UNC
THE NORTH CAROLINA TRANSLATIONAL & CLINICAL SCIENCES INSTITUTE

UNC
COLLABORATIVE STUDIES COORDINATING CENTER
DEPARTMENT OF BIOSTATISTICS

DATA MANAGEMENT SOLUTIONS FOR CTSA SETTINGS: A CASE STUDY FROM UNC

Lisa LaVange, PhD
Professor and Director, Collaborative Studies Coordinating Center (CSCC)
Department of Biostatistics, Gillings School of Global Public Health
University of North Carolina at Chapel Hill
Deputy Director, NC TraCS Institute

Outline

- Data management as a core service of NIH Coordinating Centers
- UNC's TraCS data management initiative
- TraCS DMS: Design and Development
  - System overview
  - Screen shots
- Progress/challenges to date
- Cost model for investigators

NIH Coordinating Center Model

- Key component of federally funded, multi-center studies
  - Clinical trials
  - Epidemiological/observational studies
  - Patient registries
- Services include
  - Project management and coordination
  - Data management (new data collection or existing data integration)
  - Statistical aspects of study design and analysis
  - Management of central labs/reading centers
  - Data dissemination and publications
  - QC/QA
- Scientific purpose is to ensure consistency across centers in study design and conduct
  - Critical for epidemiological studies in absence of randomization

Collaborative Studies Coordinating Center (CSCC)

- Resides in the Department of Biostatistics
- Founded in 1971 as the Lipids Research Clinics (LRC) Data Coordinating Center
- 2nd oldest and longest continuously-funded NIH coordinating center
- Currently includes 90 faculty, staff, and students
  - Matrix organizational structure
  - Functional areas: Data management; Statistics; Statistical Programming; Project management; Clinical monitoring
- Over 20 studies: clinical trials, epidemiology studies, patient registries, pharmaco-epid studies
- Disease areas: cardiovascular, pulmonary, mental health, kidney, injury prevention, periodontal, fetal health, etc.

CSCC Data Management

- Tradition of innovation in data management and statistical computing
  - First CC to implement remote data entry (1986; for ARIC)
  - One of the first CCs to implement web-based data entry (2001)
- Currently implementing data collection via ‘smart pens’ and ‘digital tablets’
  - In clinic settings (TraCS)
  - For household screening in the field (HCHS)

TraCS Data Management Service

- Key initiative of both the Biomedical Informatics and Biostatistics Cores
- Included in UNC's initial CTSA submission (Jan 2007) and resubmission (Oct 2007)
- An early milestone to show visible and meaningful progress
- Rationale:
  - In spite of a proliferation of systems and services across campus, many clinical and translational investigators have no access to systems beyond a spreadsheet for single-keyed data entry
  - Little or no coordination across campus systems exists
- Goal: Every TraCS investigator has access to a secure, high quality, data management system for data capture, processing, management, and reporting

TraCS Data Management Service

- Vision:
  - Scope: all data collected under TraCS approved protocols
  - Integration: use of data standards to promote data sharing
  - Scalability: small, single-center to complex, multi-site studies
- Planned implementation:
  - TraCS Data Management Service created and located within TraCS Institute for easy access
  - Build upon existing GIC Informatics resources in Med School
  - Hire IT manager to direct service
  - New systems development would leverage resources and experience within CSCC
  - Hire project manager/statistician in Bios Core to advise investigators on data collection modalities and serve as liaison to Data Management Service

TraCS Data Management Service

- Initial steps – conducted in parallel
  - Prototype design components based on use cases
  - Buy vs Build determination
- Commercial package evaluation
  - Analysts' reports (IDC 2004 vendor comparison)
  - Informatics consortium (CTSA Wiki; UW EDC Evaluation Project)
  - Informal CTSA networking
  - On-campus vendor meetings
- Results
  - Velos most common among centers contacted
  - OpenClinica appeared to be most viable solution

TraCS Data Management Service

- OpenClinica
  - Open source—important for state-wide reach of TraCS
  - Widespread use among clinical research centers
  - Paid service support model
  - Technology platform aligned well with internal development
  - Appears flexible for external system integration
- Short-term solution:
  - REDCap offered an immediate solution for small, less complex studies
- Long-term solution:
  - Build system around OpenClinica for use with more complex (e.g., multi-site or regulatory) studies
- Early on, realized too much customization was required to make open source solution via OpenClinica feasible

TraCS Clinical Research Data Management System (TCR-DMS)

- Development timeline: Oct 2009 – June 2010
- 5th generation CSCC DMS
  - Co-direction of Hope Bryan, CSCC IT Manager, and Brent Lamm, TraCS IT Manager
- Development goals:
  - Standards-based to facilitate data sharing: file structures, variable definitions
  - State-of-art (next generation) systems design
  - Library of re-usable modules in object-oriented environment
  - Optimization of system performance

TraCS Data Management System (TCR-DMS)

