MASCOT Integra

Data management for Proteomics

ABRF 2005
Mascot Integra: Data management for proteomics

• What is Mascot Integra?
• What Mascot Integra is not
• Security and Electronic signatures in Mascot Integra
• Instrument integration in Mascot Integra
• Designing and running proteomics experiments
• Results and reporting in Mascot Integra.

Mascot Integra is our new solution for proteomics sample and data management.

During this talk, I will be discussing these topics
**What is Mascot Integra?**

- Fully functional ‘out-the-box’ solution for proteomics workflow and data management
- Support for all the major mass-spectrometry data systems
- Powered by the Sapphire™ LIMS package from LabVantage Solutions Inc
- Oracle 9 database
- Scaleable to the largest projects.

Mascot Integra is supplied as a ready to run system. It does not require the extensive setup and customisation associated with a traditional LIMS package.

Rather than re-invent the wheel, we have partnered with LabVantage Solutions Inc, (www.lims.com). Their Sapphire LIMS package provides the sample tracking and workflow modelling functionality for Mascot Integra.

Using the Oracle database management system enables the database to scale efficiently as your data management requirements grow.
What is Mascot Integra?

• Laboratory Information Management for
  – Sample tracking
  – Isolation / fractionation / purification
  – Splitting / combining
  – 1D and 2D Chromatography
  – 1D and 2D Electrophoresis
  – Image analysis
  – Spot picking
  – Digestion / derivatisation
  – MS and MS/MS analysis
  – MS data reduction
  – Mascot database search
  – Data warehouse
  – Result reporting
  – Data mining

We aim to support all aspects of the workflows associated with Proteomics laboratory processes.
What Mascot Integra is not

• An enterprise LIMS
  – Does not require a customised configuration or extensive consultancy
  – Won’t run your entire organisation
  – No inventory management or reagent re-ordering
  – Not a substitute for instrument data systems

• As expensive as an enterprise LIMS.

Mascot Integra will supply the LIMS functionality required to run a proteomics facility. It is not an "enterprise" LIMS system, and does not contain functionality for (say) genomics experiments
LabVantage Solutions Inc

- LabVantage LIMS expertise
  - 24 year track record
  - Scale up to LabVantage Enterprise LIMS products
  - Call on World-Wide Professional Support Services
**Mascot Integra architecture**

- **3 tier system**
  - Oracle database server
  - Sybase Enterprise Application Server running a J2EE web application
  - All user functionality available through Internet Explorer

All Mascot Integra functionality is accessible through a standard web browser.
Mascot Integra complies with the requirements of FDA CFR21 part 11 ERES
Online help

• Extensive context sensitive help available on web-based help system
Security in Mascot Integra

- Role and project/study based security
- Assign roles to users
- Page access limited by user role
- Assign users to projects and studies
- Access to experiments limited by study membership.

Users can be assigned to multiple projects/studies and assigned multiple roles (e.g. a user can be both a Study Manager and an Analyst)
To illustrate this, the administrator has full access to the system and sees all the possible links from the Mascot Integra site-map. The study manager has more limited access to the system (for example, the project links are not visible to them). The technician sees only the links required to perform specific tasks.
In addition to limiting access to complete pages, individual controls on a page can be altered according to the user's role. Here we have the same page (a list of available experiments) as viewed by a technician and a study manager.
**Instrument Integration**

- Not “running” the instrument from the LIMS
- Handled via Sample / work sheet exchange
- Comes with a range of master sample sheets for the main instrument types, data systems and manufacturers
- Flexible – you can design your own master sample sheets and sample sheet templates
- Output CSV, TXT (tab delimited) and Microsoft Excel files.

Instrument integration is handled via sample sheet exchange.

Our aim is to minimise the requirement for copying data from the output of one instrument data system to the input of another. This reduces the time involved in setting up instrument runs and the opportunity for errors.
For a new instrument, there is a one-time setup to design a master sample sheet containing all the possible data elements supported by the instrument data system.

In addition to standard data types (text, numeric etc) you can specify a column as an SDC link. This means that the column defines a live link back to data held in a field in the Mascot Integra database. The data retrieved can be determined by the sample identifier associated with the row of the sample sheet.
Once the master sample sheet has been defined, a subset of the columns is selected for a sample sheet template.

