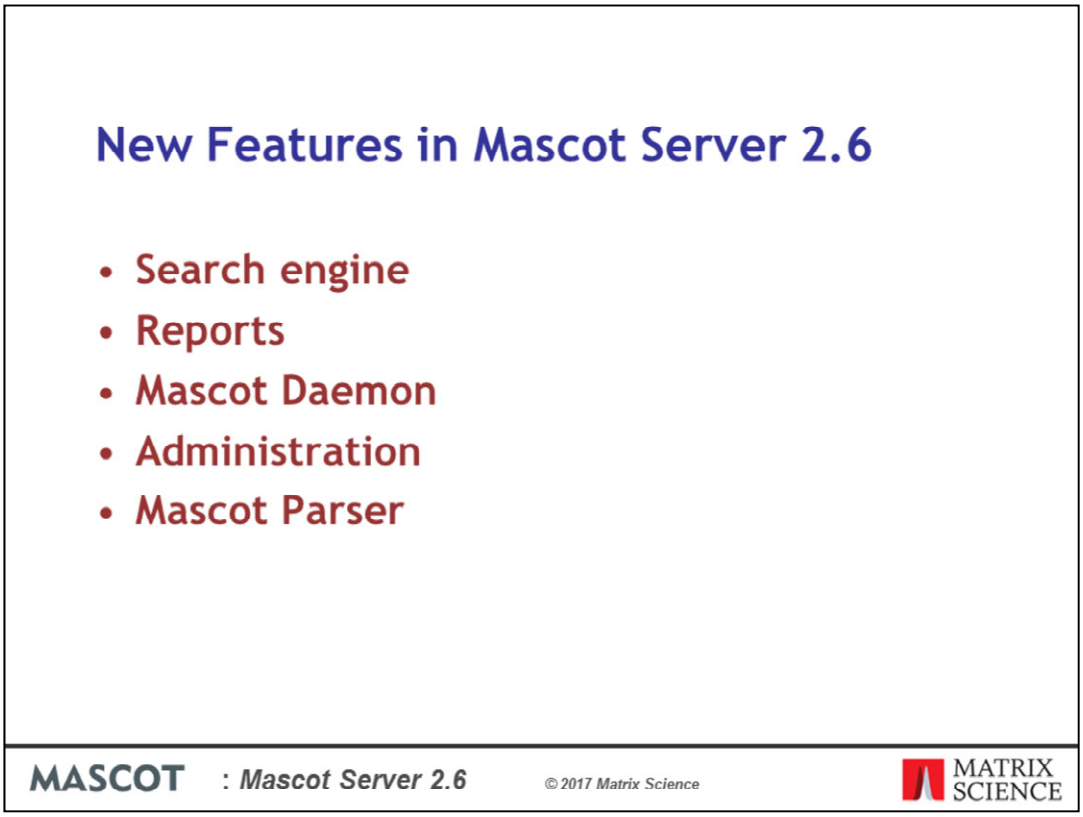
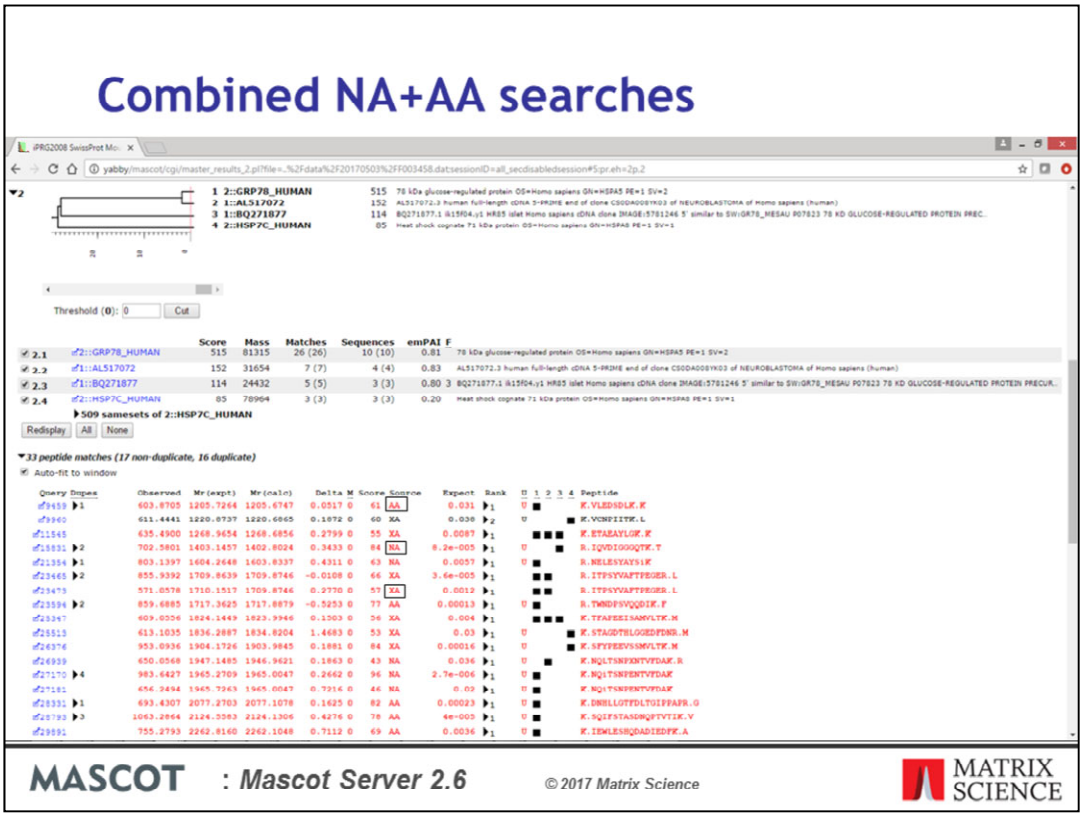
New Features in Mascot Server 2.6

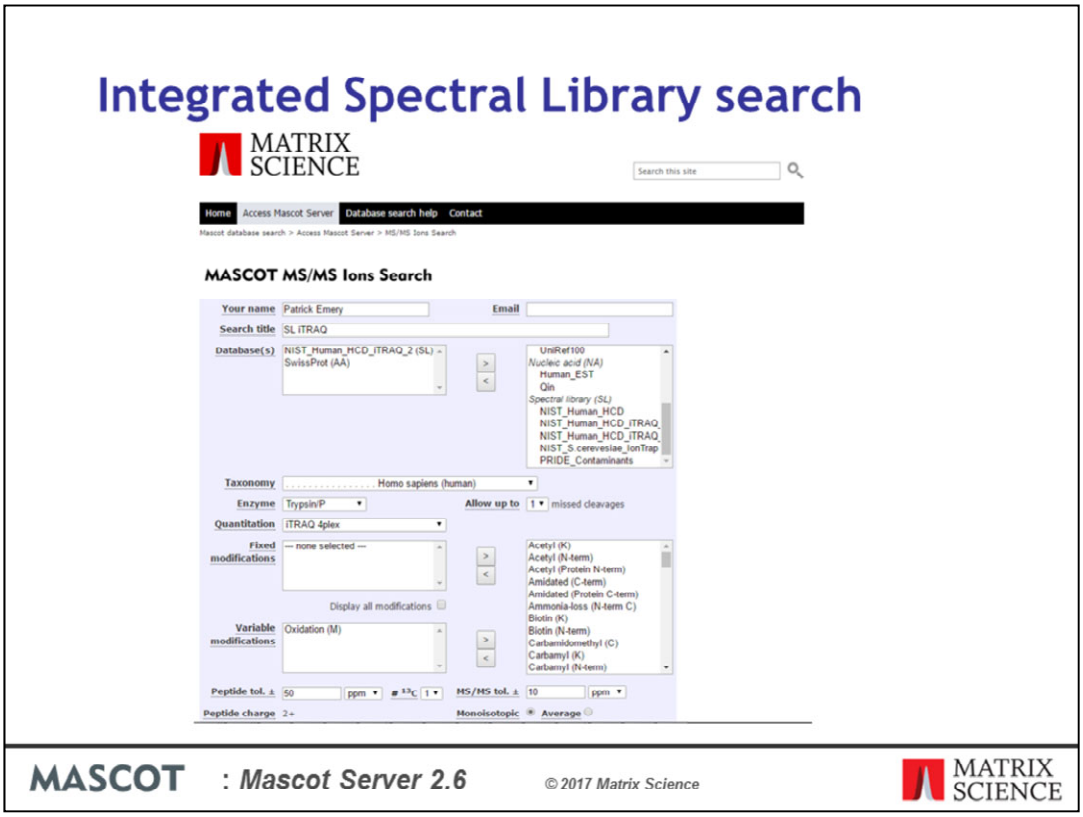


We release Mascot Server 2.6 at the end of last year. There have been a number of changes and improvements in the search engine and reports. I'll also be covering some enhancements and changes in Mascot Daemon, the Mascot server administration pages and in Mascot parser.