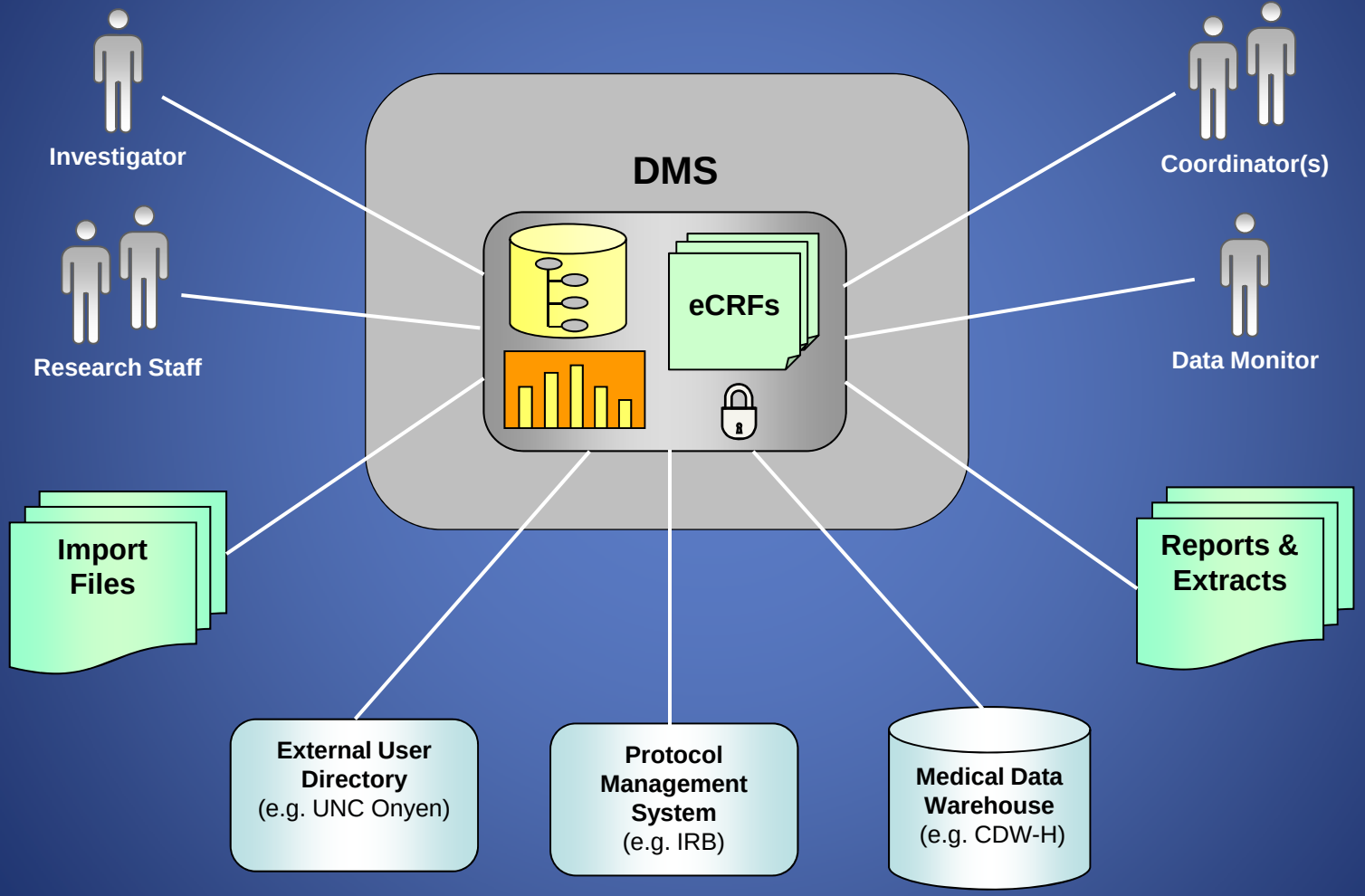
- TCR-DMS Development goals, cont.
  - Allow easy access at *low cost* for TraCS investigators
    - Minimize customization required for each protocol via use of standard CRFs
  - Features
    - Options for randomization and other algorithms
    - Real-time reporting/tracking
    - Real-time edit, queries, and resolution
  - Data security: FDA 21 CFR Part 11 compliance
- Need a new name! *Naming contest in progress…*

System Overview

TCR- DMS

- A clinical data management system, to support:
  - Clinical Trials
  - Patient Registries
  - Longitudinal & Observational Studies
- Key functionality included:
  - Study configuration - Protocol details, visit schedules, site specific definitions, eligibility determination & treatment assignment
  - *Case Report Form (eCRF) management* - online form design, custom and standard eCRFs from central library
  - *Electronic data capture* - custom field validation, complex conditional logic, automated data queries
  - *Integration of external data feeds* - laboratory and reading center imports
  - *Data extract and reporting* - CDISC ODM data interchange, standard and ad hoc reports

System Overview

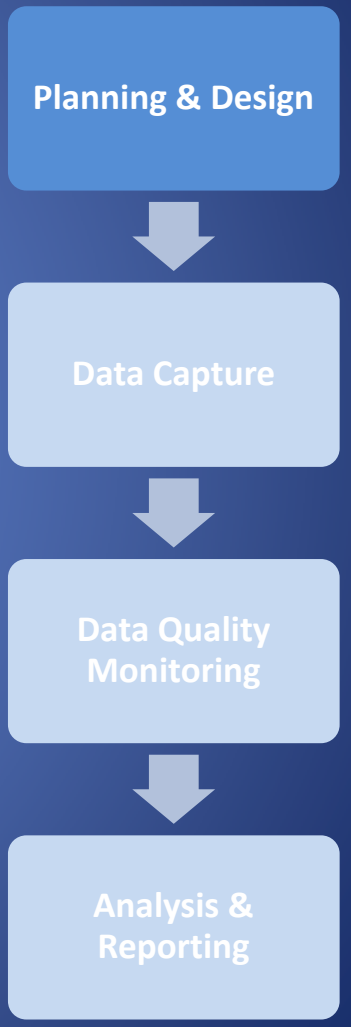


Study Phase: *Planning & Design*

- Specification of study and key protocol parameters
- Definition of sites, including primary contacts and users
- Creation and customization of case report forms
- Implementation of visit schedules with associated case report forms and data imports
- Define coding dictionaries (e.g. medications)

