Mascot Insight
Mascot Insight is a new application designed to help you to organise and manage your Mascot search and quantitation results.

Mascot Insight provides ways to flexibly merge your Mascot search and quantitation results, including MS-1 based quantitation data from Mascot Distiller, such as SILAC data, and MS2 based quantitation such as iTRAQ and TMT.

Mascot Insight allows you to further annotate your results, including areas such as Gene Ontology annotation of results, use of molecular interactions databases and manual annotation and approval of protein hits.

And Mascot Insight provides a wide range of reports covering areas such as comparing datasets, quantitation analysis, plotting charts etc, and provides exports for these to allow you to easily share those results and to export data in machine readable formats.
It is also important to mention one thing that Mascot Insight isn’t. It is not a LIMS system. Although you can use it to organise your results and report on them, and it also allows you to record details about your experiments if you wish, it will not carry out many of the functions you would associate with a LIMS such as instrument integration, sample tracking and FDA regulatory compliance.

Mascot Insight is not licensed ‘per seat’, and there is no limit on the number of users that can access the system. The price of Mascot Insight depends instead on the number of CPUs licensed for the associated Mascot server.
Mascot security provides a single login to both your Mascot server and to the Mascot Insight web application to identify users and to control user access levels.
Once you’ve logged in, this is the home page of the application. On the left we have the Data explorer tree. This is where the data available in the system is organised and displayed, with different types of node for storing different types of information. Data can be manually position on the tree, or structures can be automatically generated to your own specifications.

On the current view, we can see two main types of node – User nodes and Folder nodes. Each user is given a User node automatically, and this is for them to organise their own data and experiments. The folder node is the simplest form a grouping node and is simply used to group together other related nodes on the tree. There are other node types available in the system, including Project, Study, Experiment, Sample and Result type nodes. These different node types allow you to record different pieces of information (meta-data), and you can always add your own values.

On the right hand side of the home page, we have the ‘Search panel’ which allows us to rapidly search the tree to find a particular node or set of nodes. Security is role-based and handled by Mascot security. Exactly what you see on the dataset explorer tree when you log in depends on your users roles, and on what access settings have been given to particular nodes on the tree by their owners.
Mascot search results are automatically captured and imported into the system. For the highest level of integration, we’d recommend using Mascot Daemon to submit your searches, and Mascot Distiller for your raw data reduction and MS-1 based quantitation. Mascot Daemon comes with your Mascot server, but Mascot Distiller is a separate product. However, neither is required and you can simply submit a search through the standard search form, or 3rd party client system, if you wish.

You can also carry out a batch import of existing search results from your Mascot server. You can easily limit this to a selected set of searches, to the results for a selected user or by date range. If you have existing Mascot Distiller based quantitation datasets, Mascot Insight comes with a utility ‘XmlUpload’ which will allow you to easily import the required search results and Distiller quantitation data in a hierarchy based on the file hierarchy from the file system.

Other result types, such as data in the mzIdentML or protXML formats, can be uploaded to a selected node (or set of nodes) via the data explorer. Multiple files can be zipped together and the archive uploaded if you have multiple searches you want to import in a batch.
Here I’m carrying out a search of an Arabidopsis iTRAQ 4-plex dataset against the contaminants and SwissProt database on a Mascot server which is registered with a Mascot Insight server. I just submit the search as normal…
Shortly after the search has completed, the result will also be available in Mascot Insight. In this case the result has been imported to the default location under the ‘Unassociated Mascot search results’ node in a date folder. I could move the result to a different location on the tree by either dragging and dropping it, or by using the cut and paste options from the dataset explorer tree’s context menu. However, you can also set up filters to import the result into a tree structure of your own choosing. These filters can automatically create new folders on the data explorer tree as required. We’ll take a look at an example of this later on in the presentation.

