



LabKey Software™
PARTNERS IN SCIENCE

Statistical Center for
HIV/AIDS Research & Prevention
SCHARP

Harnessing the Web to Accelerate Clinical Research

How a LabKey Solution Helps Scientists Coordinate the Search for an HIV/AIDS Vaccine



COLLABORATION IN BIOMEDICAL RESEARCH

The biomedical research community is turning to *collaborative research* as a model for maximizing progress on major diseases. Collaborative research involves multiple teams working together to solve a problem by sharing their research plans, reagents, specimens, and preliminary data. Researchers working toward an HIV/AIDS vaccine are a prime example of this approach. Progress on an effective vaccine has been stalled by the elusiveness of the virus. In recognition of this challenge, leading research and funding organizations including the National Institutes of Health and the Bill and Melinda Gates Foundation formed an alliance dedicated to accelerating the development of a preventive HIV vaccine. One of the core strategies of this alliance is to promote collaboration by creating more efficient, faster ways for researchers to share successes and failures and avoid duplication of efforts.

To implement their collaboration strategy, NIH funded the Center for HIV/AIDS Vaccine Immunology (CHAVI) and the Gates Foundation funded the Collaboration for AIDS Vaccine Discovery (CAVD). CHAVI and CAVD are research consortia whose members comprise more than 110 institutions in 25 countries. Investigators within the member institutions apply new scientific knowledge and a diversity of cutting-edge research techniques to create and evaluate novel vaccine candidates. CHAVI and CAVD also run many different types of pre-clinical studies, including several large-scale, multi-site studies of people at risk of contracting the HIV virus. These studies collect specimens and

observational data on thousands of participants and make these resources available to other consortia members.

DEMANDS ON INFORMATION SYSTEMS

CHAVI and CAVD researchers are expected to share their project plans and experimental data with other consortium members, as a condition of their funding. This level of coordination clearly requires some support from the computing systems used by the collaborators. The broad reach of the Internet makes such large scale collaboration conceivable and increasingly achievable. It is safe to assume today that every scientific researcher has the computing resources to connect to the Internet. Beyond that basic level of commonality, however, there will be wide variation in the systems used, software installed, and networking protocols available. Because of this variation, the systems for sharing documents in a large consortium must be different from the combination of email and file servers that are commonly used for such purposes within a smaller group in a single organization.

"LabKey's people are experts in translating complex needs into practical software. They have worked closely with our statisticians and our internal IT staff to create a system that enables collaborators to analyze, validate and share datasets from labs and clinics around the globe, as they are collected. This tight collaboration enables researchers to build on each other's work much more rapidly. I think it is key to accelerating the pace of new discoveries."

— Dr. Steve Self, Executive Director of SCHARP

Sharing *experimental data* among collaborators is even more difficult than sharing project documents. For example, each of the many different experimental techniques used by CHAVI and CAVD generate different types of result data files. In some cases these files are too large to be analyzed by desktop computers and tools. In other cases the files are so numerous that they become difficult to track and compare. Ultimately, the experimental data must be analyzed and validated and then integrated with observational data and communicated in ways that allow other consortia members to find and interpret the results. As with documents, the systems that work for scientific data sharing within a single organization often do not scale well to a large-scale collaboration.

Lastly, information systems that support collaboration demand *agility* in their design and implementation. Building a system that can adapt easily to the requirements of different collaborators on a project is important to getting them to use the system effectively. The system (and the people behind it) must be flexible in order to adapt to changes in requirements and suggestions for improvements from collaborators. The urgency of the research goals for large-scale collaborations and the uncertainty of how they will be achieved also drive tight implementation schedules and a flexible and iterative development approach.

Statistical Center for
HIV/AIDS Research & Prevention
SCHARP

LABKEY SERVER: SOFTWARE SOLUTIONS FOR COLLABORATIVE SCIENCE

The Statistical Center for HIV/AIDS Research and Prevention (SCHARP) was selected to provide both data management and statistical analysis expertise to the CHAVI and CAVD consortia. SCHARP is part of the Vaccine and Infectious Disease Institute of the Fred Hutchinson Cancer Research Center based in Seattle, Washington. SCHARP formed a team to manage the project of building and maintaining the system, nicknamed "Atlas".

SCHARP needed information systems for the CHAVI and CAVD projects that would be flexible enough to adapt to rapidly

evolving research protocols and diagnostic instruments. Commercial software packages could be used for specific tasks, but the overall system required a comprehensive, custom solution. SCHARP chose to partner with LabKey Software to design and build that solution. LabKey brought development expertise in high-throughput web applications, a skill set that complemented SCHARP's strength in designing and managing biomedical research studies. The open-source LabKey Server platform provided the foundation for a customized solution, a foundation that was already proven in production use on other large research projects. The Atlas system was built as a set of plug-in modules on the LabKey Server platform.