Key Features:

- ❖ Configurable subject id assignment schemes – automated sequencing and pre-defined list support
- ❖ Customizable required fields for enrollment
- ❖ Granular authorization model – site and event level permissions assignment
- ❖ Online creation and customization of case report forms ***(* v2)***
- ❖ Support for elaborate graphically designed forms

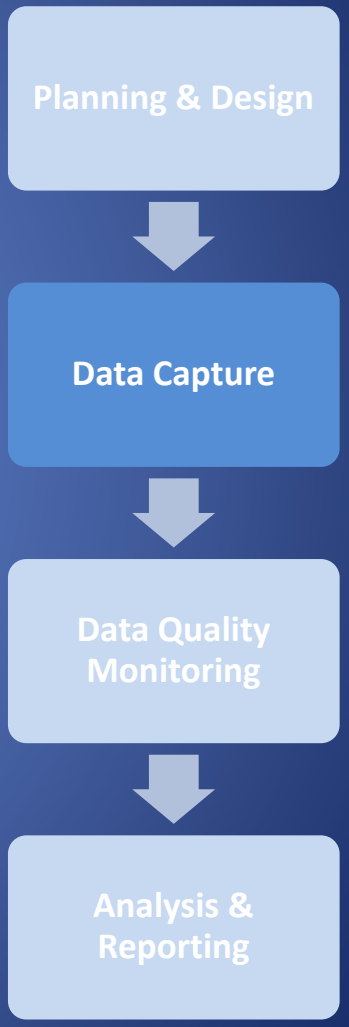


Study Phase: *Data Capture*

- Capture of research subject data via electronic case report forms
- Real-time validation against pre-defined:
  - Conditional field logic
  - Acceptable values
- Central upload of exam data from laboratory or reading center – merged with form captured data
- Central and/or geographically distributed data capture into single study

Key Features:

- ❖ Full auditing of all user interface interactions
- ❖ Auto-save of form data during data collection to prevent data loss
- ❖ Ability to override specified valid values appropriate justification

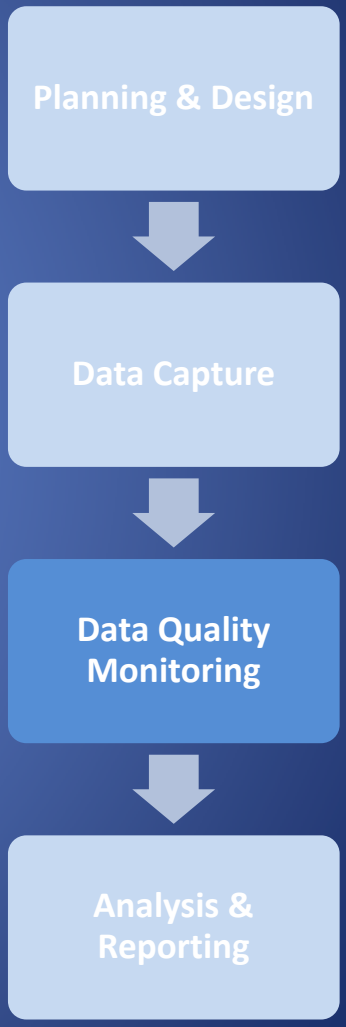


Study Phase: *Data Quality Monitoring*

- Define and execute standard reports for study specific data quality checks
- Review and resolve automatically generated data queries
- Lock completed and validated case report form data
- Ad hoc reporting to investigate data discrepancies
- Revise case report forms to accommodate study modifications

Key Features:

- Enforced separation, through authorization model, of duties between users performing data capture and data monitoring
- System restriction of modifications to locked data
- Revision tracking of updated case report forms

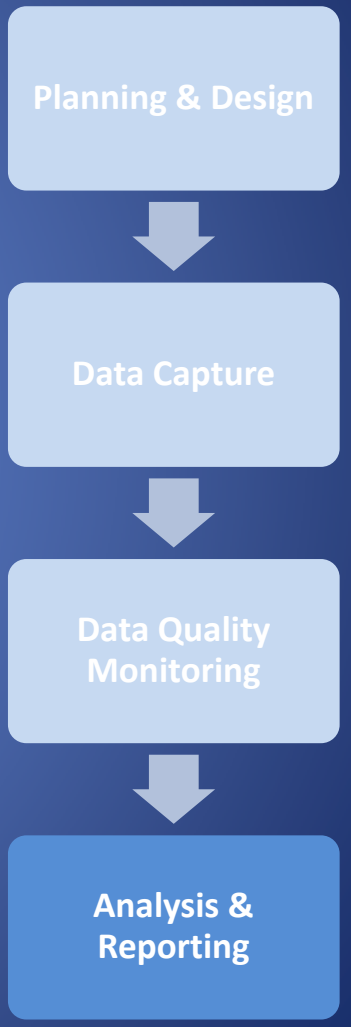


Study Phase: *Analysis & Reporting*

- Pre-defined and ad hoc reporting for study oversight (e.g. recruitment tracking)
- CDISC Operational Data Model (ODM) data extraction for input to analysis software
- Association of all data items to analysis variable names (i.e. SAS variable names)
- File download into common file types (e.g. CSV, SAS, Stata, SPSS)

