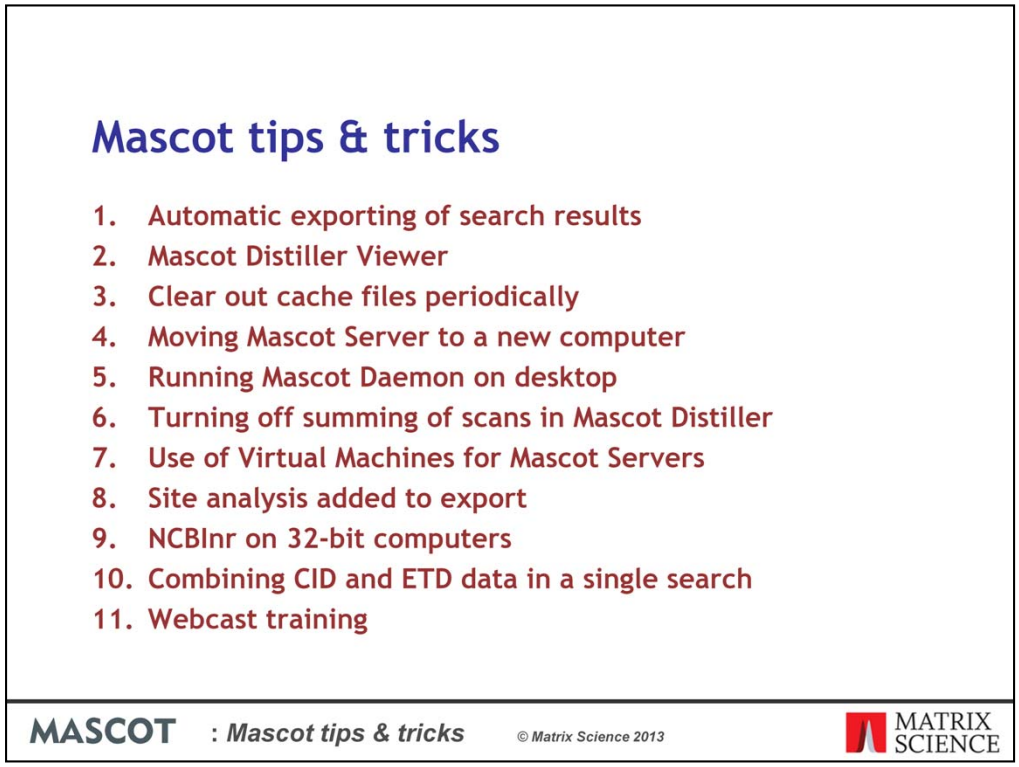
Mascot tips & tricks

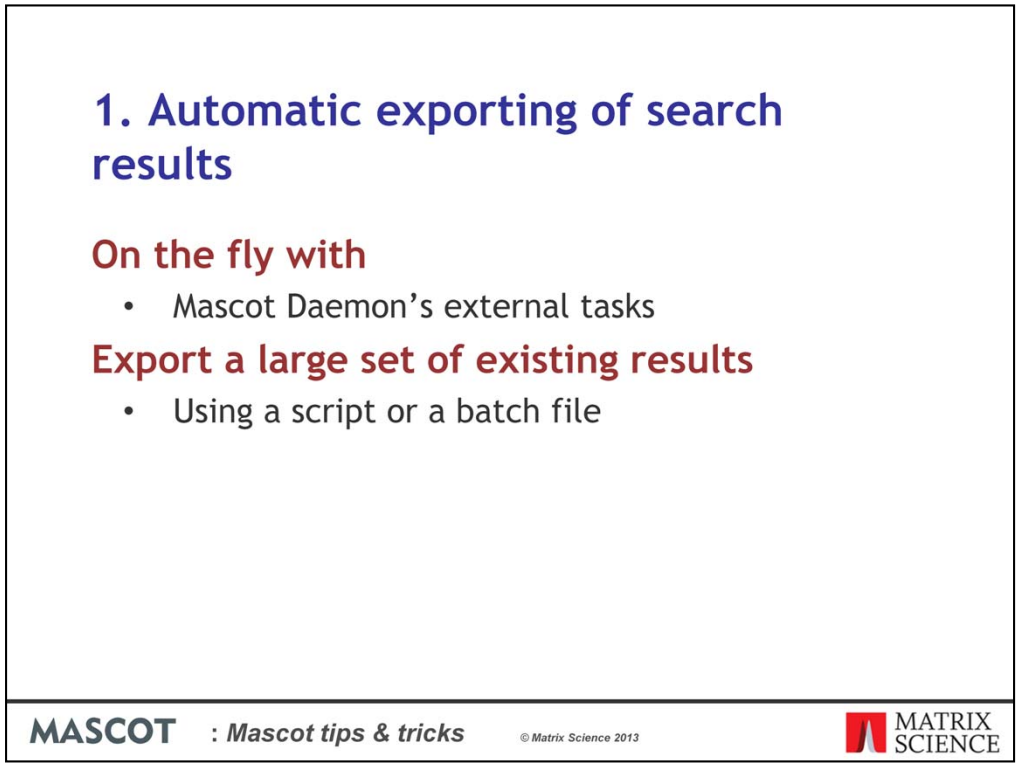
MASCOT

MATRIX SCIENCE



I am going to cover a mixture of 10 topics today. Some are questions that come up regularly in technical support conservations. There are a couple of new features that I'm going to introduce and, as a bonus feature, I am going to tell you about our online training course.

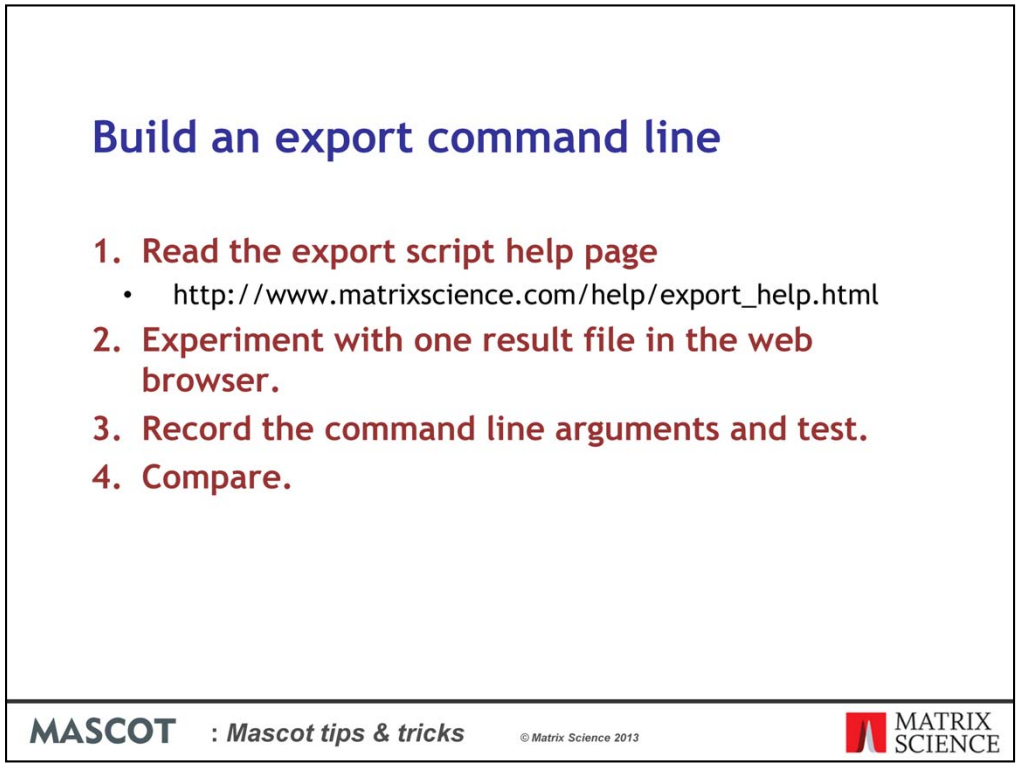
So lets get started with the first topic, Automatic exporting of search results



There are two main ways that you can automate the exporting of search results and I am going to cover two of them.

The first method is designed to export the results immediately after the search has been completed. Either on the computer where Mascot Daemon is installed or on the Mascot Server.

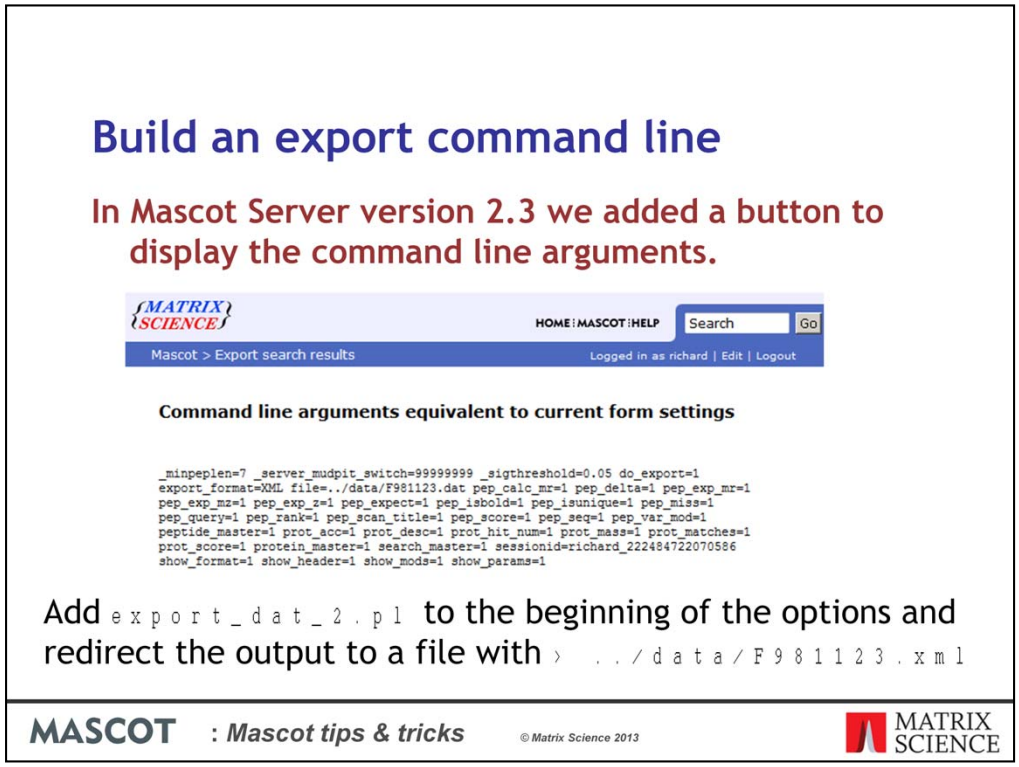
The second method is for exporting a batch of existing search results on the Mascot Server.



Before we can automate the exporting of results we need to determine which format we are going to export too and which fields we are going to export.

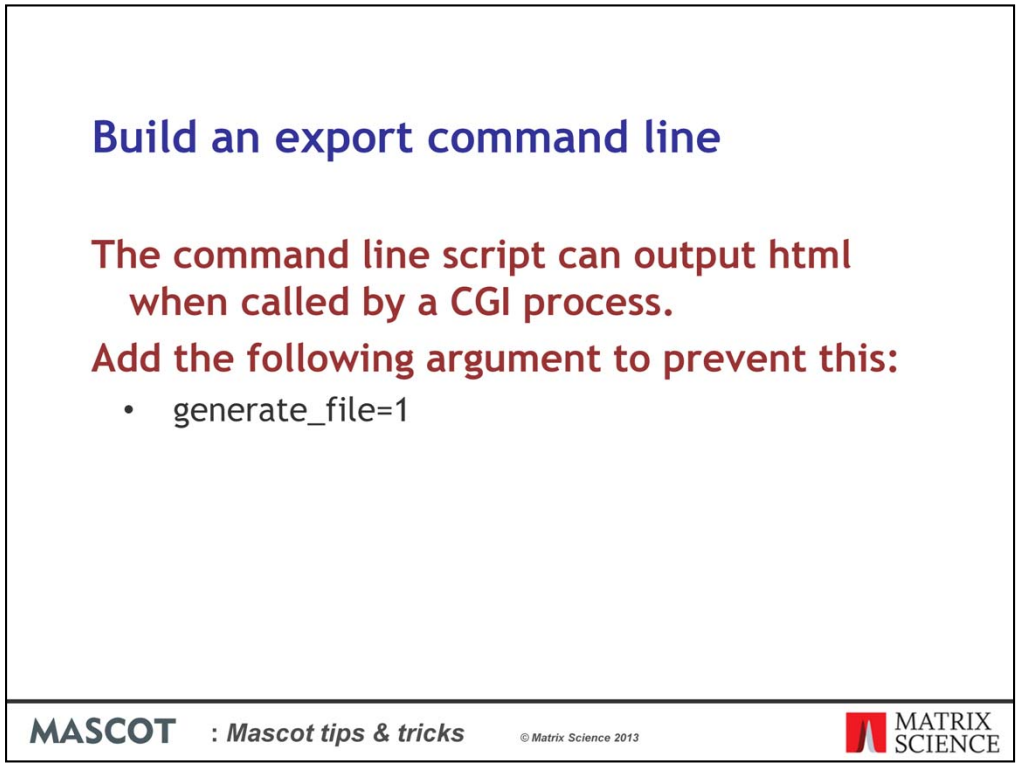
I recommend reading the export functions help page to get a better idea of how it works particularly the command line section. Then, experiment with a result file to make sure that all the information you need is included in the output. Record the command line arguments and try running the export on the command line. Compare the exported files from step 2 and 3 to make sure you have all the arguments set correctly.

