

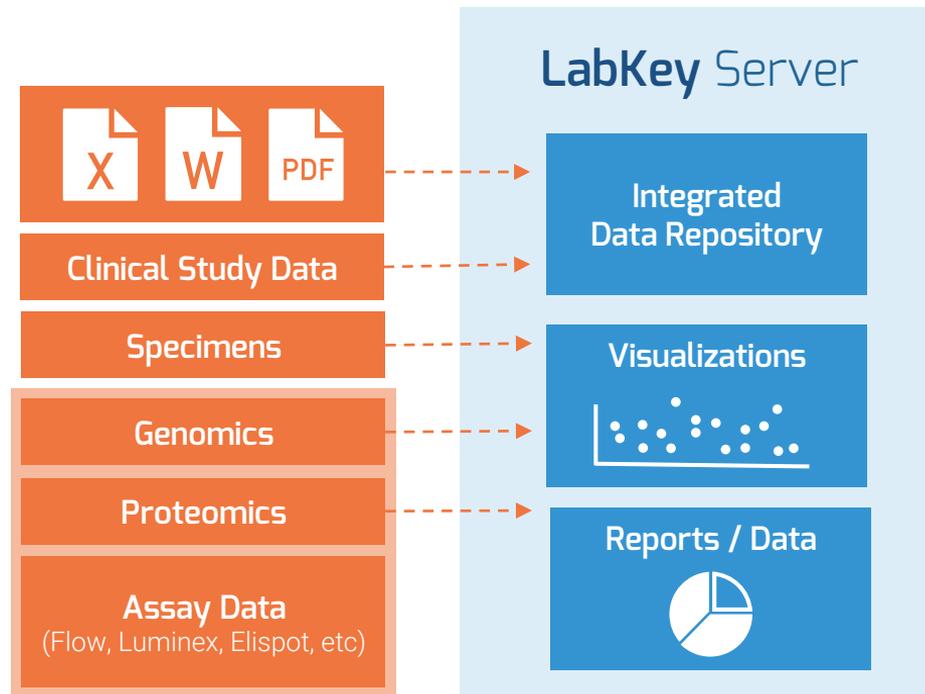
LabKey Data Models

An early challenge that LabKey users encounter is what is the best way to get their data into the server and what are the appropriate data types to use

Data in LabKey



- LabKey server offers a variety of ways to store and organize data
- Different data types offer specific features which make them more or less suited for specific scenarios
- Less structured data may import quicker but be limited in integration options
- Careful planning can improve overall LabKey experience



Factors to consider



- Nature of the data
- Usage scenarios
 - Tabular data : Lists
 - Lineage/derivations : Sample Sets
 - Complex instrument data : Assays
 - Patient/Animal data over time : Datasets
- Other integrations needed
 - Clinical or experimental data integration
 - Data quality or QC features
 - Compliance related functionality
 - Import/export requirements
 - Extensibility or API support
- Often, data structures will be used in conjunction with each other

Lists



- Simplest way to describe and input data
- Create in UI and from file
- Import, export, insert, edit and delete via Data Grid
- Import and export list archive

SAVE CANCEL

List Properties

Name	Properties
Description	
Title Field	<AUTO> ▾
Discussion Links	<input checked="" type="radio"/> None <input type="radio"/> Allow one discussion per item <input type="radio"/> Allow multiple discussions per item
Allowable Actions	<input checked="" type="checkbox"/> Delete <input checked="" type="checkbox"/> Upload <input checked="" type="checkbox"/> Export and print <input type="checkbox"/> Index each item as a separate document <input checked="" type="checkbox"/> Index entire list as a single document <input type="radio"/> Metadata only (name and description of list and file) <input type="radio"/> Data only ? <input checked="" type="radio"/> Metadata and data ?
Full-Text Search Indexing	<input checked="" type="radio"/> Standard title <input type="radio"/> Custom title ? Warning: Do not include fields that contain PHI or PII. <input checked="" type="radio"/> Index all text fields <input type="radio"/> Index all fields (text, number, date and boolean) <input type="radio"/> Index using custom template ? <input type="checkbox"/> Index file attachments

