

---

## Bridging Datafax and Labkey

David Liou  
Seattle Biomedical Research Institute

Mark Igra  
LabKey Software



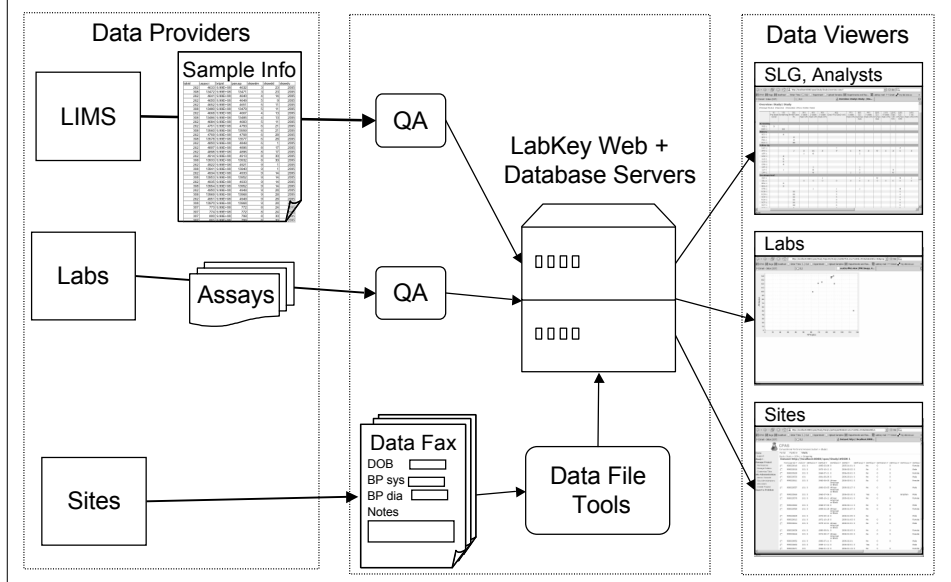
## Overview

- Delivering CRF information to researchers can be a daunting and time-consuming task. Finding a delivery system that is productive, user-friendly, and seamless is a challenge.
- The Seattle Biomedical Research Institute has deployed the DataFax clinical data management system in parallel with the open-source Labkey Study information delivery system (<http://www.labkey.org>) as a solution to this problem.
- **The following presentation is our approach to bridging the two systems.**

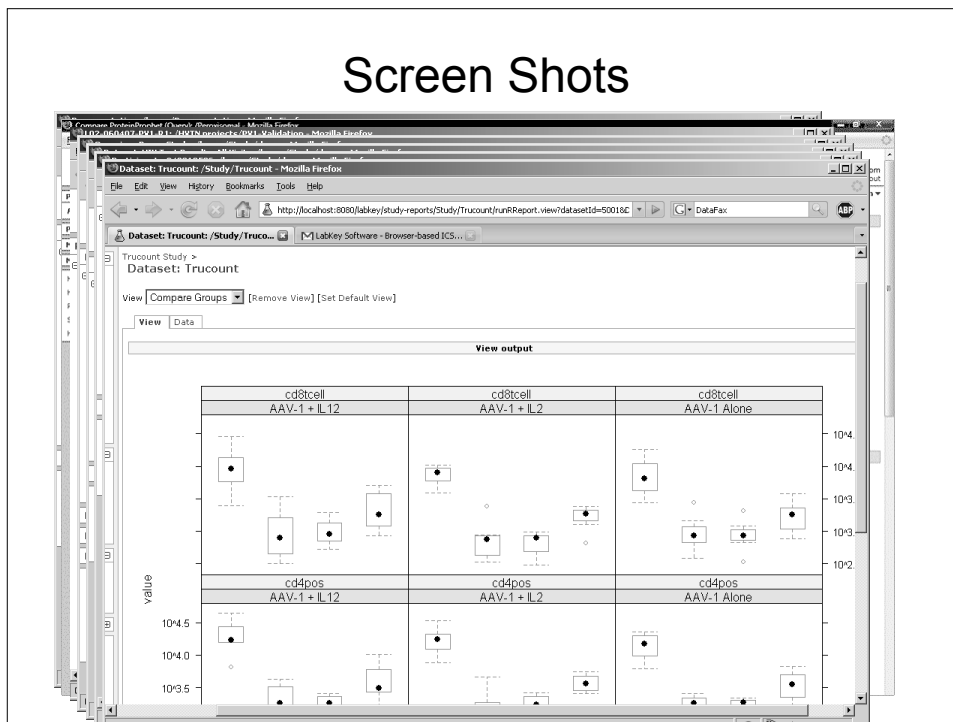
## What is LabKey

- Web-based platform processing, storing, integrating and sharing data from biomedical research
- Includes tools for storing and analyzing Proteomics, Flow Cytometry and other Assay Data
- Integrates Assay Data from Labs with CRF data from DataFax
- Secure
  - Permissions assigned to groups
  - Each study, dataset & report can be secured
- Pipeline for loading study data from DataFax