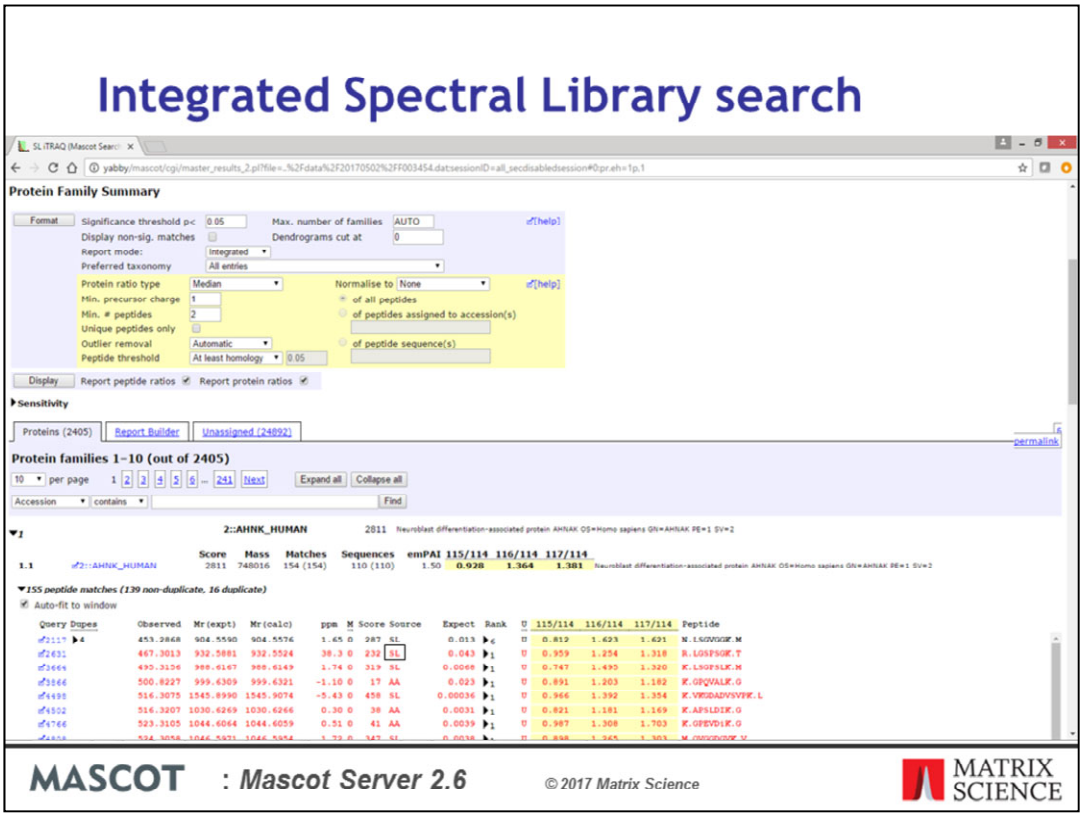


Starting with the search engine, since Mascot Server 2.3 you've been able to select and search multiple FASTA sequence databases to search at the same time. The one restriction on this is that the selected databases all had to be of the same type – e.g. all Amino Acid or all Nucleic Acid databases.

With Mascot Server 2.6, this restriction has been lifted. Here, I've searched both SwissProt and Human EST and protein family 2 contains matches from both databases. The source column in the peptides table tells you what type of database the peptide match comes from – AA, NA or, if it was identified in both database types XA.

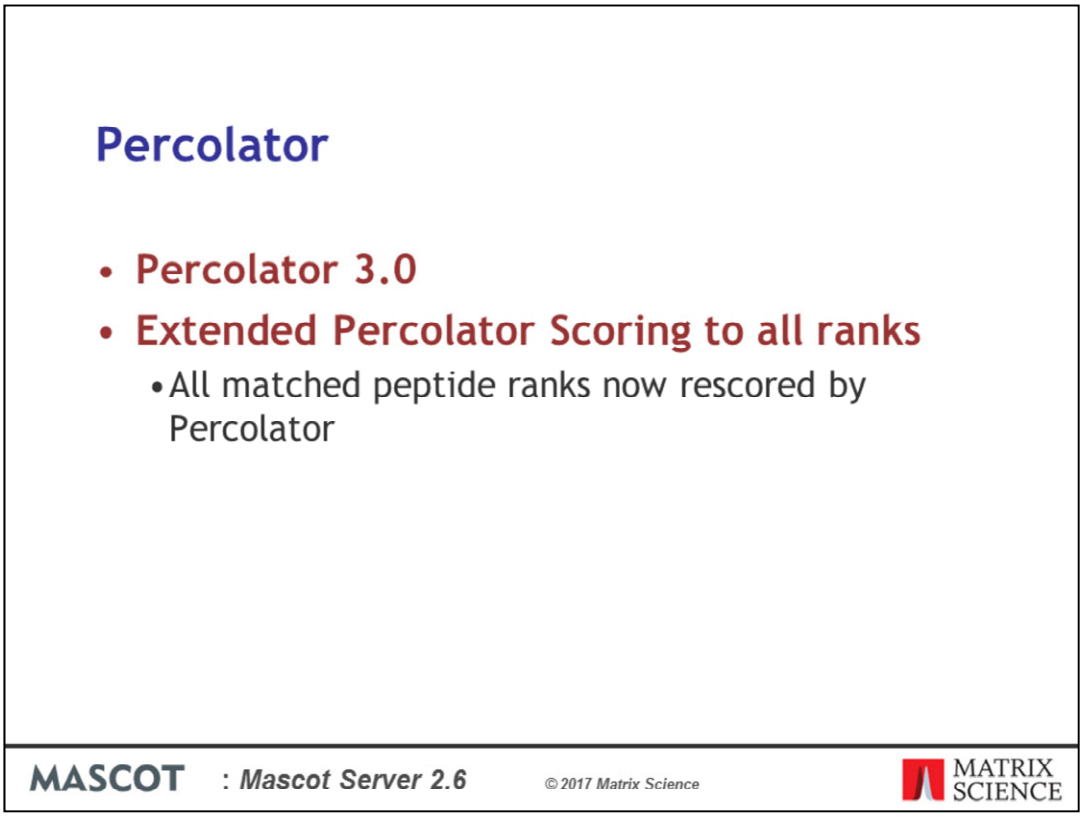


The major new feature in Mascot Server 2.6 is an integrated spectral library search using NIST MSPepSearch. Mascot and MSPepSearch library searches have been seamlessly integrated in Mascot server 2.6 and you can select any combination of Protein, Nucleic Acid FASTA databases and spectral libraries for searching. Here, I've selected the NIST Human HCD iTRAQ 2 spectral library and the SwissProt FASTA database.