PUBLISH & COLLABORATE	<ul style="list-style-type: none"> ▶ Share data among researchers with control over who can see what data ▶ Collaborate on study design and data interpretation via shared web pages ▶ Track and authorize specimen requests from remote labs
ANALYZE & COMPARE	<ul style="list-style-type: none"> ▶ Navigate from summary data to individual data and back ▶ Examine all results for a single participant over time ▶ Consolidate data to analyze aggregates and exceptions
ANNOTATE & INTEGRATE	<ul style="list-style-type: none"> ▶ Integrate diverse datasets from diagnostic assays and clinical measures ▶ Define custom metadata attributes to categorize data ▶ Attach help text and notes to data fields and data records
PROCESS & IMPORT	<ul style="list-style-type: none"> ▶ Import clinical form data from worldwide sites via DataFax ▶ Upload common assay formats (Luminex, ELISpot, NAb, microArray) ▶ Define new assay formats and custom data submission forms
STORE & QUERY	<ul style="list-style-type: none"> ▶ Filter, sort and join over standard and custom fields ▶ Search for text in data records anywhere in a study folder ▶ Upload document and image files and link to data records

Functions available in Atlas for managing observational study data

Today, the Atlas system, based on LabKey Server, provides three essential services to both consortia:

- ▶ **Secure Collaboration:** Atlas includes a set of *web-based* collaboration tools ideally suited for sharing documents and short notes. These support several common web formats, including wiki pages, message boards, and issue trackers. All of these collaboration tools are thoroughly integrated with a security model that is both thorough and simple to understand. Atlas security integrates with SCHARP's internal authentication system and provides secure user account management for external collaborators. This makes administration of groups and permissions straightforward.
- ▶ **Assay Data Management:** Atlas takes a comprehensive approach to managing data from diagnostic tests and laboratory instruments by providing customizable tools to import, analyze, validate, and integrate these datasets in a central SQL database. The system includes an assay designer that allows SCHARP staff to specify the file data format for results to be loaded into the database, as well as speedy data entry forms for capturing additional information to collect with the data file. Once the data is in the database, SCHARP researchers can employ a range of web-based tools to filter, sort, compare, search, visualize and compute statistics on the data. In the future, CHAVI and CAVD will use these tools in combination with external tools to ensure the quality of shared data.

- ▶ **Study Management:** Atlas supports the design of a *study* as an information framework for managing the data collected in the course of a research investigation on a set of subjects over time. The study framework keeps track of the participants in the investigation (the subjects) as well as the specimens to be analyzed and the time points that are relevant to the analysis. The study framework includes functions for data validation and navigation, and specimen request tracking. CHAVI, for example, runs several large observational studies consisting of hundreds of human participants around the world. The participants follow a schedule of regular clinic visits at which specimens are collected and clinical measurements are recorded. The remote clinics record their measurements on paper forms and send them via DataFax to SCHARP. Specimens are divided into vials and kept in central storage facilities until a researcher requests one or more vials to run a diagnostic assay. Assay data and metadata are linked into the same study framework. The integration of all of these datasets enables the study framework to provide built-in views that show summaries per time point, and allow SCHARP statisticians drill down by cohort and participant to find trends and anomalies.

"Partnering with LabKey accelerated the SCHARP team's ability to deliver the information systems to support CHAVI and CAVD. With LabKey supplying the core infrastructure, we had more time to focus on the integration and customization needed to make the system really fit the way we work. At the end of the day, working with LabKey gave us higher quality, faster implementation and lower risk than what we could have done on our own."

— Tom Skillman, Deputy Director of SCHARP

The remainder of this document will explore in depth these three essential services: secure collaboration, assay data management, and study management.

LABKEY SERVER SOLUTIONS FOR SECURE COLLABORATION

NEED: WEB-BASED TOOLS FOR MANAGING COLLABORATIVE RESEARCH PROJECTS

Successful research projects demand current and accurate communications among members. When all of the participants in a project are in the same hallway, such communication takes advantage of a flexible combination of means. These may include formal project meetings, informal white board and hallway discussions, email, shared project documents, phone calls, and sometimes project-specific web-sites. As the participants in the project become more numerous, or further apart geographically or organizationally, the face-to-face mechanisms become costlier and less frequent. Then the electronic means of communication become more critical.

In particular, web-based applications are often the most appropriate and reliable communication mechanisms for collaborative research projects. Web-based applications take advantage of the underlying network connectivity that is already established at all or nearly all collaborating organizations. Web-based systems also avoid the complexities of installing software locally at each site.

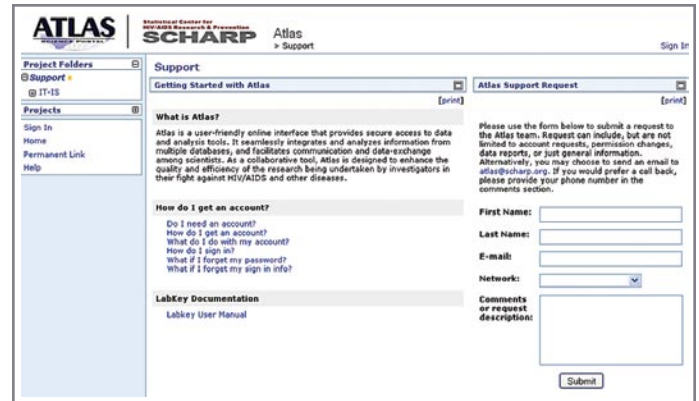
Web-based means of project communication are constantly evolving and improving. Sharing files via the web, for example, has gotten considerably easier with the advent of more capable file transfer protocols and browser-based applications that simplify the use of these protocols. Likewise, many project communications no longer have to be authored using standard desktop tools that produce proprietary format files and are optimized for printed documents. Research collaborators can now use simplified tools to create and edit project documents as HTML pages on the web, in the same place that others will read them.