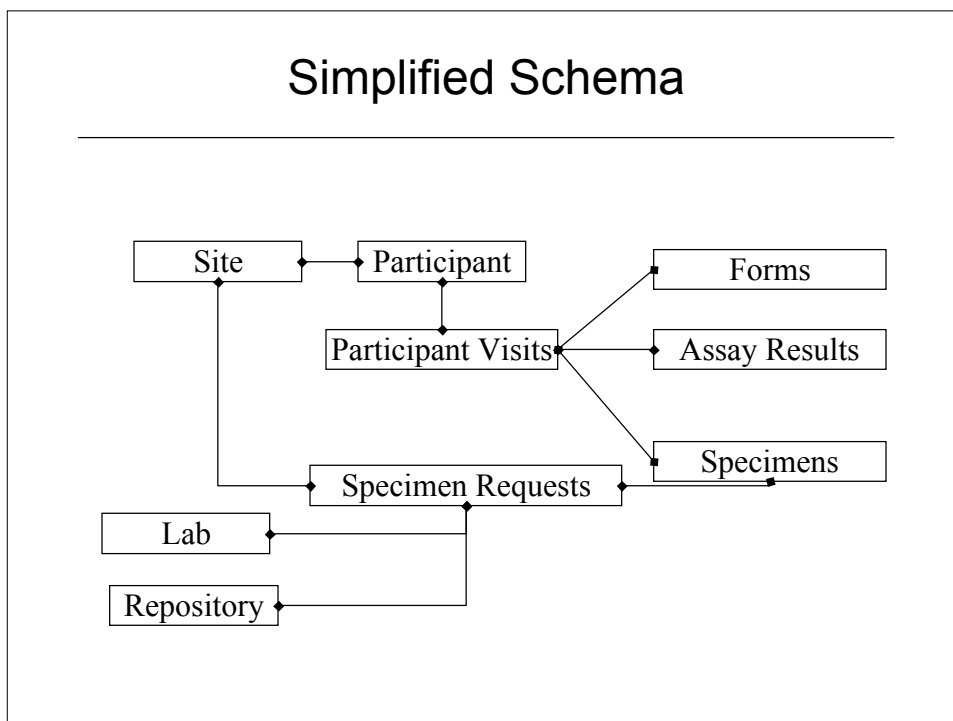
## Portal Data Flows



## Screen Shots



## Simplified Schema



## Project Goals and Requirements

---

- **Automated:** the transmission of data from DataFax to Labkey should run daily without user intervention
- **Dynamic:** the process should be aware of updates to the CRFs; the addition, modification, or removal of fields, should be reflected in the refreshed Labkey system

## Tools

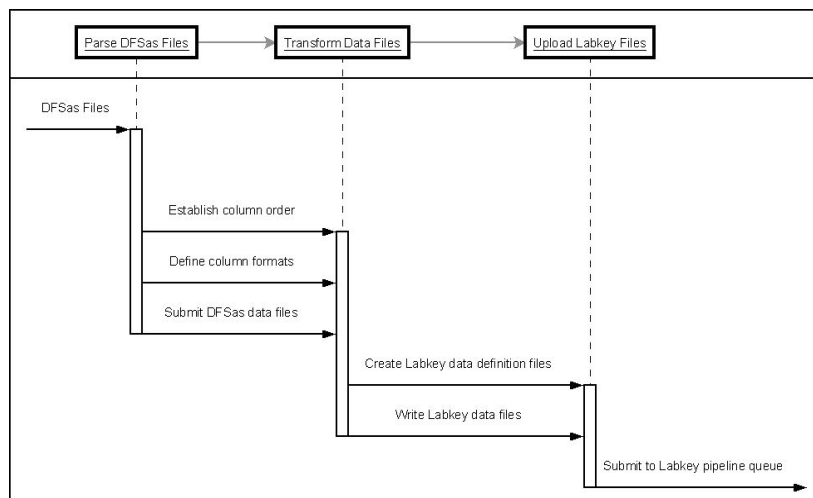
---

- DFSas
- Java
- HSQLDB

## Flow Of Process Overview

- The transmission process of data from DataFax to Labkey occurs in 3 basic steps:
  1. Parsing the DFSas Files: data definitions are formally defined based on the DFSas job file and the SAS file
  2. Transforming the Data Files: the Labkey dataset schema is created and the data files are reformatted
  3. Uploading the Labkey Files: the pipeline process to upload new data is initiated

## Flow Of Processes Diagram



## Parsing the DF SAS Files

---

- **Establish column order**
  - Column names are not included in the DFSas data files; they must be parsed out of the DFSas job file
- **Define column formats**
  - Column data types are not integrated into the DFSas data files either; these are extracted from the SAS file
  - Field codes, such as 1 – “Yes”, 2 – “No”, are also determined so that they can be replaced by their designated values
- **Submit DFSas data files**
  - The data files are imported into a temporary embedded relational database system (HSQLDB)

## Transforming the Data Files

---

- **Create the Labkey definition files**
  - Labkey requires a dataset schema
  - Labkey uses data format designations as defined by the World Wide Web Consortium (<http://www.w3.org/>)
    - xsd:int -- integer
    - xsd:double – floating point number
    - xsd:string – any text string
    - xsd:dateTime – date and time
    - xsd:boolean -- boolean
- **Write Labkey data files**
  - Data files submitted must be in a TSV (tab delimited) record format

## Uploading the Labkey Files

---

- **Submit to Labkey pipeline queue**
  - The Labkey data files are placed into the pipeline directory
  - The upload process is initiated via a secured URL triggered after the data files are in place

## Conclusion

---

- Labkey allows users to view study data in a medium that is user-friendly and flexible.
- The integration of DataFax and Labkey shows incredible promise as an infrastructure to delivery study information in a seamless and automated manner.