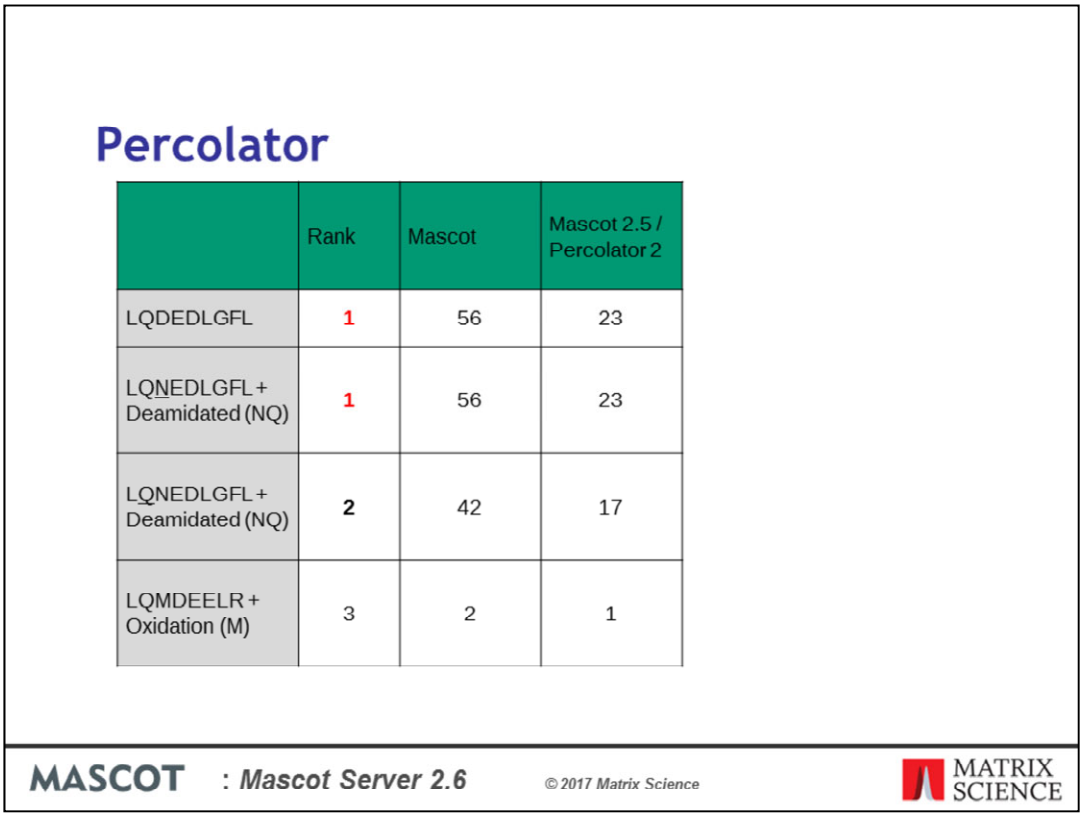


And here we have our results as an integrated report containing matches from both the Spectral Library and FASTA database searches, with the source of the match flagged up in the 'Source' column, as it was in the combined AA+NA search report – now using SL for the matches from the spectral library.

In addition to adding pre-generated spectral libraries to your Mascot server, you can get Mascot server to generate custom spectral libraries from your Mascot search results. For more details about this and for a much more detailed look at the integrated spectral library search, we gave a presentation about the integrated spectral library search at our breakfast meeting yesterday morning, the slides for which will be going up on our public website next week.



A few other changes – the version of Percolator shipped with Mascot 2.6 is now version 3.0. We've also extended percolator scoring to all matched ranks for a query. This allows percolator to re-rank the peptide matches to an individual query, allowing it to change the ranks if it scores say the rank 3 match as better than the rank 2 match.



To illustrate the changes, here we have a match from a Mascot search. As you can see, the top three matches to the MS/MS spectrum all have similar sequences and have scores above the identity threshold, which was 26 for this match. The top two matches are identical at the MS/MS level and so get the same score.

In Mascot 2.5, only the top scoring match was rescored using Percolator version 2. The rest of the top ten matches to the spectrum were then rescored scaling from the percolator score to the top match – so the relative scores and rankings always remained the same as for the original Mascot rankings

Percolator

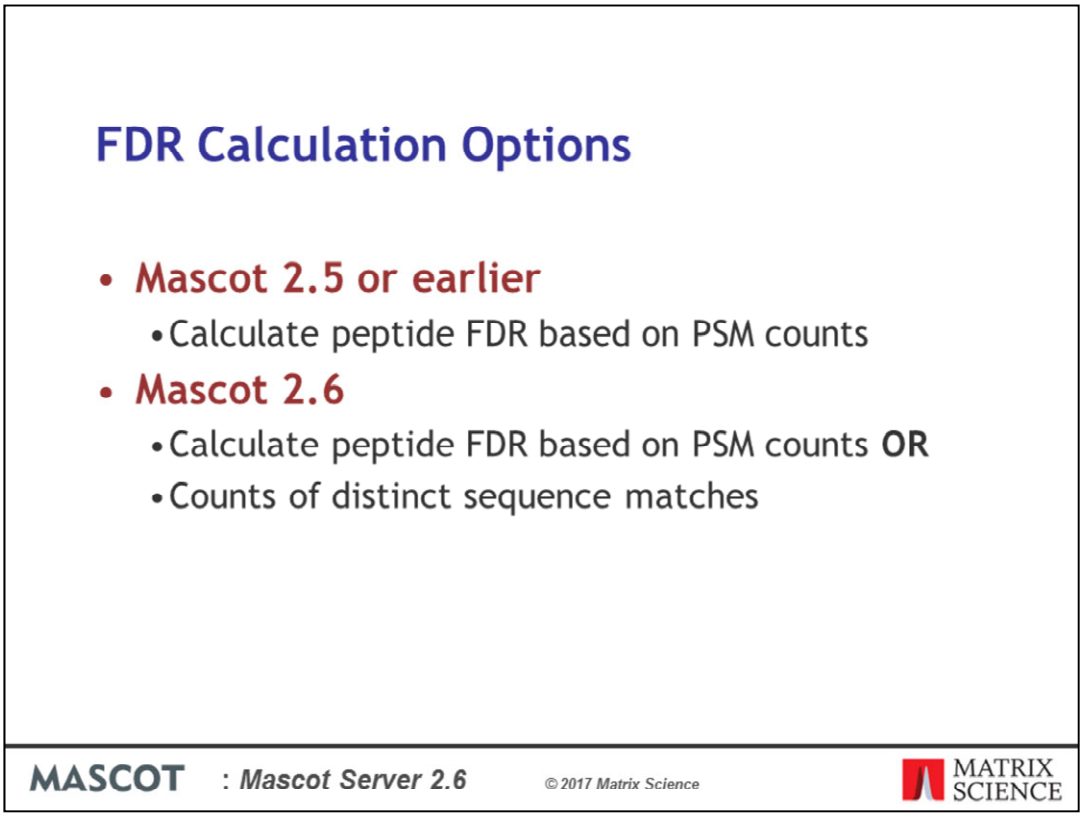
	Rank	Mascot	Mascot 2.5 / Percolator 2	Mascot 2.6 / Percolator 3	Rank
LQDEDLGFL	1	56	23	26	2
LQNEDLGFL + Deamidated (NQ)	1	56	23	26	2
LQNEDLGFL + Deamidated (NQ)	2	42	17	39	1
LQMDEELR + Oxidation (M)	3	2	1	0	3

MASCOT : *Mascot Server 2.6* MATRIX SCIENCE

In Mascot 2.6 all the matching peptide ranks to a spectrum are rescored and we've switched to Percolator 3.

The first thing to notice is that allowing Percolator to rescore all the ranks has changed the final rankings in the case – Percolator actually gives a better score to the third match which is now our rank 1 match.

Notice also that our now rank 2 matches also have a slightly higher score than they did from Percolator 2 – this is due to changes made to Percolator between releases 2 and 3 which can result in different scores for the same match. This is something you will see with your own results if you compare between Percolator 2 and 3 scoring.