Key Features:

- ❖ Field-level storage of captured data by standard types (e.g. integer, float, date) for enhanced data integrity
- ❖ Ability to perform cross-CRF and cross-study reporting due to form-independent data model



Scalability

- Designed to support full range of research studies

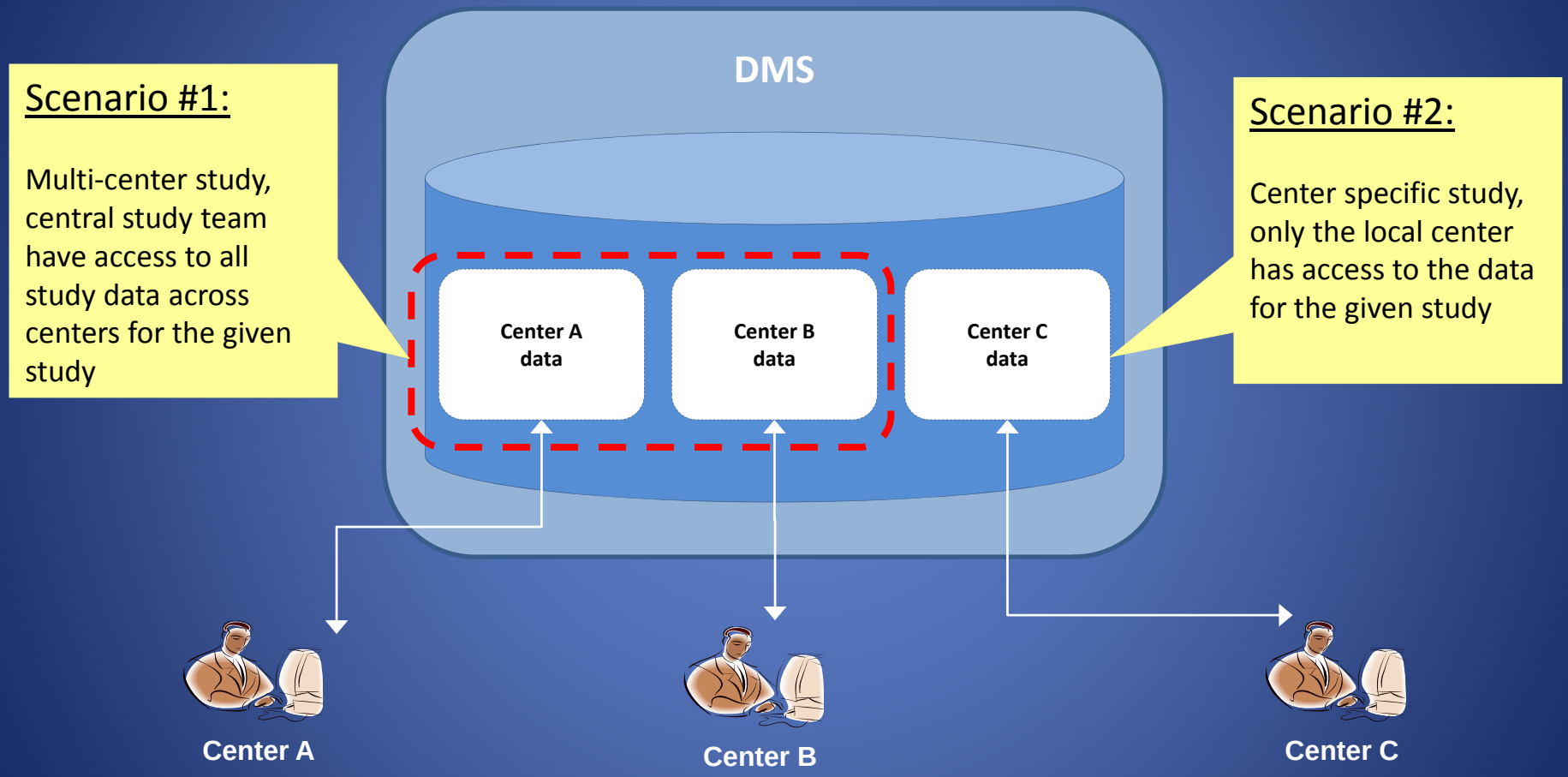
- Small / simple installation footprint to support laptop installations where network connectivity is unavailable
  - 100% web-based Java for full platform compatibility
  - Support for MySQL as database management system
  - CDISC ODM functionality to upload/merge data into larger study instance

- Scalable to support large, long-term, multi-site studies
  - Support for high-end database management systems (e.g. Oracle, DB2)
  - Capability to implement as single study instance, or multi-study instance with secure logical separation of data

