
Self-Serve DataFax

Connor McCoy
Seattle Biomedical Research Institute



Overview

- Researchers at SBRI are using data gathered from a small number of studies for a wide variety of purposes.
- Streamlining the process through which scientists build datasets provides an ongoing challenge.
- Our ideal solution: provide tools, not data.
 - Keeps programmers out of process.
 - Selfish.
 - But good!

What We've Done

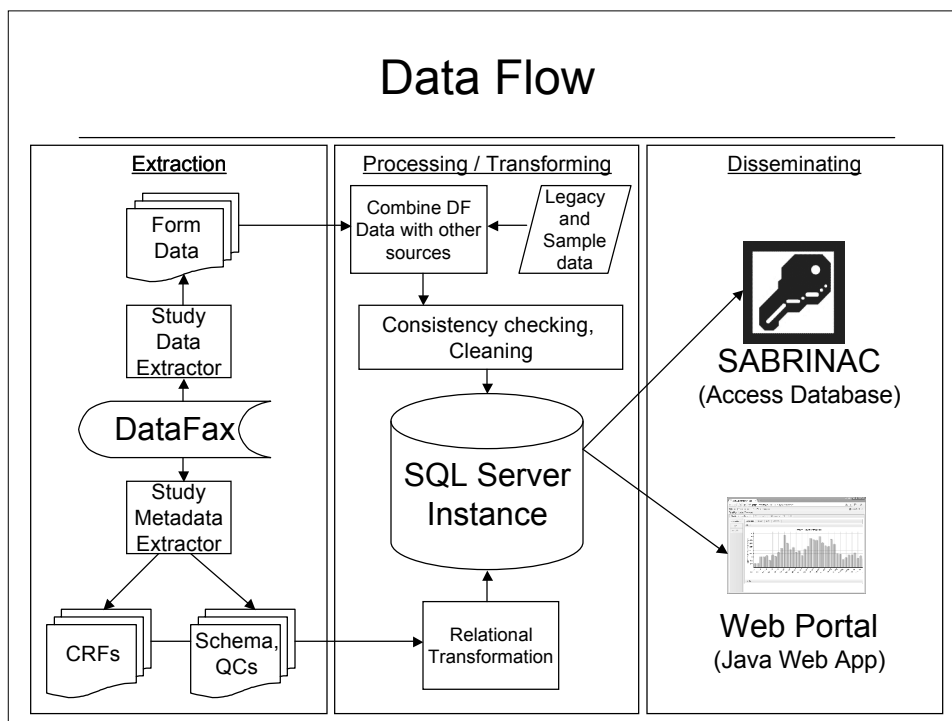
Created two end-user tools:

1. SABRINAC

- An Access Database available for download with up-to-date, cleaned study data and bundled time saving tools.

2. A Lab Portal

- One stop shop for data reports, CRF images, QC information, etc.



Extract Data From DataFax

We're interested in two things:

- 1. Actual Study data** – what's on the forms
- 2. Metadata** – QC Information, study schema, CRF images

Extract Data From DataFax

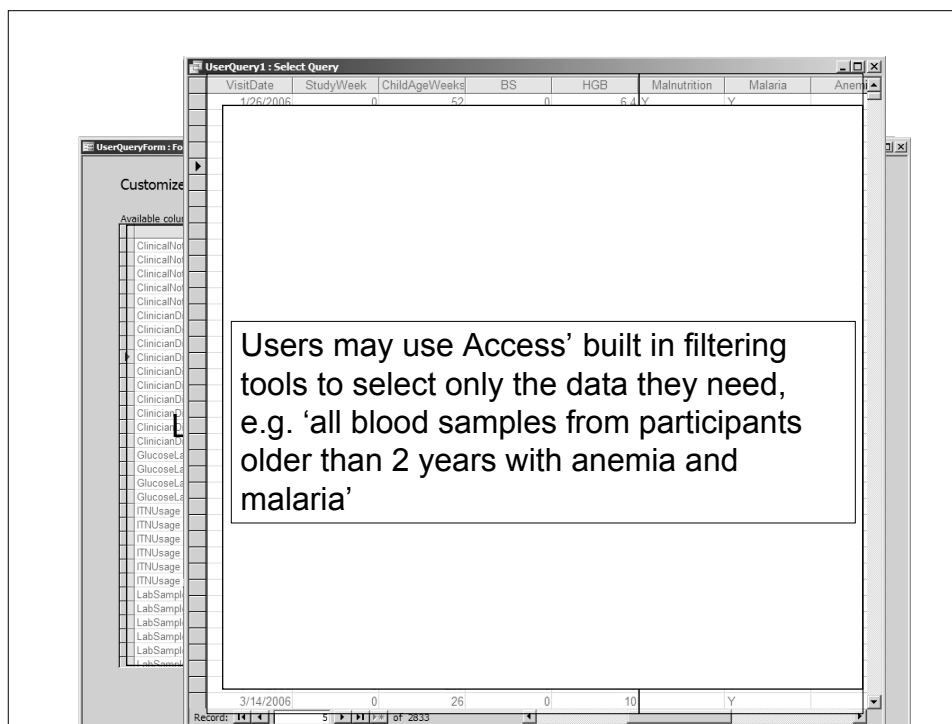
- Acquired via a pair of cron jobs, executing daily:
 - First extracts all study data via DFsas (BASH script).
 - Second extracts study metadata (Python script):
 - QC notes
 - Study schema
 - Plates
 - CRF images (these are watermarked and stored in a repository as PNGs...only new CRFs are extracted)
- Resulting data are picked up daily via rsync.

