LabKey Software™
PARTNERS IN SCIENCE

Mark Igra
LabKey Software
marki@labkey.com

Data Management for Global Health

Background

LabKey Software
PARTNERS IN SCIENCE

- Came to biomedical research from the software industry
  - Moved to Fred Hutchinson Cancer Research Center
- Spun Off into LabKey Software
  - Professional *open source* software for scientists
  - 23 Employees
  - Mostly funded by research networks (ITN, HVTN, CHAVI, nPOD).
  - Some funding from University of Wisconsin (O'Connor Lab, Primate Center)
- Focus: Data Management for Research Networks

Data Management in Global Health

LabKey Software
PARTNERS IN SCIENCE

- Many global health research projects are distributed in nature
- Integrate & Share data amongst
  - Many clinical sites
  - Many research laboratories
- Specimens are a valuable resource
  - Must be tracked & allocated for maximum scientific value
- Integrated data can be leveraged for new uses

CHAVI

LabKey Software
PARTNERS IN SCIENCE

- 7 year “Big Science” project
  - PI: Bart Haynes, Duke University
- 12 Clinical Sites
- 45 Universities/Research Institutions
- Goal: Observational studies to identify correlates of control in Acute HIV Infection
- Data management by Statistical Center for HIV/AIDS Research and Prevention (SCHARP)

Atlas – SCHARP Data Portal

LabKey Software
PARTNERS IN SCIENCE

- 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
  - 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
- Piehler et al, BMC Immunology, May 2011

Map of CHAVI

Canada
United States
Mexico
Venezuela
Colombia
Brazil
Peru
Bolivia
Chile
North Atlantic Ocean
South Atlantic Ocean
Poland
Ukraine
Spain
Turkey
Iraq
Iran
Algeria
Libya
Egypt
Saudi Arabia
Mali
Niger
Sudan
Chad
Nigeria
Ethiopia
DR Congo
Tanzania
Angola
Namibia
Madagascar
South Africa

Atlas Data Flows

LabKey Software
PARTNERS IN SCIENCE

Data Providers

LIMS

Sample Info

Labs

Assays

Forms

CRF

DOB

BP sys

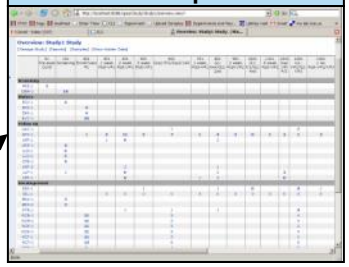
BP dia

Notes

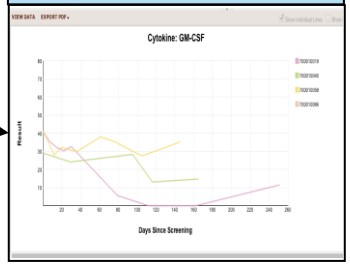
Atlas Web + Database Servers

Data Viewers

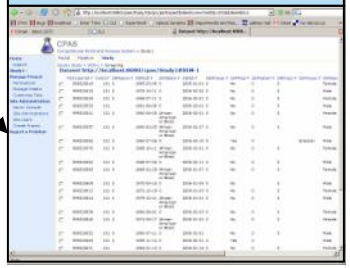
Leadership



Labs



Collaborators



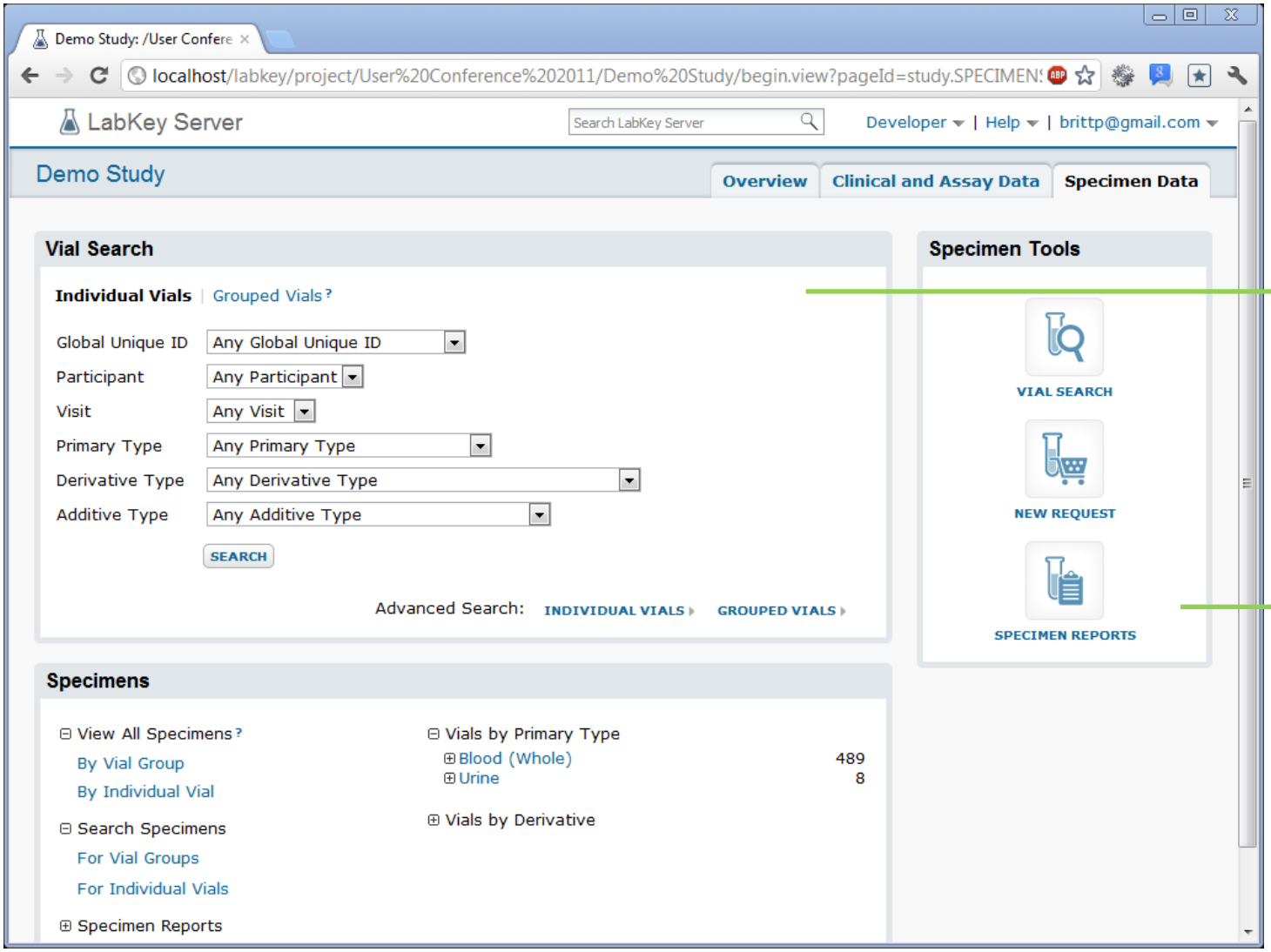
Initial Data/Sample Flows

What can you do now

LabKey Software
PARTNERS IN SCIENCE

- Clinical and Sample Data in a Single Repository
- Can ask interesting questions
  - How many samples do I have of each type
  - Where are they?
  - How many from acutely infected patients?