If you’re not sure where a search result has gone on the tree, then the tree is fully searchable using any available field – for example, on the search title
If I right click over a node, or a group of selected nodes, then I get a pop up context menu which allows me to carry out a number of operations, including editing the node, editing default search settings, cut, copy clone and paste operations. For a search result, I also get the option of viewing the original summary report in the browser, or to view the selected result in MIRA – the Mascot Insight Result Applet.
MIRA provides a tabular view of a single result or merged set of results. You can view results in either the Protein centric mode (as we are doing here), or in a peptide centric view. On the upper left we have the searches pane, which is used to change between the selected results, changing which result is being displayed in the main body of MIRA. It is also used for running search level reports from within MIRA and for setting result display parameters across multiple searches. The upper panel on the right contains the proteins table. This is the protein hit list for the currently selected result on the searches pane, and resembles the ‘Report builder’ view in the standard Mascot protein family report. If we click on a row in the Proteins table:
The peptides table, containing the list of matching peptide hits for the selected protein, and various details tabs (such as a protein view) are populated. Clicking on a table column header for either the proteins or peptides table sorts the table on that field. You can also re-order the columns by dragging them and also show and hide columns as you wish – you can save your default column selection, order and column widths if you wish; your choices are saved to the Mascot Insight database, so are available on any client PC you log in from. In the bottom left of the main pane is a panel with search functions, and options to switch between this view and a peptide (or spectrum) centric view.
Clicking on a peptide row allows you to drill down and see peptide specific information, such as an annotated spectrum.
Proteins and peptides are selected using the checkboxes in the first column of the tables. If you generate a protein or peptide level report, only the selected proteins or peptides are used. You can easily apply protein and peptide selection level filters in MIRA, which will automatically control which proteins or peptides are selected in the tables. Here, we are going to deselect out any matches to the contaminants database. Set up the filter and click ‘Apply’.
And we’ve now deselected the 11 matches to the contaminants database for this search result. I can hide unselected protein and peptide rows by checking the ‘Show selected rows only?’ checkboxes
So now I’m only viewing proteins which matched SwissProt and significant (above homology threshold) peptide matches
You can enable GO Analysis in MIRA. GO annotations are shown in the proteins table, as different coloured dots. Hover over a cell for a tooltip.
GO assignments are also presented in additional tabs as piecharts. Click on a wedge of the pie-chart, and the protein table is filtered on that GO assignment.
Now we are just looking at the 85 protein matches with the GO biological function assignment of ‘Response to stress’ and which are not matches from the contaminants database.
Once you’re happy with a particular set of protein matches, you can approve the protein hits and make manual annotations if you wish. Approving a protein hit flags the protein in the Mascot Insight database. You can use this to easily highlight and filter out just those protein hits you think are present in the search sample, or are of particular interest.
For that previous example, I simply allowed the search result to be imported into the default location on the dataset explorer. However, you can override the default result import location using an Import assignment filter. This is a filter that you can define in the system which will direct a matching search, to a specified node on the system explorer tree, overriding the default import location. In addition to that, Import assignment filters can automatically create new folders on the tree based on the values of search parameters, or by parsing out a string from a field.
Let's take a look at a real world example of how you might set up and use an Import assignment filter. The Data Explorer displays a tree structure of folders, results, and saved reports. You can make the structure as simple or as complicated as you wish, but the idea is to organise results logically so as to facilitate merging and reporting. For example, consider a quantitation experiment where 5 biological samples are analysed at 12 time intervals and each analysis is performed in triplicate (to give technical replicates). Each analysis involves 2D chromatography, requiring the mass spectrometry analysis of 8 fractions. This one experiment gives rise to 1440 Mascot search results. By grouping these results into a logical hierarchy of folders, we can minimise the work required to merge results and create meaningful reports.

In this case, the RAW data files are all in a single folder as shown, but the data file names contain all the information that we need to automatically generate a meaningful structure on the Data Explorer as the search results are imported.
So, here we have an example filename from the dataset. The name structure of the raw data file has been logically put together and breaks down like this:

Bio – Biological sample number
Time interval
R – Technical replicate number
Fraction number
Now we set up the import assignment filter. When we set up the search tasks in Mascot Daemon, we’ll run the searches as my Mascot security user, and we’ll match the results on that, so this import filter will only run on my Mascot search results. We have to select the root node on the tree to associate the results with, in this case I’ve set up a project node for the data called ‘Time course’. Now we want the system to automatically create folders and group the fractions from the same technical replicate for each time point together automatically. We can pull the required pieces of information out of the raw data filename using regular expressions, as shown above. So, under the Time course project node, we’ll generate a folder named after the task name we enter into Mascot Daemon, then under that a folder for the biological sample, then the time interval and finally the technical replicate.