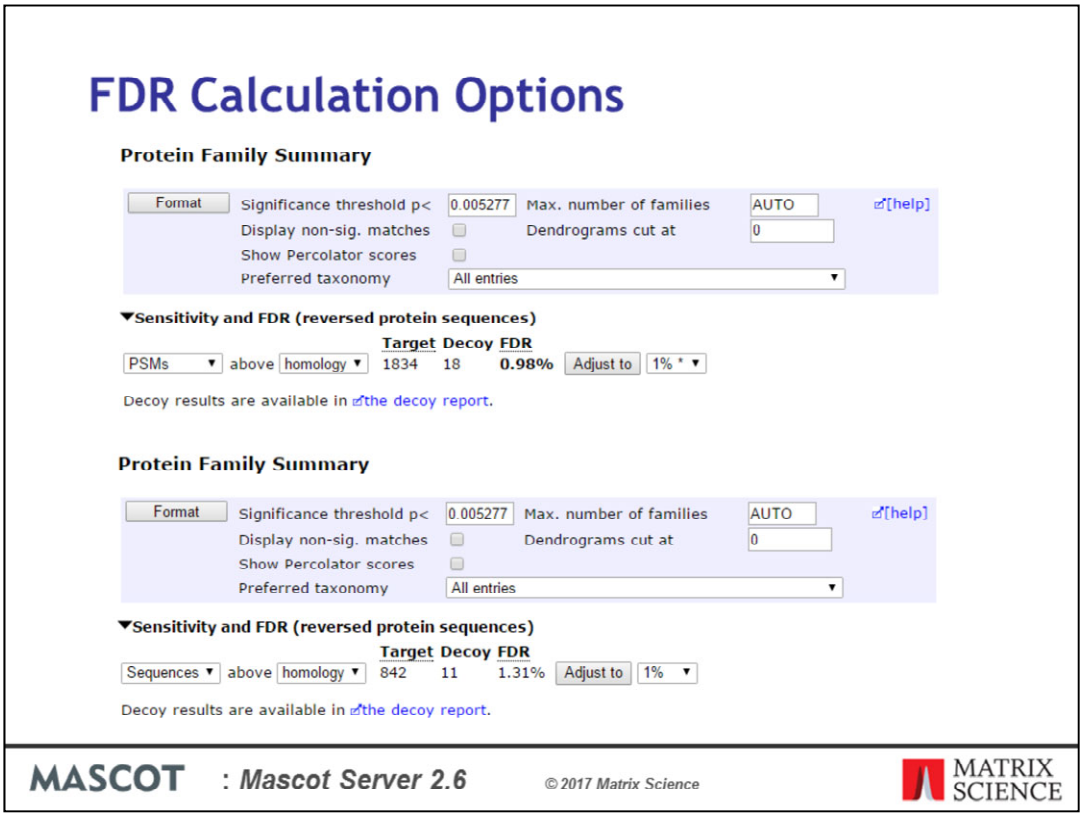


Moving on to some of the changes we've made to the reports. If you carry out an integrated decoy database search in Mascot in order to estimate your peptide false discovery rate, you'll typically then adjust the significance threshold on the report to give a target FDR of, for example, 1%. Typically, this calculation is based on the number of significant Peptide-spectrum matches against the target and decoy databases.

This allows us to obtain and control the false discovery rate at the spectrum level, and is an important step in ensuring the quality of your reports. Falsely assigning sequences to spectra could have an impact on coverage, protein grouping and ranking and quantitation calculations, for example.

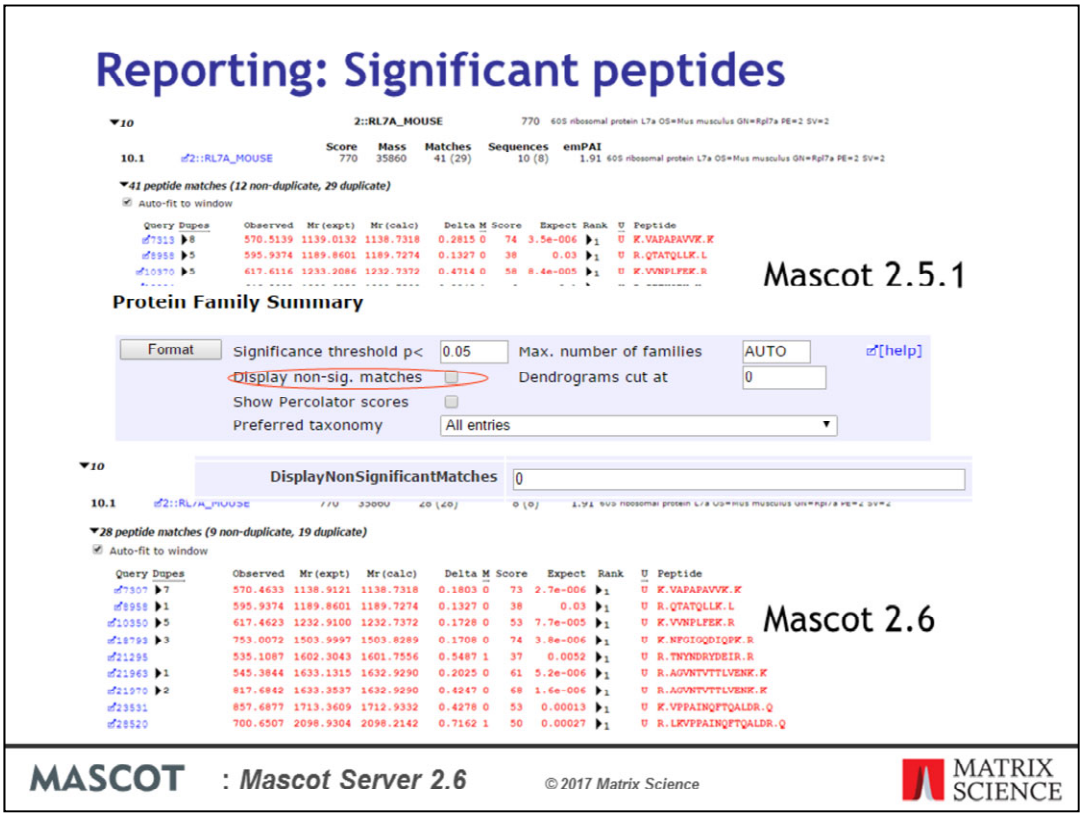
When it comes to generating an accurate, confident list of proteins present in your mixture, what really matters is not the PSM false discovery rate, however, but the peptide sequence false discovery rate - as a single false positive peptide sequence could result in a false positive protein being reported, even if the false match is only identified by a single spectrum. Therefore, calculating the false discovery rate based on distinct peptide sequences, rather than the PSMs, is of potential interest here.

In Mascot 2.6 you have the choice to calculate the peptide false discovery rate using either PSM or peptide sequence counts, switching between them using a drop down control on the protein family summary report.



Often, the two calculations will result in very similar results. This is a mouse dataset searched against SwissProt with a mouse taxonomy filter and a contaminants database. The peptide false discovery rate has been calculated using PSMs and adjusted to a target 1% FDR – in fact we have a calculated FDR of 0.98% using a calculated significance threshold just above 0.005.

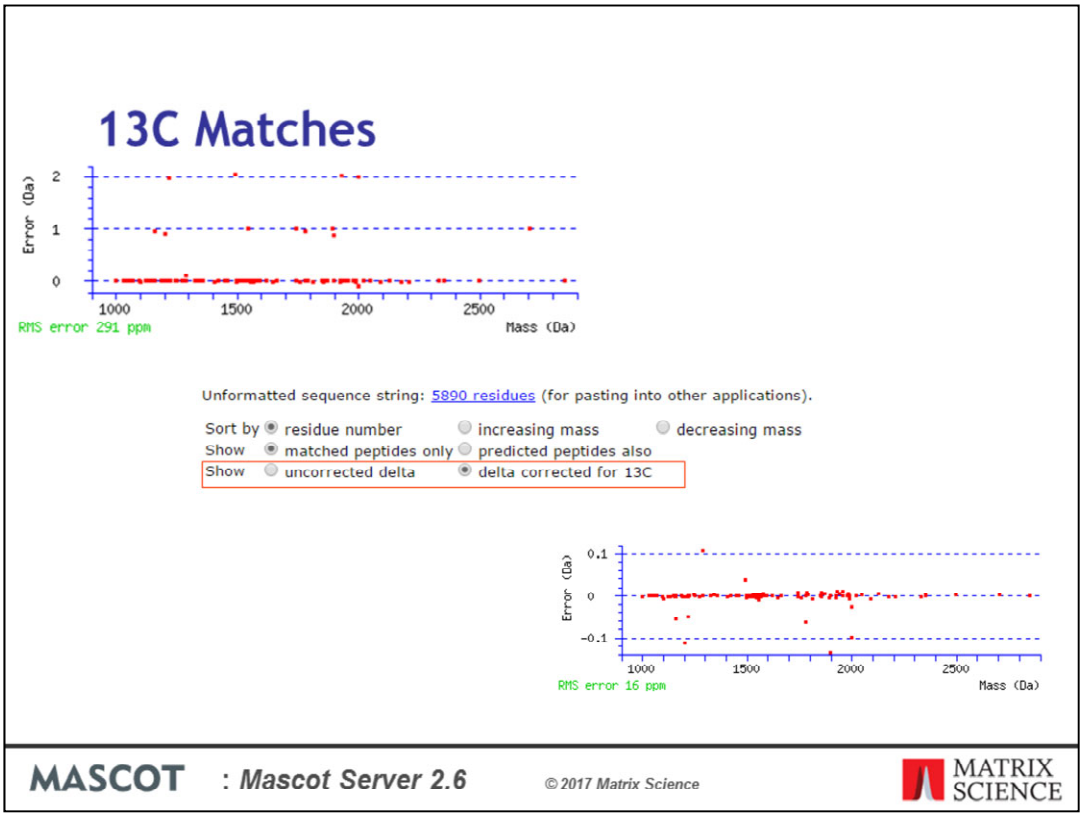
If I switch the report to calculate the FDR using Sequences and retain the same significance threshold, the calculated FDR is now 1.31% - not an enormous difference. Increasing the significance threshold to 0.004 will get us back 0.99% FDR without loosing very many target matches



Another change we've made to the default reports is that the protein family report now only shows significant peptide matches by default. Showing all matches unless you entered a score or expect value cut-off was confusing for many. On an individual report, you can choose to display non-significant matches using the checkbox on the format options at the top of the report.