Lists



- Specify fields and field types

List Fields

	Name	Label	Type
<input type="checkbox"/>	<input type="text" value="Key"/>	<input type="text"/>	Integer
<input type="checkbox"/>	<input type="text" value="StreetAddr"/>	<input type="text" value="Street Address"/>	<input type="text" value="Text (String)"/>
<input type="checkbox"/>	<input type="text" value="City"/>	<input type="text"/>	<input type="text" value="Text (String)"/>
<input checked="" type="checkbox"/>	<input type="text" value="State"/>	<input type="text"/>	<input type="text" value="Text (String)"/>
<input type="checkbox"/>	<input type="text" value="Country"/>	<input type="text"/>	<input type="text" value="Text (String)"/>
<input type="checkbox"/>	<input type="text" value="Zip"/>	<input type="text"/>	<input type="text" value="Integer"/>

Choose Field Type

Text (String)

Multi-Line Text

Boolean

Integer

Number (Double)

DateTime

Flag (String)

File

Attachment

User

Subject/Participant (String)

Lookup

Folder

Schema

Table



- Specify field attributes

List Fields

	Name	Label	Type
<input type="checkbox"/> <input type="checkbox"/> 🔍	<input type="text" value="Key"/>	<input type="text"/>	Integer
<input type="checkbox"/> <input type="checkbox"/> ✕	<input type="text" value="Name"/>	<input type="text"/>	Text (String) ▾
<input type="checkbox"/> <input type="checkbox"/> ✕	<input type="text" value="StreetAddr"/>	<input type="text" value="Street Address"/>	Text (String) ▾
<input type="checkbox"/> <input type="checkbox"/> ✕	<input type="text" value="City"/>	<input type="text"/>	Text (String) ▾
<input type="checkbox"/> <input type="checkbox"/> ✕	<input type="text" value="State"/>	<input type="text"/>	lists.State (String) ▾
<input type="checkbox"/> <input type="checkbox"/> ✕	<input type="text" value="Country"/>	<input type="text"/>	Text (String) ▾
<input type="checkbox"/> <input type="checkbox"/> ✕	<input type="text" value="Zip"/>	<input type="text"/>	Integer ▾

ADD FIELD

IMPORT FIELDS

EXPORT FIELDS

INFER FIELDS FROM FILE

Display

Format

Validators

Reporting

Advanced

Description

URL ?

Shown In Display Modes

Grid

Insert

Update

Details



- Create from file

Create new List Study 1

Import from TSV or Excel file.

People.xls

Showing first 5 rows (uncheck column checkboxes to ignore import):

<input checked="" type="checkbox"/> First	<input checked="" type="checkbox"/> Last	<input checked="" type="checkbox"/> Age	<input checked="" type="checkbox"/> Appearance	<input checked="" type="checkbox"/> Sex	<input checked="" type="checkbox"/> Height
Text (String)	Text (String)	Integer	DateTime	Integer	Number (Double)
Fred	Flintstone	37	1960-09-30 00:00	1	1.8
Wilma	Flintstone	35	1960-09-30 00:00	2	1.6
Barney	Rubble	36	1960-09-30 00:00	1	1.7
Betty	Rubble	34	1960-09-30 00:00	2	1.6
Dino	Flintstone	5	1960-09-30 00:00	1	1.0

Lists - uses



- Quick to create
- Can have lookups to other tables
- Good for spreadsheet data
- Good for list of items that other lists or datasets can reference

Assays



- Analyzes complex instrument data files
- Various assay types available in the system
 - Will differ by the experiment type and analytics performed by the server
- New assay types can be introduced
 - Simple 'file based' types
 - Java modules
- Specific visualizations/reports available
- Results can be integrated into a study by aligning participant and time information or by specimen ID



Assay data characteristics



- More structured than lists
- Participant ID and time information is required
- All assay types support custom fields at the batch and run level
- Data typically imported one run at a time
- Inherent batch-run-results organization
- Assay designs can be shared through a folder hierarchy, data can be uploaded into any folder that is in scope for the user

Run Properties?

Fields marked with an asterisk * are required.

Assay Id?

Comments?