If that works then you can take the command line arguments and use them in an automated export.



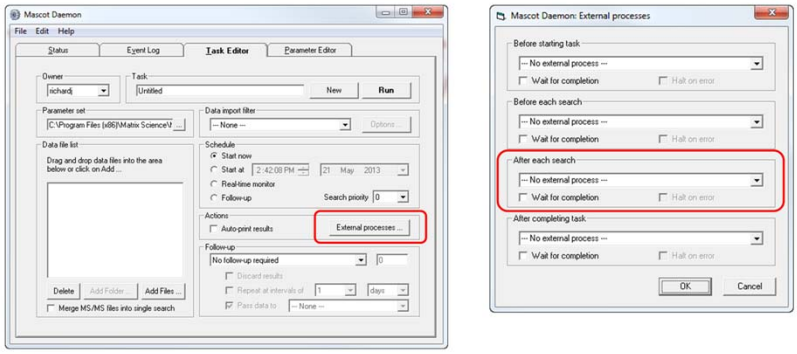
It is quite easy to make a mistake and miss out an important argument when you build the list of command line arguments by hand. Because of this we introduced the option to view the command line arguments for an export in Mascot Server 2.3.

To test these arguments add `export_dat_2.pl` to the beginning of the options and redirect the output to a file with the greater than sign



Here is a top tip: The script should not output the HTML for the interactive download button when executed, but this can happen if it is called from a CGI process. Add the generate_file=1 argument to prevent this

Automatic exporting through Mascot Daemon's external tasks



You will need ActivePerl installed on the Mascot Daemon computer.

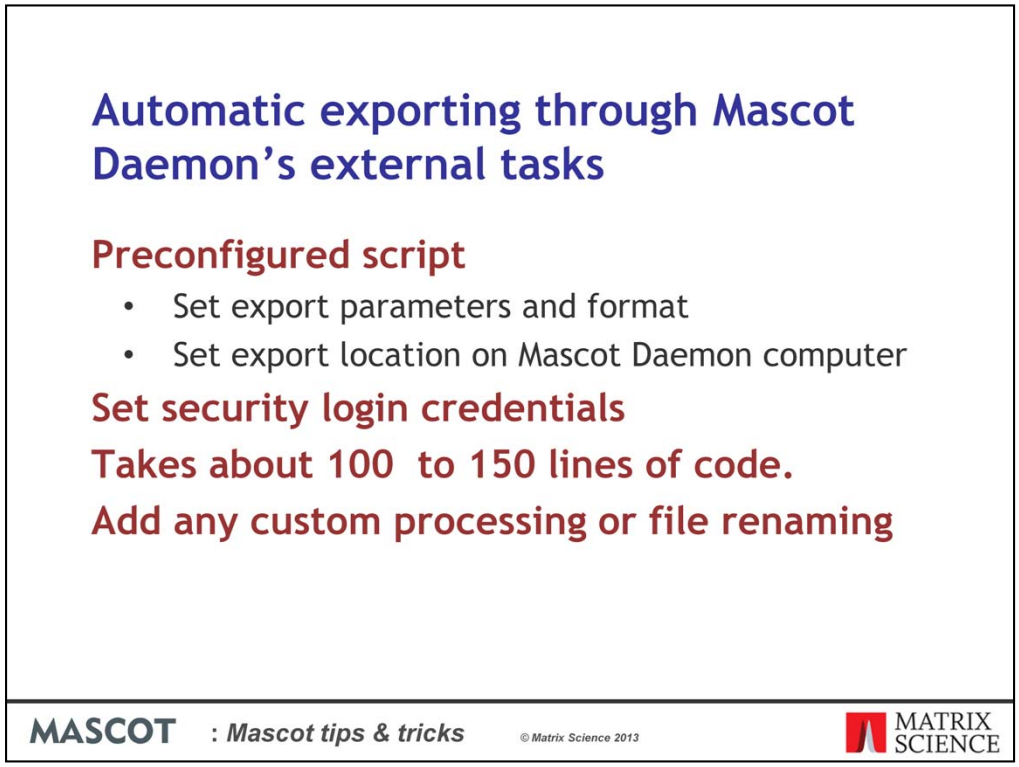
© Matrix Science 2013

Let's look at the first method to automate the exports, from the Mascot Daemon external processes.

Mascot Daemon can run an external script or program at a number of different times while processing a task. The most useful time is after each search.

Click on the external processes button and press F1 while the dialog is displayed to find out more information about the external processes feature.

I normally use the Perl programing language for small script and programing tasks but you can use any programing language as long as it accepts command line arguments. If you are going to use Perl then you will need to install ActivePerl from ActiveState on the Mascot Daemon computer.



The script or program will contact the Mascot Server through the web interface and it has to be preconfigured with the export arguments and format that you have previously worked out along with a location of where to save the resulting files.

When it runs it needs to know how to handle any Mascot Server or website security that is in place,

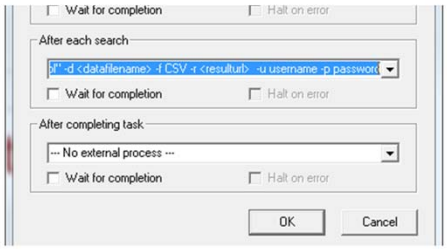
All that can be done in about 100 or so lines of code. We have an example Perl script that you can use as a starting point.

You can also add additional features to the script or program that automate other tasks you normally perform on results such as file renaming and extracting pieces of information for reports.

Automatic exporting through Mascot Daemon's external tasks

Configure Mascot Daemon to call the export script

- perl "C:/Temp/DaemonCGIExport.pl" -d <datafilename> -f CSV -r <resulturl> -u username -p password



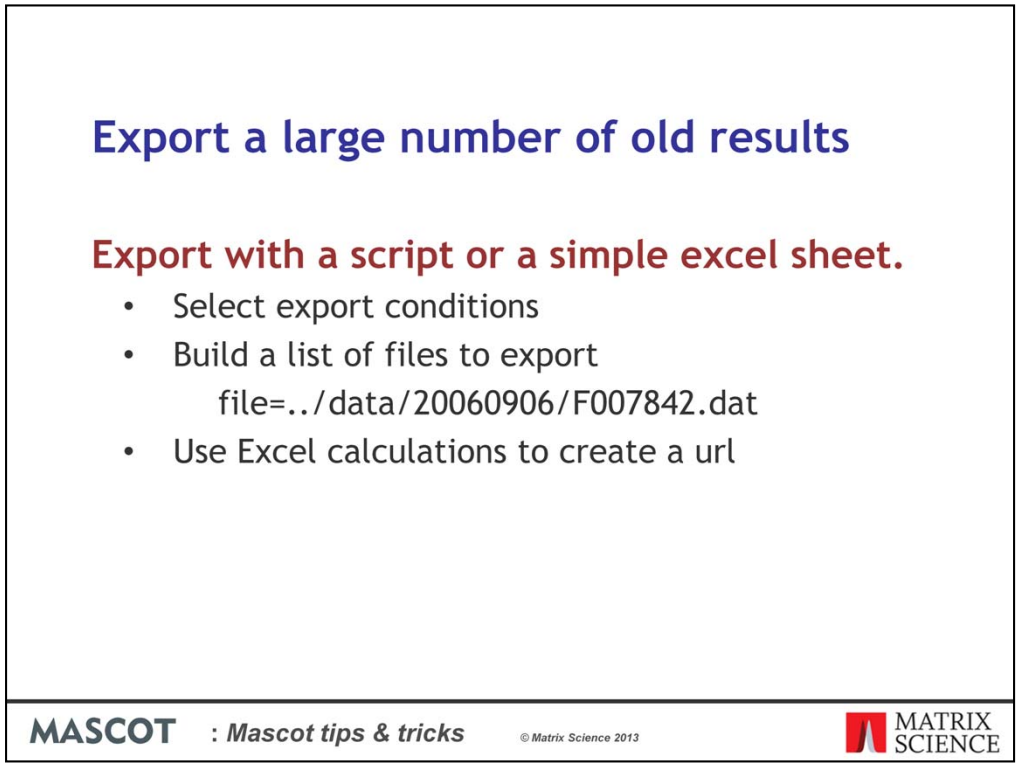
MASCOT : Mascot tips & tricks © Matrix Science 2013 MATRIX SCIENCE

Here I am going to show how to call this script. In the Mascot Daemon external processes dialog box, I fill in the After each search field. There is a character limit on the field so don't make the options too long.

Mascot Daemon supports the use of tokens that are replaced with the actual information at run time. Here I pass the input data file name and the result url to the script along with a user name and password for Mascot security.

When Mascot Daemon runs this command will be run for each search exporting the results to a location of my choosing.

Previous commands can be displayed in the drop down box by Daemon so that they can easily be selected and used again with a new Daemon task.

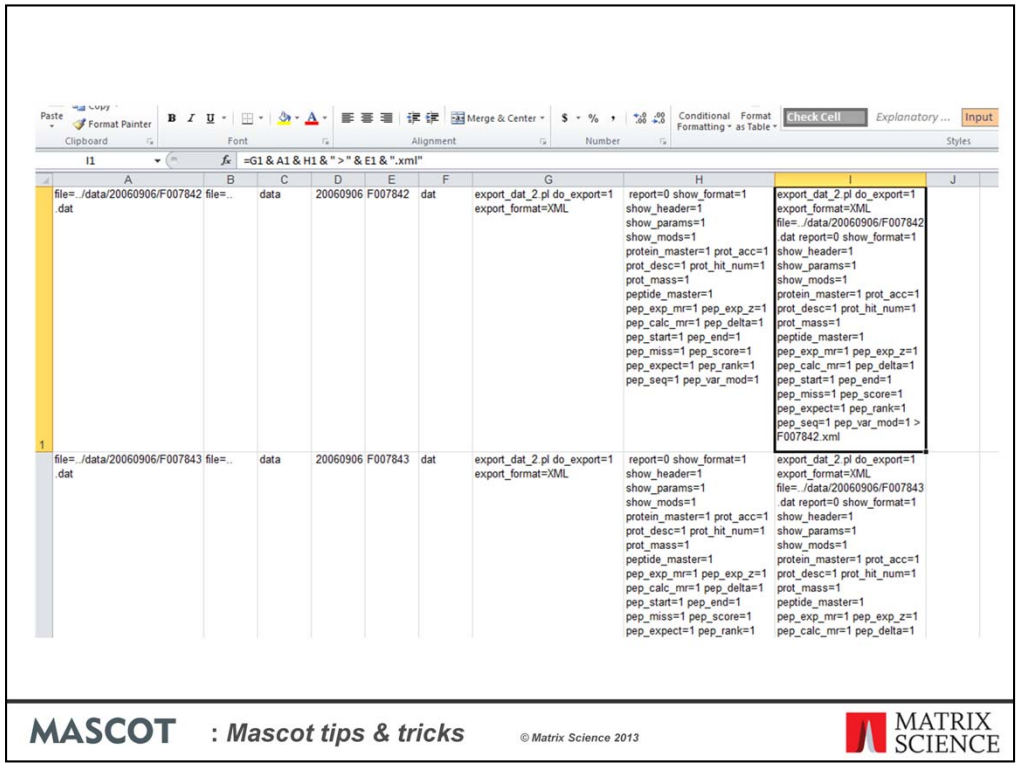


The second method of exporting search results is best suited to exporting a batch of results