SOLUTION: LABKEY SERVER FOR WEB-BASED COLLABORATION

LabKey Server was the software platform chosen by SCHARP to support CHAVI and CAVD, two large consortia focused on developing and HIV/AIDS vaccine. LabKey Server was designed from the ground up as a web-server application. It is based on Java, Tomcat and related open source components that are portable to most operating systems. It stores data in a choice of production SQL database engines, including PostgreSQL and Microsoft SQL Server. Thus, the staff responsible for installing and managing a LabKey Server have many options for where and how to run the application. The server that SCHARP manages for CHAVI and CAVD runs on Solaris or Sun SPARC hardware.



The Atlas Sign In Page



The Atlas support page uses the Wiki and Issues components of LabKey Server

LabKey Server provides SCHARP with a set of built-in tools that foster project communications:

- ▶ **Wikis** for sets of web pages that can be authored and revised in the browser by any user with appropriate permissions
- ▶ **Message boards** for a more constrained posting of announcements and schedules, in which message threads can be restricted to any subset of users
- ▶ **Issues** lists for tracking action items or defects
- ▶ **Portal pages** for displaying multiple tools on a single page
- ▶ **File upload and download tools** to exchange and share documents
- ▶ **Search tools** for quickly finding posted data and files on the site

LabKey Server makes it easy for a project team to create and maintain a web site. According to Craig Margaret, Manager of Information Systems and Scientific Computing for SCHARP:

SCHARP particularly values how LabKey makes simple things easy for their users. For example, users do not need to learn how to use an FTP application when they use LabKey. They simply use the familiar LabKey interface to upload, manage and download files. Users learn to use one interface for all their data, so they do not need to learn different software tools to work with different data types.

A number of software packages offer similar, web-based lab communication tools. LabKey's collaboration solutions are distinguished by two main attributes:

- ▶ A comprehensive and easy-to-use security model
- ▶ Tight integration with scientific data analysis and presentation tools



The Message Board component of LabKey Server hosts both public and private discussions

NEED: SIMPLE, ROBUST SECURITY THAT SCALES EASILY

The types of data handled by high-profile, collaborative projects require a system that positively identifies users on the system and restricts their access to data based on permissions granted specifically by the administrator responsible. All security models involve tradeoffs between ease of use for authorized users, ease of use for security administrators, and the difficulty of subverting the system for unauthorized users. A security model that focuses only on raising the barriers to unauthorized access may inadvertently *decrease* the overall security of the system by making it more likely that authorized users and administrators will make mistakes in configuring the system or will resort to shortcuts to get their jobs done.

“Before adopting LabKey, SCHARP had very little flexibility in collaborating online with large, external consortia. SCHARP used a complex system of SecurID® cards to manage system access. This system wasn’t well adapted to supporting large consortia because of the complexity of adding users and making changes to their permissions, and because the physical provision and replacement of the cards (via the postal service) created a potential security hole. SecurID® cards are still used for some purposes within SCHARP, but they are no longer necessary for external collaborators.”

— Sarah Ramsay, Collaborative Services Coordinator, SCHARP

SOLUTION: THE LABKEY SERVER SECURITY MODEL

The LabKey Server security model used to support CHAVI and CAVD is based on the following elements:

- ▶ **Authentication:** Users allowed to sign in to the server are identified by password-based authentication that avoids any transmission or storage of the password in clear text. Optionally, the authentication system integrates with existing authorization services to support common passwords or single sign-on.
- ▶ **Groups and projects:** Datasets and documents on a LabKey Server are organized in projects, which are top-level “folders” on the system. When a logged-in user accesses a project, the user is assigned to one or more groups. Groups can be defined at the project level or for the entire server.
- ▶ **Permissions and folders:** The permissions of groups to access data objects (read, create, edit, or delete specific data and documents) are established on a per-folder level, where a folder is either a project itself or is contained within another folder. By default, folders contained within other folders inherit the outer folder’s permissions. This makes it easy to set up the common case in which all data within the project is restricted to the same users and groups.
- ▶ **Privacy:** All traffic to and from a LabKey Server can be encrypted over the network by assigning a server certificate to the LabKey Server and configuring it to use secure protocols (https:// and ftps://).
- ▶ **Auditing:** Administrators can choose to log all changes made by users on a LabKey Server.
- ▶ **Tool-specific security:** Some tools in the product, such as message boards and study datasets, have specialized, finer-grained security configurations. For the ability to see or edit data in patient datasets, for example, can be restricted to those users associated with the patient’s clinic site, even if data from all clinics are stored under one folder for easier analysis.

ADVANTAGES AND BENEFITS

SCHARP attributes the high adoption rates of LabKey Server among its collaborators to LabKey's secure document hosting and management features.

"Atlas administrators particularly value LabKey Server's ability to lock down information (data, documents, discussion boards, etc.) such that users can access only approved data and documents. Atlas administrators appreciate LabKey's simple interface for assigning fine-grained levels of read/write permissions to both individuals and groups of users. LabKey allowed SCHARP to manage permissions from a simple, web-based interface. Alternatively, an individual staff member within a clinic could receive permissions sufficient for viewing only data she has entered."