Specimen views: home page



Simple search and browse options

Quick access to reports and tools

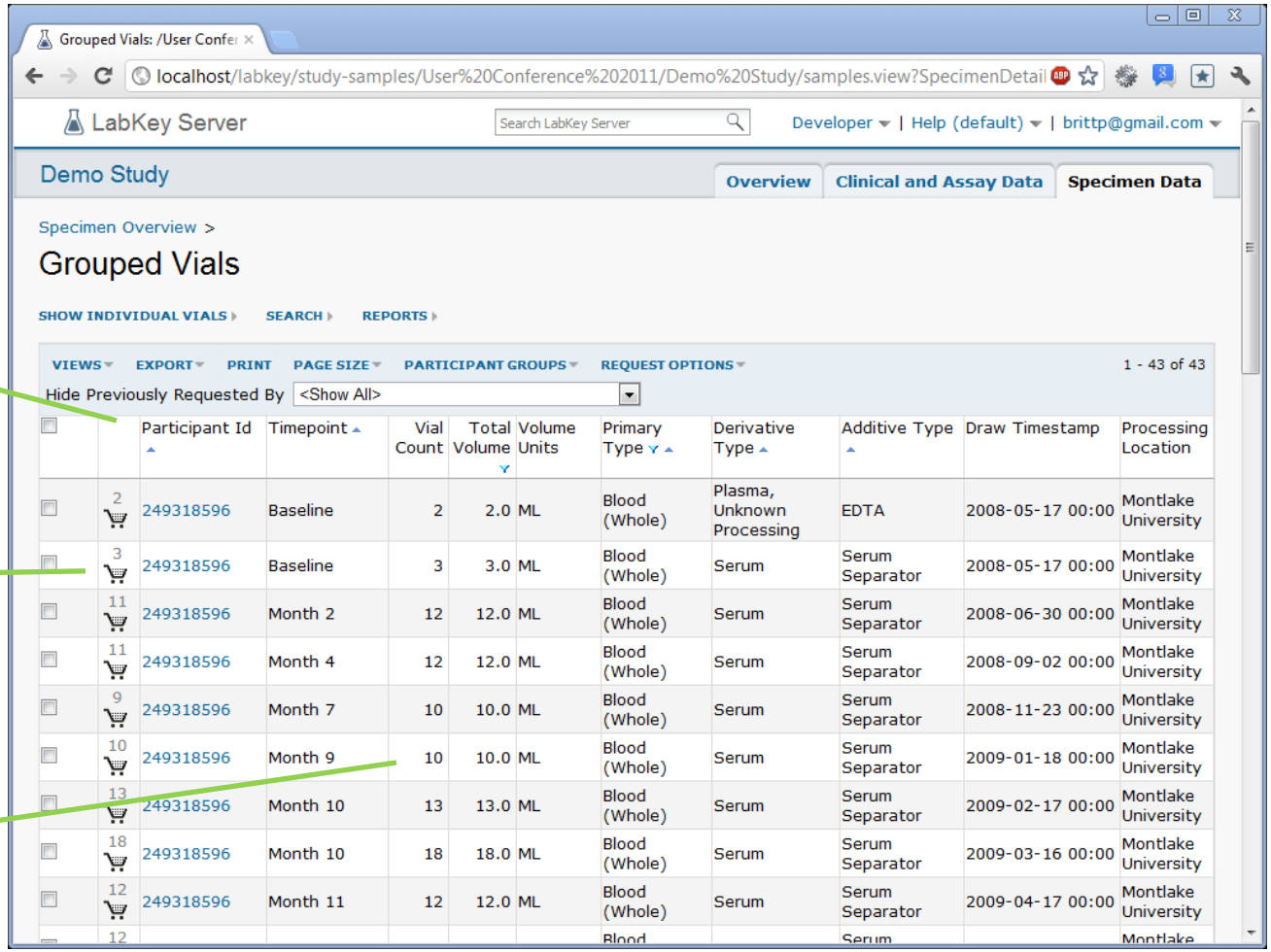
Specimen views: grouped/draw view

- Grouped view shows one row per participant/visit/type

Sort and filter via headers

Request options

Aggregate information: Vial count & total volume



		Participant Id	Timepoint	Vial Count	Total Volume	Volume Units	Primary Type	Derivative Type	Additive Type	Draw Timestamp	Processing Location
	2	249318596	Baseline	2	2.0	ML	Blood (Whole)	Plasma, Unknown Processing	EDTA	2008-05-17 00:00	Montlake University
	3	249318596	Baseline	3	3.0	ML	Blood (Whole)	Serum	Serum Separator	2008-05-17 00:00	Montlake University
	11	249318596	Month 2	12	12.0	ML	Blood (Whole)	Serum	Serum Separator	2008-06-30 00:00	Montlake University
	11	249318596	Month 4	12	12.0	ML	Blood (Whole)	Serum	Serum Separator	2008-09-02 00:00	Montlake University
	9	249318596	Month 7	10	10.0	ML	Blood (Whole)	Serum	Serum Separator	2008-11-23 00:00	Montlake University
	10	249318596	Month 9	10	10.0	ML	Blood (Whole)	Serum	Serum Separator	2009-01-18 00:00	Montlake University
	13	249318596	Month 10	13	13.0	ML	Blood (Whole)	Serum	Serum Separator	2009-02-17 00:00	Montlake University
	18	249318596	Month 10	18	18.0	ML	Blood (Whole)	Serum	Serum Separator	2009-03-16 00:00	Montlake University
	12	249318596	Month 11	12	12.0	ML	Blood (Whole)	Serum	Serum Separator	2009-04-17 00:00	Montlake University

Specimen tools: reports

- Single Participant Report

249320127 (Vial Count)								
		Baseline	Month 2	Month 6	Month 7	Month 8	Month 9	Month 10
Blood (Whole)	PBMC Cells, Viable	7	11	8	6	3	11	9
	Serum	8	6	7	7	7	12	6