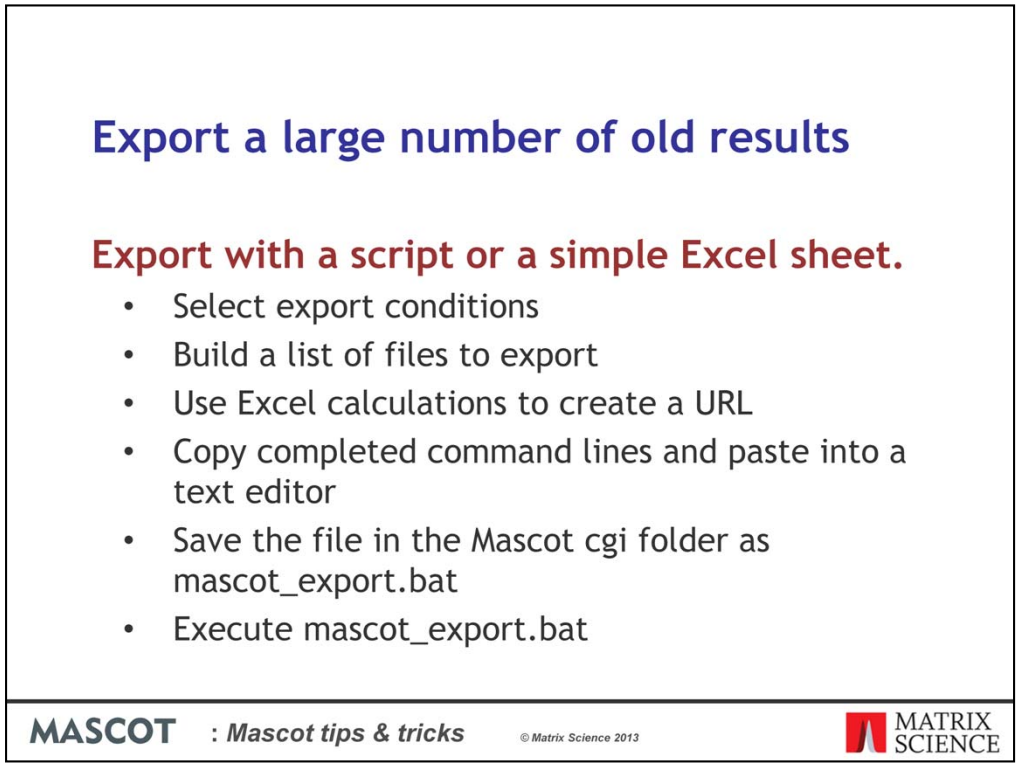
As before, select the command line arguments for the export. The next step is to build a list in Excel of all the search results we want to export. For example in Mascot Daemon, browse through the result files you want to export. For each one, click on the result link to load the result report in a browser. Copy the last bit of the URL from the browser into Excel, one URL on each row. For example

file=../data/20060906/F007842.dat

Paste this piece into Excel so that we can create a final URL for the export



I have broken up the result file path so that the filename can be used to construct the oputput path. I create the final command line by concatenating the arguments and file paths. Copy the command line column.

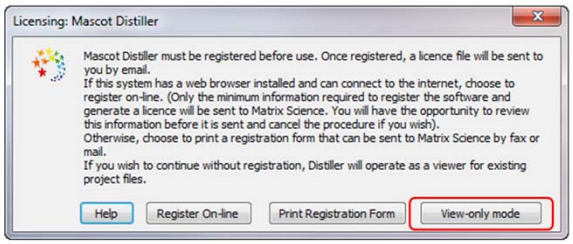


And past it into a text file. Save the file in the Mascot cgi folder as mascot_export.bat. Then, execute the batch file and each of the individual files will be exported in turn and saved to disk.

This is a great way to export hundreds of files from a project in one go.

2. Sharing results with Mascot Distiller viewer

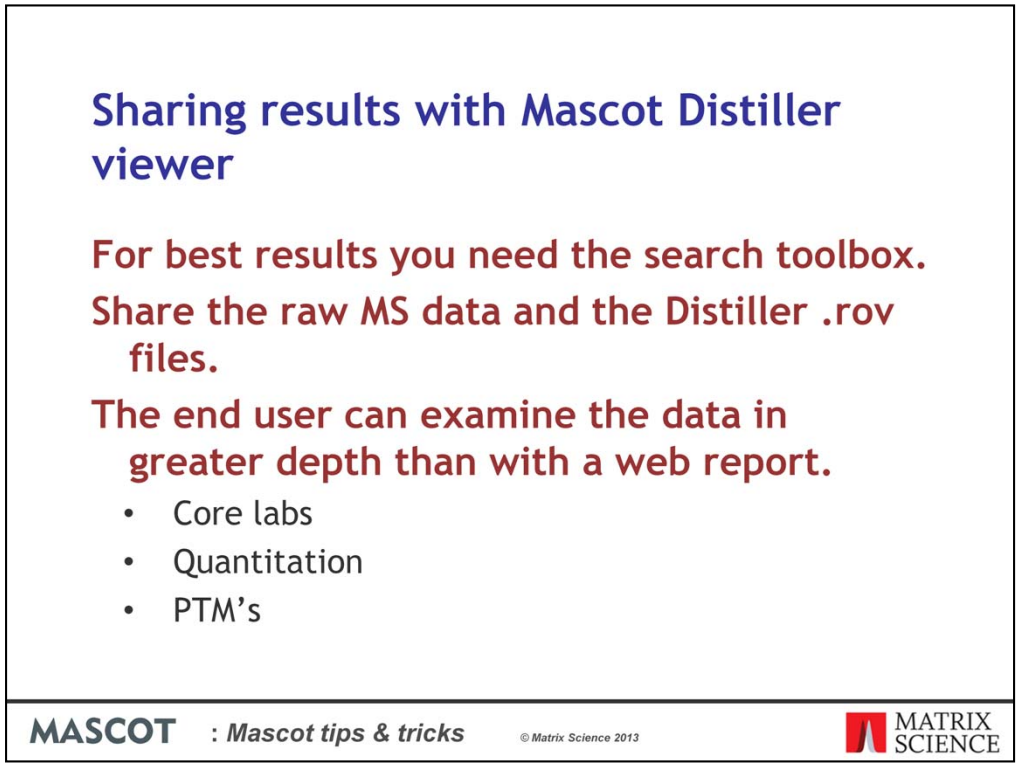
Mascot Distiller can be installed on any computer without a license and used as a viewer



© Matrix Science 2013

One often over looked feature of Mascot Distiller is the read only mode that allows collaborators to view search and quantitation results.

Mascot Distiller can be installed on any computer and, if it is not licensed, view only mode can be activated.

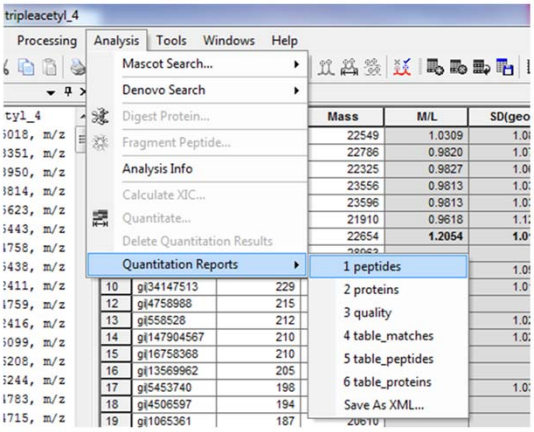


The licensed copy of Distiller used to create the project needs to include the search toolbox so that search results can be imported and saved.

You need to make both the raw MS data and the Mascot Distiller rov file available to the end user. Collaborators do not need access to the Mascot Server as everything is stored in the rov file.

With these files the end user can examine the data in greater depth than by looking at the web-based report. This is particularly useful for core labs distributing results involving quantitation or when you want to inspect hits that are evidence for PTM's

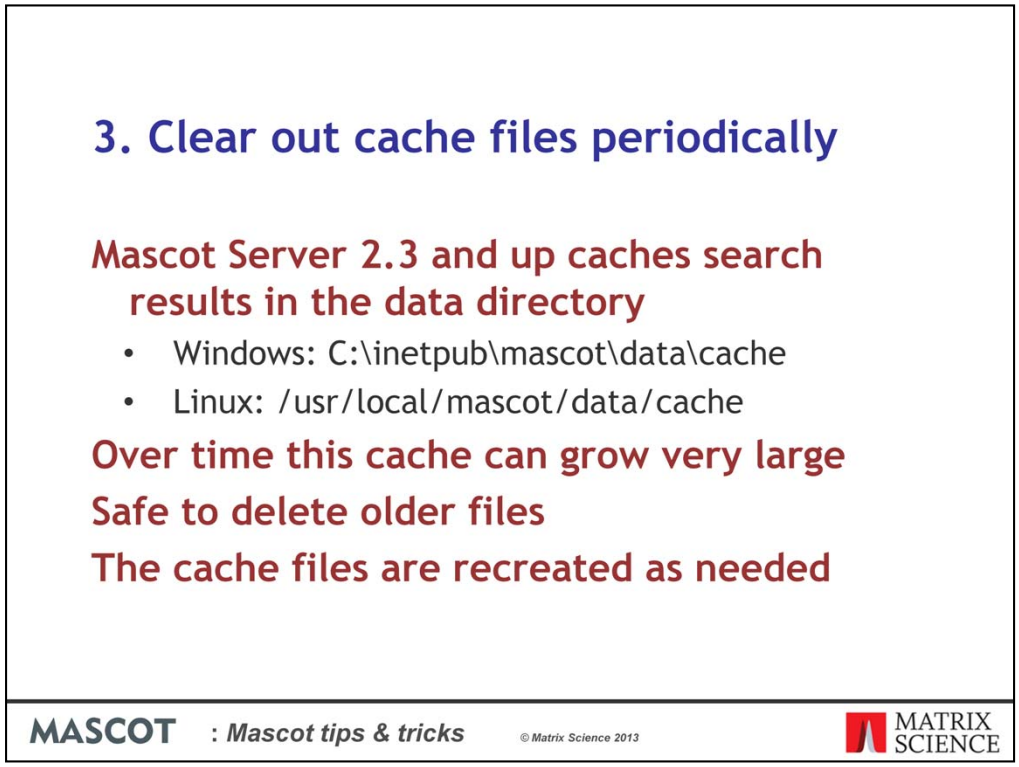
You can still save reports and copy text and images from the file but not change them.



© Matrix Science 2013

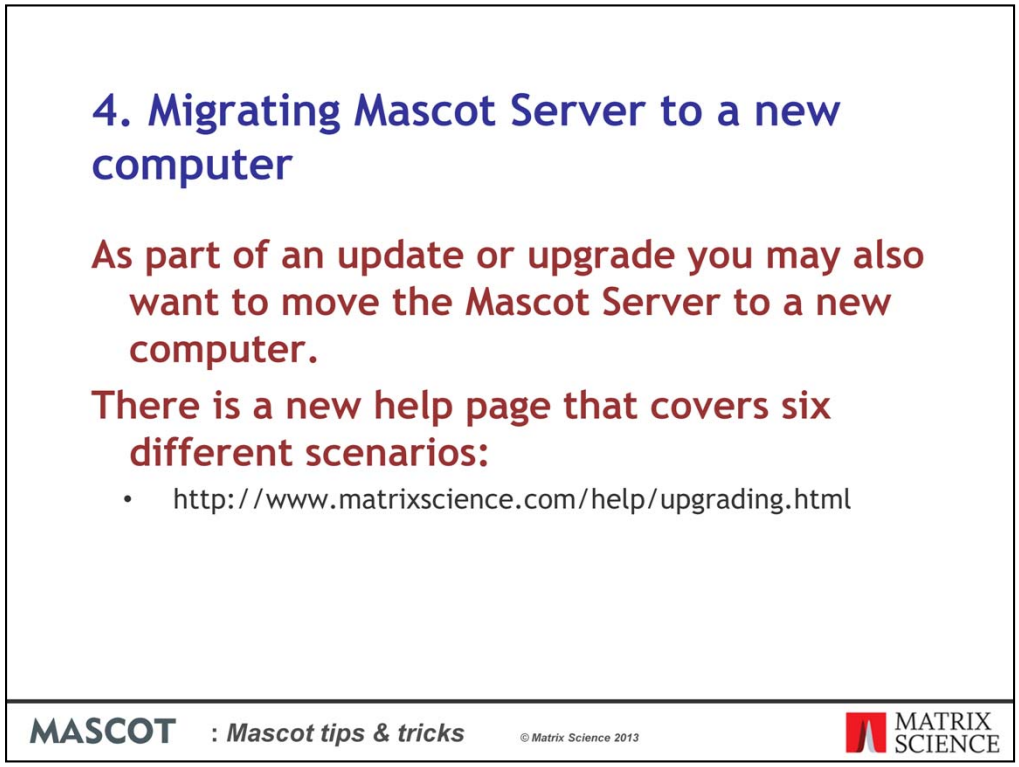
When a project is opened in the read only view mode features that can not be used are grayed out. If you were to look in the Mascot Search and de novo menus all the options would be grayed out and inactive.