— Lisa Isozaki, Senior Project Manager, SCHARP

LABKEY SERVER SOLUTIONS FOR ASSAY DATA MANAGEMENT

The set of tools listed earlier under "secure collaboration" is primarily adapted to the exchange of unstructured or semi-structured textual information. Examples include web pages of research protocols linked into a wiki and status reports presented chronologically in a message board. These communication mechanisms for unstructured text are necessary but not sufficient to meet the full needs of a collaborative research project. The collaborators also need to exchange highly structured data, such as the lab test results, clinical data, and statistical analyses that are the primary output of their research. Just like the tools for exchanging unstructured data, the mechanisms for exchanging structured data need to be web-based, secure, and integrated into a comprehensive system.

NEED: A WEB-BASED REPOSITORY FOR ASSAY DATA AND METADATA

The vaccine research conducted by CHAVI and CAVD members involves the use of several sophisticated diagnostic tests (assays) performed on the biological specimens collected from study participants. The initial set of tests that the vaccine researchers

wanted to run were plate-based assays, including ELISpot, Luminex, and Neutralizing Antibody (NAb) tests. Each of these different assay types normally stores raw data in some unique file format that can be analyzed by tools developed to read those formats. At its simplest level, the sharing of assay results among collaborators could be an application of the generalized file upload and download mechanisms described earlier under collaboration tools.

But assay datasets for large research projects are not easily managed or shared as individual result files. The person performing the assay needs to record additional information about the data that will allow others to identify who, what, when, where and how the assay data were collected. This information is called metadata, and it is used for cataloging and quickly accessing data records. Assay metadata are crucial to analyzing, finding, comparing, and validating the quality of the assay results. Despite its importance, metadata collection is prone to errors, omissions and inconsistencies, as is true for any data-entry system.

To support the CHAVI and CAVD projects, the SCHARP team needed a web-based system for assay metadata and data collection that would support the following requirements:

Simultaneous capture of data and metadata: To ensure that all the necessary assay data *and* metadata are captured together, the SCHARP team wanted to create data submission web pages that capture both assay data files and researchers' data inputs in one procedure.

Extensible, flexible assay type definition: While the three initial assays needed by CHAVI and CAVD were all based on 96-well plates, there are many differences in the data file formats and the metadata collected for each assay type. The metadata requirements also tend to change over time as researchers gain insights into important factors to capture. The SCHARP team needed a tool that allowed them to select from a few template types for different assay file formats and then customize the metadata collected.

Tools to design and deploy usable, efficient assay data entry

pages: The design of a data entry web page has a direct effect on the accuracy and completeness of the data it collects. SCHARP needed a system that would allow them to design data entry pages that support fast and accurate data entry with minimal duplication of effort.

An architecture that reinforces data quality: Ensuring the quality of data collected is of primary importance to the success of large research studies such as HIV/AIDS vaccine research. The experts at SCHARP have defined a rigorous set of procedures for validating the quality of data collected. These procedures include the establishment of strong standard operating protocols for performing assays and capturing their results.

SOLUTION: LABKEY SERVER ASSAYS

LabKey built an assay management system into LabKey Server to meet the SCHARP team's requirements. The assay management features are powerful and extensible, and have been refined and improved over several releases based on the feedback of frequent users.

Assay Designer: LabKey assay management begins with the assay designer, the tool used by a SCHARP team member for specifying exactly which type of data and metadata are to be captured and uploaded. The user determines the data to be collected by selecting one of the built-in types as a starting point. The user can then customize the built-in assay definition to add or remove properties, select lists, captions and more. Most changes in the configuration of data to be collected can be made via the designer without changing the programming code. If a new assay has a markedly different data file format, LabKey or a sophisticated user can create a new assay type definition and add it to the system.

Generated Assay data entry forms: The end result of the assay design process is a data entry web page that lab researchers use when uploading assay results. The web pages generated by the assay designer support the following features of good user interface design:

- ▶ Data entry fields are type-specific (numbers are distinguished from text and from dates). The formats of data elements that vary by locale (such as dates and numbers) can be customized.
- ▶ The user is only prompted once for values that are definitely going to be consistent across a set of assay runs uploaded at the same time (for example, the lab location). For data elements that may or may not change with each run, the last value entered becomes the default for the next run unless explicitly cleared. There is also a way to let the user quickly apply a value entered once to multiple data entry elements.
- ▶ The navigation between data entry fields is simple, efficient and consistent. Users can tab between the fields in the expected order. The distinctions between upload set properties and individual run properties are clear, as are the steps to achieve "save and continue" or "finish the set".
- ▶ When the choices are confined to a small set of known values, the user is presented with a dropdown list of those choices. Option choices can be easily added or dropped from the user pages by an administrator.
- ▶ The labels next to each data entry field are customizable for clarity. The designer can also add context-sensitive help on each data entry field. The help text that appears when the user moves the mouse pointer over a help symbol next to the field label.
- ▶ The data submission process can take advantage of special features for handling unique identifiers. For example, bar code readers can enhance the speed and accuracy of recording the sample used. LabKey Server also supports the mapping of user-defined identifiers across datasets. If a number of assays were performed and submitted before the sample and participant datasets were established, for example, the ID mapping feature would allow an Atlas administrator to correctly link the temporary IDs used in the assay data to the official participant and sample IDs.

