LabKey Server: An open source platform for scientific data integration, analysis, and collaboration

Mark Igra
Partner, LabKey Software
marki@labkey.com
Presentation Topics

- Why Scientific Data Integration
- Examples of Usage
  - Common requirements for scientific data integration
- Architecture
- Future Directions
Why Scientific Data Integration

- **Data Volume**
  - Hundreds of millions of results from thousands of high throughput assay runs

- **Data Variety**
  - Clinical, Demographic, Assay & Specimen Data
  - Scientific data annotation

- **Collaboration**
  - Investigators aren’t all in the same place
  - Need secure, selective data sharing

- **Broad Range of Analyses**
  - Queries, Reports, Domain Specific Tools
  - Custom Novel Applications
Core team came from software industry, not bioinformatics

Moved to Hutchinson Center to work on proteomics for biomarker discovery

Expanded from there into

- More assays
- Study data management
- Lab Data Management
- Network Data Management
Problems Faced
- Number of reads from MS2-based proteomics assay were exploding & difficult to handle on existing tools
- Analysis Pipeline on cluster was difficult to optimize

Solution
- The original LabKey Server (CPAS)
- Pipeline to run MS2 analysis jobs via web browser and load results into Database
- Web based analysis tools to view, combine & share results
- Lots of Data
  - More than 90,000 MS2 Runs
  - More than 700,000,000 peptide identifications

### Proteomics Examples

**Blast**

<table>
<thead>
<tr>
<th>m/z</th>
<th>Intensity</th>
<th>Charge</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>55.532</td>
<td>116.0965</td>
<td>2+</td>
<td>13</td>
</tr>
<tr>
<td>100.1105</td>
<td>215.2291</td>
<td>+1</td>
<td>14</td>
</tr>
<tr>
<td>136.645</td>
<td>272.2810</td>
<td>+3</td>
<td>13</td>
</tr>
<tr>
<td>180.1836</td>
<td>359.3592</td>
<td>+4</td>
<td>13</td>
</tr>
<tr>
<td>349.8530</td>
<td>698.6980</td>
<td>+5</td>
<td>13</td>
</tr>
<tr>
<td>370.3739</td>
<td>755.7499</td>
<td>+6</td>
<td>13</td>
</tr>
</tbody>
</table>

**Quantitation**

Currently showing elution profile for 2+. Show: 1+ 3+ 4+ 5+ 6+ Edit elution profile selection

**Scale Graph**

- X Start: 200
- X End: 1800

**Graph**

- Various peaks and ions labeled with charges (b^+, y^+)
- Mass values and intensities indicated for each peak.
Flow Cytometry

- **Problem**
  - Lab using Flow Cytometry to measure intracellular cytokines in many samples
  - Per-run quantity of data increasing
  - Need consistent analysis within an experiment type
  - Cross-run analyses
  - Each experiment type may have different statistics

- **Solution: LabKey Flow**
  - High-throughput flow analysis engine
  - Loading of flow statistics
  - Adaptable data model for varying analysis types
  - Query tools for analyzing data