You can export quantitation results, as shown here, but not change any of the processing parameters or format options.



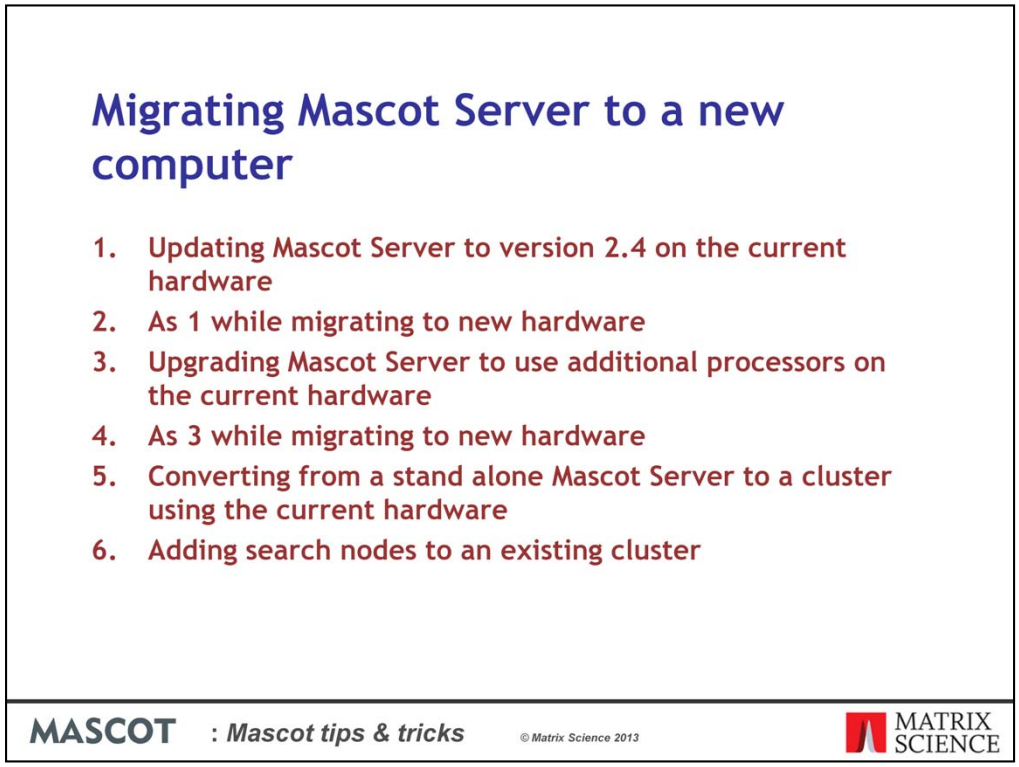
In Mascot Server 2.3 we adding a caching feature to Mascot Server. This makes it a lot quicker to view the search results once the cache has been created. Combined with the protein family report it allows you to view large result sets produced by today's MS instrumentation. However the cache files will grow with time and are a bout a third of the size of the raw results files.

It is safe to delete older cache directories to free up disk space. If you open one of the older reports at a later data the cache files will be recreated.

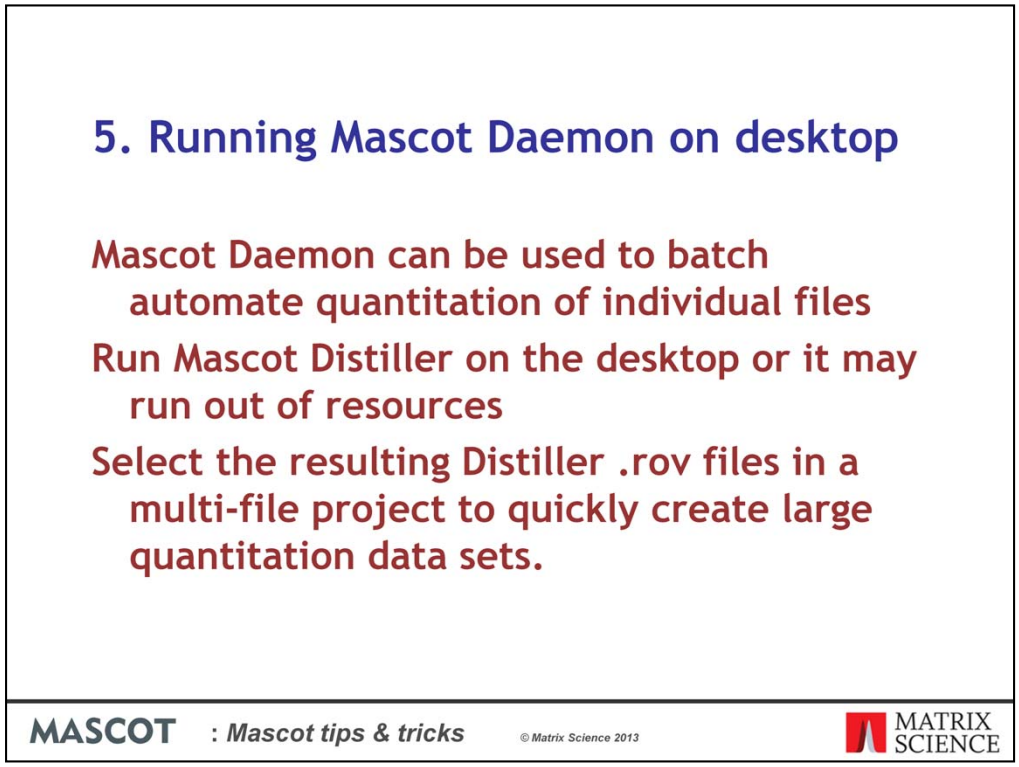


There is a new help page that covers six different scenarios: moving, updating, upgrading, or combinations of moving updating and upgrading

If you want to make a trial installation, do not register using your permanent Mascot 2.4 product key. Instead email Matrix Science to request a temporary license key



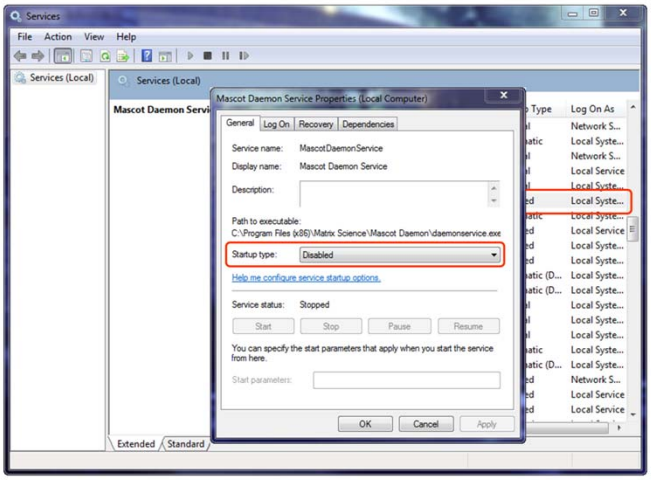
The different scenarios are as follows:



If you are using Daemon to batch automate Distiller quantitation, you need to run the Daemon service component on the Windows desktop, in the system tray. Otherwise, it may run out of resources. This is particularly important for datasets that used with Mascot Distiller multi-file quantitation.

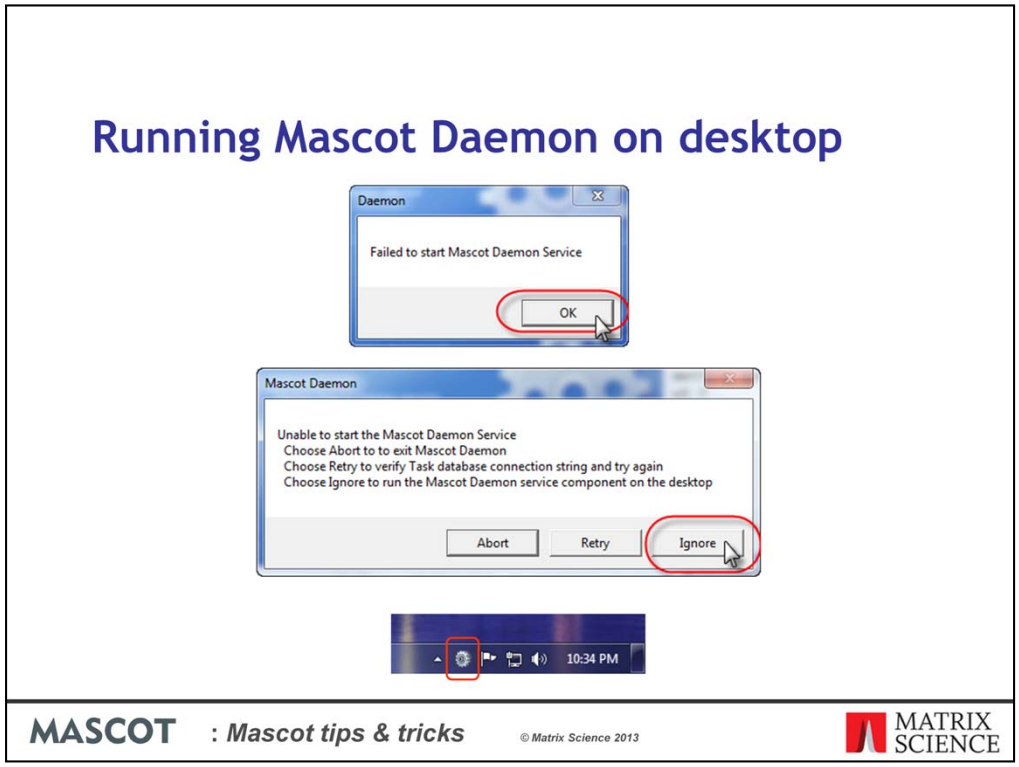
Batch processing individual files in Daemon is efficient because each project then contains the individual search and quantitation results. To create a multi-file project, select the individual project files in Distiller. This is much faster than creating a multi-file project from the raw files.

Running Mascot Daemon on desktop



MASCOT : *Mascot tips & tricks* © Matrix Science 2013 MATRIX SCIENCE

To change the settings for Mascot Distiller go to the Start menu->Settings->Control panels->Administrative tools->Services panel and double click the Mascot Daemon service. Set startup type to Disabled and save the changes.



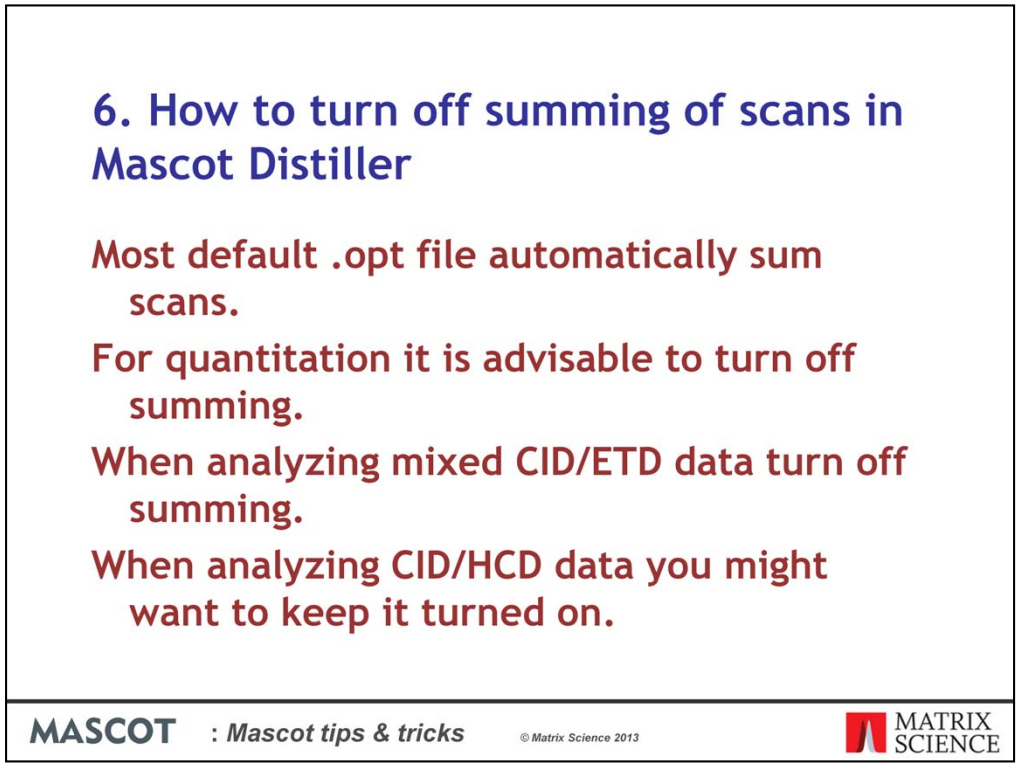
When you restart Mascot Daemon it will report that it has failed to start the Mascot Daemon service. Click Ok.