At this stage, you can specify default values for the columns.
A wide variety of MALDI target plates are used by different instruments. To accommodate additional or specialist plate types, Mascot Integra even allows you to define new plate types.
The format is self explanatory. The definition on this slide produced the plate shown in the next slide...
<table>
<thead>
<tr>
<th>Sample ID</th>
<th>Description</th>
<th>Total Content</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>S-050129-0001</td>
<td>Salmon Digest</td>
<td>89</td>
<td>ul</td>
</tr>
<tr>
<td>S-050129-0002</td>
<td>Yeast extract</td>
<td>50</td>
<td>ul</td>
</tr>
<tr>
<td>S-050129-0003</td>
<td>S. pombe culture</td>
<td>0</td>
<td>ul</td>
</tr>
<tr>
<td>S-050129-0004</td>
<td>S. pombe lysate resuspended in 2X buffer</td>
<td>0</td>
<td>ul</td>
</tr>
<tr>
<td>S-050129-0005</td>
<td>S. pombe lysate resuspended in 2X buffer</td>
<td>0</td>
<td>ul</td>
</tr>
<tr>
<td>S-050129-0006</td>
<td>S. pombe lysate resuspended in 2X buffer</td>
<td>40</td>
<td>ul</td>
</tr>
<tr>
<td>S-050129-0007</td>
<td>Whole cell extract 6.5mg/ml 130 concentration</td>
<td>130</td>
<td>ul</td>
</tr>
<tr>
<td>S-050129-0008</td>
<td>Whole cell extract 6.5mg/ml 150 concentration</td>
<td>150</td>
<td>ul</td>
</tr>
<tr>
<td>S-050129-0009</td>
<td>Whole cell extract 6.5mg/ml 150 concentration</td>
<td>150</td>
<td>ul</td>
</tr>
<tr>
<td>S-050129-0010</td>
<td>S. pombe cells</td>
<td>0</td>
<td>ul</td>
</tr>
<tr>
<td>S-050129-0011</td>
<td>S. pombe culture resuspended in 2X buffer</td>
<td>0</td>
<td>ul</td>
</tr>
<tr>
<td>S-050129-0012</td>
<td>S. pombe culture resuspended in 2X buffer</td>
<td>0</td>
<td>ul</td>
</tr>
<tr>
<td>S-050129-0013</td>
<td>Purified S. pombe whole cell extract 49.5 ul</td>
<td>49.5</td>
<td>ul</td>
</tr>
<tr>
<td>S-050129-0014</td>
<td>Purified S. pombe whole cell extract 139.5 ul</td>
<td>139.5</td>
<td>ul</td>
</tr>
<tr>
<td>S-050129-0015</td>
<td>Purified S. pombe whole cell extract 149.5 ul</td>
<td>149.5</td>
<td>ul</td>
</tr>
<tr>
<td>S-050129-0016</td>
<td>Purified S. pombe whole cell extract 149.5 ul</td>
<td>149.5</td>
<td>ul</td>
</tr>
<tr>
<td>S-050129-0017</td>
<td>S. pombe culture resuspended in 2X buffer</td>
<td>2 mm</td>
<td></td>
</tr>
</tbody>
</table>
Experiment design:

- Experiments are broken into a series of experimental tasks
- Experimental tasks have inputs and outputs (e.g. Sample, Plate)
- Experimental tasks can be linked together based on their inputs and outputs
- Each task is then associated with a series of pages which model and setup the physical task
  - While running a task, the user can add task and experiment notes (free text) at any time
- A comprehensive library of flexible proteomics tasks is provided.

Individual tasks in an experiment represent discrete laboratory processes e.g. running a 2D gel, loading a 96 well plate, enzyme digestion etc.

For the experimental task of loading a MALDI target plate, the input would be a a sample list, and the output would be a loaded target.

Experimental tasks can have multiple inputs and outputs.
A complete workflow can be saved as a new experimental task, which is available from the task list menu bar. Any default values which were entered against the experimental tasks would also be saved in this new template. This provides a simple method for setting up the default workflows for your laboratory.
Running experiments:

• Now we'll go through setting up and running a common type of proteomics experiment
  – 2D Gel analysis
    • Image analysis with Nonlinear Dynamics Phoretix™ Software
    • MS run
    • Defining Raw data processing with Mascot Distiller
    • Setting up a batch search with Mascot Daemon
    • Running and importing Mascot searches
    • Examining the results
Rather than running through the entire work flow, we’ll just be looking at the circled tasks. These define the Mascot Integra interfaces to the NonLinear Dynamics Phoretix package, to the mass-spec data system, and to the Mascot search engine.
In this workflow, we are tracking the loading of the gels with samples, and the capture of the stained gel images. The information is transferred between Mascot Integra and Phoretix using the Phoretix XML Exchange Format.