- Shulman et al, Cytometry A, Sep 8 2008
### Flow Dashboard > labkey-analysis >

**labkey-demo.xml analysis**

**Run Comment:** [Type to enter a comment]

### SHOW GRAPHS

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>DETAILS 119142.fcs</td>
<td>0</td>
<td>9,764</td>
<td>97.64</td>
<td>3,713</td>
<td>1,917</td>
<td>3,711</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DETAILS 118813.fcs</td>
<td>0</td>
<td>9,770</td>
<td>97.7</td>
<td>2,357</td>
<td>1,888</td>
<td>2,357</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DETAILS 118981.fcs</td>
<td>0</td>
<td>9,807</td>
<td>98.07</td>
<td>2,338</td>
<td>1,848</td>
<td>2,337</td>
<td>47</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DETAILS 118547.fcs</td>
<td>0</td>
<td>9,555</td>
<td>95.55</td>
<td>2,065</td>
<td>1,703</td>
<td>2,065</td>
<td>2</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DETAILS 119199.fcs</td>
<td>0</td>
<td>9,728</td>
<td>97.28</td>
<td>2,336</td>
<td>1,907</td>
<td>1,977</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DETAILS 119154.fcs</td>
<td>0</td>
<td>9,543</td>
<td>95.45</td>
<td>2,541</td>
<td>1,629</td>
<td>2,541</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DETAILS 118836.fcs</td>
<td>0</td>
<td>9,496</td>
<td>94.96</td>
<td>504</td>
<td>466</td>
<td>503</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DETAILS 119010.fcs</td>
<td>0</td>
<td>9,817</td>
<td>98.17</td>
<td>3,804</td>
<td>1,824</td>
<td>3,801</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DETAILS 118762.fcs</td>
<td>0</td>
<td>9,743</td>
<td>97.43</td>
<td>5</td>
<td>131</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DETAILS 118756.fcs</td>
<td>0</td>
<td>9,753</td>
<td>97.53</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DETAILS 119171.fcs</td>
<td>0</td>
<td>9,662</td>
<td>96.62</td>
<td>2,200</td>
<td>2,196</td>
<td>1,661</td>
<td>5</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DETAILS 118801.fcs</td>
<td>0</td>
<td>9,706</td>
<td>97.06</td>
<td>2,251</td>
<td>2,087</td>
<td>2,251</td>
<td>2</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DETAILS 118754.fcs</td>
<td>0</td>
<td>9,787</td>
<td>97.87</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
View Graphs

Flow Dashboard > labkey-analysis > labkey-demo.xml analysis

Run Comment: Type to enter a comment

[Hide Graphs] [Large Graphs] [Medium Graphs] [Small Graphs]

Details > 119142.fcs 9,764 97.64 3,713 1,917 3,711 1 0

Details > 118613.fcs 9,770 97.7 2,357 1,886 2,357 1 0
labkey-demo.xml analysis

Run Comment: [Type to enter a comment]

Available Fields:
- Flag
- Created
- Run
- Analysis Script
- Compensation Matrix
- Statistic
- Graph
  - {<APC-A>}
  - {<Alexa 680-A>}
  - {<FITC-A>}
  - {<PE Cy55-A>}
  - {<PE Cy7-A>}

Selected Fields:
- Name
- Flag
- S:Count
- S:%P
- 4+:Count
- 8+:Count
- Lv/L:<PE Cy55-A>:<PE Tc RD-A>:
- 3+:<PE Cy55-A>:<FITC-A>:
- 4+:<SSC-A:<APC-A>:
- 4+:<SSC-A:<Alexa 680-A>:
- 4+:/(1IFNg+&IL2+):Count
- 4+:/(1IFNg+&IL2+&IL4+&TNFa+):Count
- 4+:/(1IFNg+&IL2+&IL4+&TNFa+):Count

Editing an unsaved view.

DELETE | VIEW GRID | SAVE

Name | Flag | S:Count | S:%P | 4+:Count | 8+:Count | 4+:/(1IFNg+&IL2+):Count | 4+:/(1IFNg+&IL2+&IL4+&TNFa+):Count | 4+:/(1IFNg+&IL2+&IL4+&TNFa+):Count

DETAILS | 119142.fcs | 9,764 | 97.64 | 3,713 | 1,917 | 3,711 | 1 | 0 | 1 | 0

S/Lv/L | S/Lv/L/3+ | S/Lv/L/3+/4+ | S/Lv/L/3+/4+
Problems faced

- Combine many data types for HIV Vaccine studies
  - Clinical Response Forms (CRF), Specimens, Many Assays
- Enable secure collaboration for scientists worldwide
- Allocate & distribute valuable specimens

Solution

- Secure web portal for HIV Vaccine Enterprise Data
- Used by several networks to share data
  - CHAVI, CAVD, HVTN, HPTN (3000 Users Worldwide)
- Core software was written by LabKey
- SCHARP runs Atlas
  - Defines available data and relationships
  - Manages security and permissions
  - Manages data loading
  - Builds custom modules