And then choose Ignore to run the Mascot Daemon service from the desktop. You will then see the Mascot Daemon cog icon in the taskbar notification area.

There are two points to note about the change:

Mascot Daemon Service is now running in your user account and will be able to access any network drives that you can normally access making it easy to search data from remote computers.

When you log out of the computer, Mascot Daemon will stop in the middle of what it is doing but will continue a task when restarted.



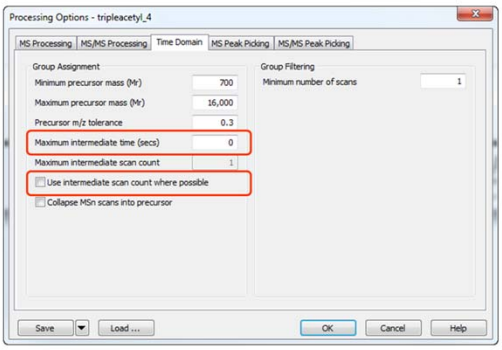
Most of the default processing options or .opt files that ship with Mascot Distiller have scan summing turned on. In some cases, this can improve signal to noise and result in higher scoring matches.

There are other cases where you will want to make sure scan summing is turned off. For example, when analyzing data sets containing alternate CID/ETD.

Alternatively, you might want to keep summing turned on for alternate CID and HCD data.

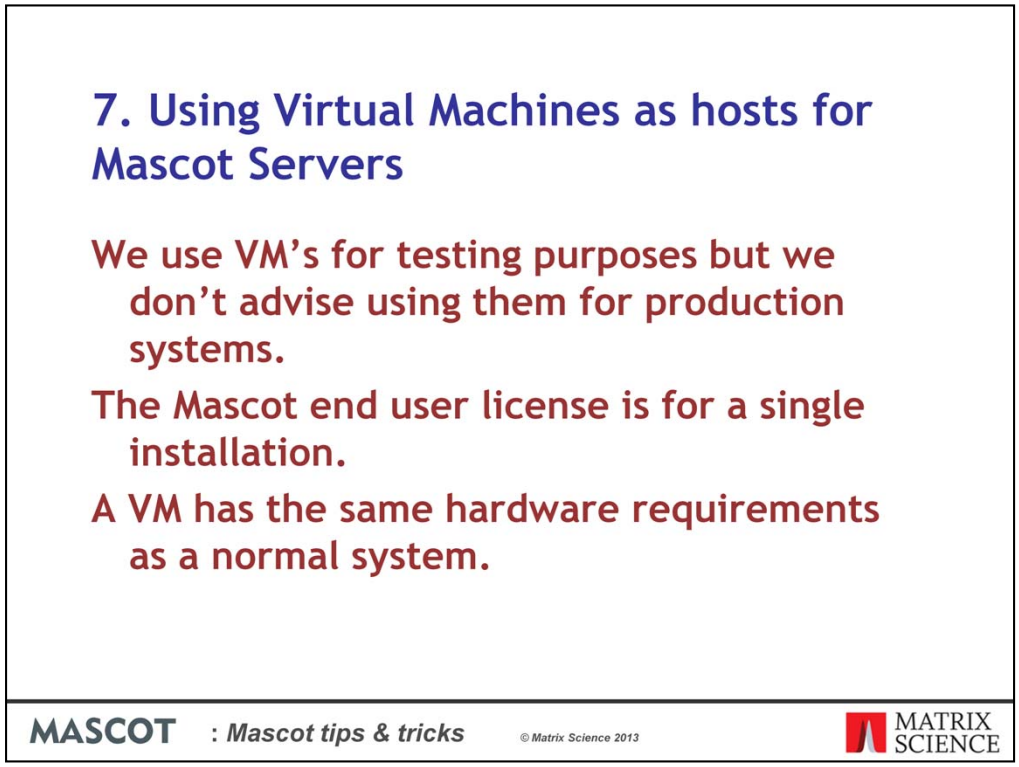
How to turn off summing of scans in Mascot Distiller

- Set the maximum intermediate time to 0
- Uncheck the "Use intermediate scan count where possible" check box



MASCOT : *Mascot tips & tricks* *© Matrix Science 2013* MATRIX SCIENCE

You can turn off summing in the Mascot Distiller processing options in the Time Domain tab by Setting Maximum intermediate time (sec) to zero and unchecking the Use intermediate scan count where possible box. Save the changes and use the options to process the data.

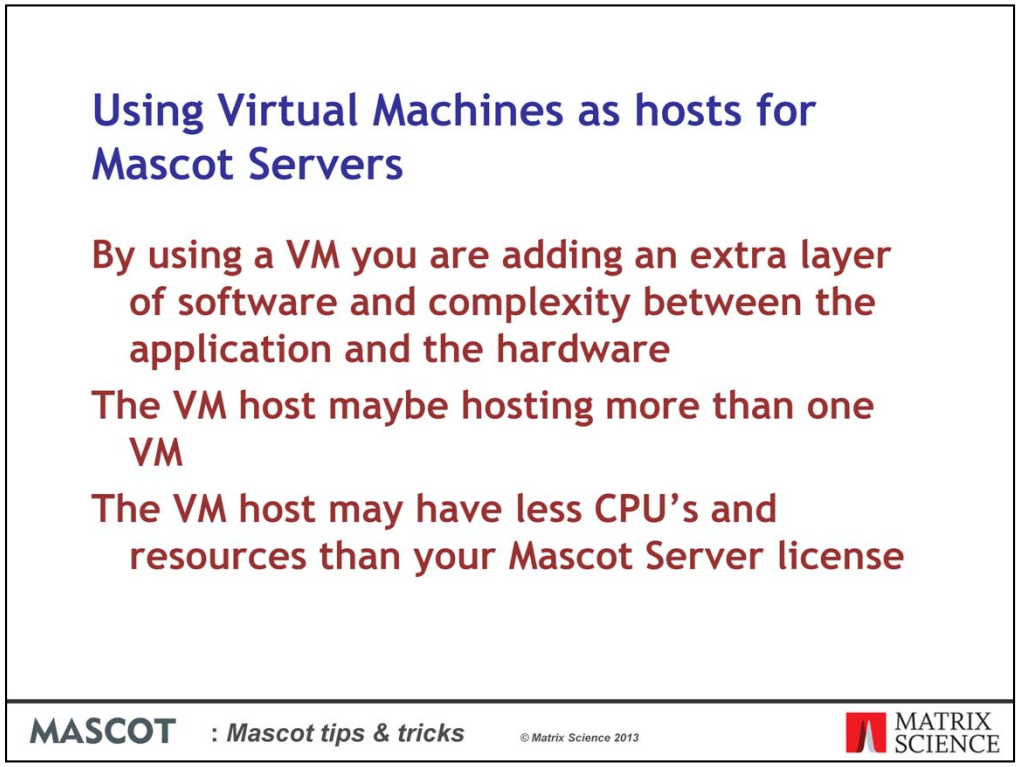


Questions about running Mascot in a virtual machine are coming up quite a lot in technical support.

At Matrix Science, we use VM's for testing purposes, particularly when testing a new version of the server on different operating systems. However we do not recommend there use for productions environments.

The Mascot Server end user license is for a single instance of the software, whether in a VM or in the host OS. If you use a VM you are only licensed to run one instance at a time. The Mascot Server software is locked to the underlying hardware so moving the VM to another host will mean that you need to reregister the license again.

One important thing to remember is that the Mascot Server VM will have the same requirements as a non-VM system. The Mascot Server web site has a help page on PC Hardware for Mascot Servers



One of the reasons why we don't recommend using VM's is because it adds an extra layer of software and complexity between the application and the hardware. This may degrade performance.

The VM host may be running multiple VM's all sharing the same resources. This means that you may be sharing the CPU time with other applications.

It is quite possible to configure a VM to have 2 quad core CPU's, which would be suitable for a 2 CPU Mascot Server license, but then run that image on a computer with fewer cores. This means that your Mascot Server looks like it has suitable hardware but the performance will be less than you might expect.

In general the Mascot Server software license is more expensive than the computer hardware up to about 8 CPU's and if you wish to obtain maximum bang for your buck dedicated hardware is the way to go.

7. Exporting Site analysis results

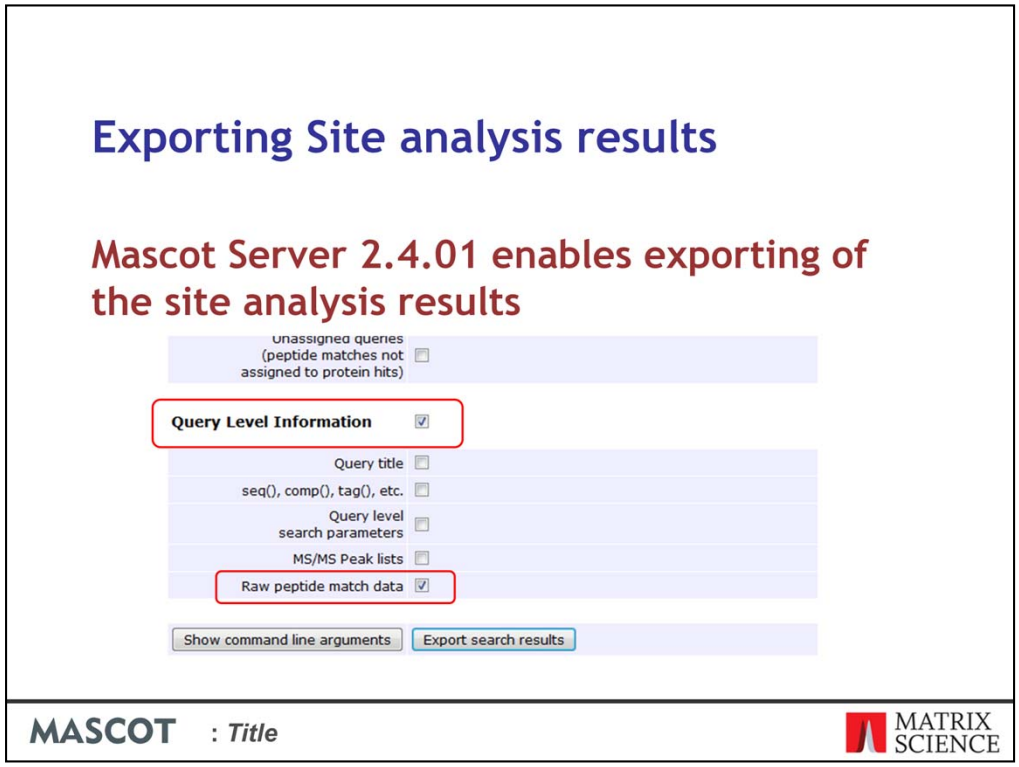
Introduced with Mascot Server version 2.4.0 in the peptide_view

Score	Mr(calc)	Delta	Sequence	Site Analysis
83.4	1846.7179	0.1889	DIGSESTEDQAMEDIK	Phospho S4 84.56%
75.8	1846.7179	0.1889	DIGSESTEDQAMEDIK	Phospho S6 14.73%
62.7	1846.7179	0.1889	DIGSESTEDQAMEDIK	Phospho T7 0.72%
26.9	1846.7808	0.1261	KLNSNPENYCESELK	
22.8	1846.7729	0.1339	KMEDSVGCLETAEEVK	
15.5	1846.9230	-0.0161	GAYTIEQHPVLGLEIK	
14.2	1846.7729	0.1339	KMEDSVGCLETAEEVK	
13.9	1846.8754	0.0315	YVKGIYENLPSIDEK	
13.8	1846.8866	0.0202	QLIEAPDPVPSFEVAR	
13.3	1846.9052	0.0016	KIDFSNIAMLFGGVQK	

MASCOT : *Title* MATRIX SCIENCE

Site analysis was introduced in Mascot Server 2.4 in the Peptide View report and is based on the score differences between matches with different arrangements of modifications as shown here in this example.

