

Putting the pieces together: Integrated Research Data Management Using the LabKey Server

Jeanne Holden-Wiltse, MPH

Co-Director of Data Management for RPRC, NYICE, CBIM, CIBEM

Andy Straw, MS

BLIS Leader Developer/Administrator

Center for Integrative Bioinformatics and Experimental Mathematics

University of Rochester Medical Center



Agenda

- Background
- Motivating Problems
- Bio-Lab Informatics Server (***BLIS***)
- Integration – Value added
- LabKey Extension Mechanisms
- Considerations

Background

- Center for Integrative Bioinformatics & Experimental Mathematics, Dept of Biostatistics & Computational Biology, UPMC (Hulin Wu - PI)
 - Provide data management and integration for infectious disease studies
 - Perform statistical and bioinformatics analyses
 - Develop novel statistical methodology
 - Mathematical modeling of host responses to influenza incorporating systems biology approaches

Motivation

- 2005 CBIM
- 2007 NYICE
- 2007 URM C D-CFAR, 2012 CFAR
- 2011 RPRC
- Several R01's

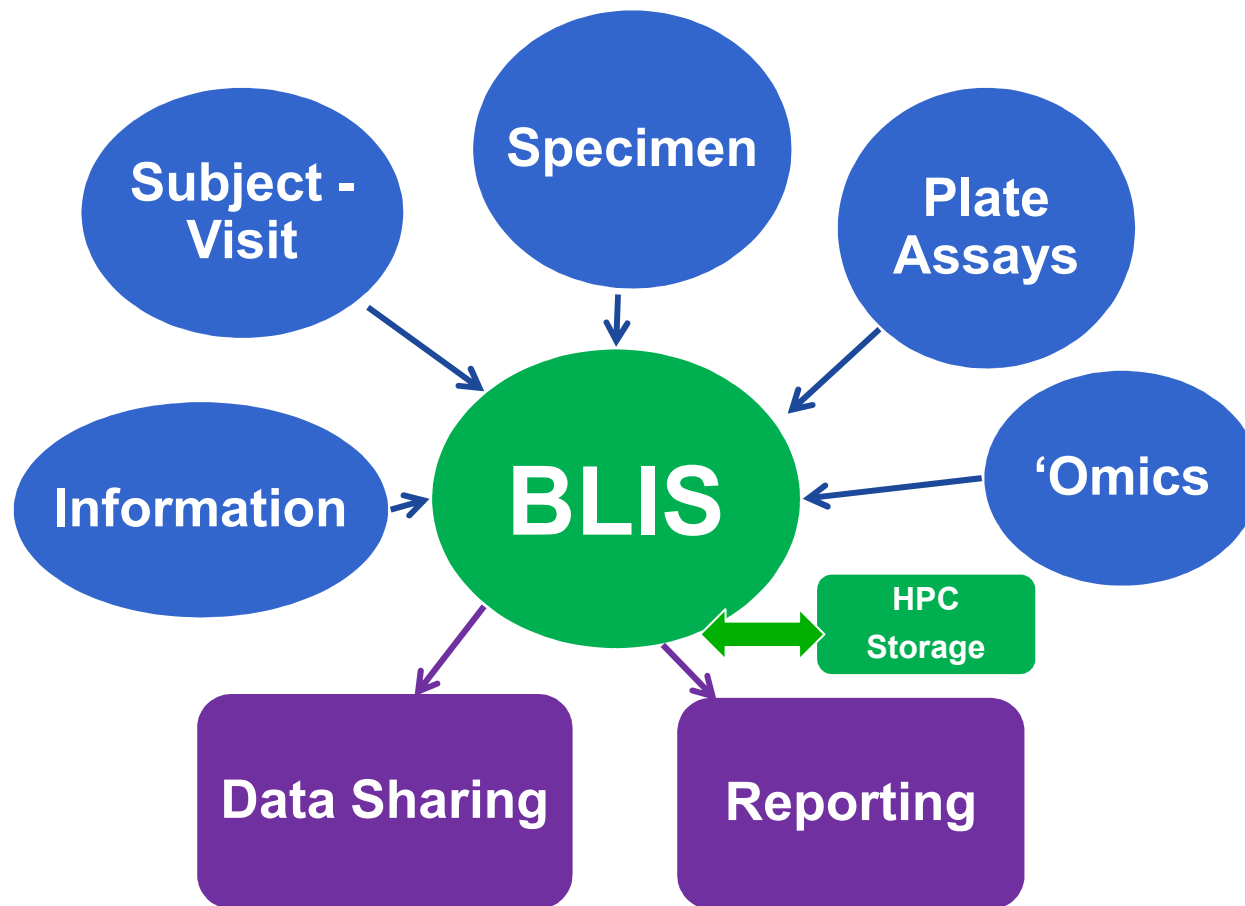
Bio-Lab Informatics Server (*BLIS*)