Standards Supported

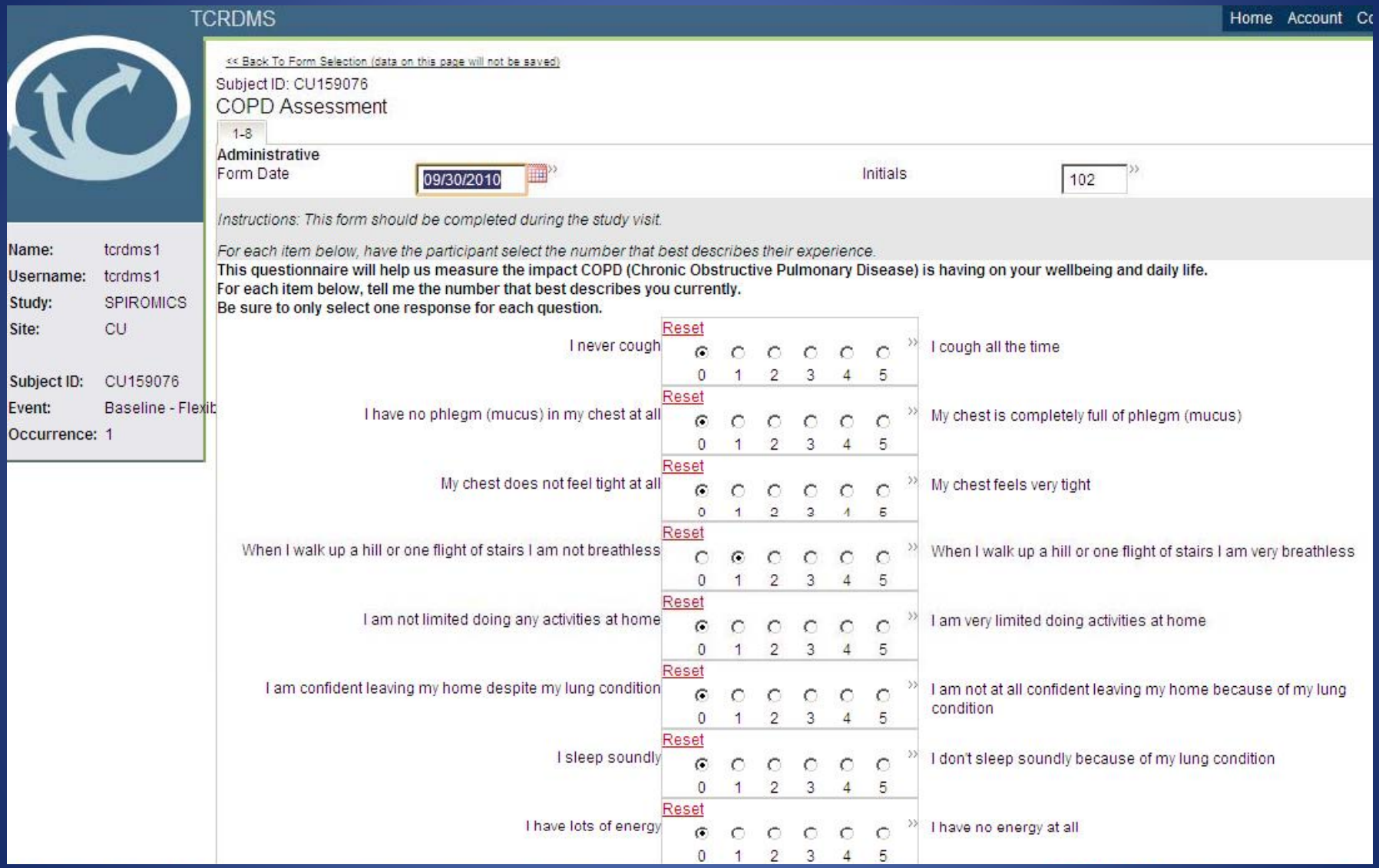
- CDISC Operational Data Model (ODM)
  - Data format for interchange and archival of data
  - Standard XML file includes study data, meta-data, audit information
  - Compatible with many software systems
  - Adopted by FDA and potentially by NIH
- CDISC Clinical Data Acquisition Standards Harmonization (CDASH)
  - Basic standards for data collection including best practices for CRF design
  - Recommended set of data collection elements for 16 domains
- FDA 21 CFR Part 11 - information security *(Version 2)*
  - System validation
  - Audit trail
  - System access controls
  - Electronic signatures

Data Isolation & Sharing

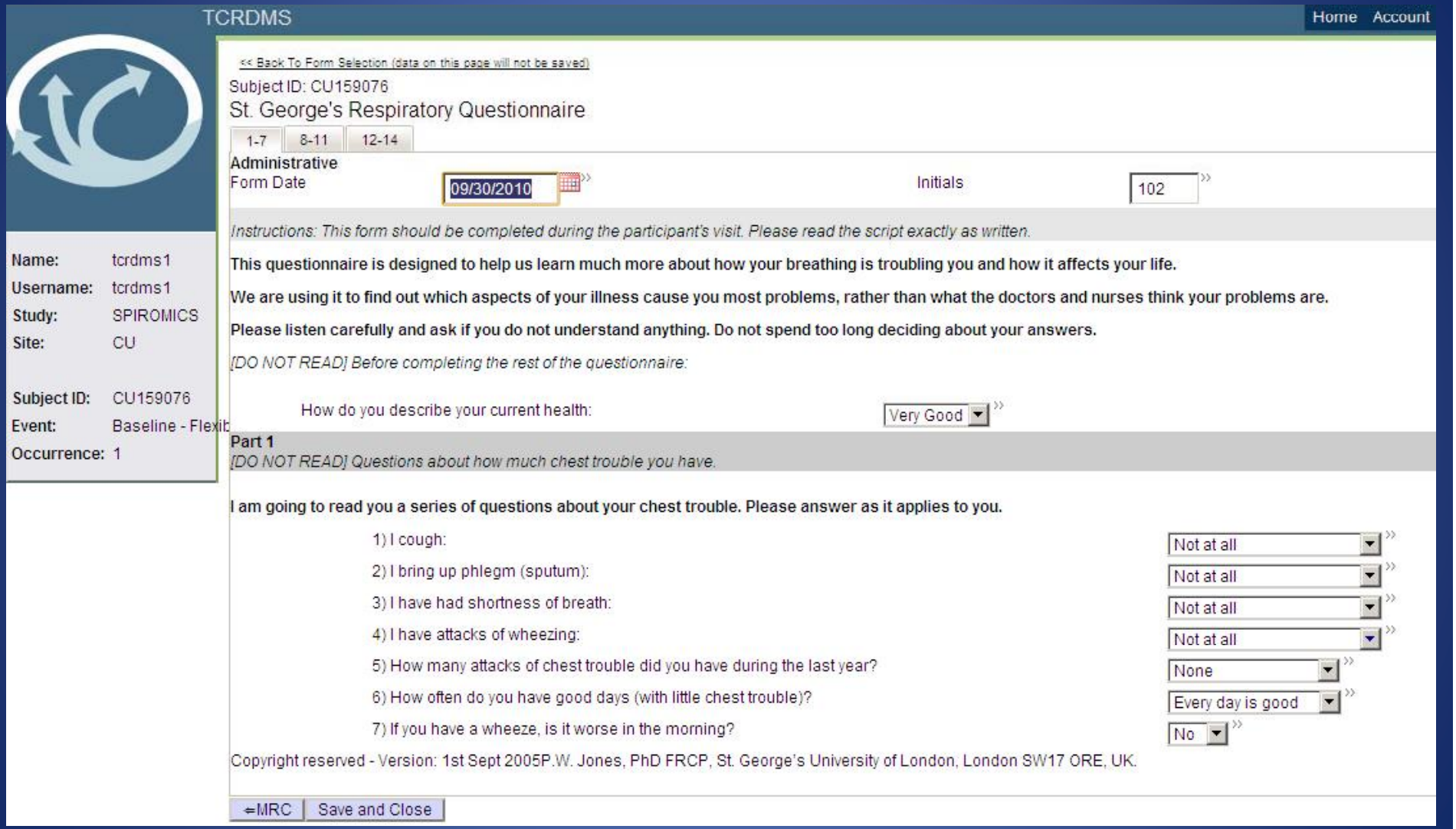


Screen Shots

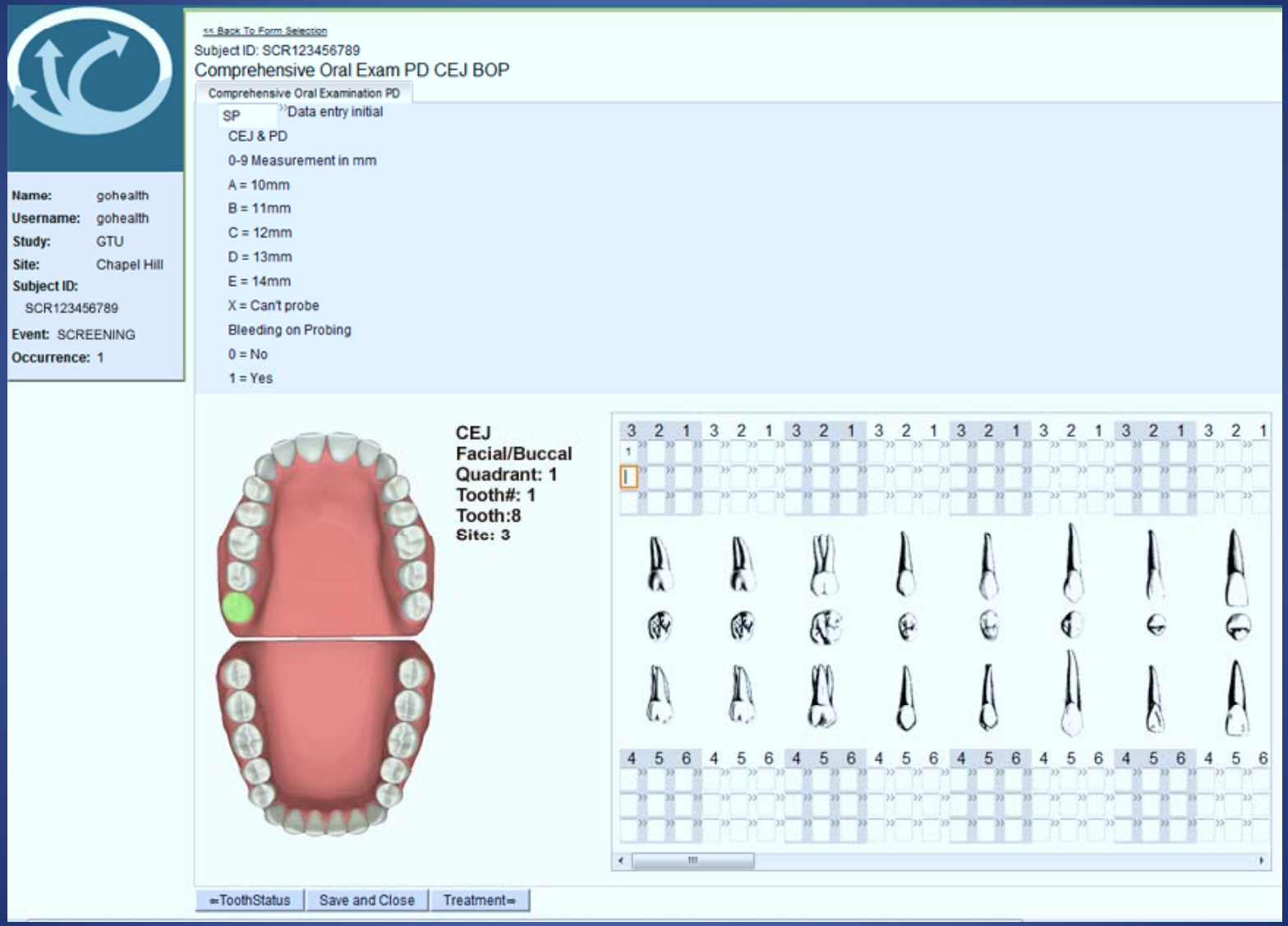
COPD Assessment Form



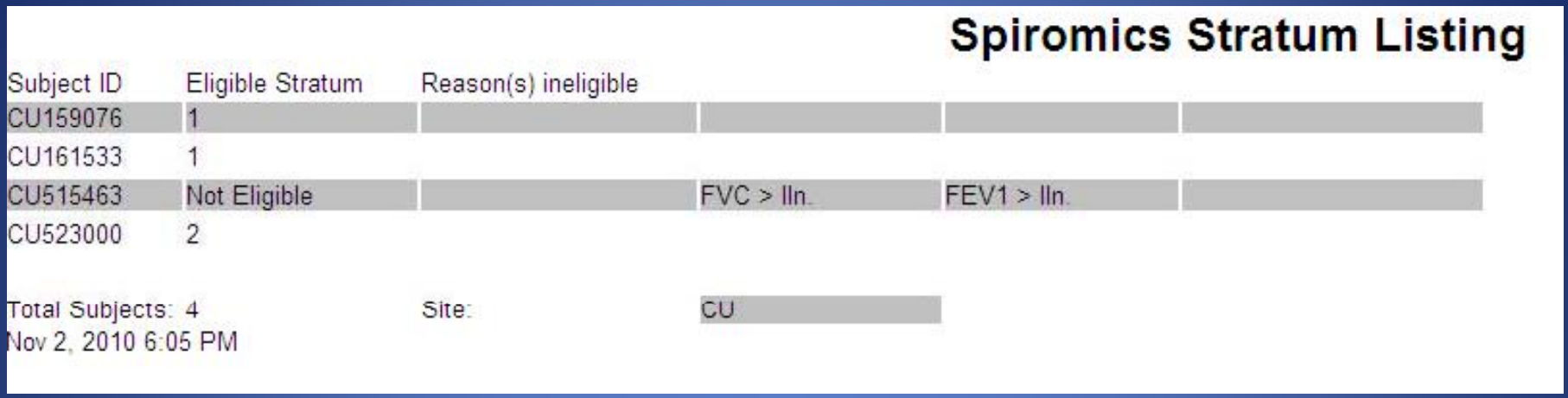
Respiratory Questionnaire



Oral Exam Form



Stratum Report



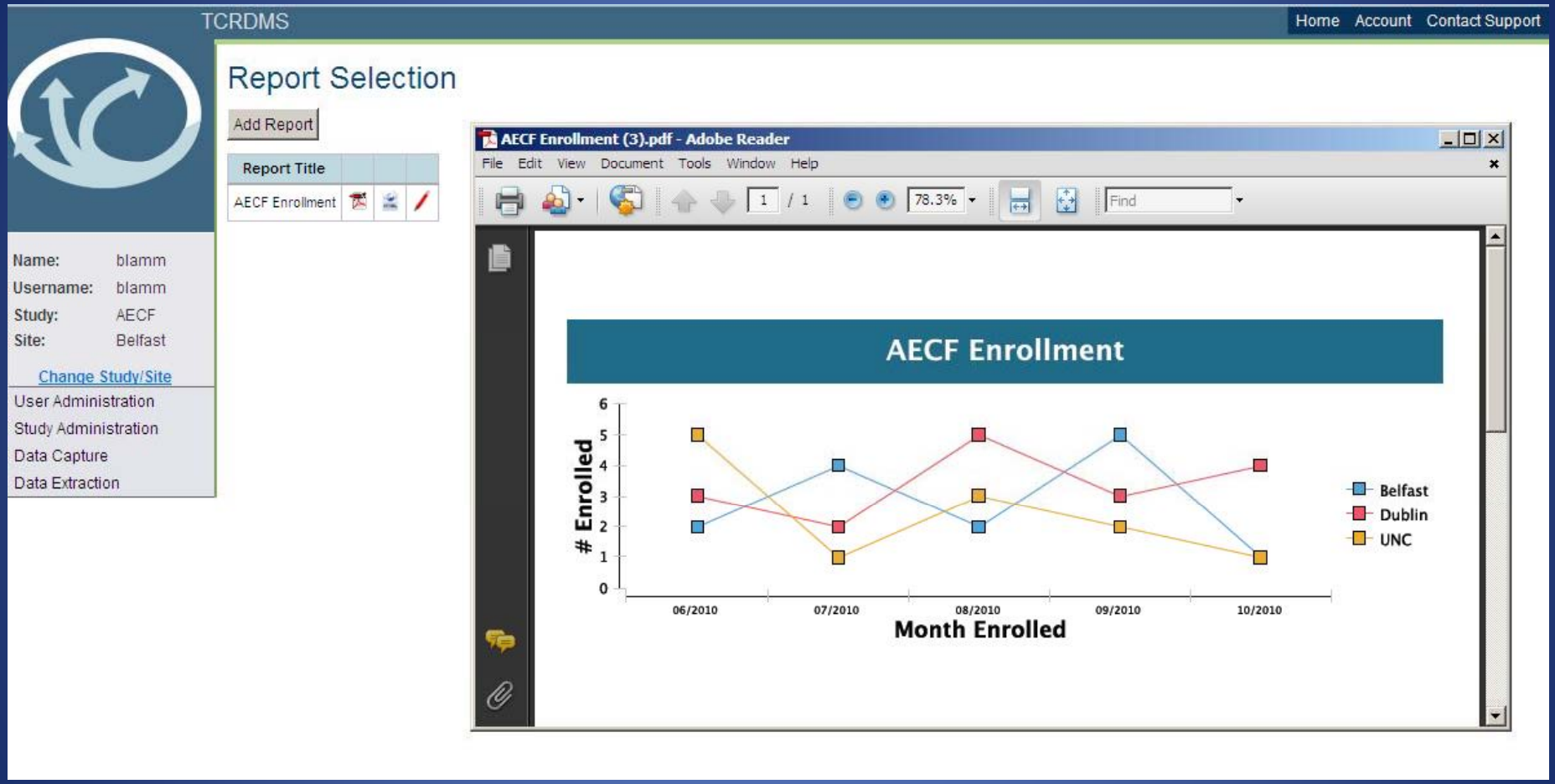
Spiromics Stratum Listing

Subject ID	Eligible Stratum	Reason(s) ineligible			
CU159076	1				
CU161533	1				
CU515463	Not Eligible		FVC > lln.	FEV1 > lln.	
CU523000	2				

Total Subjects: 4 Site: CU

Nov 2, 2010 6:05 PM

Enrollment Report



Progress

- Functionality
  - Screen painter in testing
  - Automatic updates for query resolution in development
- Staffing
  - Delays in hiring project/data manager position in Biostatistics Core
- Usage
  - REDCap used, in use, or planned use in ~40 studies
  - TCR-DMS in use for 4 large, multi-site studies (3 CSCC studies and 1 TraCS study)

Challenges

- Cost recovery
  - Ongoing challenge to convince investigators with GCRC experience (informatics services at no charge) to include data management costs in grant applications
  - Secure funding for system maintenance and enhancements
    - Recharge center
    - Include budget for specific enhancements in grant applications