Originally, this information could not be exported, but this has been fixed in Mascot 2.4.1.



To include site analysis, the export needs to include the top ten matches for each query. We have to include the following arguments:

- Query level information
- Raw peptide match data

Exporting Site analysis results

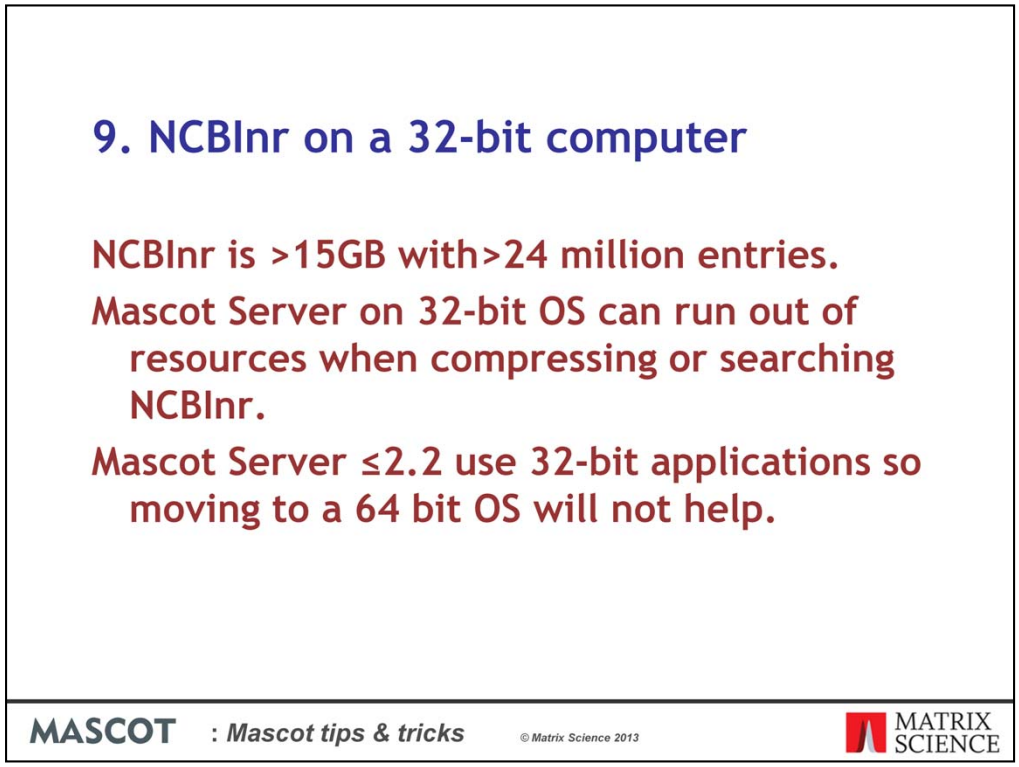
Export site analysis results to XML, CSV and mzIdentML

CSV in the pep_var_mod_conf column in the last section of the report

pep_score	pep_seq	pep_var_mod	pep_var_mod_pos	pep_var_mod_conf	pep_scan_title
67.78	DGARPDVTESESGSPEYR	Methyl (DE) [+14.02]	0.0000000000X0000000.0	22.20%	29: Sum of 8 sca
66.1	DGARPDVTESESGSPEYR	Methyl (S) [+14.02]	0.000000000X00000000.0	15.08%	29: Sum of 8 sca
66.1	DGARPDVTESESGSPEYR	Ser->Thr (S) [+14.02]	0.000000000X00000000.0	15.08%	29: Sum of 8 sca
65.09	DGARPDVTESESGSPEYR	Methyl (S) [+14.02]	0.00000000000X000000.0	11.95%	29: Sum of 8 sca
65.09	DGARPDVTESESGSPEYR	Ser->Thr (S) [+14.02]	0.00000000000X000000.0	11.95%	29: Sum of 8 sca
64.2	DGARPDVTESESGSPEYR	Methyl (DE) [+14.02]	0.000000000X000000000.0	9.73%	29: Sum of 8 sca
64.15	DGARPDVTESESGSPEYR	Methyl (T) [+14.02]	0.0000000X0000000000.0	9.62%	29: Sum of 8 sca
64.15	DGARPDVTESESGSPEYR	Thr->Asp (T) [+13.98]	0.0000000X0000000000.0		29: Sum of 8 sca
57.74	DGARPDVTESESGSPEYR	Methyl (S) [+14.02]	0.00000000000000X0000.0	2.20%	29: Sum of 8 sca
57.74	DGARPDVTESESGSPEYR	Ser->Thr (S) [+14.02]	0.00000000000000X0000.0	2.20%	29: Sum of 8 sca
11.31	GEMAYYVVSDGTNKAYR	Acetyl (N-term)	1.00000000000000000.0		29: Sum of 8 sca

MASCOT : *Mascot tips & tricks* © Matrix Science 2013 MATRIX SCIENCE

The site analysis information can be exported to the XML, CSV and mzIdentML export formats. In the CSV file the information is reported like this with the site confidence reported with the query level information at the end of the report in the pep_var_mod_conf column.

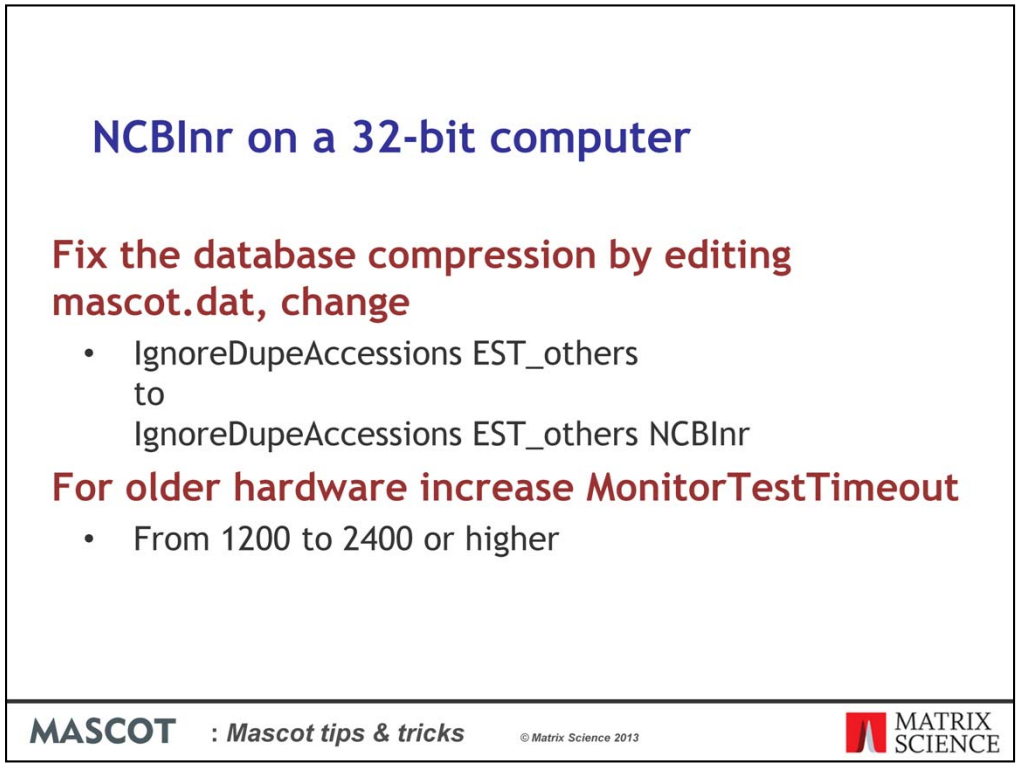


Another support question that is coming up quite a lot at the moment with users of older versions of Mascot Server is the use of NCBInr on 32-bit operating systems.

NCBInr has been doubling roughly every 18 months for quite a few years now and is 15GB in size and more than 24 million entries.

Mascot Servers on 32-bit operating systems can run out of resources when compressing or searching NCBInr. Older versions of Mascot Server, version 2.2 and earlier use 32 bit applications to compress and search so moving them to a 64 bit OS will not help much.

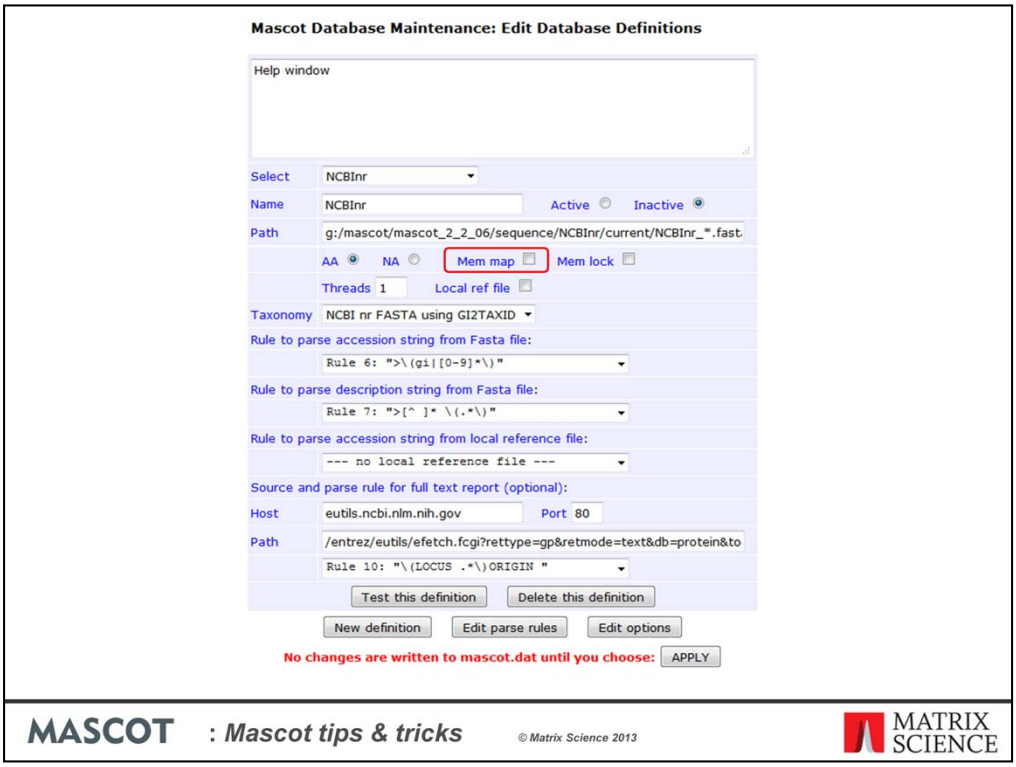
The problem normally presents itself after the database has been updated and the status does not change to “In use”



There are a couple of things we can do to enable Mascot to search NCBInr on these systems.