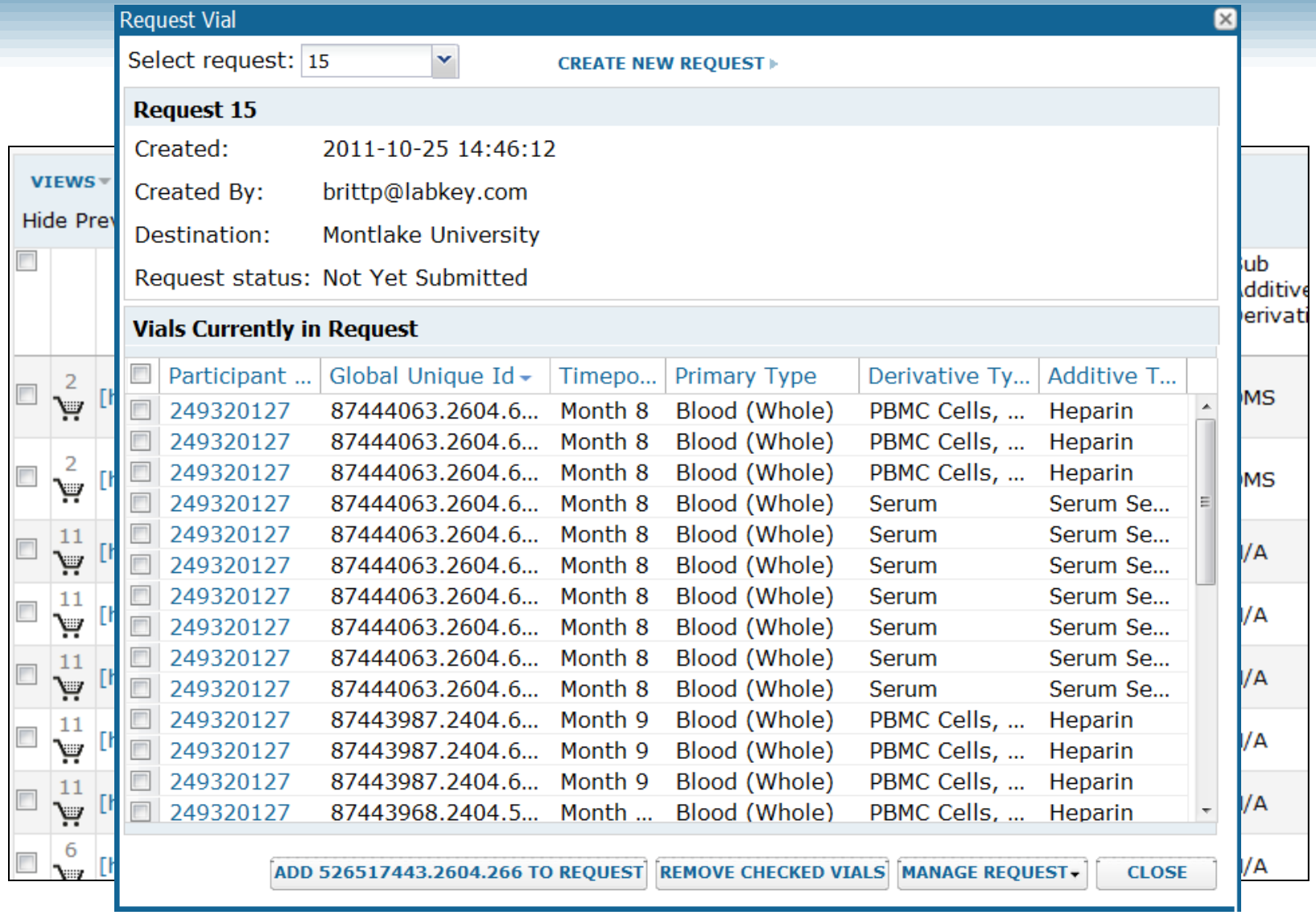
- Participant collections by type

Blood (Whole) (Vial Count)										
	Baseline	Month 1	Month 2	Month 3	Month 4	Month 5	Month 6	Month 7	Month 8	Month 9
249318596	5		12		12			10		10
249320107	12		22		10				14	
249320127	15		17				15	13	10	23
249320489	18	17	14			19	17			
249320897	10		6	8	6		12			12
249325717	14	18		24				6		

