BASE & LCB’s Data Warehouse

Adam Ameur
The Linnaeus Centre for Bioinformatics

November 5\textsuperscript{th}
Lund, Sweden
The LCB DWH Team
Jan Komorowski, Project leader
Herman Midelfart, Project manager
Adam Ameur, Developer
Jakub Orzechowski-Westholm, Developer
Ola Spjuth, Former member
Vladimir Yankovski, Database administrator

WCN Microarray Platform
Hanna Göransson
Anders Isaksson

BMC Computer Department, UU
Nils-Einar Eriksson
Emil Lundberg
Background

Why BASE?

- Microarray experiment-oriented collection of data,
- MIAME compliant,
- Open source,
- Good extendibility.

GOALS:

- integration with gene annotation profiling,
- support of genomic data mining processes,
- on-line services for microarray users.
Conceptual Design

Advantages:
- separation of hardware,
- separation of users,
- independent development,
- connection to many LIMS'.

Problems:
- keeping data consistent
Splitting functionality

LIMS (BASE I)
- Creation of BIOASSAYSets
- Application Runs
- Experiment Export

LCB DWH (BASE II)
- Uploading files
- Changing information about samples, hybridizations, rawBIOASSAYs
Implementation

The EXPORT of data

<table>
<thead>
<tr>
<th>Field</th>
<th>Op</th>
<th>Value</th>
<th>Buttons</th>
<th>Translated value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Add/Update</td>
</tr>
</tbody>
</table>

Presets: Save current as new preset: [ ] Ok

<<prev next>> 1 of 2 hits, 30 per page

<table>
<thead>
<tr>
<th>Name</th>
<th>Date added</th>
<th>Owner</th>
<th>Group</th>
<th>World</th>
<th>Raw data sets</th>
<th>BioAssaySets</th>
<th>BioAssays</th>
<th>Channels</th>
</tr>
</thead>
<tbody>
<tr>
<td>v_Exp-ALL-BIT 2003-06-11 vox</td>
<td>MAcourse (mv)</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>w_Experiment 2003-06-14 voxk</td>
<td>none (...)</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Export only one marked to Data Warehouse. [GO]

- SampleAnnotationType
- Sample
- SampleAnnotation
- Extract
- Label
- LabeledExtract
- Hybridization
- HybridizedLabeled

- ImageAcquisition
- Reporter
- RawBioAssay
- Experiment
- ExperimentRawBioAssay
- RawBioAssayData

THE LINNAEUS CENTRE FOR BIOINFORMATICS
http://www.lcb.uu.se
The LCB Data Warehouse is up and running on SUN Fire V880:

- 8 UltraSPARCIII processors
- 16 GB of RAM
- 600 GB of disk space

Operating system: SUN Solaris
Web server: Apache 1.3.26
Database server: MySQL 3.23.45
PHP interpreter: PHP 4.3.3
Plugins

R - wrapper

- Arrayplots
- Normalizations*
- Normalization between arrays*
- B-statistics*
- B-statistics with dupcor*

* - Bioconductor functions
Future development

- BASE (LIMS)
  - Local/Remote
- Unigene
- Locuslink
- Gene Ontology
  - Web Resource

Processing:
- reconcile
- derive
- export

Processing:
- parse
- reconcile
- derive
- transform
- export

- BioAssay Mart
- Reporter Mart
- GO Mart

Ad hoc query tools
- Report writer
- Bioinformatics applications
- Plug-ins
- Visualization Tools

Results
Future development

- Integration with gene annotation profiling
- Good connectivity to Bioinformatics tools
- Implementation of more plugins
- Representation of data analysis results
- More secure, reliable and user-friendly service