  - Cost model –*see budget document*
- Data security issues – move towards model where patient level data are stored on TraCS servers with secure remote access from desktops/laptops

Future

- Interoperability
  - Electronic transfer of UNC patient chart data from EMR warehouse to automatically populate data entry screens
    - Pilot tested in breast cancer clinics; in development for statewide cancer registry
    - Also in development for UNC patients as part of Bronchiectasis patient registry
  - Eventual expansion to download patient level data from EMR warehouses at other institutions(?)
  - Could facilitate data sharing among CTSA institutions
- TCR-DMS as proprietary software -
  - Share with other CTSA sites (e.g., REDCap model)?
  - Commercialize?

Background Material

Technical Details

- Web-based user interface, support all major browsers
  - JEE platform: JSF (MVC), Hibernate (ORM), JQuery (+JavaScript), Spring, AspectJ
- Database-independent, support all major relational database mgmt systems
  - No vendor-specific code, Java Persistence Architecture (JPA) implementation of data model
- JEE-compliant, support all major web containers
  - Developed and tested with Apache Tomcat, other commercial-grade containers could be utilized
- Support for LDAP integration for single sign-on (SSO), including multiple LDAP integrations per installation
  - Spring security model - external and/or local authentication support available

Key Technical Challenges

- Addressing the diversity of study specific data requirements without the need for custom software development
  - Commonly used techniques sacrifice strong data typing / validation
  - Decompose forms into atomic typed data elements; recomposed dynamically for UI rendering
- Supporting complex conditional logic (e.g., skipped fields, cross-form dependencies and outcomes)
  - Typically only available in ultra high-end (i.e., expensive) commercial packages or through costly custom development
  - Chaining of logical expressions to implement infinitely complex dependencies without custom programming
- Ensuring robust data integrity, while providing real-time validation with support for situational overrides
  - Traditionally implemented via post-submission validation with no ability to override
  - Client-side validation / override (real-time); full validation redundancy server-side