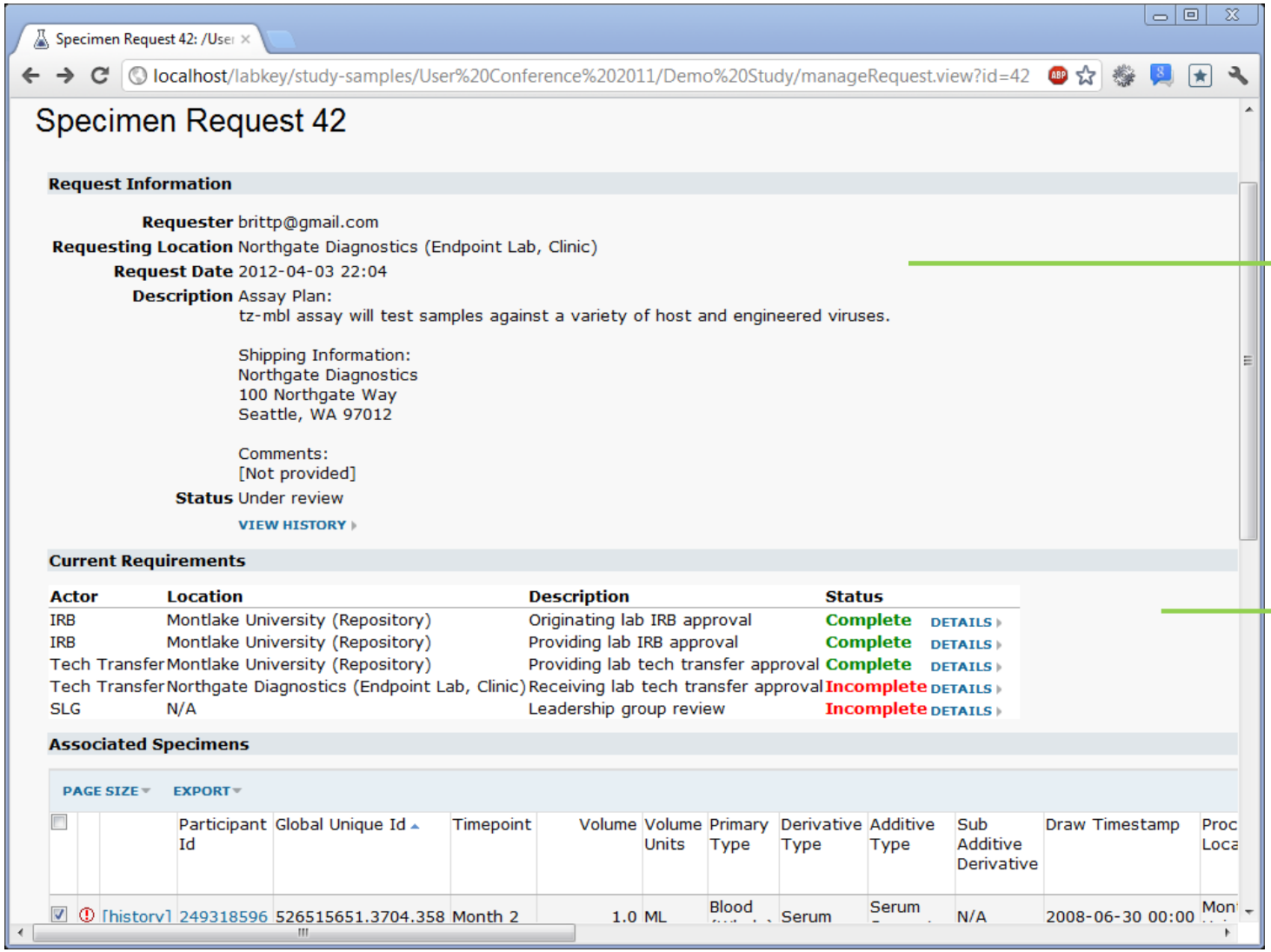
Research Lab Specimen/Data Flows

LabKey Software
PARTNERS IN SCIENCE

Specimen requests: request creation



Specimen requests: request status



Read-only details, history, and status

Transparent request approval workflow

Lab Data Flows

LabKey Software
PARTNERS IN SCIENCE

Lab Data Process

Labs

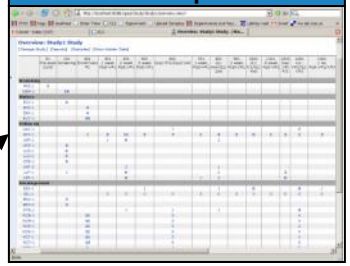
Assays

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

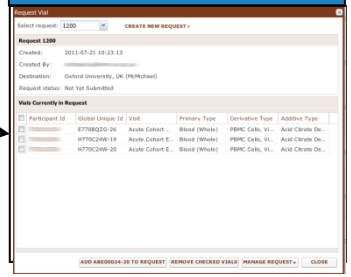
Atlas Web + Database Servers

Data Viewers

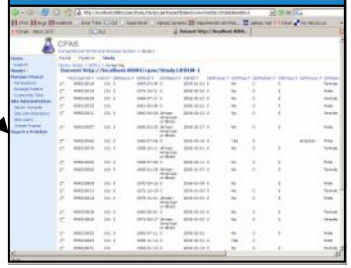
Leadership



Labs



Collaborators

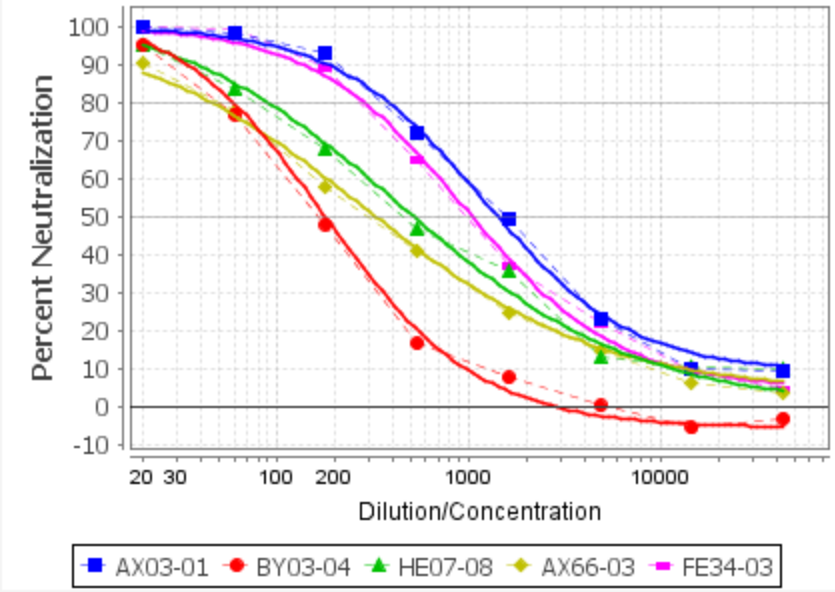


Assays: upload example

LabKey Software
PARTNERS IN SCIENCE

Run Summary: m0902051;3997-1.xls

Assay Id	m0902051;3997-1.xls	**Experiment Performer**	Britt
Created	2012-03-13 12:00:03.731	**Curve Fit Method**	Five Parameter
Created By	brittp@labkey.com	**Batch**	2012-03-13 batch 3



Range	108120
Virus Control	109095 ± 5%
Cell Control	975 ± 6%

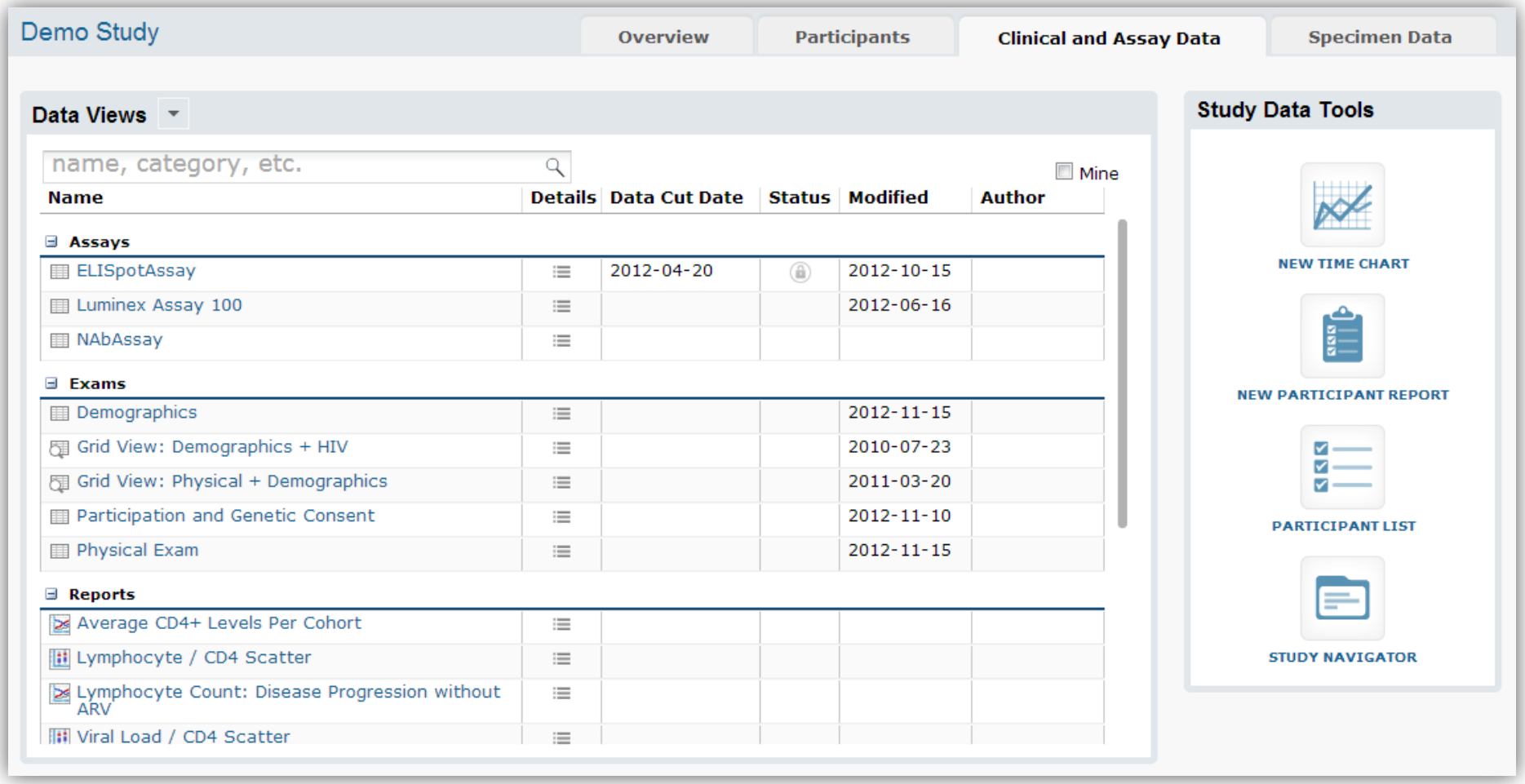
	Cutoff Dilutions			
	Curve Based		Point Based	
	50%	80%	50%	80%
AX03-01	1353	357	1547	354
BY03-04	188	55	165	50
HE07-08	492	93	456	77
AX66-03	306	45	294	48
FE34-03	988	301	963	270

Specimen ID	Dilution Factor	Initial Dilution	Method	Fit Error	AUC	PositiveAUC
AX03-01	3.0	20.0	Dilution	2.2	0.570	0.570
BY03-04	3.0	20.0	Dilution	2.5	0.270	0.286
HE07-08	3.0	20.0	Dilution	3.4	0.437	0.437
AX66-03	3.0	20.0	Dilution	2.1	0.391	0.391
FE34-03	3.0	20.0	Dilution	2.2	0.525	0.525

Demographic Data | Assay Data | Specimen Data

Participant	Gender	CD4+	Hemoglobin	Specimen	Available?
249318596	Male	420	12.2	5390.204.36	Yes
249320107	Female	344	13.2	2493.404.56	Yes
249320127	Male	897	21.0	2493.323.33	Yes
249320489	Male	234	16.4	2493.789.44	No
249320897	Female	893	12.9	5266.440.61	Yes

Clinical and assay data homepage



Dataset

LabKey Software
PARTNERS IN SCIENCE

CHAVI 001 > Study Overview >

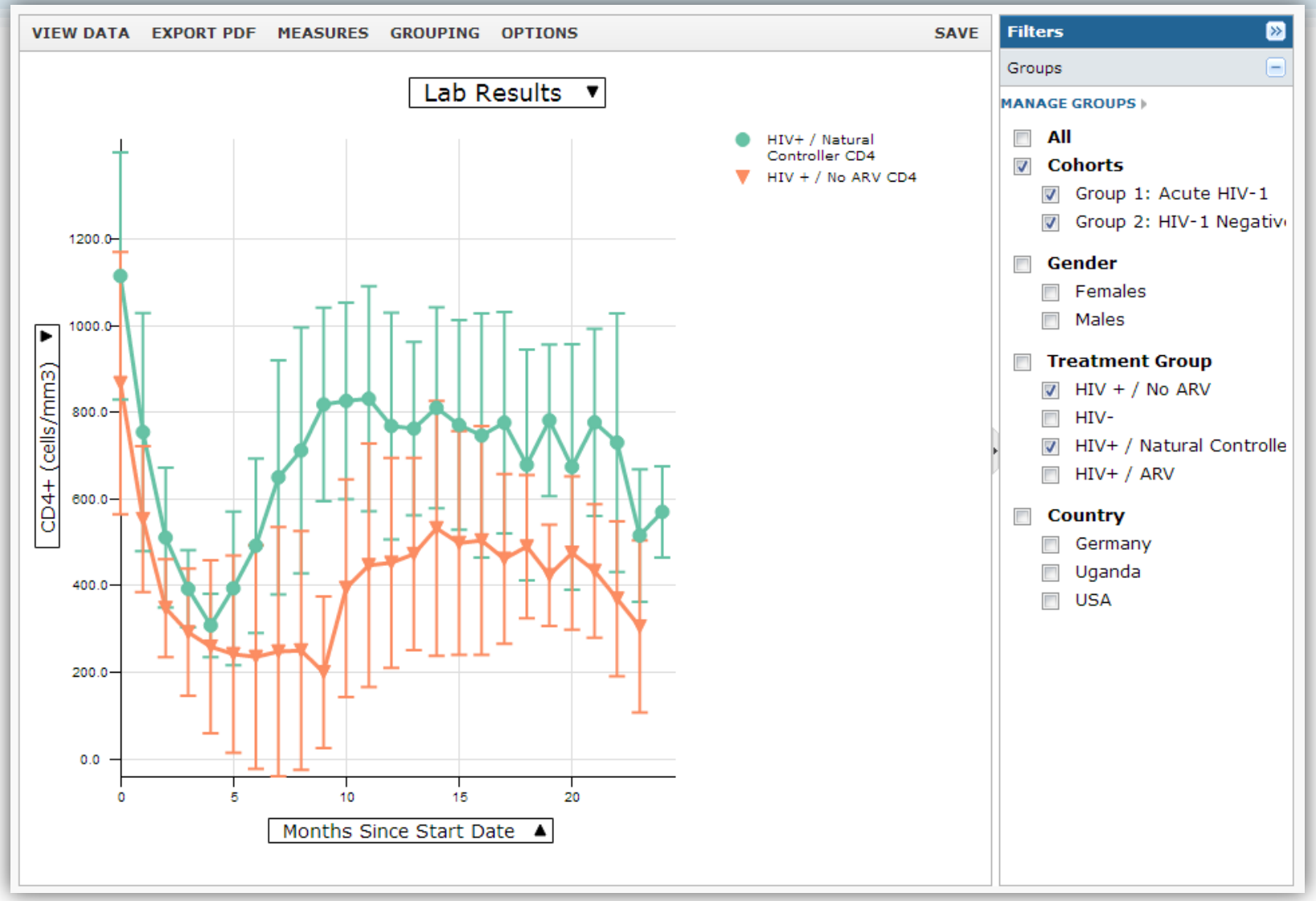
Dataset: Binding Antibody, All Visits

VIEWS | PARTICIPANT GROUPS | EXPORT | PAGE SIZE | VIEW SPECIMENS

1 - 100 of 282 Next › Last

Participant ID	Sequence Num	Draw Date	Enrollment date	Assay ID	Isotype	Antigen	Coat Lot	Dilution	Concentration	Reading	St. Dev.	No Antigen	No Ant. St. Dev.	QC Titer 50	Plate ID	Network	Protocol	Assay Type
[redacted]	101.0	20-Jun-06	20-Jun-06	JTL 61-38	IgA	CONT 140 CF	#2	10		0.394	0.013			151450.0	3	CHAVI	1	ELISA
[redacted]	102.0	27-Jun-06		JTL 61-38	IgA	CONT 140 CF	#2	10		0.387	0.034			151450.0	3	CHAVI	1	ELISA
[redacted]	103.0	05-Jul-06		JTL 61-38	IgA	CONT 140 CF	#2	10		0.408	0.016			151450.0	3	CHAVI	1	ELISA
[redacted]	104.0	12-Jul-06		JTL 61-38	IgA	CONT 140 CF	#2	10		0.417	0.0070			151450.0	3	CHAVI	1	ELISA
[redacted]	105.0	20-Jul-06		JTL 61-38	IgA	CONT 140 CF	#2	10		0.436	0.014			151450.0	3	CHAVI	1	ELISA
[redacted]	106.0	14-Aug-06		JTL 61-38	IgA	CONT 140 CF	#2	10		0.52	0.0020			151450.0	3	CHAVI	1	ELISA
[redacted]	107.0	07-Sep-06		JTL 61-38	IgA	CONT 140 CF	#2	10		0.634	0.03			151450.0	3	CHAVI	1	ELISA
[redacted]	108.0	10-Oct-06		JTL 61-38	IgA	CONT 140 CF	#2	10		0.227	0.076			151450.0	3	CHAVI	1	ELISA
[redacted]	109.0	29-Nov-06		JTL 61-38	IgA	CONT 140 CF	#2	10		0.574	0.01			151450.0	3	CHAVI	1	ELISA
[redacted]	110.0	26-Feb-07		JTL 61-38	IgA	CONT 140 CF	#2	10		0.413	0.084			151450.0	3	CHAVI	1	ELISA
[redacted]	101.0	20-Jun-06	20-Jun-06	JTL 61-38	IgA	GP41	J15/6J13R	10		0.833	0.034			19790.0	4	CHAVI	1	ELISA
[redacted]	102.0	27-Jun-06		JTL 61-38	IgA	GP41	J15/6J13R	10		0.531	0.024			19790.0	4	CHAVI	1	ELISA
[redacted]				JTL 61-														

Time chart


VIEW DATA
EXPORT PDF
MEASURES
GROUPING
OPTIONS
SAVE
Lab Results
HIV+ / Natural Controller CD4
HIV + / No ARV CD4
CD4+ (cells/mm3)
1200.0
1000.0
800.0
600.0
400.0
200.0
0.0
0
5
10
15
20
Months Since Start Date
Filters
Groups
MANAGE GROUPS
All
Cohorts
Group 1: Acute HIV-1
Group 2: HIV-1 Negativ
Gender
Females
Males
Treatment Group
HIV + / No ARV
HIV-
HIV+ / Natural Controlle
HIV+ / ARV
Country
Germany
Uganda
USA

Participant View

LabKey Software
PARTNERS IN SCIENCE

CHAVI 001 > Study Overview > Dataset: Binding Antibody, All Visits >

Participant - [redacted]

NEXT PARTICIPANT | SEARCH FOR '[redacted]'

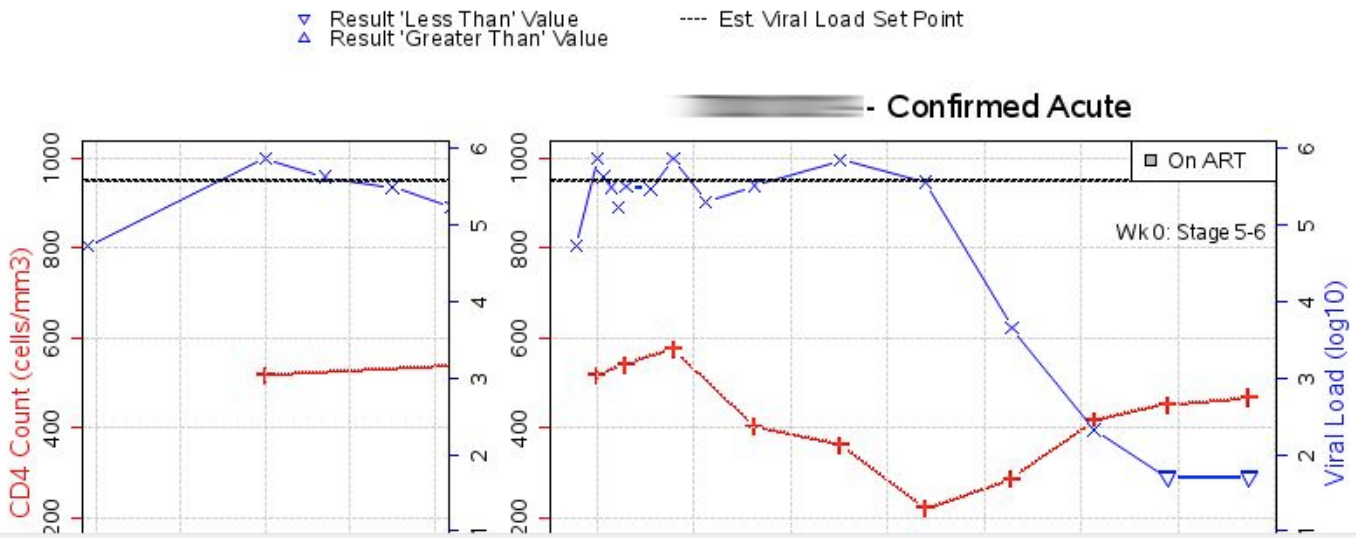
All Datasets | Specimen Timeline | Clinical Data | Requested Specimens | All Specimens | Binding Antibodies | Cytokines

Clinical Data

CRF Data

EXPORT TO EXCEL

Cohort:	Confirmed Acute
Age:	18
Gender:	Female
Race:	White
Tribe:	
Clade:	



Building on the Data

LabKey Software
PARTNERS IN SCIENCE

- Dozens of small applications needed for different groups
  - Most used data already in LabKey System
  - But custom workflows, reports & analysis required
- LabKey provided an API and simple application building tools to leverage this data

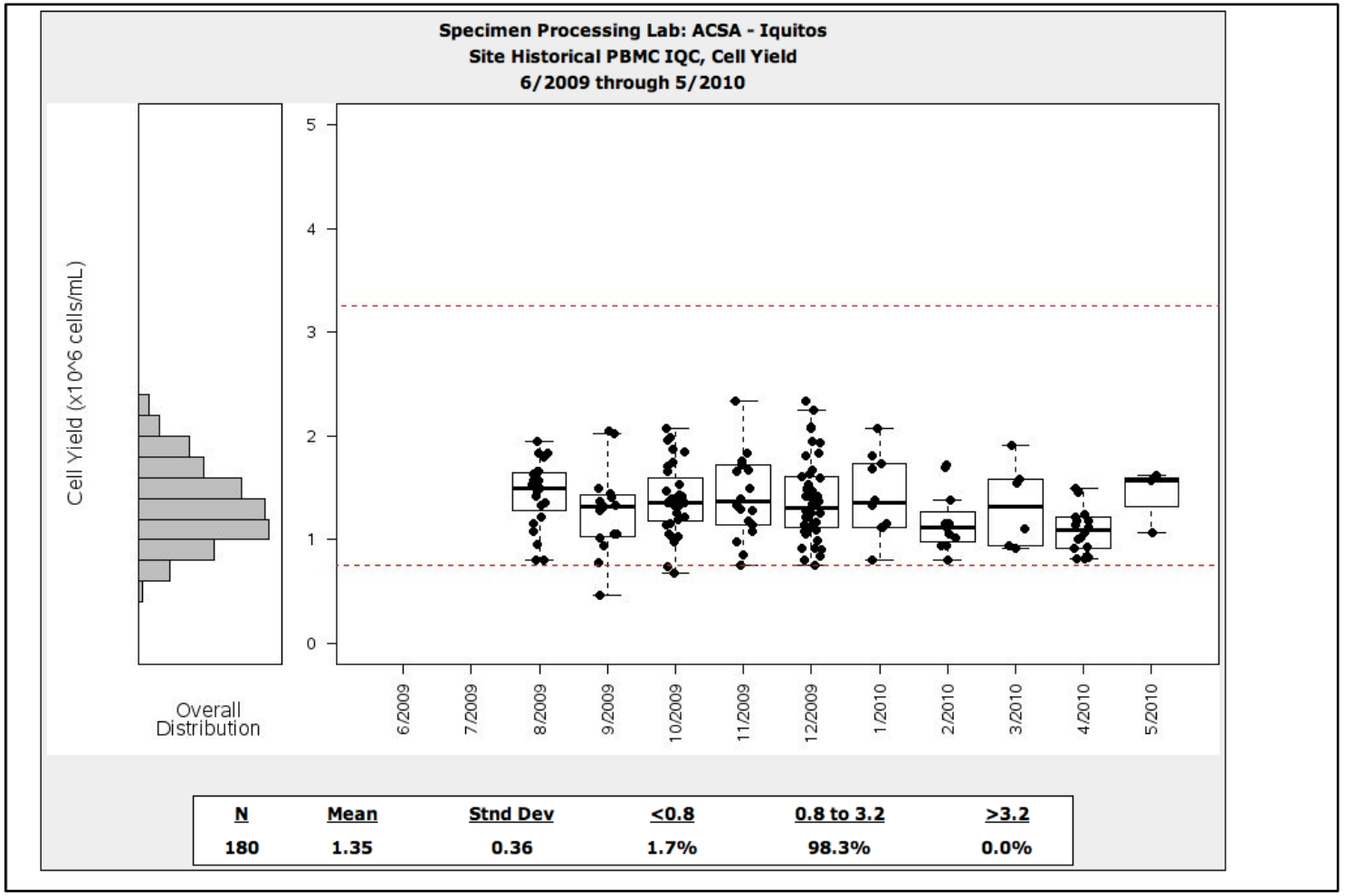
HVTN Quality Control

LabKey Software
PARTNERS IN SCIENCE

- HIV Vaccine Trials Network
  - ~ 50 Sites
  - ~ 50 Protocols, ~10 currently active
- PBMCs are generated at site-affiliated labs
  - Yields and quality were an issue
- Report yields across sites
  - Each site can see their own yields
  - Central lab can evaluate all sites

HVTN IQC

- Internal Quality Control for Sample Processing



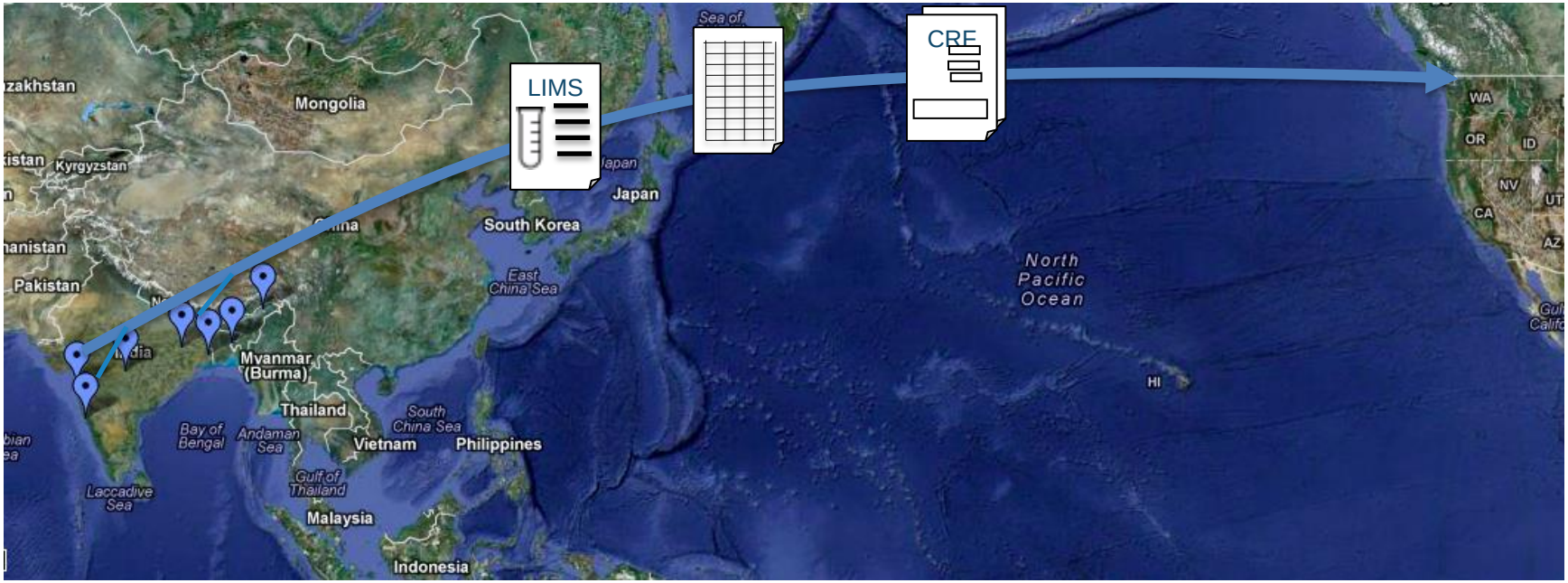
ICEMR

LabKey Software
PARTNERS IN SCIENCE

- International Center of Excellence for Malaria Research
  - Studying Malaria Evolution in India
  - PI: Pradeep Rathod
- Collecting Clinical, Specimen and Assay Data
  - Different infrastructure
  - But data coordination needs are very similar
- Specimens cannot leave India

ICEMR - MESA

International Center of Excellence in Malaria Research
Malaria Evolution in South Asia
PI: Pradip Rathod, University of Washington



Data Reuse & Ancillary Studies

LabKey Software
PARTNERS IN SCIENCE

Study Execution

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

Study Initiation

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

Nelson et al, Ancillary study management systems a review of needs
BMC Med Inform Decis Mak Jan 2013

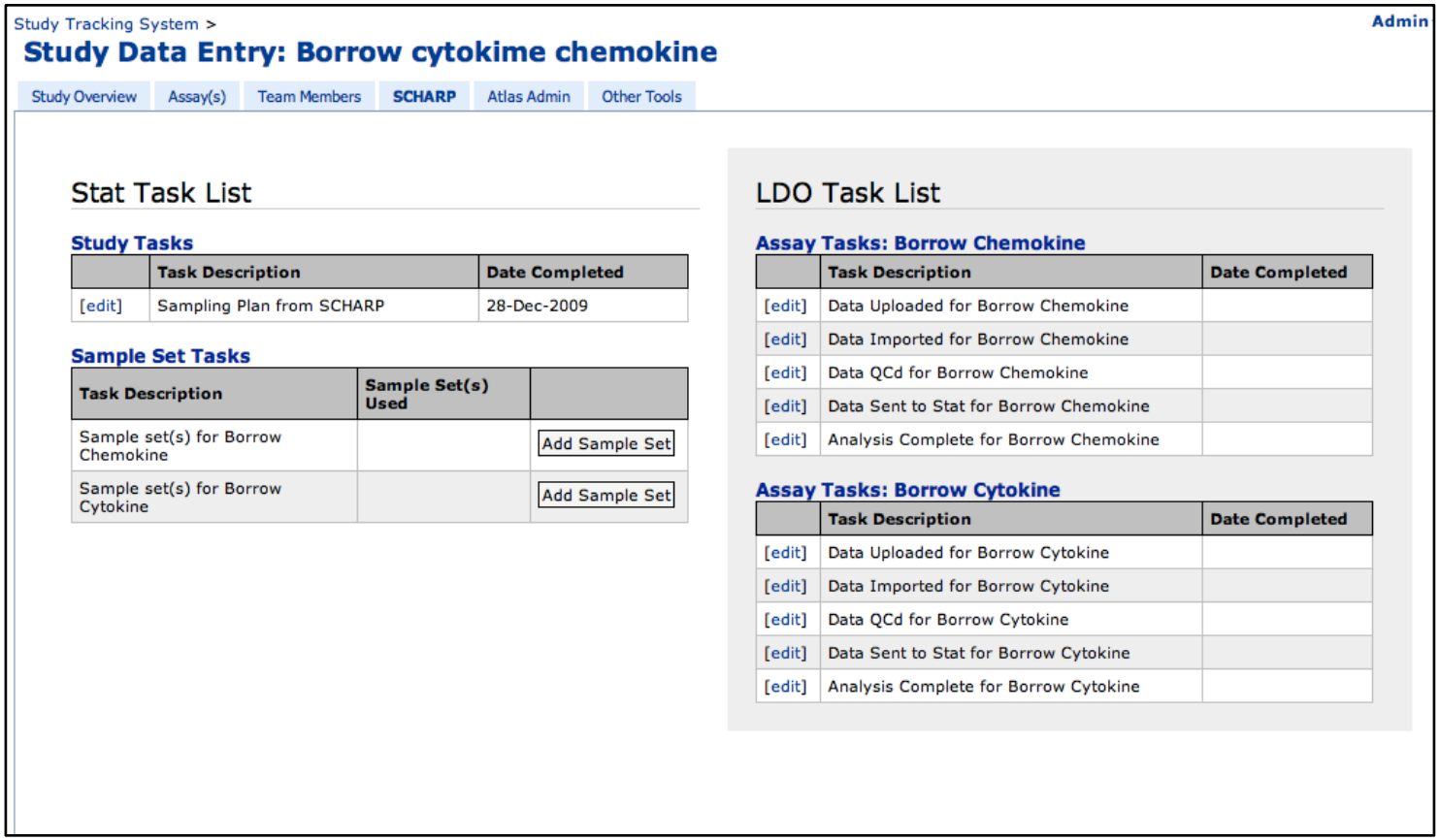
Results Sharing

8. Data/specimen repatriation
9. Publication

RV144 Followup Study Tracking System

LabKey Software
PARTNERS IN SCIENCE

- Tools for tracking information about studies



Study Tracking System >

Admin

Study Data Entry: Borrow cytokime chemokine

Study Overview | Assay(s) | Team Members | SCHARP | Atlas Admin | Other Tools

Stat Task List

Study Tasks

	Task Description	Date Completed
[edit]	Sampling Plan from SCHARP	28-Dec-2009

Sample Set Tasks

Task Description	Sample Set(s) Used	
Sample set(s) for Borrow Chemokine		Add Sample Set
Sample set(s) for Borrow Cytokine		Add Sample Set

LDO Task List

Assay Tasks: Borrow Chemokine

	Task Description	Date Completed
[edit]	Data Uploaded for Borrow Chemokine	
[edit]	Data Imported for Borrow Chemokine	
[edit]	Data QCd for Borrow Chemokine	
[edit]	Data Sent to Stat for Borrow Chemokine	
[edit]	Analysis Complete for Borrow Chemokine	

Assay Tasks: Borrow Cytokine

	Task Description	Date Completed
[edit]	Data Uploaded for Borrow Cytokine	
[edit]	Data Imported for Borrow Cytokine	
[edit]	Data QCd for Borrow Cytokine	
[edit]	Data Sent to Stat for Borrow Cytokine	
[edit]	Analysis Complete for Borrow Cytokine	

Caveats

- Scientific Leadership Low/Late Adopters
- Assay Pipelines Need to be very Low Overhead
  - Can work for high throughput assays
- Data sharing is fraught with issues
  - Frequently data & specimen use restrictions
  - “Ownership” of specimens and data is disputed
  - Investigators don’t trust others to interpret data properly

Conclusion

LabKey Software
PARTNERS IN SCIENCE

- Software is Relevant for Global Health Projects
- Coordination
- Reproducibility
- Data Reuse

LabKey Software
PARTNERS IN SCIENCE

If you use LabKey Server for your research, please reference these publications about the platform:

General Use: Nelson EK, Piehler B, Eckels J, Rauch A, Bellew M, Hussey P, Ramsay S, Nathe C, Lum K, Krouse K, Stearns D, Connolly B, Skillman T, Igra M. LabKey Server: An open source platform for scientific data integration, analysis and collaboration. BMC Bioinformatics 2011 Mar 9; 12(1): 71.

Ancillary Studies: Nelson EK, Piehler, B, Rauch A, Ramsay S, Holman D, Asare S, Asare A, Igra M. Ancillary study management systems: a review of needs. BMC Med Inform Decis Mak Jan 2013

Proteomics: Rauch A, Bellew M, Eng J, Fitzgibbon M, Holzman T, Hussey P, Igra M, Maclean B, Lin CW, Detter A, Fang R, Faca V, Gafken P, Zhang H, Whitaker J, States D, Hanash S, Paulovich A, McIntosh MW: Computational Proteomics Analysis System (CPAS): An Extensible, Open-Source Analytic System for Evaluating and Publishing Proteomic Data and High Throughput Biological Experiments. Journal of Proteome Research 2006, 5:112-121.

Flow Cytometry: Shulman N, Bellew M, Snelling G, Carter D, Huang Y, Li H, Self SG, McElrath MJ, De Rosa SC: Development of an automated analysis system for data from flow cytometric intracellular cytokine staining assays from clinical vaccine trials. Cytometry 2008, 73A:847-856.

LabKey Software™

PARTNERS IN SCIENCE

Mark Igra

marki@labkey.com

Any questions?

Who we are

LabKey Software
PARTNERS IN SCIENCE

- LabKey is a “professional open source” software company
  - 23-person team (and growing)
- Background is in commercial software development
  - Lead technical roles on SQL-Server, Visual Basic, Excel, Access, BEA WebLogic, and Amazon’s Data Warehouse
  - Team members hold 25+ software patents

Our open source approach

LabKey Software
PARTNERS IN SCIENCE

- LabKey solutions are built around LabKey Server
  - Open source software (Apache 2.0 license)
  - Freely downloadable, no license fees
  - Over $10 million invested in the platform to date

- Open source is more than a licensing model
  - **Cost**: past work is leveraged by all
  - **Operations**: best-practices are built-in and shared
  - **Grants**: funders appreciate sustainable infrastructure
  - **Publications**: methods papers/sections are straightforward

Data Challenges

LabKey Software
PARTNERS IN SCIENCE

- Modern researchers face growing data challenges
  - Increased data volume
    - Millions of results from thousands of high throughput assay runs
  - Increased data variety
    - Clinical, demographic, assay, and specimen data
  - Increased requirements for collaboration
    - Investigators are specialized and distributed
- LabKey partners with researchers to address these issues
  - Data management
  - Data integration
  - Distributed collaboration