Assay types currently supported: As of version 8.3, the LabKey assay management system includes support for the following assay types:

- ▶ **ELISpot:** ELISpot support includes an online plate-layout tool and recognizes the file formats of three different plate readers.
- ▶ **Neutralizing Antibody:** This assay type (more precisely, TZM-bl Neutralization) best exemplifies how good UI design can reinforce standardization of an assay. The run summary view for this assay shows all of the relevant results in a concise combination of graphical and table elements.
- ▶ **Luminex:** This type illustrates LabKey’s ability to parse the machine-generated output file (Excel format) for most of the metadata needed.
- ▶ **General:** When the entire dataset for an assay can be represented by a small tab-separated text file, the General assay type allows researchers to quickly define a standard form for uploading the file data with any elements of metadata deemed necessary by the designer.
- ▶ **Gene Expression (microarray):** This assay type was developed at the University of Washington and illustrates the benefits of an open source project that welcomes outside contributions.

Currently, SCHARP uses Neutralizing Antibody, Luminex and General assays. It plans to evaluate adoption of LabKey’s tools for Gene Expression in the near future.

LabKey Server also manages the upload and analysis of the complex datasets produced in flow cytometry and mass spec proteomics. LabKey supports high-volume, standardized validation of flow cytometry results. Likewise, LabKey Server supports researchers doing high-throughput proteomics experiments using mass-spectrometry. A detailed description of LabKey’s support for these data types, however, is beyond the scope of this paper.

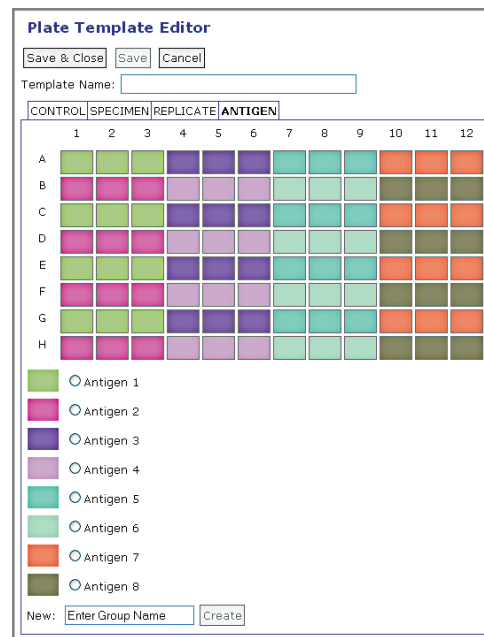
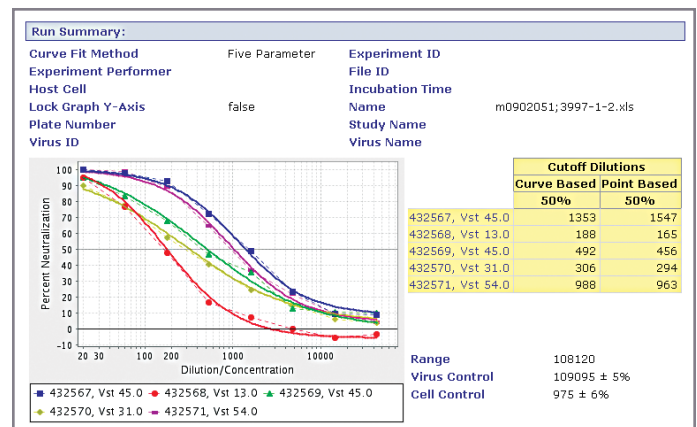


Plate template editor



Summary page for an uploaded Neutralizing Antibody assay

ADVANTAGES AND BENEFITS

Here’s how members of the SCHARP team describe the assay management features that resulted from their partnership with LabKey:

Lexicon Standardization: LabKey’s support for custom drop-down lists in upload forms reinforces the use of standardized terminology across projects.

“We use the Atlas system to share data from neutralizing antibody assays across the CAVD consortium. The process of uploading the assay data enforces the use of a standard lexicon by supplying drop-down menus in the online submission forms. The drop-downs allow entry of only the standard form of any given term. For example, when the name of a reagent is provided via a dropdown menu, any user entering the ingredients of an experiment will use the name standardized across all projects. This makes it easier to compare procedures and results across experiments and labs.”

— Michael Seaman, Director, CAVD Pre-clinical Neutralizing Antibody Laboratory

Data Integration and Standardization: SCHARP found LabKey’s ability to integrate many different types of data particularly helpful in gaining efficiency, uniformity and predictability.

“The unified architecture of the LabKey platform gave us more control over the data workflow, facilitating standardization in data models and user support, eliminating inconsistencies and reducing errors.”

— Blake Wood, Senior Project Manager, SCHARP

Extensibility and rapid implementation: LabKey’s platform allows integration of new data types as technologies change. SCHARP looks forward to easily adding new data types (particularly Microarrays and SNPs) to its repository.

“SCHARP values the ability to extend LabKey’s existing data management tools instead of having to implement new data management systems for new data types as they arise.”

— Craig Margaret, Manager of Information Systems and Scientific Computing, SCHARP

LABKEY SERVER SOLUTIONS FOR STUDY MANAGEMENT

The SCHARP team needed secure collaboration tools and an assay repository in order to support observational research studies run by the CHAVI and CAVD consortia. The types of studies conducted by the two consortia are markedly different, but SCHARP manages both sets of studies with LabKey Server.