If you prefer the old behaviour, this can be re-enabled by editing a Mascot.dat setting, 'DisplayNonSignificantMatches' and setting the value to 1.

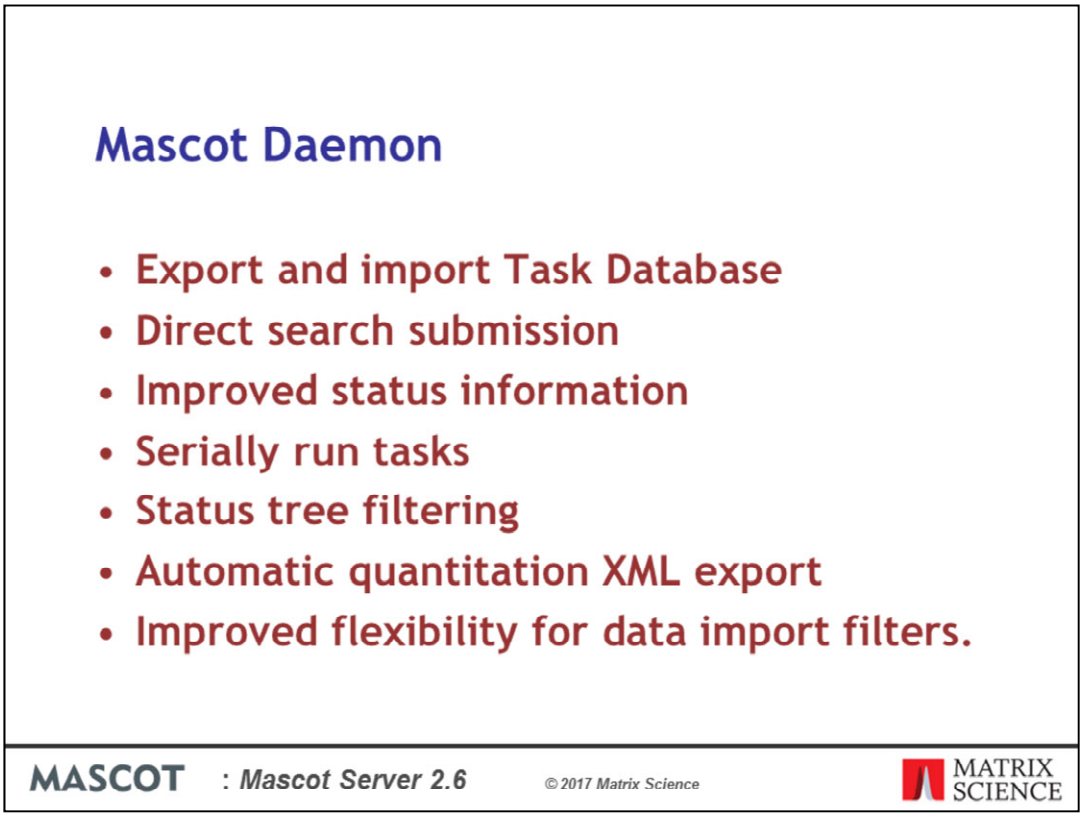
The eagle eyed amongst you will have spotted that this slide also demonstrates another minor change we've made to the Protein Family report in Mascot 2.6 – duplicate peptide matches are now ordered by the expect value and not the raw ions score, so for this first peptide [VAPAPAVVK], the top reported query number is different between Mascot 2.5.1 and Mascot 2.6, because query 7307 has a lower expect value than query 7313.



If you've selected a 13C value of 1 or 2 and you look at the protein view report on a Mascot 2.5 or earlier, you'll see that the error graph at the bottom of the report looks something like this…

This is because it is the uncorrected precursor delta values of the matched peptides that are being reported, so you have clusters of mass errors around the 0, 1 and 2 Da mark. While this view makes it very clear that we have a 13C peak picking issue, it means that the actual spread of delta values is almost impossible to look at, so we can't get a very good impression of what our mass accuracy actually is, and the calculated RMS error value is completely wrong.

In Mascot 2.6, we've added these radio button control to the protein view report. This allows us to switch between the delta uncorrected and corrected for 13C. If I switch to show the corrected value, the graph now looks like this and we can see that we've actually got a very good mass accuracy and our RMS error value has dropped from 291ppm to 16ppm.



Moving on to Mascot Daemon. Daemon has received a number of upgrades in release 2.6. We've added options to export and import your Task Database, making it easier to backup an instance of your Task Database and restore it if required.

If Mascot Daemon is installed on your Mascot server (Windows only), then it can submit searches to Mascot directly from the commandline. This gets around the 4Gb upload limit some of you may have encountered if you're running Mascot on IIS.

The status information about running tasks and searches has been improved, so you'll get better feedback about what is happening with a task.

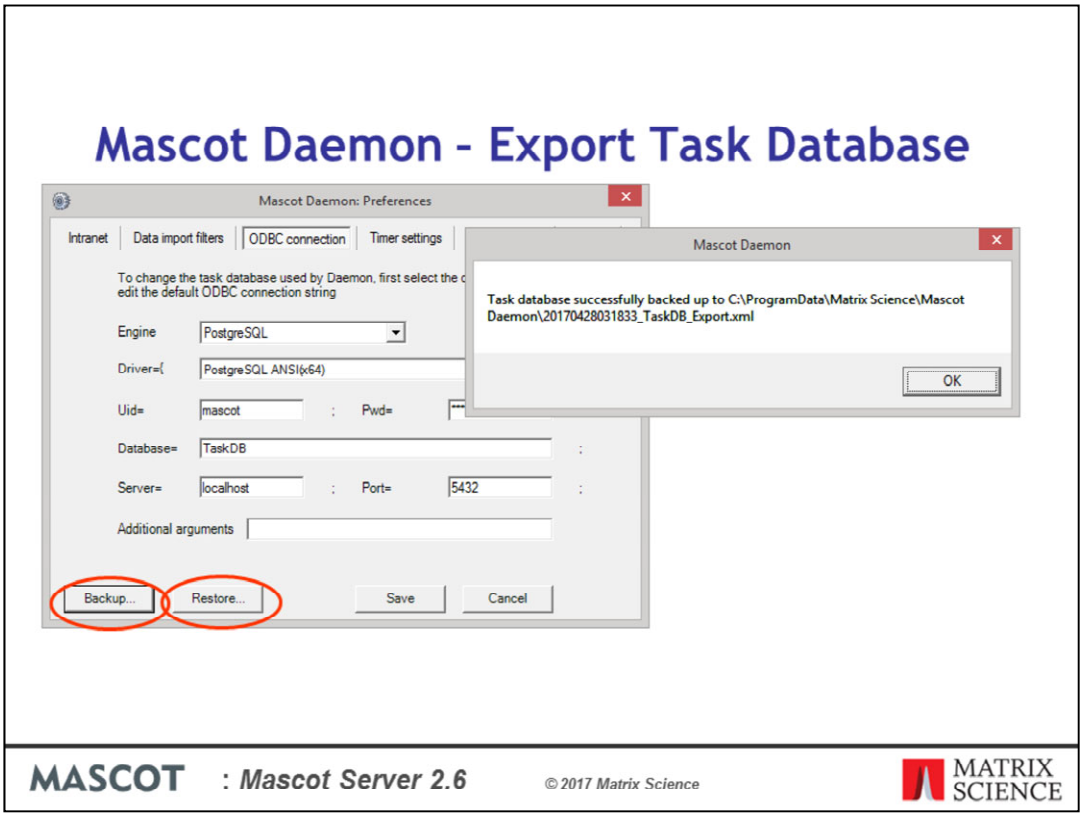
We've added the ability to run tasks serially – that is to start a task only when a previous task has completed, so you can now avoid setting up 10 tasks and having them all submit searches at the same time.