Nelson et al, BMC Bioinformatics, March 2011
Atlas Data Flows

Data Providers
- LIMS
- Labs
- Forms

Data Viewers
- Leadership
- Labs
- Collaborators

Sample Info

Atlas Web + Database Servers

CRF
- DOB
- BP svs
- BP dia
- Notes

Labid USpeci Txtpid Parusp Drawdm Drawdd Drawdy

Data Providers

Atlas Web + Database Servers

Data Viewers
## Combining Data

### Demographic Data

<table>
<thead>
<tr>
<th>Participant</th>
<th>Gender</th>
<th>CD4+</th>
<th>Hemoglobin</th>
<th>Specimen</th>
<th>Available?</th>
</tr>
</thead>
<tbody>
<tr>
<td>249318596</td>
<td>Male</td>
<td>420</td>
<td>12.2</td>
<td>5390.204.36</td>
<td>Yes</td>
</tr>
<tr>
<td>249320107</td>
<td>Female</td>
<td>344</td>
<td>13.2</td>
<td>2493.404.56</td>
<td>Yes</td>
</tr>
<tr>
<td>249320127</td>
<td>Male</td>
<td>897</td>
<td>21.0</td>
<td>2493.323.33</td>
<td>Yes</td>
</tr>
<tr>
<td>249320489</td>
<td>Male</td>
<td>234</td>
<td>16.4</td>
<td>2493.789.44</td>
<td>No</td>
</tr>
<tr>
<td>249320897</td>
<td>Female</td>
<td>893</td>
<td>12.9</td>
<td>5266.440.61</td>
<td>Yes</td>
</tr>
</tbody>
</table>