- Customized LabKey Server (2010)
- Data Management Operations
 - Developers
 - Statisticians/Analysts
- 150 Users
- 8 Projects
- 27 studies (mostly longitudinal)
 - 3885 subjects, 14803 visits, 345 datasets, 129817 specimen vials, 48323 .fcs files

Integration/Management Goals

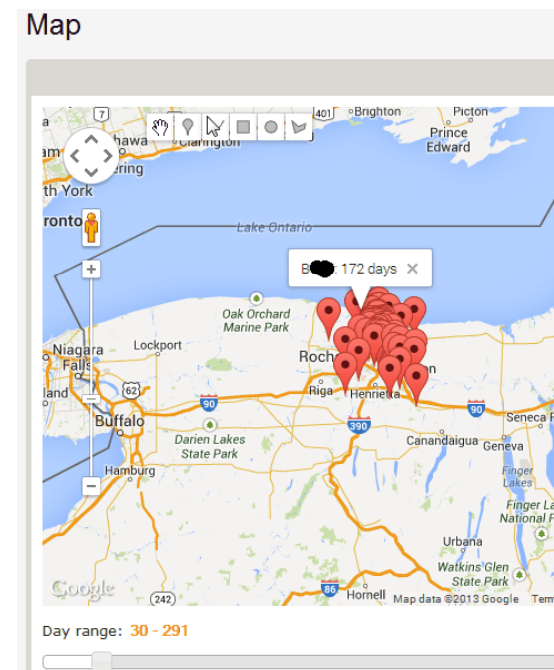
- Manage information/data from multiple sources, formats
- Timely, complete and clean
- Feedback
- Data Sharing
- Flexible, Adaptable, Responsive to changing research needs
- BLIS as active research member

Integration Requirements



Integration - Information

- Study/Lab Portals
 - Protocol Docs, SOPs
 - Sharepoint
 - iCalendar
 - Google Maps
- Collaborative workspace
 - Manuscript development



Integration – Subject/Visit

- Human, Animal
- BLIS eCRFs
 - validation controls
- txt, csv, xls from
 - REDCap
 - CROs
 - Labs

Start Page > eCRF Dashboard >
Enrollment

Enroll New Participant

Cohort:

birth
Choose Cohort
birth
hospital
B105

Assigns Next Participant ID

Enroll New Subject

EDIT MANAGE HI

Demographics

Subject Information Volunteer Number: 10-0011- []

Gender: ☐ Male ☐ Female DOB: [] (mm/dd/yyyy)

Ethnicity: ☐ Hispanic or Latino ☐ Not Hispanic or Latino

Race: (may check more than one)

☐ American Indian / Alaskan Native
☐ Asian
☐ Native Hawaiian or Other Pacific Islander
☐ Black or African American
☐ White

Comments: []

Eligibility Checklist

Volunteer Number: 10-0011- []

Screen date: [] (mm/dd/yyyy)

Visit 1 date: [] (mm/dd/yyyy)

Inclusion Criteria: All answers must be YES or N/A for the volunteer to be eligible

1. Aged between 18 and 32 years, inclusive, or 60 years or older. ☐ No ☐ Yes

2. No prior history of laboratory documented infection with novel H1N1 virus or immunization with novel H1N1 vaccine. ☐ No ☐ Yes

Integration - Specimen

- Biospecimen Inventory
 - Requests
 - Tracking/Reporting
 - Contract closeout
- Unique specimen ID
- LDMS (LIMS)
- xls (small studies)