The status tree can now be flexibly filtered, so if you've got a very large number of tasks and searches in your Task Database, you can easily find the old result you've been looking for

If you're using Mascot Daemon to automate Mascot Distiller quantitation for methods such as SILAC, you can now get Daemon to automatically export the quantitation results

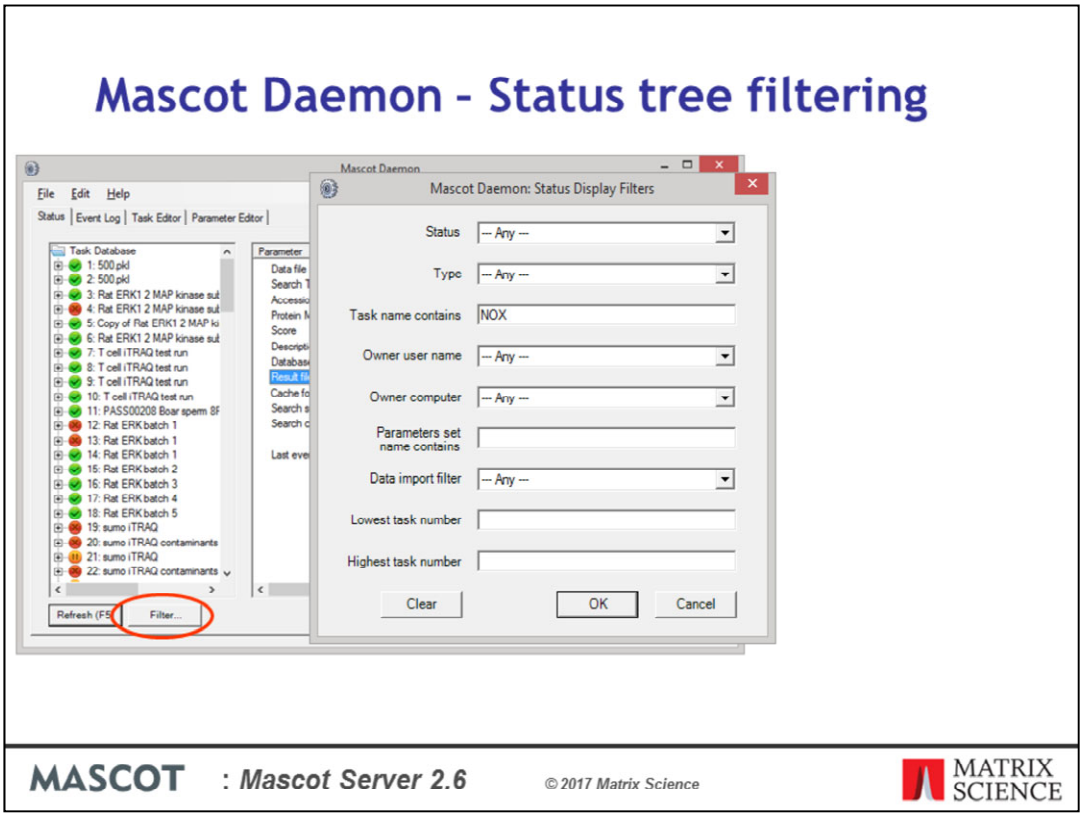
in our XML format, ready for use in other software such as Scaffold Q+S

And we've improved the flexibility of the data import filters used to process raw data files to peaklists, so you can now control where the peaklist files are saved on your PC.



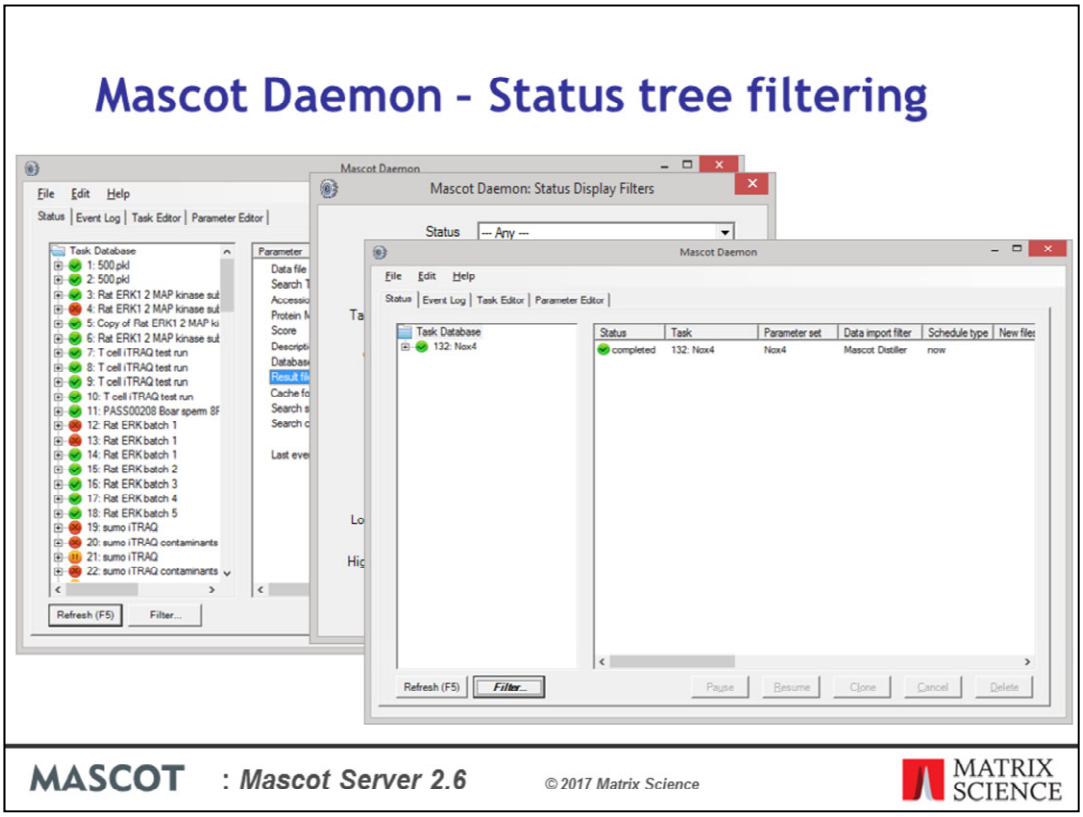
To export the Task Database, open the preferences dialog in Mascot Daemon. On the ODBC connection tab you'll find a new 'Backup…' button. Click on this and it'll allow you to backup the contents of your Task Database to an XML file. The 'Restore' button on the same page will allow you to restore your Task Database to the state recorded in a selected XML export file.

This facility also simplifies the process of changing the database engine you're using for Mascot Daemon. All you need to do now is use the export facility to export your current task database, switch over to the new database engine and then restore from your export.



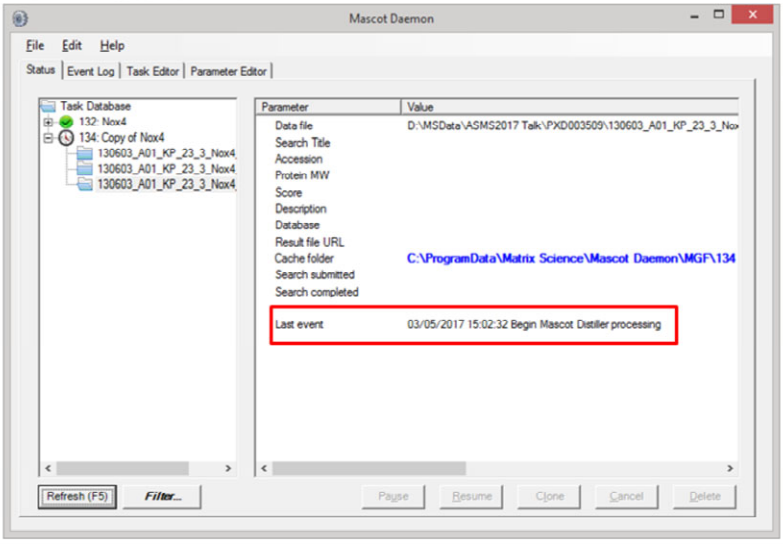
Taking a look at the filtering options we've added to the status tab of Mascot Daemon. Here is a screenshot of Mascot Daemon running on my laptop – I've got a reasonably large number of old tasks in my Task database and finding one that I want to perhaps look at the results of, or clone as the basis of a new task can be quite time consuming. I want to search for a particular task were I searched a dataset which was looking for human Nox4 interacting proteins.