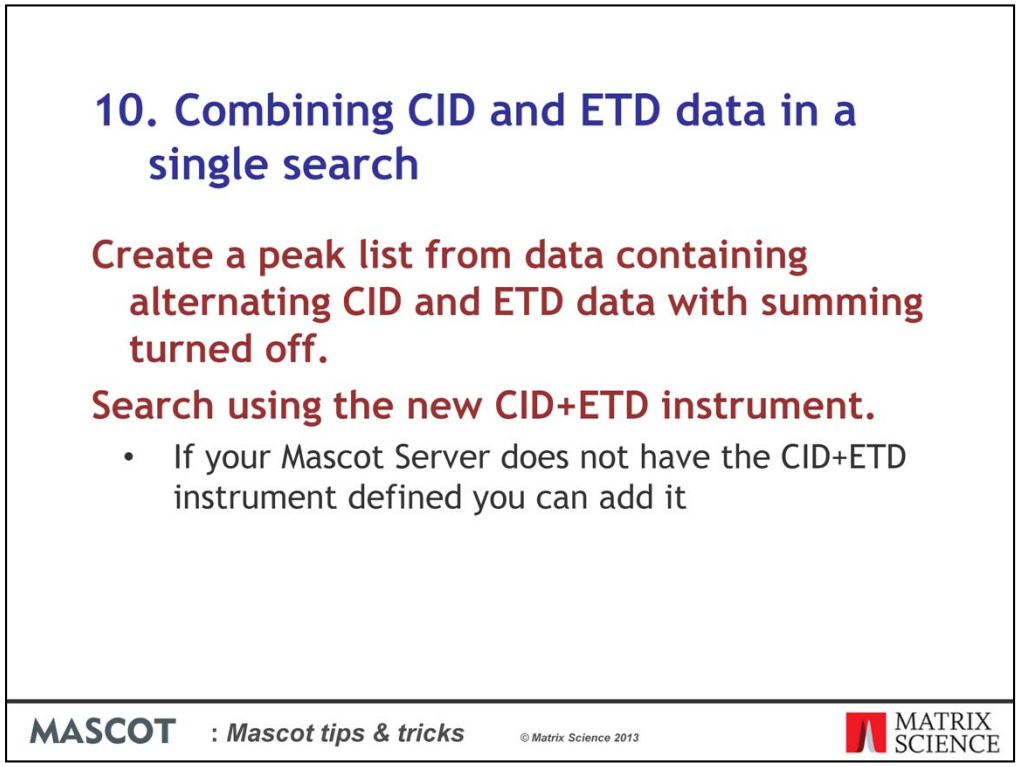
First is to edit the mascot.dat file and add NCBInr to the IgnoreDupeAccessions option. This means that, if there are duplicate accessions in the database, matches will be misreported. NCBInr has been very reliable, and we are not aware of this ever happening, but we do not recommend adding other very large databases to the IgnoreDupeAccessions directive unless you can verify by other means that there are no duplicate accessions.

For Mascot Server running on older computers the time taken to run the test search can be longer than the default MonitorTestTimeout option. Doubling or tripling this value in mascot.dat may be required.



After making those changes the database will sometimes halt with a “failed to memory map error”. Or, the database reaches “In Use” and you can start searches against it but they crash. If so, you will have to make this additional change: Go to the database definition for NCBInr and make sure that the “Mem lock” option is unchecked, which it is by default, and uncheck the “Mem map” option and apply the changes. This should enable you to search NCBInr but both searches and opening of the results files will be a slower than normal.

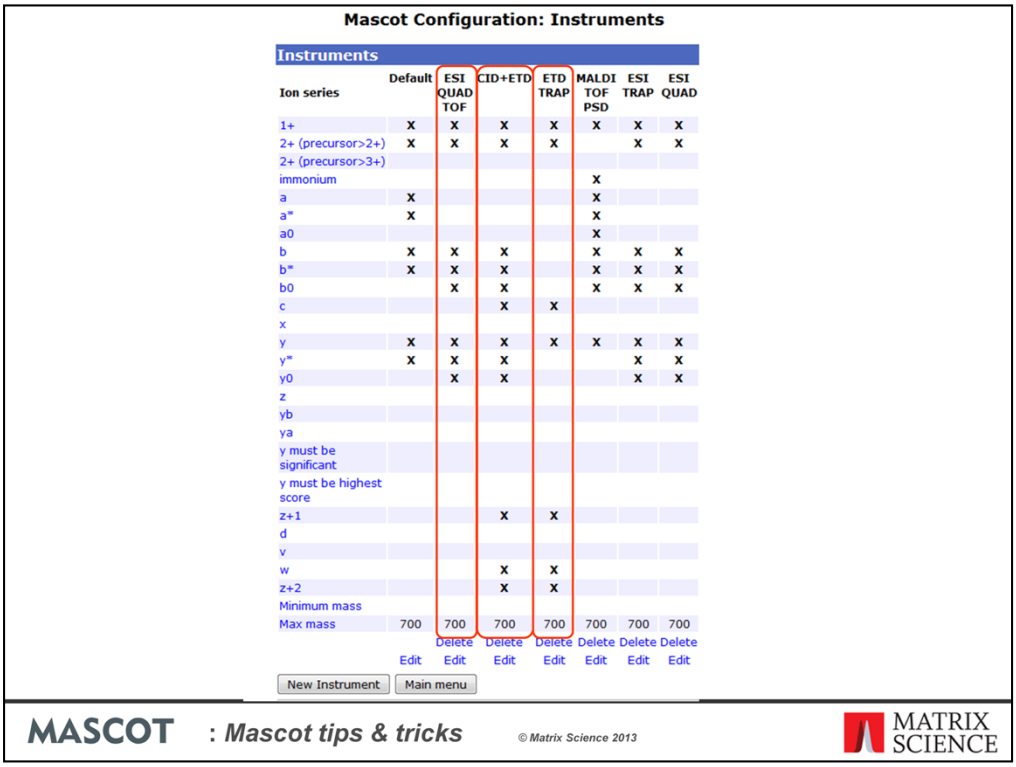
If you are using an older version of Mascot Server on older hardware it is probably a good time to consider updating both the software and hardware.



It is very easy to search data with mixed CID/ECD or ETD data. First make sure that scan summing is turned off. The results will be clearer to analyze that way. The peak list can contain a mixture of spectra.

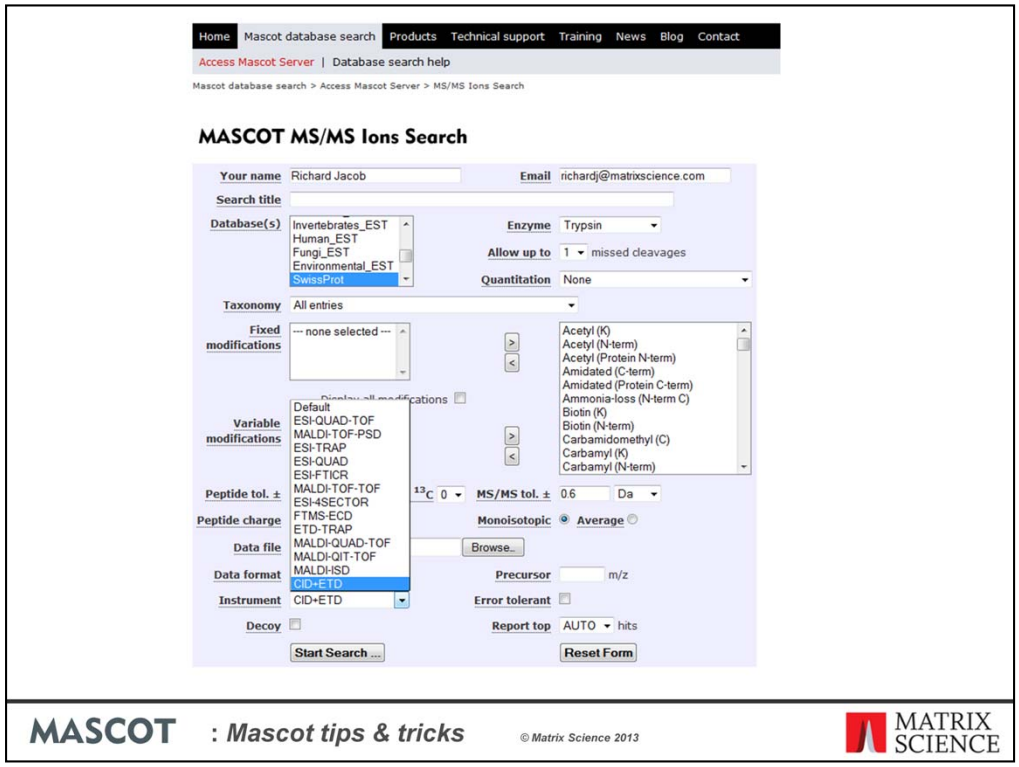
When you search the peak list, select a CID+ETD instrument type.

If your Mascot Server does not have a CID+ETD instrument setting you can add it through the configuration editor.



As you can see we take the CID and the EDT entries and combine the ion series that we search in to a new entry.

The CID+EDT instrument will give near identical results to searching each spectra with the relevant CID or ETD instrument type.



We can select the instrument type from the drop down list of instruments at the time of the search.

That's it! Peptides will be identified with predominately CID ion series or ETD ion series. There may be one or two peaks that match an ion series for the other type of fragmentation but they will have little effect on the score.

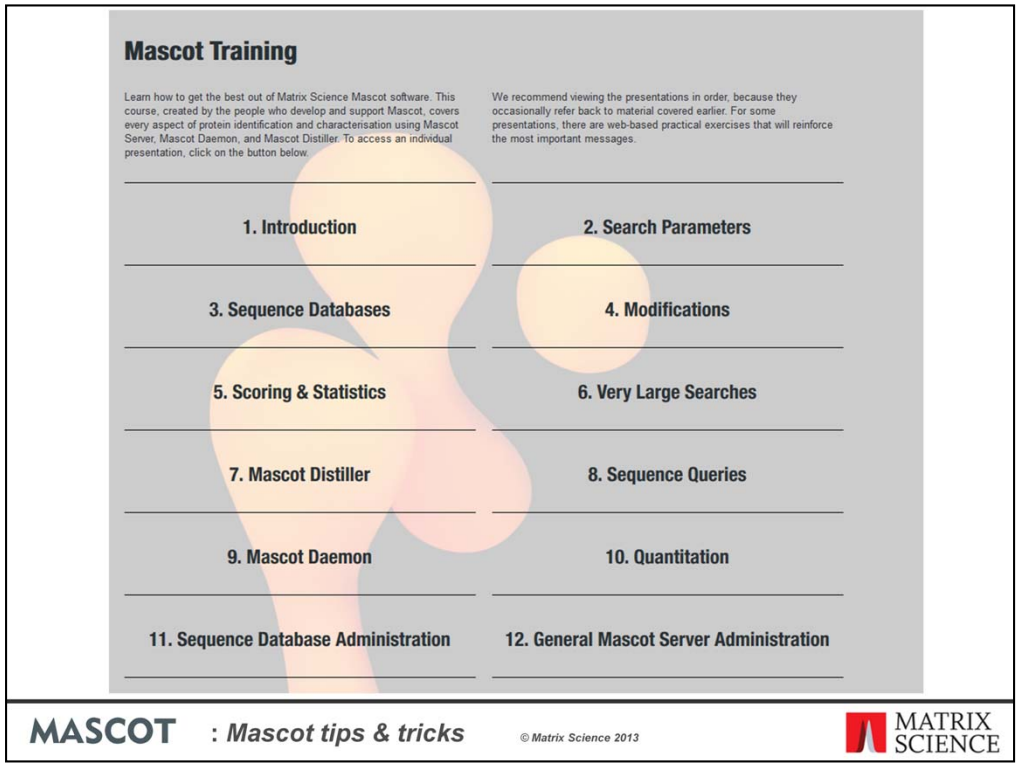
11. Mascot training

© Matrix Science 2013

At around this time of year we are normally contacted by someone looking for a Mascot training summer camp! Sadly we have to direct them elsewhere. This is the Minnesota Vikings mascot Viktor the Viking who is based here in Minneapolis.

For the last 6 years we have held training courses one or twice a year on the east and west coasts of the States, in London UK and other European and worldwide locations. We realized that it was not always possible for everyone that wanted to attend the course to attend so at the end of last year we released the complete course on our website.

The course is completely free and requires no registration to view it. There are the same exercises that we used in the live training courses and, if you complete the multiple choice exam at the end of the course, you will receive a certificate. The course is available from this link on our public website.