Run Data * The ELISA data files must be in the BioTek Microplate Reader Excel file format (.xls or .xlsx extension).
Upload a data file
 No file chosen

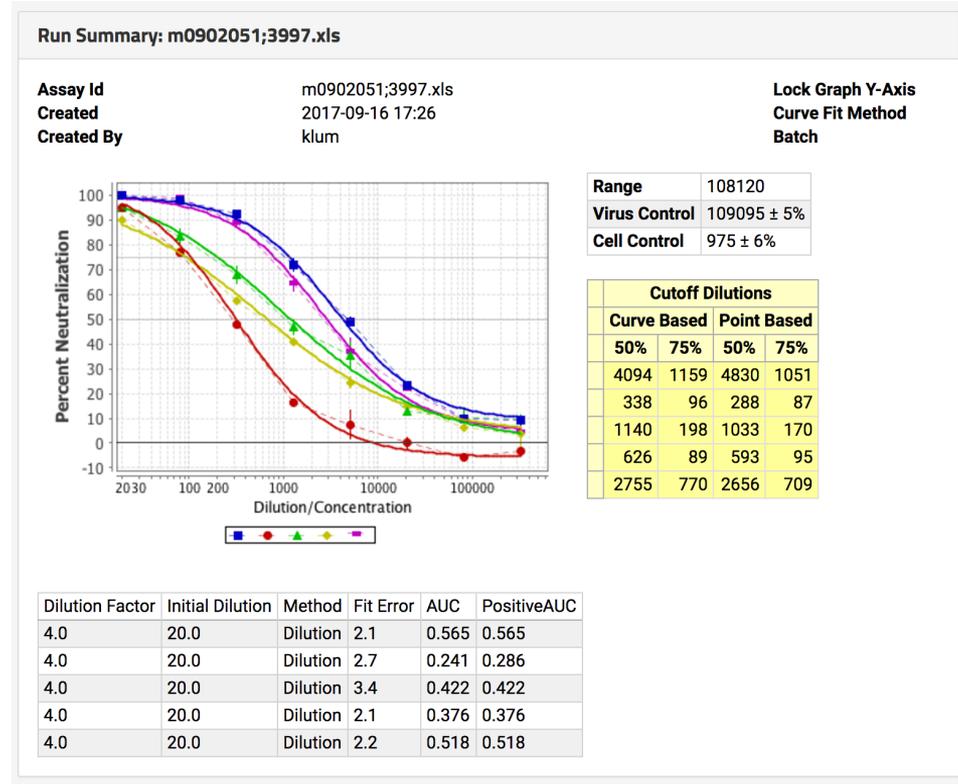
Same? **Specimen ID?** **Participant ID?** **Visit ID?**

Specimen 1	<input type="text"/>	<input type="text"/>	<input type="text"/>
Specimen 2	<input type="text"/>	<input type="text"/>	<input type="text"/>
Specimen 3	<input type="text"/>	<input type="text"/>	<input type="text"/>
Specimen 4	<input type="text"/>	<input type="text"/>	<input type="text"/>
Specimen 5	<input type="text"/>	<input type="text"/>	<input type="text"/>

Assay types available



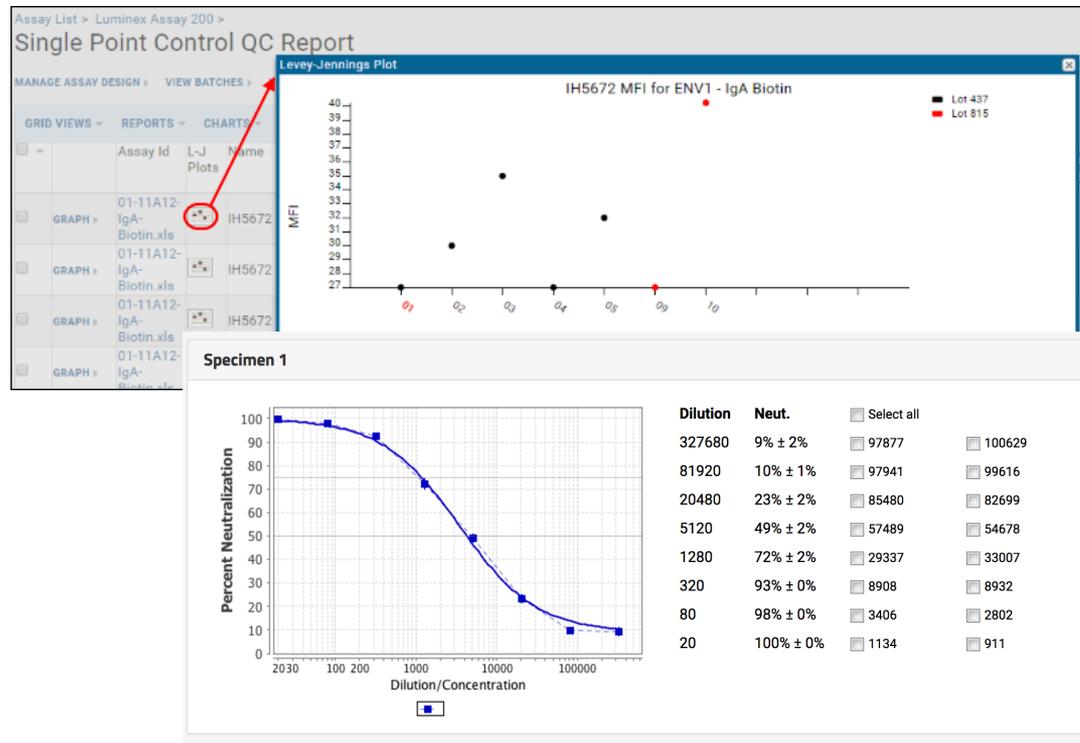
- General (GPAT)
- Luminex
- Flow
- NAb
- ELISpot
- Fluorospot
- Elisa
- Plus others...