If I click on the 'Filter' button at the bottom of the pane this dialog window will open, allowing me to search on a number of fields associated with the tasks. For example, I could easily use the 'Status' filter drop down so that we'd only see 'Running' tasks on the Status tree, or I could easily pull out all the tasks where I've used Mascot Distiller as the Data import filter. In this case, I'm going to search for Nox4 in the task title. Click on OK



And now the Task Status tree is filtered and there is my task. Notice that the font on the 'Filter' button has changed to bold italic, showing me that I've got a filter applied.

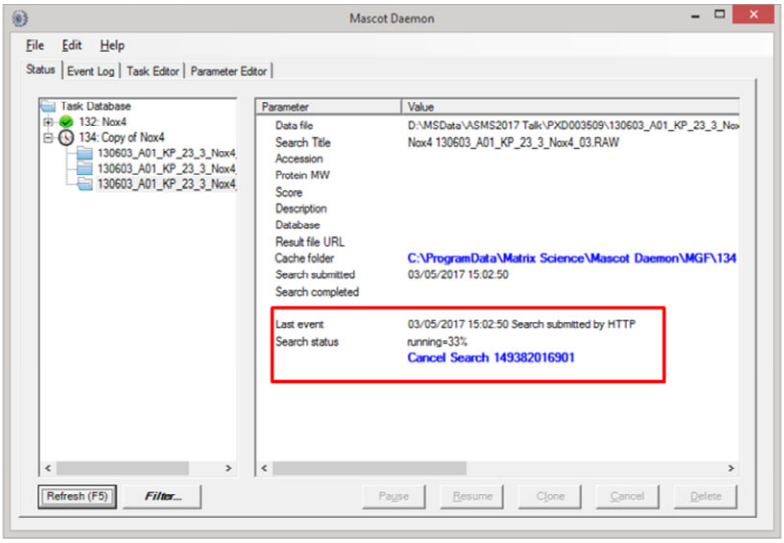
Mascot Daemon - Improved Status Information



MASCOT : Mascot Server 2.6 © 2017 Matrix Science MATRIX SCIENCE

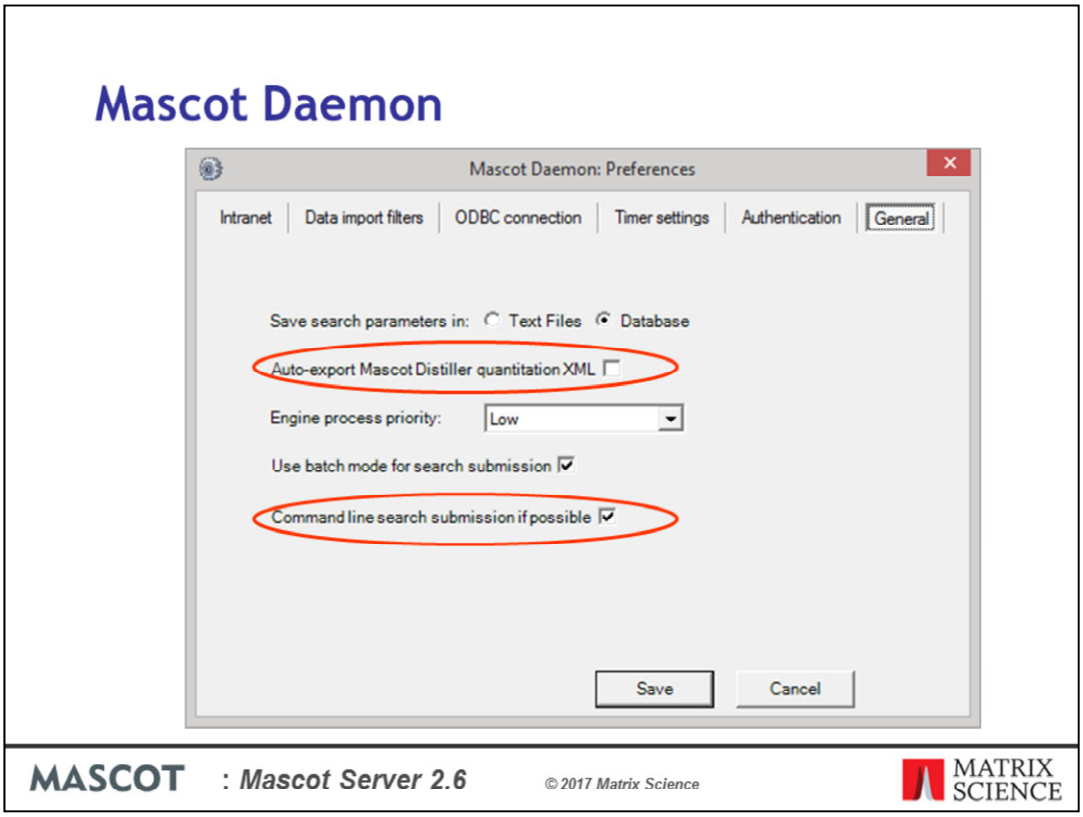
Here is an example of the improved status information from a running Mascot Daemon Task – I'm using Mascot Distiller as a data import filter here to process the raw data into a peaklist, and even before we have a search result, Daemon now adds a node to the tree for the file so that we can get feedback information and to persist any errors that are encountered – in previous versions of Daemon, much of this information was only available from the Daemon event log and was much less clear. For this file, you can see that we have a last event field telling us that Mascot Distiller processing has started.

Mascot Daemon - Improved Status Information



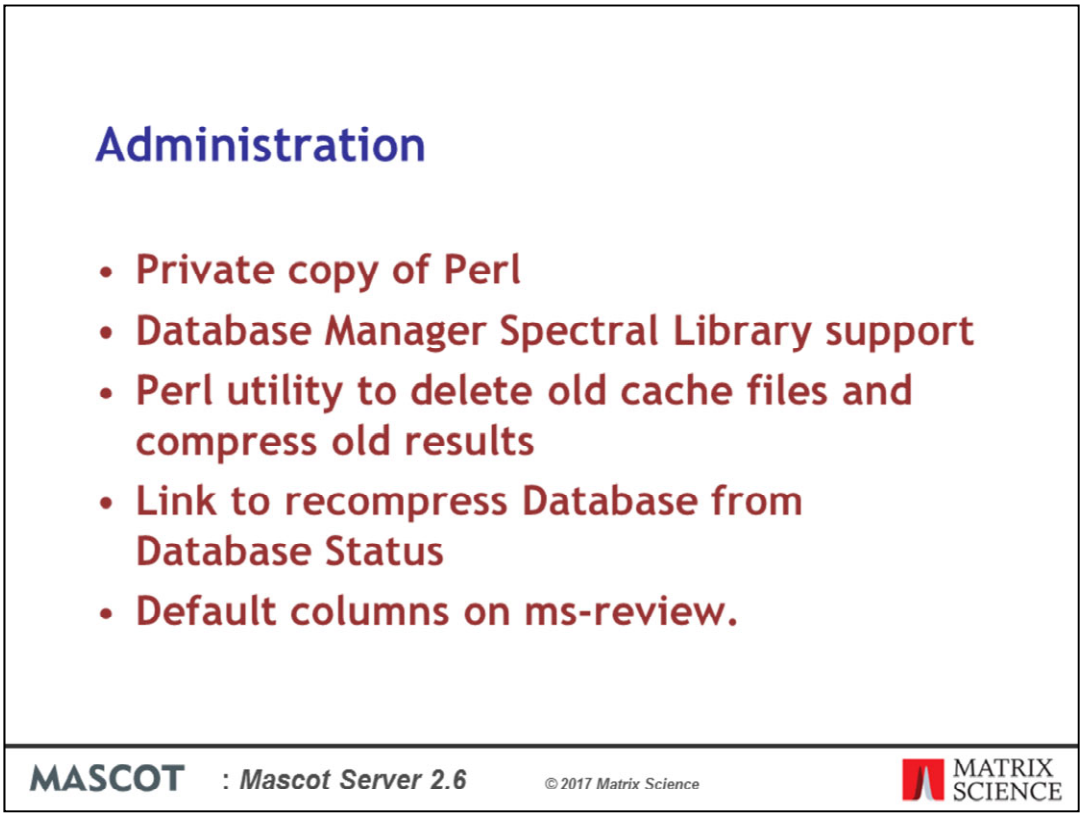
MASCOT : Mascot Server 2.6 © 2017 Matrix Science MATRIX SCIENCE

A little later, and the processing has completed and, as you can see, we're being informed that the search has been submitted to the Mascot server (over HTTP, rather than directly submitting at the command line in this case). We also have search status information and can see that the search is 33% completed, and we have a link to cancel the search if we need to.