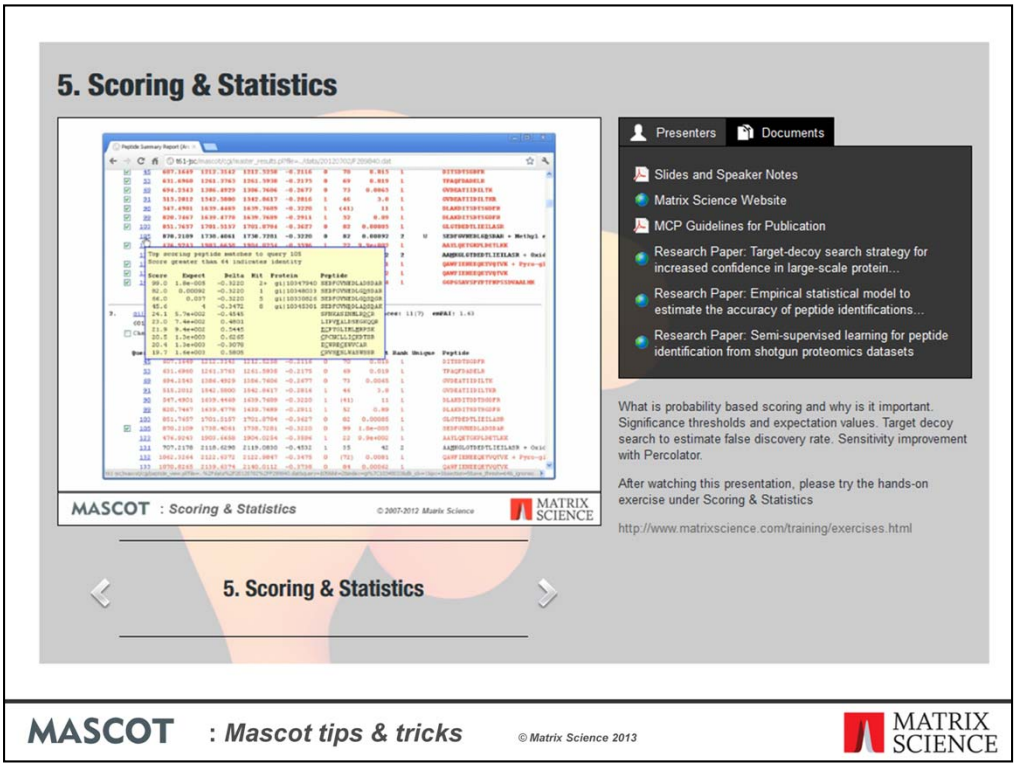


The course has the same content as our live course and has been updated for the latest version of Mascot Server.

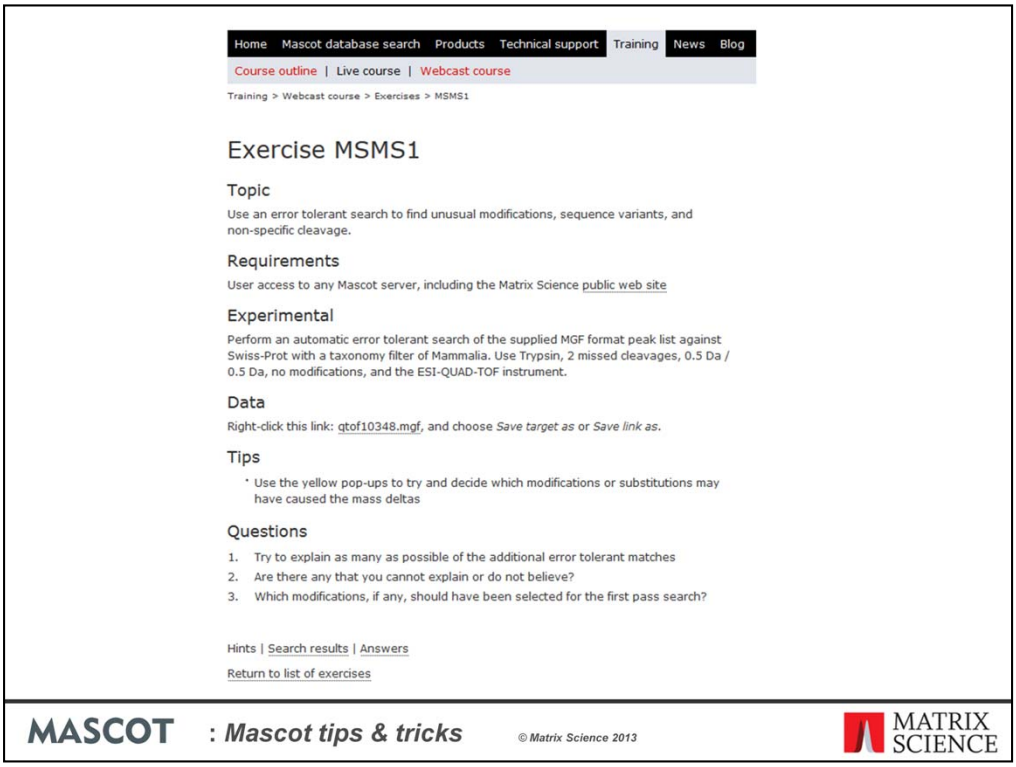
We cover everything from a basic introduction to bring everyone up to speed, more in depth training on scoring and statistics and specialist subjects like Mascot Server administration.

One nice thing about the course is that you can follow it in your own time

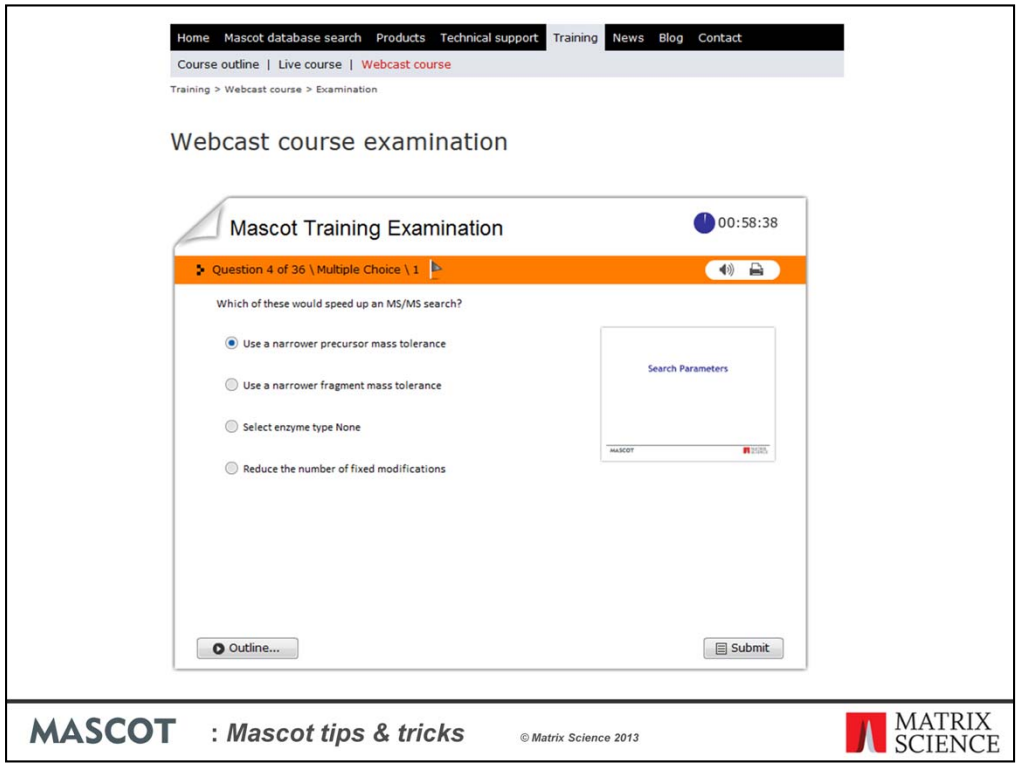
Each presentation includes a short Video introduction and audio narration over the slides. You can pause and repeat any slides that interest you.



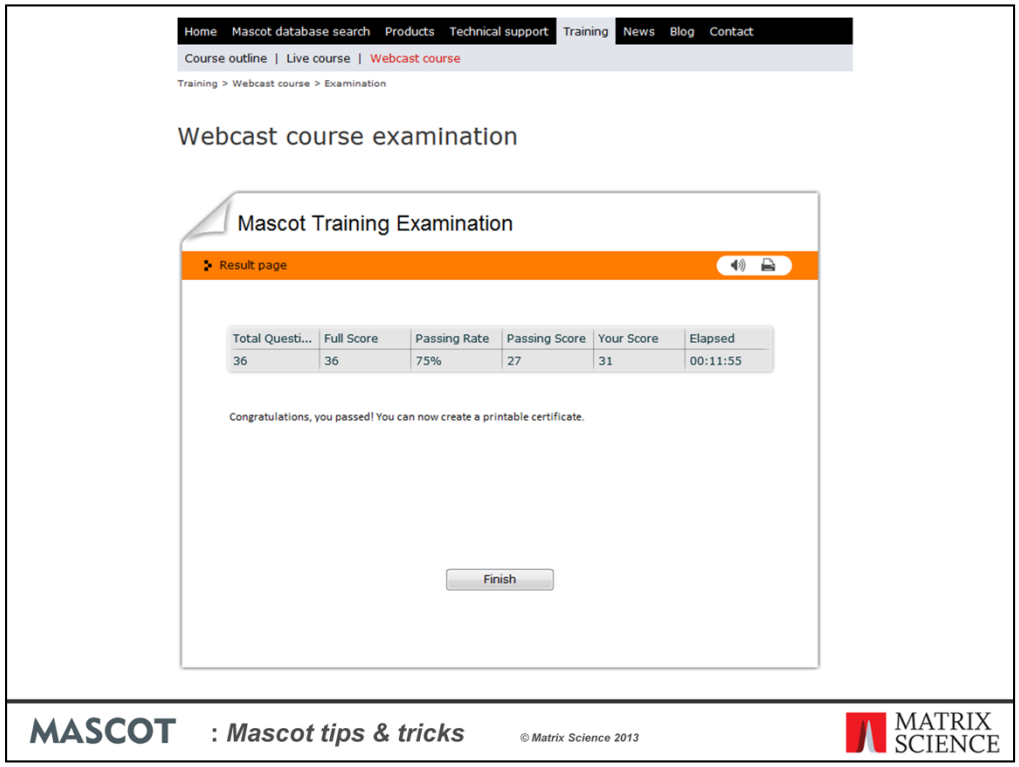
Each presentation includes a PDF of the slides and speakers notes plus links to relevant publications and websites



After many of the lessons there are one or more exercises to put your knowledge into action. There are instructions, links to the data, and a set of questions to answer. If you need it there are hints to help you along and links to the search results and answers.



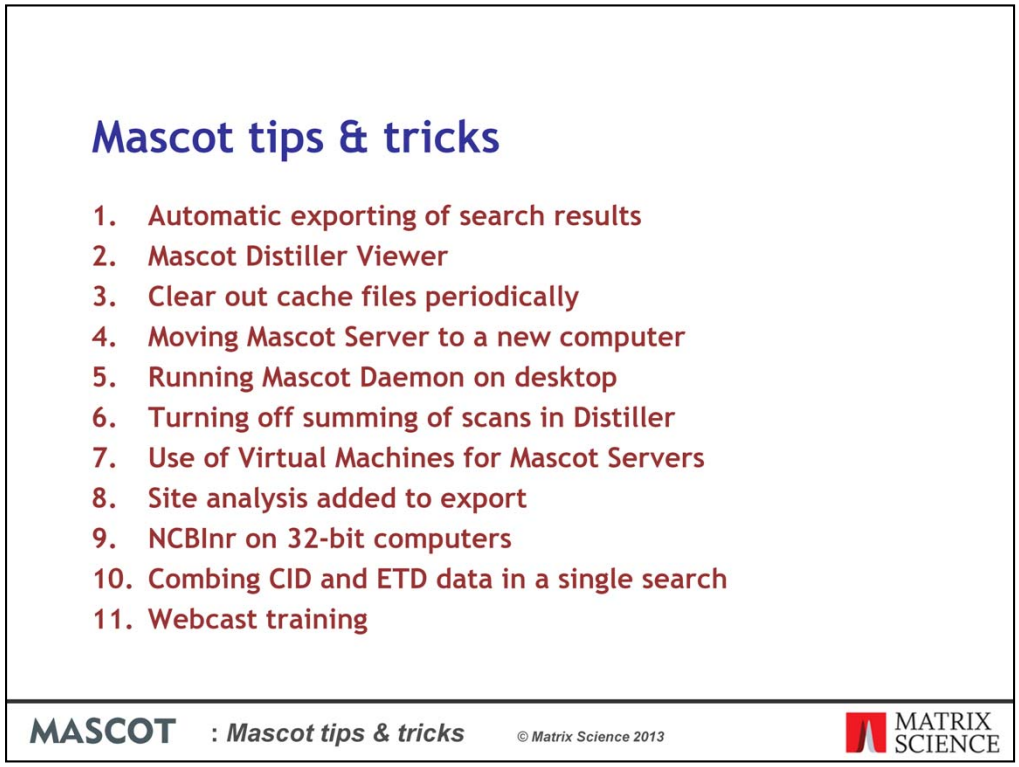
After you have finished the course and exercise take the test to earn a certificate. It is a multiple choice exam and you have an hour to complete it.



At the end you can see your score and if you passed

And then print your certificate.

We do hope that you will take the course as it contains a lot of knowledge and information that make you a more proficient user of our software.



I have covered 11 topics in this mascot tips and tricks presentation and I hope that you all heard something useful for your research.