### Assay Data

### Specimen Data
### Dataset: Binding Antibody, All Visits

<table>
<thead>
<tr>
<th>Participant ID</th>
<th>Sequence Num</th>
<th>Draw Date</th>
<th>Enrollment Date</th>
<th>Assay ID</th>
<th>Isotype</th>
<th>Antigen</th>
<th>Coat Lot</th>
<th>Dilution</th>
<th>Concentration</th>
<th>Reading</th>
<th>St. Dev.</th>
<th>No Antigen</th>
<th>No Ant. St. Dev.</th>
<th>QC Titer</th>
<th>Plate ID</th>
<th>Network</th>
<th>Protocol</th>
</tr>
</thead>
<tbody>
<tr>
<td>101.0</td>
<td>20-Jun-05</td>
<td>20-Jun-05</td>
<td>JTL61-38</td>
<td>IgA</td>
<td>CONT</td>
<td>140 CF</td>
<td>#2</td>
<td>10</td>
<td>0.394</td>
<td>0.013</td>
<td>0.015</td>
<td>151450.0</td>
<td>3</td>
<td>CHAVI</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>102.0</td>
<td>27-Jun-05</td>
<td></td>
<td>JTL61-38</td>
<td>IgA</td>
<td>CONT</td>
<td>140 CF</td>
<td>#2</td>
<td>10</td>
<td>0.387</td>
<td>0.034</td>
<td>0.015</td>
<td>151450.0</td>
<td>3</td>
<td>CHAVI</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>103.0</td>
<td>05-Jul-06</td>
<td></td>
<td>JTL61-38</td>
<td>IgA</td>
<td>CONT</td>
<td>140 CF</td>
<td>#2</td>
<td>10</td>
<td>0.408</td>
<td>0.015</td>
<td>0.015</td>
<td>151450.0</td>
<td>3</td>
<td>CHAVI</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>104.0</td>
<td>12-Jul-06</td>
<td></td>
<td>JTL61-38</td>
<td>IgA</td>
<td>CONT</td>
<td>140 CF</td>
<td>#2</td>
<td>10</td>
<td>0.417</td>
<td>0.0070</td>
<td>0.015</td>
<td>151450.0</td>
<td>3</td>
<td>CHAVI</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>105.0</td>
<td>20-Jul-06</td>
<td></td>
<td>JTL61-38</td>
<td>IgA</td>
<td>CONT</td>
<td>140 CF</td>
<td>#2</td>
<td>10</td>
<td>0.436</td>
<td>0.014</td>
<td>0.015</td>
<td>151450.0</td>
<td>3</td>
<td>CHAVI</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>106.0</td>
<td>14-Aug-06</td>
<td></td>
<td>JTL61-38</td>
<td>IgA</td>
<td>CONT</td>
<td>140 CF</td>
<td>#2</td>
<td>10</td>
<td>0.520</td>
<td>0.0020</td>
<td>0.015</td>
<td>151450.0</td>
<td>3</td>
<td>CHAVI</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>107.0</td>
<td>07-Sep-06</td>
<td></td>
<td>JTL61-38</td>
<td>IgA</td>
<td>CONT</td>
<td>140 CF</td>
<td>#2</td>
<td>10</td>
<td>0.634</td>
<td>0.015</td>
<td>0.015</td>
<td>151450.0</td>
<td>3</td>
<td>CHAVI</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>108.0</td>
<td>10-Oct-06</td>
<td></td>
<td>JTL61-38</td>
<td>IgA</td>
<td>CONT</td>
<td>140 CF</td>
<td>#2</td>
<td>10</td>
<td>0.227</td>
<td>0.015</td>
<td>0.015</td>
<td>151450.0</td>
<td>3</td>
<td>CHAVI</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>109.0</td>
<td>29-Nov-06</td>
<td></td>
<td>JTL61-38</td>
<td>IgA</td>
<td>CONT</td>
<td>140 CF</td>
<td>#2</td>
<td>10</td>
<td>0.574</td>
<td>0.015</td>
<td>0.015</td>
<td>151450.0</td>
<td>3</td>
<td>CHAVI</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>110.0</td>
<td>26-Feb-07</td>
<td></td>
<td>JTL61-38</td>
<td>IgA</td>
<td>CONT</td>
<td>140 CF</td>
<td>#2</td>
<td>10</td>
<td>0.413</td>
<td>0.015</td>
<td>0.015</td>
<td>151450.0</td>
<td>3</td>
<td>CHAVI</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>101.0</td>
<td>20-Jun-06</td>
<td>20-Jun-05</td>
<td>JTL61-38</td>
<td>GP41</td>
<td>J1S/6J13R</td>
<td>10</td>
<td>0.833</td>
<td>0.034</td>
<td>19790.0</td>
<td>4</td>
<td>CHAVI</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>102.0</td>
<td>27-Jun-06</td>
<td></td>
<td>JTL61-38</td>
<td>GP41</td>
<td>J1S/6J13R</td>
<td>10</td>
<td>0.531</td>
<td>0.024</td>
<td>19790.0</td>
<td>4</td>
<td>CHAVI</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Notes:**
- Table contains data on participant IDs, draw dates, enrollment dates, assay IDs, isotypes, antigens, coat lots, dilutions, concentrations, readings, standard deviations, no. antigens, no. antigen standard deviations, QC titers, plate IDs, networks, and protocols.
<table>
<thead>
<tr>
<th>Derivative Type</th>
<th>Additive Type</th>
<th>Enrollment</th>
<th>Week 2</th>
<th>Week 3</th>
<th>Week 4</th>
<th>Week 8</th>
<th>Week 12</th>
<th>Week 16</th>
<th>Week 24</th>
<th>Week 36</th>
<th>Week 48</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blood (Whole)</td>
<td>Tempus</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>PBMC Cells, Viable</td>
<td>Acid Citrate Dextrose</td>
<td>5</td>
<td>2</td>
<td>5</td>
<td>1</td>
<td>6</td>
<td>2</td>
<td>7</td>
<td>3</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Serum</td>
<td>Serum Separator</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Whole Blood Pellet, Specify Methodology</td>
<td>EDTA</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plasma</td>
<td>Acid Citrate Dextrose</td>
<td>24</td>
<td>10</td>
<td>7</td>
<td>6</td>
<td>5</td>
<td>1</td>
<td>11</td>
<td>3</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>Plasma</td>
<td>EDTA</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Tear Flo Strips</td>
<td>None</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cryopreserved primary cells from Non-Blood Spec Type, Viable</td>
<td>None</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Fluid Portion from a Non-Blood Specimen Type</td>
<td>None</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td></td>
<td></td>
<td>3</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>None</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Lab Data Flows