CHAVI conducts multi-year, multi-site studies of hundreds of people who have or who are at risk of contracting HIV/AIDS. CHAVI study participants agree to visit a clinic on a regular schedule. At clinic visits, participants are examined, specimens are collected, and health status forms are filled out. A substantial number of participants and clinics are in Africa. CHAVI runs several such studies simultaneously.

In contrast, CAVD runs a larger number of smaller-scale studies that involve a small number of human or animal subjects, mostly at a single site, and mostly with much shorter time frames. Like CHAVI studies, these studies involve collecting specimens from subjects over time. Also like the large CHAVI studies, the small-scale CAVD studies need to integrate many types of measurements – assays, exams, and forms—by study specimens from study subjects over time.

NEED: INTEGRATION, ANALYSIS AND PRESENTATION OF STUDY DATA

At a most basic level, all of these studies involve similar steps in managing data.

- ▶ **Identification and characterization of the source of samples:** Samples can be collected from people (“participants”) or animals (“subjects”), or simply acquired (as in the purchase of a standard protein mixture, for example).
- ▶ **Data capture:** Datasets come in many formats. Many are in the form of spreadsheets, but their row and column layouts vary widely. Other diagnostic tests produce large XML files. Medical history forms filled out by participants and

examination forms completed by clinicians are usually paper documents, and the clinics themselves often lack reliable web access. In these cases fax machines and email accounts are often used for the initial step in submitting data.

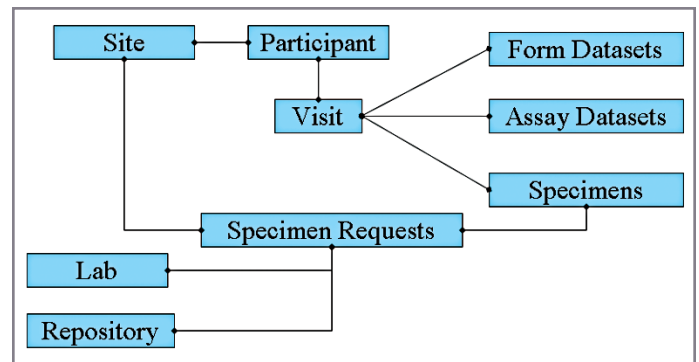
- ▶ **Data analysis and interpretation:** Much of the data captured requires some form of pre-processing before a researcher can make use of it. Mass spectrometry output files, for example, have little biological meaning until they have undergone a complex set of calculation steps performed by software.
- ▶ **Data validation:** Sometimes the quality assurance process is not uniform across labs, so additional quality assurance is necessary to prepare data for analysis by statisticians. Data collected can be invalid for any number of reasons, including researcher error, instrument defects, software defects, and experimental conditions.
- ▶ **Data integration:** The datasets being analyzed and validated do not exist in isolation. They are pieces of a whole picture (the subject) taken at different times.
- ▶ **Data sharing and presentation:** Once the data sets have been thoroughly analyzed, validated and integrated, they are ready to be shared and formatted to best communicate some result.

SOLUTION: LABKEY SERVER STUDY FRAMEWORK

SCHARP chose to partner with LabKey to build the study management application for these needs. The study module that resulted from this partnership is a standard part of the LabKey Server open-source project. It handles both the large, highly structured CHAVI studies as well as the multitude of smaller-scale CAVD studies.

LabKey studies start with a definition of the subject or “participant” of the study. This entity is the focal point of data integration. It might be a human study participant, or an animal subject, or a genetically consistent strain of yeast cells. Participants are identified uniquely and can be assigned to groups (“cohorts”) for the purpose of comparisons between

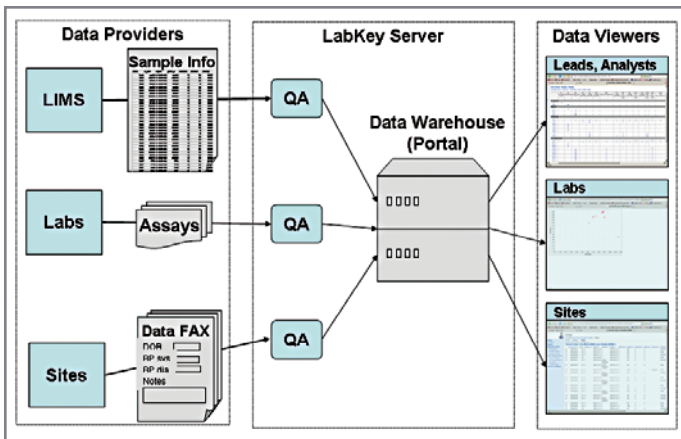
populations. Participants are described by characteristic properties (“demographics”) that can be drawn from either standard vocabularies or any custom property the researcher wants to track. Participants are the source of biological tissue specimens of particular tissue-type designations (blood, tumor, etc.).



The core entities in the LabKey study framework.

The other key to study data in LabKey is the time dimension. LabKey supports pre-planned, highly-structured studies that are based on a schedule of “visits,” the points in times when data on participants is recorded and/or specimens collected. LabKey also supports more ad-hoc studies in which the time value is a date, and can be either the date on which a specimen is collected or the date on which an assay is processed. Together, the subject identifier and the time point value (plus an additional key if needed) are used to identify the source of a dataset. This enables LabKey Server to integrate data into powerful analysis views that track subjects across time.

Besides the assay data integration features previously described, LabKey Server also supports the loading of form data, where forms are either paper or electronic data entry pages that are imported into the study. LabKey integrates tightly with DataFAX to support the display of data submitted by fax from remote clinics.