Assay QC tools



- Type validation
- Required fields
- Lookups
- Default values
- Field validators
 - Range or regex checks
- Missing value indicators
- Transform scripts
- Assay type specific QC tools
 - Luminex : Levy-Jennings
 - NAb : Well exclusion tool



Additional assay considerations

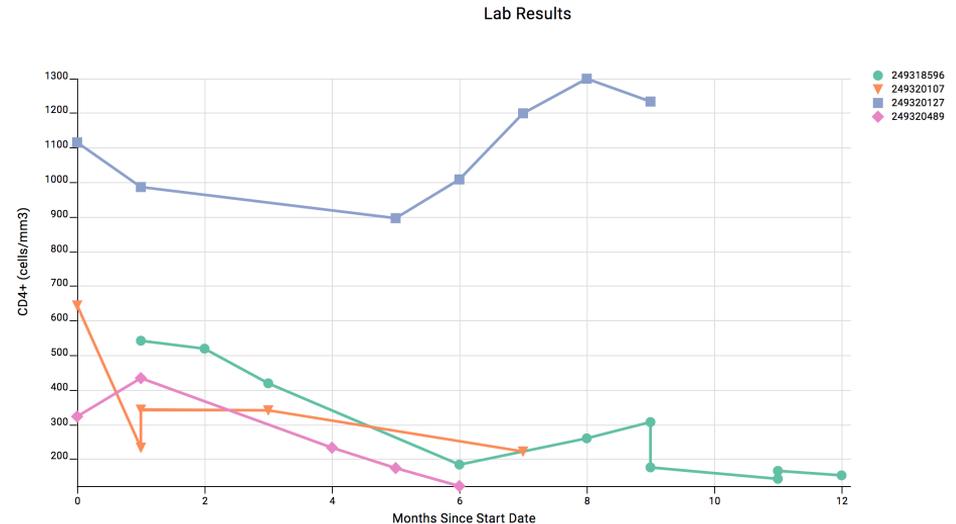


- Assay data can be 'copied' to a study and data is aligned using either participant ID and time or by specimen ID
- Export/Import
 - Run/result data can be downloaded from the data grid
 - API access through LABKEY.Query
 - Assay designs can be exported and imported
 - Assay data and designs (for GPAT) can be copied via folder export/import
- APIs exist to programmatically upload runs
 - LABKEY.Experiment.saveBatch
 - LABKEY.Assay.importRun
- Editable run data
 - Editable results data for the general assay type

Datasets



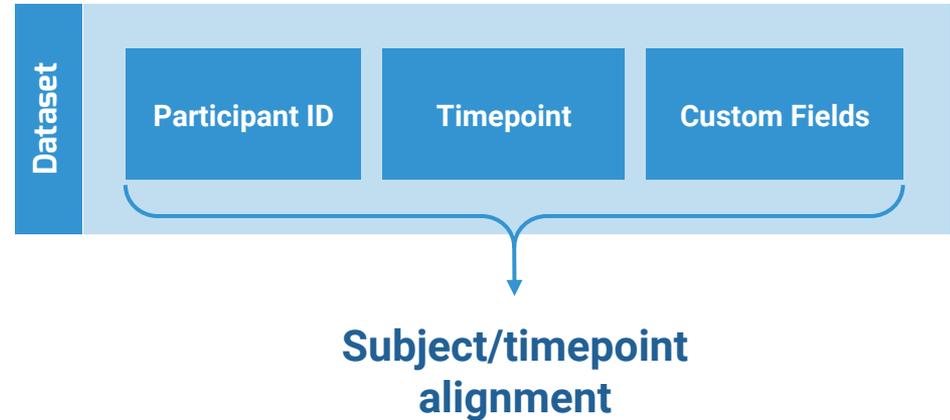
- Integrated as part of a study
 - Designed to track participant data over time
- Can automatically join/lookup to other datasets
 - Create intelligent visualizations and reports
- Can be populated by assay data
- Usually scoped to a single study although in some contexts can be shared
- Query snapshots can be used to populate datasets from other sources
- Specialized visualizations available
 - Time chart, participant report, study overview