Spot detection and picklist generation (for spot cutting robots) is handled by Phoretix. This information is the imported back into Mascot Integra.
Nonlinear Dynamics Phoretix integration 2
We can choose to import all the detected spots or just those spots which we have chosen for picking.
Next, we select a pre-defined template for a Bruker TOF instrument samplesheet

Some values are automatically filled in from the database

Export as CSV or xls format file for import into Bruker’s FlexControl data system
Peak detection with Mascot Distiller

- Mascot distiller used for automated peak detection
- Used in conjunction with Mascot Daemon for automated search submission
- Import of peak lists created by other software is supported.
Mascot Distiller offers optimised peak detection for all the popular raw data formats.
Peak detection with Mascot distiller continued...

The Mascot Distiller library is used as a data import filter by Mascot Daemon
Mascot search submission

- Automated search submission using Mascot Daemon
- Multiple Mascot daemon clients running as services
- Parameter sets are held centrally in the Mascot Integra database, not in local parameter files.

Mascot search submission is handled by Mascot Daemon.

A common arrangement is to have a Mascot Daemon client on each instrument data system, all communicating with the Mascot Integra database. Users assign processing and search tasks to the relevant Mascot Daemon client via the Mascot Integra interface.

Parameter sets are held in the central database tables, ensuring that the parameter sets used by all the Mascot Daemon clients are identical.
This looks like Mascot Daemon, but it is actually a Java Applet hosted in the web browser. This can be used for monitoring and managing all of the search tasks on all of the Mascot Daemon clients.
The Java applet is also used to manage the central sets of search parameters.
Results are automatically parsed and imported into Mascot Integra as each search completes.
All the standard Mascot reports can be generated from the data held in the database. In addition, individual protein and peptide matches can be annotated and approved.
Approval information is flagged up on the reports, such as Protein View...
... And Peptide View
One advantage of tracking workflows in Mascot Integra is that all of the relationships between parent and child samples are available. Here, we drill down from a gel spot sample...
... to the corresponding mascot protein hit
Filtering reports

- Flexible report filters
- Ships with a range of useful filters
  - E.g. limit returned results to significant matches
- User definable
- Requires some knowledge of SQL.

Because all of the results data is held in relational database tables, we can define flexible filters to select results for a report.
This filter will only display protein hits containing one or more peptide matches that include the consensus sequence for phosphorylation (serine or threonine followed by proline).
Another report uses BLAST to cluster protein hits from multiple searches into groups according to their peptide matches.
**Custom reports**

- Supports generation of Microsoft Excel reports
- Requires some knowledge of SQL query language
- Example reports will be supplied, users will be able to design and upload their own custom reports
- Database product – other reporting tools can be used

The interface for custom reports is Microsoft Excel. Almost everyone has access to and some familiarity with Excel. It is a very powerful tool that allows you to put together and format all types of reports.

Mascot Integra comes with example reports and tutorials for building your own reports.
In this example of a custom report, a MudPIT dataset has been searched against the IPI Human database and also the same database with all the protein sequences reversed.

Any matches to the reverse database should be treated as false positives. Peptide matches are then extracted and plotted against their expectation values in Excel, displaying the true and false positive rates as a function of expectation value.
Mascot Integra ships as a turn-key system on IBM hardware

Shipped as turn key system running on IBM xSeries 226

• Dual 3.2 GHz Xeon processors
• 2Gb RAM
• 4 x 146.8 GB SCSI RAID
• DDS/5 tape drive
• Windows Server 2003

Pricing based on
• number of named users
Summary

- Laboratory Information Management for proteomics
- *Not* an enterprise LIMS
- Oracle database
- All user functionality through a simple, clear web-browser interface
- Role and project/study membership based security
- Intuitive graphical experiment design
- Flexible custom reporting using Microsoft Excel.
Acknowledgements

• Mark Weeks, John Sinclair, Richard Jacob
  – Ludwig Institute for Cancer Research, London UK
• Katheryn Resing
  – University of Colorado, USA
• LabVantage Solutions Inc