Processing & Transforming

- **Two Java Programs**
 - First uploads raw study data to SQL Server Database.
 - Second uploads metadata, including image index.
- **Database transformation**
 - Stored procedures to:
 - Combine data from 3 observational studies into single, large dataset.
 - Combine data from DataFax with data from our in-house sample tracking system.
 - Convert to a more relational format.
 - Link QC data with images and participants.
 - Clean data, removing inconsistencies, dirty records, error records, etc.

Providing Researchers with Data: SABRINAC

- **SaBRI Nightly Automated Compilation**
- **Microsoft Access Database available for download.**
 - Why? Many users already familiar with Access, good Excel integration.
- **When first opened, creates tables based on most recent data from SQL Server.**
- **Provides researchers with a sandbox for building datasets.**
- **Distributed with tools designed to make common tasks easier.**



Results

Advantages

- Easy for researchers to create their own dataset.
- Users have access to a copy of the full database, if they are comfortable using it.

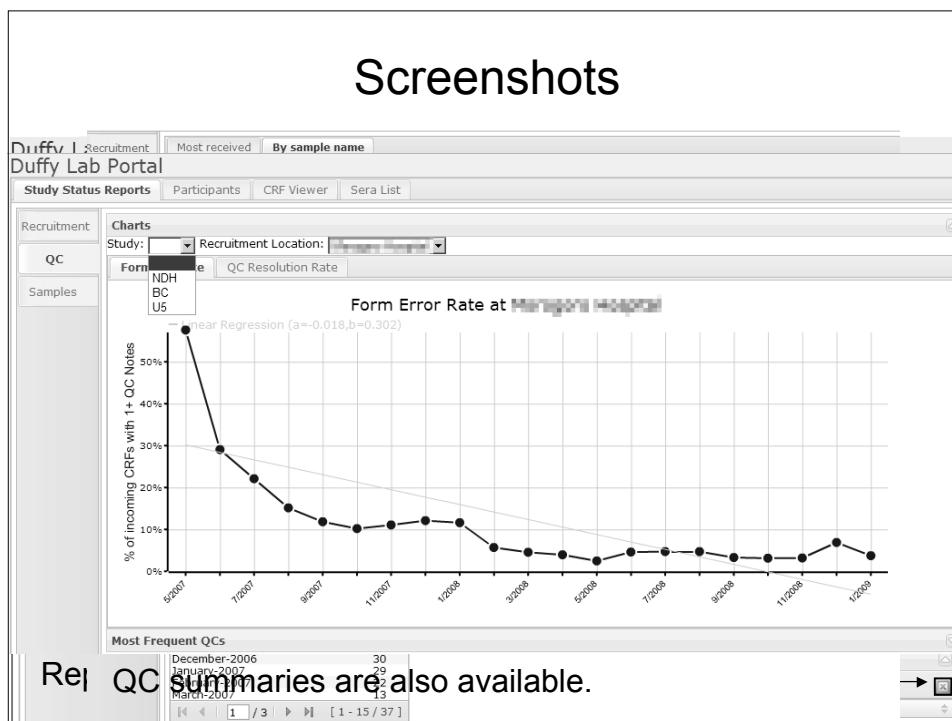
Disadvantages

- Access is not designed with security in mind.
- Users must re-download the database to stay up to date.
- Only updated once daily.

Providing Reports: Our Lab Portal

- Developed to meet need for reporting on study status.
- Java Web Application built on the ZK Framework.
 - Easy, desktop-like web applications in Java.
- Secure:
 - Authenticated against LDAP Server.
 - Role management .
 - SSL.
- Read only access to data.

Screenshots



Screenshots - CRF Viewing

The screenshot displays a web interface for viewing CRF information. A modal window titled "CRF Information" is open, showing details for a study. Below this, a table lists "Quality Control Notes for this CRF" with two entries. The background shows a "Users ma" window and a "Can view info" button.

Status	Field	Problem	QC	Created	Resolved
✘	2c. Do children sleep under nets?	Missing value	Question '2c. Do children sleep under nets?' is a required field based on the response to '2. Does family have mosquito nets?' but currently it is blank. Please clarify.	jarthur 20/NOV/2007	
✘	5. Income apart from farming	Missing value	Question '5. Income apart from farming' is a required field but currently is blank. Please clarify.	jarthur 20/NOV/2007	

Benefits

- Allows study PIs to quickly see summary reports relevant to study performance.
- Provides researchers with images of faxed CRFs.
 - Researchers can check values outside the norm, make sure there wasn't a data entry error (common request).
 - Saves data management staff time.
 - CRFs are watermarked as copies.

Conclusions

- Very active interest and use of these new tools.
- Identifying trends in Quality Control is significantly easier.
- Streamlining the link between data consumers and data producers makes everyone happier.

Questions?

Links

ZK Web Framework: www.zkoss.org

Python: www.python.org

Open Flash Chart: <http://teethgrinder.co.uk/open-flash-chart-2/>