Dataset data characteristics



- Participant and time information required
- Custom fields supported
- Demographics datasets can have one row per subject
- Non-demographic datasets can have one row per participant/visit combination or participant/visit/third key combination
- Can integrate with specimen data



Dataset QC tools



- Type validation
- Required fields
- Lookups
- Default values
- Field validators
 - Range or regex checks
- Missing value indicators
- Trigger scripts
- Dataset QC checks

QC States: All data

default This grid view has been modified. REVERT EDIT SAVE

QC STATE

- All data
- Needs Review
- Verified
- Update state of selected rows
- Manage states

	Participant ID	Date	Weight (kg)	Temperature (C)	Systolic Blood xxx/		Pulse	Respi
<input type="checkbox"/>	249318596	2008-06-30	86	36.0			76	64
<input type="checkbox"/>	249318596	2008-08-04	84	37.0			79	67
<input type="checkbox"/>	249318596	2008-09-02	83	37.0			85	68
<input type="checkbox"/>	249318596	2008-11-23	80	36.7		135	85	70
<input type="checkbox"/>	249318596	2009-01-18	79	36.8		142	90	75
<input type="checkbox"/>	249318596	2009-02-17	79	37.3		129	87	79
<input type="checkbox"/>	249318596	2009-03-16	79	38.0		122	80	64
<input type="checkbox"/>	249318596	2009-04-17	78	40.0		122	76	63
<input type="checkbox"/>	249318596	2009-05-16	77	39.0		121	78	65
<input type="checkbox"/>	249318596	2009-06-15	75	39.0		123	76	64

Dataset import/export



- Dataset data can be populated through a variety of pathways:
 - Bulk import : copy and paste tsv or file upload
 - Single row insert/update
 - API
 - ETL
 - Study archive
 - Study reload
- Tools available to publish or create ancillary studies

The screenshot shows a web application window titled "Publish Study". On the left is a navigation menu with the following items: General Setup, Participants, **Datasets** (highlighted), Timepoints, Specimens, Study Objects, Lists, Grid Views, Reports and Charts, Folder Objects, and Publish Options. The main content area is titled "Datasets" and contains the instruction "Choose the datasets you would like to use from the source study:". Below this is a list of datasets with checkboxes: Label, Demographics, ELISpotAssay, GenericAssay, HIV Test Results, Lab Results, LuminexAssay, MicroarrayAssay, NAbAssay, Participation and Genetic Consent, Physical Exam, and Status Assessment. At the bottom of the list, there is a "Data Refresh:" section with three radio buttons: "None" (selected), "Automatic", and "Manual". In the bottom right corner of the window, there are two buttons labeled "PREVIOUS" and "NEXT".

Dataset specific integrations



- Clinical data from external systems can be synchronized
 - REDCap
 - Medidata
 - Datstat
- PHI support added in 17.3
 - Annotated columns can be hidden/blanked if PHI level of user is below configured level
 - PHI annotated columns can be omitted from export





- Manage specimens as part of a study
- Integrate with other study types
- Divide specimens into vials/aliquots
 - Tracks availability and location
 - Handles requests
 - Part of Study
 - Export/Import
 - Publish (including hiding PHI)

Specimen management - import



- Simple repository: import from spreadsheet
- Advanced repository: import archive via pipeline
 - Describes events about new & changed specimens & vials
 - Specifies locations & primary, derivative & additive types
- Imported event data expanded into three tables
 - Specimen Event: all events kept over time
 - Vial: current state of each vial
 - Specimen: current state of each specimen