Although in this example, the required data for the import filter is contained in the raw data filename, it could be extracted from any other available field. For example, if the raw files were organised in a meaningful directory structure on your hard drive, the same structure could be re-created in the Explorer tree, automatically.
Before we run the searches, this is what the explorer tree in Mascot Insight looks like
We set up the task in Mascot Daemon as normal, making sure we choose Mascot Distiller as the import filter and, since this is an N15 quantitation dataset, we’ll set it to automatically carry out quantitation using Mascot Distiller as well. When a search is completed, the result will automatically be imported into Mascot Insight, and the Import assignment filter we set up will direct it to the desired location on the explorer tree. When Daemon completes quantitation for a dataset, the quantitation results are also automatically uploaded and imported.
So that when the searches have completed, the explorer tree now looks like this - with a total of 181 folders automatically created under the ‘Time course’ project node, and under each of those folders the eight fractions for each technical replicate at each time interval, all without any manual intervention after the import assignment filter had been set up.
Now we need to merge each of those groups of eight fractions into merged search results. Because the results are grouped together on the tree in the way we want to merge them, to do this we just need to right click over the ‘Time course’ project node (the common parent node) and select the ‘Merge sibling results’ option from the context menu.

The system will then run through the dataset, merging together search results at the same level of the tree so, all of the fractions for each technical replicate will be merged together. More advanced, manual merge options are available if you need to do some additional processing when carrying out the merge — for example, if you have rotated the samples around iTRAQ report ions channels between replicates.
When the merge is completed, the organisation of the search results on the tree will have changed, so that each of the eight fraction search results for each technical replicate of each time interval is now a child node of a new ‘Merged searches’ result node. Now that we have all of the merged search nodes generated, we can use them just like any ‘normal’ search result node and view them in MIRA, or generate reports from them.
Now that we’ve merged the results, we want to set the peptide false discovery rate to 1% for all of the searches. Again, we can do this for all the searches by right clicking over the ‘Time course’ project node and selecting ‘Adjust search settings’
On the form that opens we can set the target FDR and any other search and quantitation settings that we wish. I’m sure you’ll appreciate how powerful it is to be able to change settings like these across tens or hundreds of files from a single form - the changes you make here are saved in the database, so you don’t need to keep adjusting them every time you want view the dataset.

That was a fairly complex example of how you can use an import assignment filter, and often the desired folder structure would be much simpler. However, it does illustrate that you can easily and automatically create a complex, organised folder structure in Mascot Insight.
The system ships with some 30 reports designed specifically for proteomics data. These cover a wide range of areas such as result comparison, quantitation and quantitation clustering/grouping, Gene Ontology analysis, Interactions database analysis and general graphing reports such as scatter plots. You can also write your own reports. In a later presentation, Richard will be showing you some examples of the available reports.
So, in summary, Mascot Insight is a new server based application – it can be installed on the same server as your Mascot server, but currently only if you’re running Mascot on a Windows system (we are planning to release a Linux version). It allows you to easily organise your search results, and allows automatic import and assignment of Mascot search results. It allows you to merge Mascot search results, including all types of quantitation dataset.

The system ships with a large number of reports built into it, which cover areas such as protein and peptide level comparisons between datasets, various quantitation based reports including reports for comparison, ratio clustering and quality control. You can carry out Gene Ontology analysis of your data, including looking for enriched and depleted GO terms compared with a reference proteome and generate reports based on molecular interactions databases in the PSI-MITAB format.

The results of these reports are easily exported from the system, and there are various other export options.

And, if you have a specific report you want to be able to generate, you can write your own reports in Java using a common API across different result formats. Richard will be giving an example of this in his presentation later on.

Finally, Mascot Insight is not expensive to license – licensing cost is related to the number of CPUs licensed for the associated Mascot server, not ‘per seat’.