If you want to auto export Mascot Distiller quantitation XML results after search and quantitation has completed, the checkbox to enable this can be found on the 'General' tab.

This is also where you can control whether or not Mascot Daemon will try and submit the search to Mascot directly on the command line if possible



The next section is of interest if you're the administrator of your local Mascot server. We've made a number of changes to (hopefully) make your life easier.

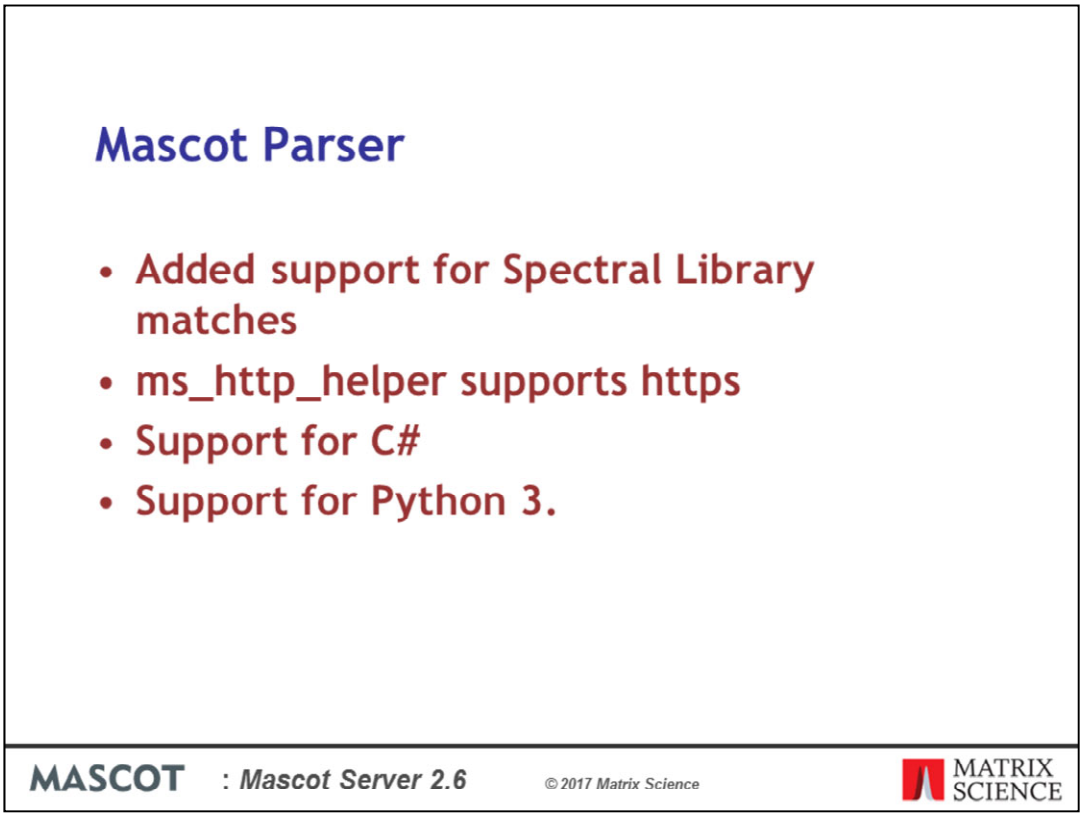
One major change we've made is that Mascot 2.6 installs and uses it's own, private, copy of Perl. Having to deal with annual releases of Perl, each binary incompatible with the previous, was becoming a nightmare. So now, if some other software needs the latest version, or an old one, there is no problem as you can install the required version without upsetting your Mascot server installation.

We've made a number of changes to the Database Manager to support spectral libraries – both enabling pre-defined definitions and creating your own local libraries from your Mascot search results (more details in yesterdays presentation)

One issue that many Mascot server administrators encounter is that of running out of disk space. When you view a search result, Mascot creates a number of cache files to speed up report generation. These cache files can take up quite a lot of space and they can be safely deleted to free up disk space as the system will simply recreate them as required when you next open the source search result. To automate this, and compress old search results, we've added a Perl script, tidy_data.pl, which will do this automatically using a time cut-off you specify. You can easily run tidy_data.pl automatically using a CRON job on Linux or a scheduled task on Windows.

We've added a link on the Database Status page to recompress a database – there is no longer any need to stop ms-monitor and delete the .stats file manually

And, something that has been on the wish list for a long time – you can now set the default columns displayed by the ms-review utility.

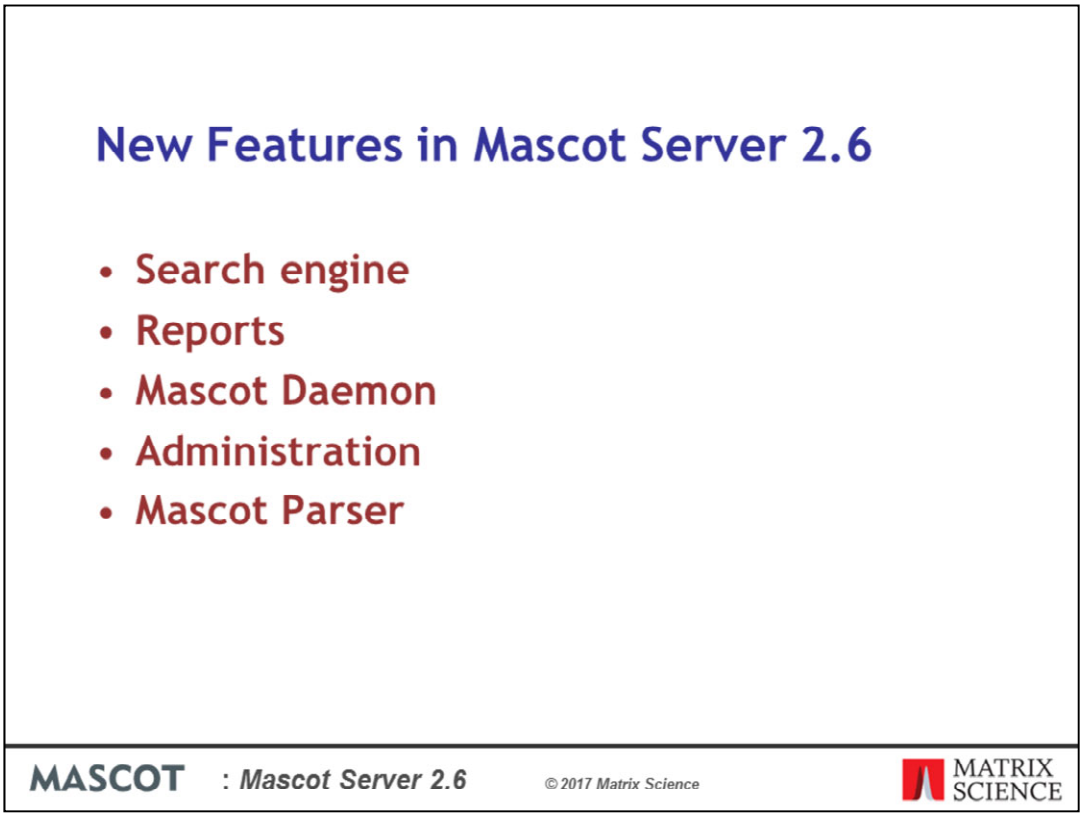


This slide is only of interest to developers who are accessing the Mascot server and search results from third party applications, but it is quite important. There have been a number of changes made to Mascot Parser, which is our API for accessing and parsing Mascot search results. There are too many minor changes and bug fixes to list here, but these are some of the major changes which might affect you if you're writing client software to access Mascot search results:

The first and most important is that Mascot Parser supports the new sections in the Mascot results file for Spectral Library matches, and will carry out protein grouping using the results from a spectral library search alongside any results from a Mascot database search.

If you're using the ms_http_helper utility class in your code to access your Mascot server, this now supports https as well has http

And we've added language support for both C# and Python 3, alongside C++, Perl and Java.



So, I've outlined some of the key new features in Mascot Server 2.6, including the integrated spectral library search, changes to the reports, enhancements to Mascot Daemon, improvements on the administration side and extended language support for Mascot Parser. Of course I've had to miss out many more small improvements and bug fixes which may be of interest to you, and a fuller list of changes can be found on the support page for Mascot 2.6 on our website or on your local server.