Specimens - events



Vial History Study 1

Vial Summary

Globally Unique ID 106443199.2404.501
Participant 249320107
Visit Vst 2.0080716E7
Volume 1.0 ML
Collection Date 2008-07-16 00:00
Collection Location Rainier Lab Storage (Repository)
Comments and QC [UPDATE](#) >

[RETURN TO VIAL VIEW](#) >

Vial History



[Compact View](#)

<input type="checkbox"/>	Vial Id	Location	Storage Date	Ship Batch Number	Ship Date	Specimen Condition	Participant Id	Draw Timestamp	Primary Type	Derivative Type	Additive Type	Visit Value	Protocol Number	Volume	Volume Units
<input type="checkbox"/>	106443199.2404.501	Rainier Lab Storage	2005-11-09 00:00	3	2005-11-10 00:00	SAT	249320107	2008-07-16 00:00	Blood (Whole)	PBMC Cells, Viable	Heparin	20080716.0000	845	1.0	ML
<input type="checkbox"/>	106443199.2404.501		2005-11-09 00:00	3	2005-11-10 00:00	SAT	249320107	2008-07-16 00:00	Blood (Whole)	PBMC Cells, Viable	Heparin	20080716.0000	845	1.0	ML
<input type="checkbox"/>	106443199.2404.501		2005-11-09 00:00	3	2005-11-10 00:00	SAT	249320107	2008-07-16 00:00	Blood (Whole)	PBMC Cells, Viable	Heparin	20080716.0000	845	1.0	ML

Specimens - vials



Vials Study 1

MANAGE STUDY > GROUP VIALS > SEARCH > REPORTS >

GROUPS > REQUEST OPTIONS > ENABLE COMMENTS/QC > IMPORT SPECIMENS

1 - 20 of 20

Hide Previously Requested By

PrimaryType.Description = 'Blood (Who...)'
 DerivativeType.Description = 'PBMC Ce...'
 AdditiveType.Description = 'EDTA'
 [CLEAR ALL](#)

Compact View

<input type="checkbox"/>		Participant Id	Global Unique Id	Volume	Volume Units	Primary Type	Derivative Type	Additive Type	Locked In Request	Site Name	At Repository	Available	Availability Reason	Quality Control Flag	Draw Timestamp
<input type="checkbox"/>	2 [history]	249320489	87443215.2604.529	18,432,000.0	CEL	Blood (Whole)	PBMC Cells, Viable	EDTA	false	Alki University	true	true		false	2008-12-03 00:00
<input type="checkbox"/>	2 [history]	249320489	87443215.2604.528	18,432,000.0	CEL	Blood (Whole)	PBMC Cells, Viable	EDTA	false	Alki University	true	true		false	2008-12-03 00:00
<input type="checkbox"/> ?	2 [history]	249320489	87443215.2604.527	18,432,000.0	CEL	Blood (Whole)	PBMC Cells, Viable	EDTA	false	In Transit	false	false	This vial is unavailable because it is not currently held by a repository.	false	2008-12-03 00:00
<input type="checkbox"/> ?	2 [history]	249320489	87443215.2604.526	18,432,000.0	CEL	Blood (Whole)	PBMC Cells, Viable	EDTA	false	In Transit	false	false	This vial is unavailable because it is not currently held by a repository.	false	2008-12-03 00:00

Specimens - requests



Specimen Request 12 Study 1

Request Notes

Attachments

This request's requirements are complete. Next steps include:

- Email specimen lists to their originating locations: [ORIGINATING LOCATION SPECIMEN LISTS](#) ▶
- Email specimen lists to their providing locations: [PROVIDING LOCATION SPECIMEN LISTS](#) ▶
- Update request status to indicate completion: [UPDATE REQUEST](#) ▶

Unsubmitted Request

This request has not been submitted.

Request processing will begin after the request has been submitted.

[SUBMIT REQUEST](#)

[CANCEL REQUEST](#)

Request Information

Requester: dave
Requesting Location: Montlake University (Repository)
Request Date: 2017-09-20 12:55
Description: Assay Plan: Plan 1

Shipping Information:
617 Eastlake Ave E
Seattle, WA 98109

Comments:
[Not provided]

Status: Not Yet Submitted

[VIEW HISTORY](#) ▶ [UPDATE REQUEST](#) ▶ [ORIGINATING LOCATION SPECIMEN LISTS](#) ▶ [PROVIDING LOCATION SPECIMEN](#)