1. Request Specimens
2. Run Assay
3. Upload Spreadsheet and Metadata
4. Automated Transform and Analysis (optional)
5. Review
6. Copy to Study
Assays: upload example

Piehler et al, BMC Immunology, May 2011
Dozens of small applications needed for different groups

- Most used data already in LabKey System
- But custom workflows, reports & analysis required

SCHARP has detailed knowledge of the requirements

LabKey provided an API and simple application building tools
Internal Quality Control for Sample Processing

Specimen Processing Lab: ACSA - Iquitos
Site Historical PBMC IQC, Cell Yield
6/2009 through 5/2010

<table>
<thead>
<tr>
<th>N</th>
<th>Mean</th>
<th>Stnd Dev</th>
<th>&lt;0.8</th>
<th>0.8 to 3.2</th>
<th>&gt;3.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>180</td>
<td>1.35</td>
<td>0.36</td>
<td>1.7%</td>
<td>98.3%</td>
<td>0.0%</td>
</tr>
</tbody>
</table>
RV144 Followup Study Tracking System

- Tools for tracking information about studies

![Study Tracking System](image_url)

**Study Data Entry: Borrow cytokine chemokine**

**Stat Task List**

- **Sample Set Tasks**
  - Task Description: Sample set(s) for Borrow Chemokine
  - Date Completed: 28-Dec-2009

- **Sample Set Tasks**
  - Task Description: Sample set(s) for Borrow Cytokine

**LDO Task List**

- **Assay Tasks: Borrow Chemokine**
  - Task Description: Data Uploaded for Borrow Chemokine
  - Date Completed: [edit]

- **Assay Tasks: Borrow Cytokine**
  - Task Description: Data Uploaded for Borrow Cytokine
  - Date Completed: [edit]
Problem
- 30 years of daily data on thousands of animals
- Clinical staff & vets need health record
- Researchers need scientific data
- Colony Management is an ongoing problem

Solution
- EHR is a specialized, ongoing “study”
- LabKey enhanced scalability of study solution
- LabKey enhanced API
- LabKey wrote tools to transfer data into new EHR within minutes of entry into old EHR
  - Now LabKey EHR is only one in use
- Wisconsin Primate Center built custom views & reports to analyze the data
## Primate EHR

### Table

<table>
<thead>
<tr>
<th>Id</th>
<th>Current Room</th>
<th>Current Cage</th>
<th>Problem Number</th>
<th>Date Observed</th>
<th>Date Resolved</th>
<th>Category</th>
<th>Remark</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td></td>
<td></td>
<td>2011-10-04</td>
<td>2011-10-07</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td>2011-05-16</td>
<td>2011-05-20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td>2010-11-22</td>
<td>2010-11-30</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td>2008-05-22</td>
<td>2008-05-23</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td>2007-09-05</td>
<td>2007-09-07</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td>2007-06-05</td>
<td>2007-06-08</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td>2005-01-12</td>
<td>2005-01-26</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td>2003-12-03</td>
<td>2003-12-03</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td>2003-03-17</td>
<td>2003-04-10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td>2002-11-19</td>
<td>2002-11-19</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td>2002-07-22</td>
<td>2002-07-22</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td>2002-03-07</td>
<td>2002-03-19</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Weight

![Weight Chart](chart.png)

**Weight (kg)**

**Min Date:**

**Max Date:**

[Refresh]
Problem
- Sharing clinical trial amongst distributed teams
- Publish data in support of studies

Solution
- ITN TrialShare
- Publish Study Capabilities
  - Support for point in time access to study data
  - Copy protected data
  - Interactive access to publication data
- Improved Visualization
Public access to bio-repository samples is now available.