Specimen Report: By Specimen Type

Cohort filter: 1
 Participant Group: All Groups
 Availability status: All vials
 Base view: Base report on all vials
 Specimen type: One report per primary type
☐ Hide Empty Columns
☒ Vial Counts
☒ Total Volume

[REFRESH](#) [PRINT VIEW](#) [EXPORT TO EXCEL](#)

BLD (Vial Count/Total Volume)

	Visit 0 (Day 0)	Visit 1 (Day 3)	Visit 2 (Day 7)	Visit 3 (Day 10)	Visit 4 (Day 28)
1	18/5.0000008E7	21/6.7600008E7		24/8.0000008E7	20/6.1200007E7
	23/7.8000008E7	21/6.7600008E7		28/1.0200008E8	21/6.7600007E7
	15/3.5000008E7	15/3.5700008E7		19/5.5000008E7	22/7.0980007E7
	14/3.3000008E7		24/8.1600008E7	22/7.4200008E7	21/6.5000008E7
	14/3.0000007E7	13/2.6E8		22/7.0000008E7	24/8.0000007E7
	5/2800000.5	1/3600000.0		4/1.2600001E7	4/540000.5
	12/2.0400007E7	2/2.4E7			14/3.3000007E7
	14/3.0000007E7			28/1.0400008E8	22/7.3500007E7
	11/2.2400006E7			8/8.0E7	22/7.2800007E7
	12/2.1500008E7	5/8.2E7		16/4.4480008E7	
	22/7.0000008E7				
	17/4.6800008E7			15/3.6400008E7	16/4.0800008E7
	4/1.0800001E7	3/1900000.5		4/5300001.0	4/5600001.0
	11/1.7700008E7	10/1.0300008E7		14/3.1200008E7	1/4.37E7

Integration – Assays

- ELISA
- ELIspot
- HAI, MN
 - Preload specimen IDs
- qtPCR
- Specimen Processing/extraction
- Frequent plate design changes

ELIspot Assay Pipeline

Assay Dashboard

Zand Lab Home

Data Views

- Vaccine Elispot Runs with CTL download file links
- Vaccine Elispot Well Data with CTL Download Link
- Vaccine Elisa Well Data with Links
- Vaccine Elisa Runs with Links

Tools

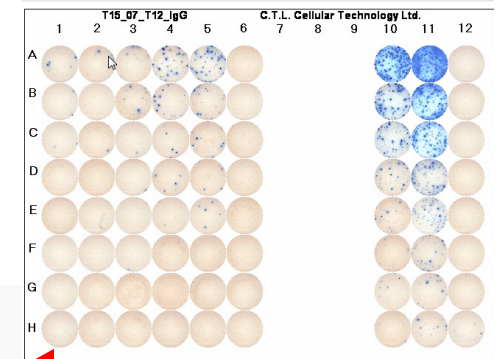
- Upload Files
- Add Review Assay QC Info

Processed Data

- Estimated Antibody Titers for Vaccine Elisa

"Raw" Data Views

VaccineElispot_CTL_Link



QUERY	VIEWS	EXPORT	PRINT	PAGE SIZE
Flag	Plate ID	File Download	Plate Image	Created
	CBIM DAY 7.tsv	File Download	Plate Image	2011-07-22 14:17
	T01_01_VAX.tsv	File Download	Plate Image	2011-07-22 14:18
	T01_03_VAX.tsv	File Download	Plate Image	2011-07-22 14:18
	T01_05_IGG.tsv	File Download	Plate Image	2011-07-22 14:18

"Processed" Data Views

QUERY ▾ VIEWS ▾ EXPORT ▾ PRINT PAGE SIZE ▾					1 - 100 of 264 Next ▸ Last ▹		
Specimen ID	IgA	IgG	IgM	Capture Antibody	id	day	Raw Data
F0600C-03	4344.133767	245103.5523	7823.973572	Flu	6	0	Raw Data
F0601C-03	3324.345772	204259.5111	6378.824536	Flu	6	1	Raw Data
F0602C-03	3731.383903	210172.9144	6180.524461	Flu	6	2	Raw Data
F0603C-03	3499.503257	195785.2407	6209.227121	Flu	6	3	Raw Data
F0604C-03	4008.103719	207829.9171	6771.131168	Flu	6	4	Raw Data
F0605C-03	2986.968298	196418.0812	6918.295646	Flu	6	5	Raw Data
F0606C-03	4136.821791	230601.0944	10919.94324	Flu	6	6	Raw Data

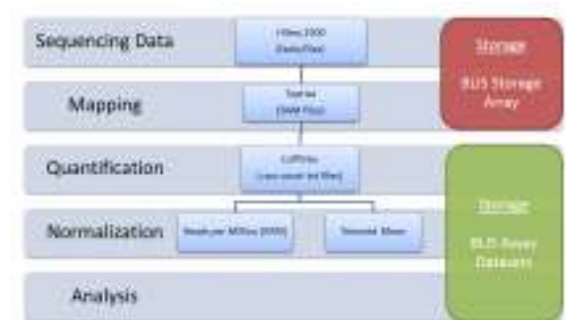


ROCHESTER
MEDICAL CENTER



Integration - Omics

- Flow cytometry
- Microarray
- RNAseq
- 16S rRNA
- Imaging (coming)
- Files pipeline (raw and processed)
 - HPC
 - Large file archive



Integration - Reporting

- Monitoring/ Visit Check Lists
- QC checking
- NIAID reports
- Visualization - Charting
- Messaging
- Issue Tracker for data discrepancies

Participant - ****

SEARCH RESULTS: 1 | CONFIRMED VISIT: 1

	Enrollment	1 Bath	1 Month	RSV Screening	Positive Eligibility Visit (Day 1)	Positive Eligibility Visit (1 Week)	Positive Eligibility Visit (4 Weeks)
RSV Diagnostic							
Demographics							
Microbiome Extraction							
Sample collected by							
Via Audit Log							
RSV Diagnostic Consent							
RSV Diagnostic DOB							
RSV Diagnostic Eligibility							
Visit Age							
Consent							
Eligibility							
Demographics							
Microbiome							

Integration – Data Sharing

- Datasets for analysis
 - SAS, R, Excel
- NIH data repositories
 - Specific formats
- Security –
 - Fine grain data sharing
 - FISMA requirements

LabKey Extension Mechanisms We Use

- LabKey SQL
- LabKey import/export features
 - To/from Excel, TSV, CSV
 - Archive formats (Specimen; Flow Analysis; etc.)
- LabKey client APIs
 - Java (and HTTP) for external apps (used by developers)
 - JavaScript on wiki pages (used by end users)
- LabKey module mechanism
 - File-based
 - Java-based (use LabKey server APIs)

What We Have Implemented Using LabKey Extension Mechanisms

1. Framework for electronic case report forms
2. Automated import of specimen data from LDMS
3. Excel templates to minimize specimen data entry
4. QC dashboard for flow cytometry file metadata
5. Import of custom flow cytometry analysis file formats
6. Java-based custom assay modules
7. Extracting, formatting data for publishing to public databases

Where We Use Extension Mechanisms

	Solution	Extension Mechanisms				
		SQL	Im/ Ex	Java API	JS API	Mod
Subject / Visit	1. eCRF Framework	X			X	
Specimen	2. Import from LDMS		X	HTTP		
	3. Excel Templates	X	X	X		
Assay / Omics	4. QC dashboard for flow file metadata	X			X	
	5. Import custom flow analysis file formats		X			
	6. Java-based custom assay modules	X				X
Data Sharing	7. Extract, format, publish to public DBs	X		X		

1. Framework for Electronic Case Report Forms

Problem: Collect large amount of data from clinic/hospital visits

- Data is “similar but different” across multiple visits, multiple cohorts (need to re-use datasets, pieces of forms)
- Must be easy for nurses to collect data “live” via iPads
- Data validation is complex, but critical
 - Can’t “go back” to subjects later if data are missing or incorrect
 - eCRF data is source of record – no paper forms
- LabKey’s built-in “Insert New” and “Edit” forms are not enough