Associated Specimens

											EXPORT	SPECIMEN SEARCH	UPLOAD SPECIMEN IDS	REMOVE SELECTED	
<input type="checkbox"/>		Participant Id	Global Unique Id	Sequence Num	Timepoint	Volume	Volume Units	Primary Type	Derivative Type	Additive Type					
<input type="checkbox"/>	[history]	249320489	87443205.2604.509	20090115	Month 6	14,356,000.0	CEL	Blood (Whole)	PBMC Cells, Viable	EDTA					
						Count (non-blank) ? : 1					Sum : 14,356,000.0				

Specimens - reports



Specimen Report: Participant 249318596 Study 1

Availability status

Base

Type breakdown

ParticipantId

Hide Empty Columns

Vial Counts

Total Volume

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

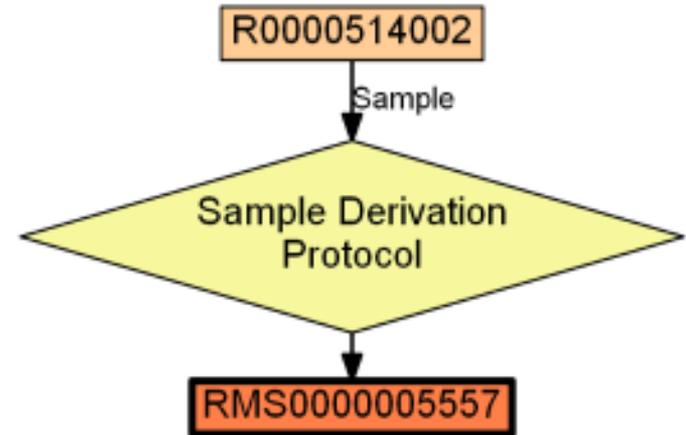
249318596 (Vial Count)

		Baseline	Month 1	Month 2	Month 3	Month 4	Month 5	Month 6	Month 7	Month 8	Month 9	Month 10	Month 11	Month 12	Month 13
Blood (Whole)	Plasma, Unknown Processing	2													
	Serum	3		12		12			10		10	31	12	12	18
Urine	Urine	1							1			2		1	1

Sample Sets



- Represent the materials that are analyzed in experiments
- Each sample set can have a different set of customizable fields
- Supports derivation/lineage by defining a parent sample
- Natively tracks Protocols, Parents, Children materials
- Created by pasting tabular or uploading files and identifying the key fields
 - Lineage is established by uploading the parent identifier



Sample Sets - scoping



- Each folder can have many sample sets
- Each sample set can have many samples
- Sample set definition visibility
 - Current folder
 - Current project
 - Shared project

Sample Sets - lineage



- Connects parent and child samples
 - Modeled as derivation experiment runs
- Supports many-to-many relationships
 - Pooling
 - Aliquoting
- For TSV import, use special “Parent” column to specify lineage

Sample sets – use cases



- Linking to Downstream experiments:
 - Create a lookup from the assay design to the sample set
 - Add a “Name” column to the sample set, to make this work
- Derivation/lineage tracking
 - Record the subdivision of a single sample into a set of smaller samples, aliquots
 - Chemical recipes
- Quick LIMS = Add a lookup to an inventory table



- Schemas and queries
 - `exp.SampleSets` – all available sample sets
 - Samples schema – separate queries for each sample set
 - `exp.Materials` – union of all samples
- JavaScript APIs
 - `LABKEY.Query` – insert, update, delete via sample schemas
 - `LABKEY.Experiment.saveBatch()` – connect samples to assay data

Two systems: specimens and sample sets



- Specimen Repository (more specialized for studies)
 - Only exists in a study
 - Seamlessly coupled to participant and assay data
 - “Shopping Cart” request system for access to outside labs, tracks ownership
 - Has a large set of required fields
- Sample Sets (more general purpose)
 - Can exist outside a study, but can integrate with a study too
 - Concepts of material parentage: precursors and derivatives
 - Not just tissues, covers materials, derivatives, media, formulations
 - User has control over fields

Summary

Data type	scenarios	constraints	visualizations	QC
Lists	Tabular data	few	standard	standard
Sample sets	Sample information including derivation	Key field(s) identifier	standard	standard
Assays	Lab instrument data	Patient ID and time required	app specific	programmatic, app specific
Datasets	Clinical or experimental data	Patient ID and time required, uniqueness constraints	app specific	programmatic, app specific
Specimens	Specimen inventory	special fields required	standard	app specific