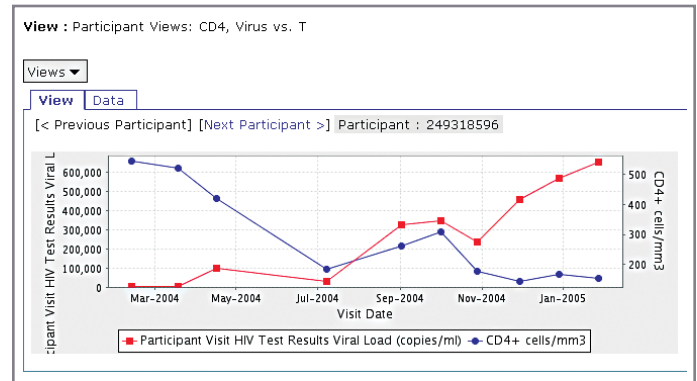


Data integration in LabKey can be done in either a “top down” or a “bottom up” fashion. Participants, specimens and visits can be pre-defined and assigned unique identifiers at the beginning of the study such that they are tied to all datasets when they are collected. Alternatively, the assays and form data can be tracked using some alternate key at data collection time and then mapped to the true study identifiers at upload time.

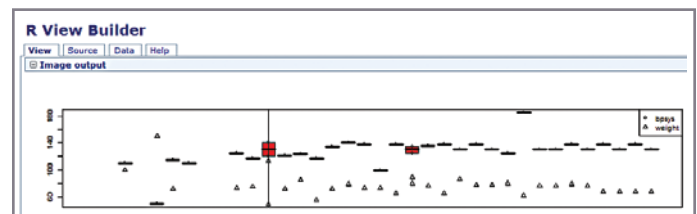
LabKey Server supports a number of powerful analysis tools that operate on the loaded data.

- ▶ The LabKey **Data Grid View** supports interactive filtering and sorting of any single dataset, as well as export of any dataset to Excel or TSV files.
- ▶ The **Study Navigator** is a web page component that supports interactive roll-up and drill-down of assay results, form data, and sample characteristics. Drilling into a participant view shows all of the data for one subject, whereas drilling into a time point shows data from all of the subjects corresponding to that point in time or that scheduled visit.
- ▶ The LabKey **Query** facility connects related datasets together. This allows the user to compose and answer complex questions about the data without having to know how to do a SQL join.

- ▶ **Support for R scripting:** The R open-source language enables sophisticated statistical analysis and charting. Additionally, it allows interactive data visualization that can lead to deep insights into the validity and interpretation of the data.



The Study Navigator supports drill-down to view and chart data per participant



Example of the use of R within LabKey Server. The plot shows blood pressure and weight progressing over time for all participants.

The SCHARP team takes advantage of all of these study management tools in support of data quality. CHAVI studies, for example, use a work flow along these lines:

1. A participating clinic collects data in the field, often in Africa.
2. The clinic uploads its data via DataFax, which converts the data from a fax image into the records and fields of a data entry form. The accuracy of this conversion step is checked by data coordinators at SCHARP.

3. The scrubbed form data is posted on Atlas for use by selected CHAVI collaborators.
4. At the same time, lab technicians upload lab data via Atlas (vs. DataFax). They do quality control before sharing this data with the broader team.
5. SCHARP biostatisticians analyze both lab and field data.
6. Collaborating principal investigators review the analyses and possibly request further analyses.
7. After finalization, collaborators write up their conclusions for publication.
8. After publication, study data are shared publicly as part of a Public Use Dataset (PUD). Anonymity of participants is maintained.

As the datasets are subjected to increasing levels of review, they can be shared with an increasingly broader circle of consortium members and funding agencies. Users can build data-centric reports and charts to communicate their interpretations of the consolidated data to the extended team. LabKey Server's data-centric tools are completely integrated with the more general collaboration tools described earlier. This allows users to author a wiki, for example, that combines data presentation with free-form text that interprets and explains the data.

The programmability and customization features of LabKey Server allow SCHARP to take full advantage of the high-quality data in their repository. In the CHAVI studies, for example, it became clear that some of the specimens were in extremely high demand to participating researchers—in particular the specimens collected from participants who contracted HIV/AIDS while participating in the study. The SCHARP team quickly realized they needed a process to allocate these valuable specimens to the researchers who could make the best use of them. LabKey built a specimen request tracking system into the study module to meet this need. The system supports a configurable series of sign-off actions for specifically authorized

users across multiple sites. The SCHARP and LabKey teams have also defined and built into LabKey Server an easier and more flexible programming model for creating data-centric web pages. The SCHARP team will use this programming interface to improve the usability and functionality of the study management pages, enabling users to gain deeper insights into the data.

ADVANTAGES AND BENEFITS

The LabKey Server study features enabled the SCHARP team to meet the needs of the CHAVI and CAVD projects quickly and comprehensively. As they describe it,

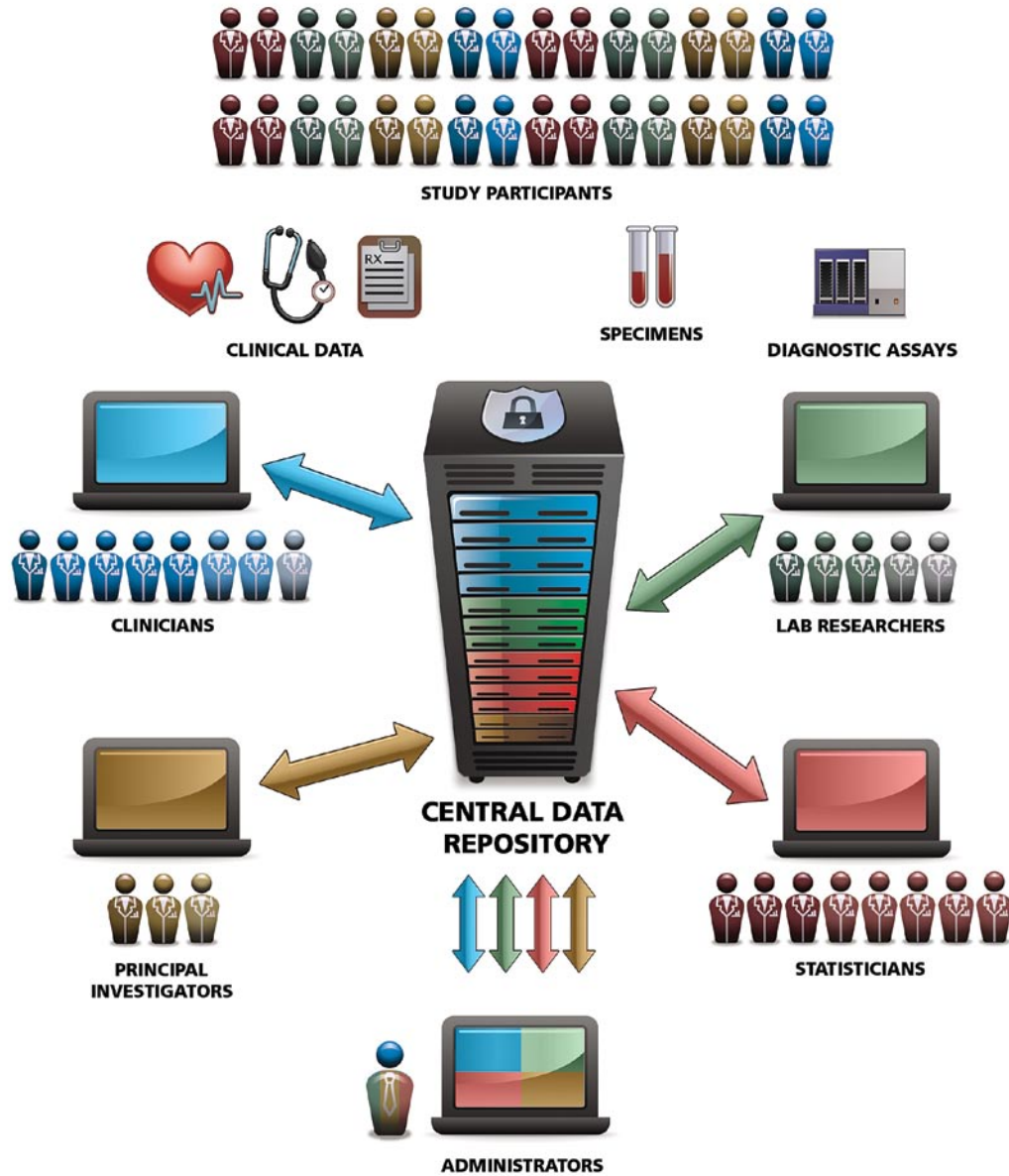
“Before adopting LabKey, SCHARP lacked an online staging area for performing quality control on uploaded data. LabKey will allow SCHARP to quickly visualize and analyze uploaded data, then share the results widely only after finishing quality control. Technicians can review the results of an assay before copying assay data to a study folder shared by the rest of the team.”

— Sarah Ramsay, Collaborative Services Manager, SCHARP

“With LabKey in place, SCHARP is comfortable taking on more clients and additional data types. It is easy to train new users. And users have the ability to “mine” across all the data in the system once it is published to collaborators. Altogether, the system is clearly a powerful research accelerator.”

— Tom Skillman, Deputy Director, SCHARP

The study management features, combined with secure collaboration tools and assay repository support, make LabKey Server a powerful platform for enabling collaborative research. LabKey Server has become a key element in SCHARP's strategic plans and organizational capacity to support large collaborative projects.



LabKey Server facilitates collaboration by providing secure web access to a central data repository

THE BIG PICTURE: ACCELERATING RESEARCH PROGRESS

SCHARP has realized many advantages in partnering with LabKey to build the Atlas information system for CHAVI and CAVD. These advantages include:

- ▶ **Flexibility:** The system can incorporate new requirements quickly.
- ▶ **Scalability:** The system can handle very large studies and high-throughput experiments.
- ▶ **Security:** A simple and powerful security system limits data access to authorized users only.
- ▶ **Integration:** The system combines all the relevant information in one place.

- ▶ **Data Quality:** The system enables SCHARP to ensure higher quality data.

Ultimately, partnering with LabKey enhanced SCHARP's organizational capacity to manage large collaborative projects.

All of the Atlas features built or enhanced to support CHAVI and CAVD have become part of the LabKey Server platform, and are thereby freely available for other collaborative projects. And the LabKey Software team is available to tailor the platform to the unique needs of those projects. 🧪