1. Framework for Electronic Case Report Forms

Solution: eCRF JavaScript framework that supports:

- Building sets of forms composed of reusable sections
- Dynamic showing/hiding based on visit and/or cohort
- Intuitive navigation
- Complex validation

Approach:

- LabKey DataSets – one per “section”
- LabKey JavaScript client API + wiki pages
- 3rd-party JavaScript libraries (e.g., Backbone for MVC)
- Framework of base “classes” => new forms require minimal code

1. Framework for Electronic Case Report Forms

Study Visit

SECTIONS

- Visit Date (& Time)
- Vital Signs**
- Blood Draw
- PLEX Therapy

eCRFs ▾

Vital Signs Participant ID: [REDACTED]

Systolic Pressure: mm Hg ☐ Not Available

Diastolic Pressure: mm Hg ☐ Not Available

Heart Rate: beats/min ☐ Not Available

Respiration: breaths/min ☐ Not Available

Height: inches ☐ Not Available

Weight: pounds ☐ Not Available

Temperature: °F ☐ Not Available

1. Framework for Electronic Case Report Forms

1 Month (Baseline) Visit

SECTIONS

- Visit Information
- Samples Collected
- Adverse Events
- Diary Booklet

SOP ⓘ ADD NOTE 🗨

UNDO CHANGES ⚠ SAVE 💾

Visit Information Participant ID: [REDACTED] Visit: 10

Evaluation Date
Format: mm/dd/yyyy

Evaluation Time
Military time. Format: hh:mm

Evaluation Location ☐ Home ☒ Clinic ☐ Hospital

Physical Findings

General appearance ☐ Well ☒ Mild ☐ Moderate ☐ Severe

If Moderate or Severe illness: ☒ Respiratory ☐ Non-Respiratory ☐ N/A

IF RESPIRATORY ILLNESS CHECKED, FINISH BASELINE VISIT AND THEN GO TO RSV ILLNESS SCREENING (FORM 2)

Weight: kg ☐ Not Available?

2. Automated Import of Specimen Data from LDMS

LDMS:

- Sophisticated specimen inventory management system
- Used by several of the studies we support

Problem: LabKey studies use internal specimen repository

- Need to export data from LDMS, import to LabKey (multiple studies)
- Data in LabKey must be “reasonably” current

2. Automated Import of Specimen Data from LDMS

Solution: Combination of small manual steps, plus custom application to automate export/import process

Approach:

- Use LDMS reporting and export features to export data (manual)
- Use custom Java application to:
 - Transform exported data into LabKey specimen archive format
 - Split data into one archive per study
 - Move archive files to LabKey using WebDAV Java API
 - Start pipeline import job for each study using LabKey HTTP interface
- Done daily by developer (5 minutes)

3. Excel Templates to Minimize Specimen Data Entry

Background:

- Some studies we support do NOT use LDMS, but still require use of LabKey specimen repository.
- Sometimes: specimen Ids encode info

Problem: Tedious and error-prone for lab personnel to properly enter all required specimen info, in proper format.

3. Excel Templates to Minimize Specimen Data Entry

Solution:

- Agree with labs on how many of which type of specimens will be generated for each subject-visit combination
- Use custom Java application to generate partially-filled-in Excel templates, including specimen id, and other known info - one row per expected vial
- Lab personnel manually fill in rest (mainly: volume and location)
 - Lab personnel delete rows for any vials not actually generated
- Use custom Java application to transform filled-in Excel files into LabKey specimen archive format
- Use LabKey specimen archive format to import data into LabKey

[illegible]

4. QC Dashboard for Flow Cytometry File Metadata

Background:

- Flow (FCS) files contain metadata in header keywords
 - Useful for storing specimen Id, experimental conditions (panel, stimulation, etc.)
- LabKey Flow module imports FCS files, reads and stores keyword metadata
 - One row per file, one column per keyword
- Labkey Flow module can use keyword value as specimen Id
 - LabKey automatically links to specimen repository of target study

Problem: Some metadata values are entered manually – errors result that must be detected and corrected

4. QC Dashboard for Flow Cytometry File Metadata

Solution: Dashboard page showing all detected errors

Approach:

- LabKey SQL queries (or views) to detect metadata errors
- JavaScript API to invoke queries, display errors on wiki page
- Shared, unambiguous communication of errors (no files, emails)

Downside:

- Metadata errors not detected until files imported =>
Files must be re-imported after errors are corrected

4. QC Dashboard for Flow Cytometry File Metadata

FCSFiles QC

EDIT NEW MANAGE HISTORY PRINT

Ignore Case: ☐
Unchecked: values like "u", "F", "Tmat" will be reported as errors. (Should be "U", "F", and "TMAT")
Checked: values like that will not be reported as errors.

Bad Value for Panel (should be "ICS", "TMAT", or "TPHE")

ANALYSIS FOLDER QUERY VIEWS CHARTS EXPORT PRINT PAGE SIZE

View: default

Parameters: IgnoreCase = 0

Run	Name	Panel
No data to show.		

Bad Value for Stim Condition (should be "U", "S", "ANTICD3", or "LPS")

ANALYSIS FOLDER QUERY VIEWS CHARTS EXPORT PRINT PAGE SIZE

View: default

Parameters: IgnoreCase = 0

Run	Name	Stim Condition
No data to show.		

Bad Value for Sample Type (should be "DC", "UCB", or "HD")

ANALYSIS FOLDER QUERY VIEWS CHARTS EXPORT PRINT PAGE SIZE

View: default

Parameters: IgnoreCase = 0

Run	Name	Sample Type
No data to show.		

Bad Value for Subject Category (should be "F", "HD", "P", or "T")

ANALYSIS FOLDER QUERY VIEWS CHARTS EXPORT PRINT PAGE SIZE

View: default

Parameters: IgnoreCase = 0

Run	Name	Subject Category
No data to show.		

Sample Type DC, but Global ID not found in [REDACTED] study

ANALYSIS FOLDER QUERY VIEWS CHARTS EXPORT PRINT PAGE SIZE

Run	Name	Sample Type	Specimen ID
<input type="checkbox"/>	PROP_20130808_TMAT Specimen_001_DC 030.fcs	DC	J160LBHY-11
<input type="checkbox"/>	PROP_20130808_TMAT Specimen_001_12MO 016.fcs	DC	J160LBHY-11
<input type="checkbox"/>	PROP_20130808_TMAT Specimen_001_12MO 030.fcs	DC	J160LBHY-11
<input type="checkbox"/>	PROP_20130808_TMAT Specimen_001_DC 016.fcs	DC	J160LBHY-11

Sample Type HD, but Global ID not found in [REDACTED] study

ANALYSIS FOLDER QUERY VIEWS CHARTS EXPORT PRINT PAGE SIZE 1 - 5 of 5

Run	Name	Sample Type	Specimen ID
<input type="checkbox"/>	PROP_20130816_ICS Specimen_001_HD 131 UNSTIM.fcs	HD 131	J160PYQ8-04
<input type="checkbox"/>	PROP_20130816_ICS Specimen_001_HD 131 STIM.fcs	HD 131	J160PYQ8-04
<input type="checkbox"/>	PROP_20130816_ICS Specimen_001_UNSTAINED.fcs	HD 131	J160PYQ8-04
<input type="checkbox"/>	PROP_20130815_TMAT Specimen_001_HD 131.fcs	HD 131	J160PYQ8-04
<input type="checkbox"/>	PROP_20130815_TMAT Specimen_001_UNSTAINED.fcs	HD 131	J160PYQ8-04

ANALYSIS FOLDER QUERY VIEWS CHARTS EXPORT PRINT PAGE SIZE 1 - 5 of 5

Sample Type UCB, but Global ID not found in [REDACTED] studies

ANALYSIS FOLDER QUERY VIEWS CHARTS EXPORT PRINT PAGE SIZE

Run	Name	Sample Type	Specimen ID
<input type="checkbox"/>	PROP_20130808_TMAT Specimen_001_CB P4 50.fcs	UCB	J160LBHY-11

Mismatch between Panel and End of Folder Name

ANALYSIS FOLDER QUERY VIEWS CHARTS EXPORT PRINT PAGE SIZE

Run	Name	Panel
No data to show.		

5. Import of Custom Flow Cytometry Analysis File Formats

Background:

- LabKey Flow module supports multiple formats for importing analysis results (primarily: #cells of various types, per sample)
 - FlowJo workspace (exported to XML)
 - FCS Express
 - A LabKey flow analysis archive (zip file) format
- We have investigators using other tools to analyze flow data
 - FLOCK (<https://import.niaid.hih.gov>)
 - SWIFT (<http://www.ece.rochester.edu/projects/siplab/Software/SWIFT.html>)
 - FlowJo, stats exported to Excel, aggregated from multiple workspaces

Problem: How to import these other analyses?

5. Import of Custom Flow Cytometry Analysis File Formats

Solution: Small amount of reusable custom Java code to transform other formats into LabKey's archive format.

- Store analyses in native format in LabKey “Files” web part, for reference
- Transform data from native format to LabKey format via external Java program
- Create analysis archive (zip file), import via LabKey Flow module

6. Java-Based Custom Assay Modules

Background: LabKey supports import of experiment (assay) data

- Concepts of Batch, Run, Results
- General Purpose Assay supports:
 - Tabular data file (Excel, TSV, CSV)
 - One run per file
 - One row per result (i.e., experiment data per sample)
- Other LabKey assays support certain specific file formats

Problem: We need support for:

- Other file formats (e.g., XML)
- Files containing batch and run properties, in addition to results
- Ability to include calculated properties, not just those entered by user or found in file

6. Java-Based Custom Assay Modules

Solution: Use LabKey Module concept to add custom assay types

- File-based: less/no code, less flexible
- Java-based: mostly code, more flexible

Approach:

- Primarily Java-based modules
- Built set of our own base classes that use, extend LabKey server classes; Added some utility classes
- Development of new assay module requires minimal additional code (e.g., developed one in ½ day, < 200 lines of code – mostly config data)
- *Caveat: Assays must fit within certain restrictions (though we have been progressively removing restrictions)*

7. Extracting, Formatting Data for Publishing to Public Databases

Problem: Grants/Contracts require some study data be published to external databases in very specific formats

Solution: Custom external Java application

Approach:

- Use Java Client API to extract data from LabKey
- Use 3rd-party Java APIs to fill in Excel templates provided by external database
- Zip files, then upload via external database web site

Considerations

Lots of hooks provided by LabKey

- Significant learning curve: What they all are, what each is good for
- Significant power and flexibility
- Choice of hook may depend on:
 - The expertise your team members have
 - Where you want to touch the data (outside or inside LabKey)
 - E.g., prefer to fix assay metadata errors at source, but easier to detect them after import into LabKey
- Other hooks we have NOT fully explored:
 - Transformation scripts
 - File-based modules
 - UI customizations (e.g., adding buttons)
 - Study/Folder templates

Considerations

Required amount of software development expertise varies

- Very little:
 - Transforming flow analysis data from one TSV format to another (Java, Perl, Python, R – any will do)
 - QC dashboard for flow metadata (basic SQL + basic JavaScript/HTML)
- A bit more:
 - Automated import of specimen data from LDMS (Java, Swing, LabKey Java client API, 3rd-party Java APIs)
- Even more + LabKey source code + help from LabKey personnel:
 - Java-based custom assays (documentation of server APIs is minimal) (Each developer has own LabKey server, built from source)

Acknowledgements

Director: Hulin Wu, PhD

Co-Director: *Jeanne Holden-Wiltse*, MPH

Database & Software Developers

Andy Straw, MS (Lead, Admin)

Anthony Corbett

Jennifer Dutra

Jeffrey Williams

Alicia Tyrell

University Support

Center Integrated Res Computing

Primary Data Center

Biostatisticians/Bioinformaticians

Hongmei Yang, PhD

Joe Guido, MS

Lu Wang, MS

Sanjukta Bandyopadhyay, MS

Shuang Wu, PhD

Hongyu Miao, PhD

Xing Qiu, PhD

Le Zhang, PhD

Consultant

LabKey Software, Seattle, WA

ALL BLIS USERS

Funding