- TrialShare presents information to the scientific community about ITN clinical samples, including important trial information and sample availability. Review ITN samples for closed studies, where investigators from the scientific community are invited to submit research proposals through this website. Proposals for use of ITN samples will undergo scientific review by the ITN and does not guarantee its selection. Additional information on procedures for application and review can be obtained by downloading the ‘Request for Proposals’.
Visualization with Participant Groups
nPOD – Network for Pancreatic Organ Donors

**INFORMATION**

**DEMOGRAPHICS**
- Age/Gender: 22 year old Male
- Race: White
- BMI: 24.09

**DIABETES**
- Donor Type: T1D
- HLA A*01/24, B*08/18, DR*03/06, DQ*01/02
- C Peptide Levels: 0.025
- Duration (Years): 14

**INVESTIGATOR REPORTS**

Network for Pancreatic Organ Donors with Diabetes (nPOD): developing a tissue biobank for type 1 diabetes
Diabetes/Endocrinology Research and Reviews, 2012
PMID: 22655477

Expression of the enteroviral capsid protein VP1 in the islet cells of patients with type 1 diabetes is associated with induction of protein kinase R and downregulation of MIP-1
Richardson SJ, Leete P, Bone AJ, Foulis AK, Morgan NG
Diabetologia, 2013
PMID: 23226423

Pancreas organ weight in individuals with disease-associated autoantibodies at risk for type 1 diabetes
Campbell-Thompson M, Wasserman C, Montgomery EL, Atkinson MA, Kaddis J.
Journal of the American Medical Association, 2012
PMID: 22233891

**ACTIVE PROJECTS**

- **Anderson**
  - Investigator: Dr. D Anderson
  - Description: Virus detection in pancreas and other tissues

- **Clark**
  - Investigator: Dr. H Clark
  - Description: Non-Study Related Lab Processing

**TISSUE INVENTORY**

- PanHead: 3
- PanTail: 3
- PLN-A: 1
- PLN-C: 1
- Serum: 1
- Spleen: 1

**PATHOLOGY IMAGES**

- Various images of tissue samples and pathology slides.
Ancillary Studies

Study Initiation
1. Hypothesis generation
2. Proposal review
3. Creation of protocol or plan

Study Execution
4. Consent verification or acquisition
5. Retrieval of existing data
6. Delivery & analysis of specimens
7. Data integration

Results Sharing
8. Data/specimen repatriation
9. Publication

Nelson et al, Ancillary study management systems a review of needs
BMC Med Inform Decis Mak Jan 2013
Data integration is the unifying theme
- Samples to Assays
- Assays to Subjects
- Subjects to Studies
- Translational Medicine involves integrating data from “Molecules to Populations” (Kuhn et al 2008)

Data types to integrate are constantly evolving
- Both file and “structured data types”
- Can’t rebuild new systems for new data types

Tools grow as fast as data types

Scientists need to share data
- Central lab doing work for distributed clients
- Distributed group of labs contributing
- Need to manage many users
Basic Connectivity

LabKey Server

File System

LabKey Database
(PostgreSQL/MS SQL)

LabKey Schemas
Major Feature Areas

- Data Model
- Security
- Extensibility
- File Handling
- Assays
- Data pipelines to get data in
- Reports & visualizations to get data out
- Auditing
Data Model Goals & Requirements

- Consistent
  - User interface
  - Reporting & Querying
- Seamless data integration
  - Select columns across study, assay, specimen data
- Extensible
  - Data types are dynamic & user defined
  - Existing data stored in external systems
- Rich Column Types
  - Concepts, Lookups, Formatting, URL, Validation
  - Out of Range (e.g. < 30), SAS-style missing values
- Secure: See security slide
Data Exposed via Relational Model

Virtual Schema

- Modules expose a “User Schema”
- User Schema maps virtual tables/columns to queries on underlying database
- User Schema depends on user’s permissions

Users typically query by editing grid views

- Add columns from related tables
- Filter, Sort

Full SQL Queries Available

- SQL translated to underlying relational database
Query Processing

Custom Application (Perl, R, JavaScript) Using API

SQL Query or Table + Column List

Tabular Data (including additional metadata)

LabKey Server

LabKey Query Service

Translated SQL

Relational DB

Tabular Data
Major Feature Areas

- Data Model
- **Security**
- Extensibility
- File Handling
- Assays
- Data pipelines to get data in
- Reports & visualizations to get data out
- Auditing
Security Model Requirements

- Pervasive
  - Applies to data no matter how you get at it
  - Modules, Search, Reporting, Queries, API, Files
- Easy Partitioning of Data by security context
- Easy Administration
  - Uniform administration across applications
  - Integrates with existing authentication systems
- Easy implementation by custom modules
- Extensible where necessary
Data and files are visible in folders
Major Feature Areas

- Data Model
- Security
- Extensibility
- File Handling
- Assays
- Data pipelines to get data in
- Reports & visualizations to get data out
- Auditing
Extensibility: Developers, Admins & Users

- **Users:**
  - Create custom views of their data using graphical tools

- **Administrators**
  - Configure LabKey Server
  - Set up projects and security.
  - Customize schema via data sets, assays, etc.
  - Create SQL queries and reports

- **Application Module Developers**
  - Build Custom Modules solving particular problems
  - Modules may be private or shared, open source or not
  - File-based modules (HTML + Script + SQL + XML)
  - Java Modules (usually built by LabKey)

- **Platform Developers:**
  - Build the core LabKey Server and key modules
  - Java code that is distributed in the LabKey Server distribution.
  - LabKey Software employees with few exceptions
- Needs to be available on all data in the system
- Need flexible reporting infrastructure
  - LabKey cannot predict reports needed
  - Integrate R
- Interactive reports are preferable
  - Grids are interactive now
  - R reports allow filtering/redisplaying data
  - Interactive graphs under development
File Handling Requirements

- Files are the fundamental unit of data exchange
  - Contain structured & unstructured data
  - Sometimes huge
  - Sometimes many little files

- Need to find the files
  - Full text search + properties

- Need rapid & flexible file transfer

- Need to analyze & load structured data within files
  - Extensible actions including “load data”
Audit service available to all modules

Modules responsible for auditing actions
  - Easy implementation for basic data auditing
  - Would be nice if this were automatic
  - Many modules have custom audit events
    - (e.g. login, password reset etc)

Each audit event type can have custom associated data
  - E.g. list updates store old values in audit trail
Futures

- Data Integration Improvements
  - Data Loading
  - ID Management
  - Smart Union of related datasets
- Lab Data Management
- Data Integration for EMRs
Data Integration is a common need across scientific endeavors

LabKey Server is a platform for addressing these challenges

- Secure
- Built for specialized challenges of scientific data
- Extensible
- Open

We continue to work hard to advance the platform capabilities and usability
Partners & Funders
- Martin McIntosh (FHCRC)
  - NCI
  - Canary Foundation
- Steve Self (SCHARP)
  - CHAVI (NIH)
  - HVTN (NIH)
  - CAVD (Gates Foundation)
- David O’Connor (Wisconsin)
  - Primate EHR (ARRA)
  - Genotyping Tools (NIAID)
- Parag Mallick (USC)
- Michael Katze (UW)
- Immune Tolerance Network
- nPOD

LabKey Development
- Matthew Bellew
- Adam Rauch
- Britt Piehler
- Josh Eckels
- Kevin Krause
- Brendan MacLean (now UW)
- Nick Shulman (now UW)
- Karl Lum
- Nick Arnold
- Cory Nathe
- Ben Bimber
- Alan Veniza

Documentation
- Elizabeth Nelson
- Steve Hanson

Test
- Trey Chaddick
- Elizabeth Van Nostrand

Management & Operations
- Britt Piehler
- Kristin Fitzimmons
- Peter Hussey
- Ren Lis
Any questions?

Mark Igra
marki@labkey.com
If you use LabKey Server for your research, please reference one of these publications about the platform:

