DATA MANAGEMENT

Why you need it, and where to get it
Goofus: “I want to read the funnies.”

Gallant: “Are you finished with the funnies?”
Highlights for Health Researchers

Now with Data Management too!
“I don’t have time to go to your dumb old conference – I’m cleaning my data.”

“My data was ready for analysis so quickly, I’ve got three posters for the conference!”
“Wait, you can’t have one of those! We had an invalid enrollment and the randomization scheme has not been adjusted.”

“Congratulations! Our online, real-time randomization page says you can have one of these.”
“Will you try to re-interview these subjects? Their answers don’t make sense!”

“Your interview data made a lot of sense. Let’s do a follow-up!”
“It will take me weeks to perform this analysis – these data have to be restructured.”

“This data is set is just like I specified – you’ll have your results in a jiffy.”
OVERVIEW

- Data and databases
- Uses of databases
- Database alternatives
- Data systems
- Data management best practices
RELATIONAL MODEL OF DATA

- Attribute: Value
- Entities with shared attributes = relation
- Set of attributes = relation definition
- Attributes to identify = key
- Instances of definition = tuple
- Set of tuples = relation
“A logical system for generating true statements about the world.”
Attributes have a “type”

- Set of allowable values
- Set of valid operators for other types
- Logic, including default values and relationships to other attributes
Relational algebra/calculus

- Sub-sets of relations
- Product of multiple relations
- Relations of other relations

Relations = abstract sets

- Unordered tuples
- Unordered attributes
RELATIONAL DATABASES

- Relations = Tables
- Attributes = Columns
- Tuples = Rows
- Types = approximated by data types
- Relational calculus/algebra = approximated by Structured Query Language (SQL)
DATABASE FUNCTIONS

- Persistence
- Currency
- Consistency
- Efficiency
- Access/Security
1. Reduce errors

2. Reduce work
OTHER TYPES OF DATABASES

- Attribute-based
- Object-relational
- Hierarchical/XML
DATABASE ALTERNATIVES

- Paper
- Text files
- Spreadsheets
- Statistical software
DATABASE VS. DATA SYSTEM

- Relational Database Management System (RDBMS)
  - Storage/backup
  - Consistency/transaction handling
  - Access/security
  - Manipulation via SQL
DATABASE VS. DATA SYSTEM

Data Systems

- User interface
- Access/security
- Interaction with other systems/people
- Output: reports, tables, data extracts
Types of Data Systems

- Office macro applications
- Desktop databases
- EDC systems
- Clinical trial management systems
- Custom applications
Order Form

Name: Susan Harkins
Date: May 23, 2012
Ext #: 677
Dept: Choose an item
Part: Choose an item
Quantity: Executive, Accounting, Customer Service, IT
Price: text.
OFFICE MACRO APPLICATIONS
OFFICE MACRO APPLICATIONS

**Advantages**
- Flexible
- Widely accessible
- Free

**Disadvantages**
- Flexible!
- One user at a time
- Limited options
- Customization requires skill
DESKTOP DATABASES
DESKTOP DATABASES

**Advantages**
- Rigorous if designed correctly
- Broadly customizable
- Single-file versions easy to maintain
- Interfaces and output in one application
- Wizards, etc., for basic functionality

**Disadvantages**
- Limited multi-user ability
- Security can be complex
- Prone to corruption in larger applications
- Complex functionality requires programming expertise
Front Desk Sign-In

1. Enter your PSC Student ID Number.

2. Enter your first name.

3. Enter your last name.

4. Enter your gender.
   - Male
   - Female

5. Enter your classification.
   - Freshman
   - Sophomore
   - Junior
   - Senior
EDC SYSTEMS
EDC SYSTEMS

**ARDS/Sepsis Registry**

Displayed below is a table listing all existing records/responses and their status for every data collection instrument (and if longitudinal, for every event). You may click any of the colored buttons in the table to open a new tab/window in your browser to view that record on that particular data collection instrument. Please note that if your form-level user privileges are restricted for certain data collection instruments, you will only be able to view those instruments, and if you belong to a Data Access Group, you will only be able to view records that belong to your group.

<table>
<thead>
<tr>
<th>Study ID</th>
<th>Demographic Patient Information (Arm 1: Arm 1)</th>
<th>Medical History Patient Information (Arm 1: Arm 1)</th>
<th>ICU Admission (Arm 1: Arm 1)</th>
<th>ICU Daily Assessments ICU Day 0 (Arm 1: Arm 1)</th>
<th>Mechanical Ventilation ICU Day 0 (Arm 1: Arm 1)</th>
<th>Other Interventions ICU Day 0 (Arm 1: Arm 1)</th>
<th>Laboratory ICU Day 0 (Arm 1: Arm 1)</th>
<th>Specimens - TEST ICU Day 0 (Arm 1: Arm 1)</th>
<th>Medications Test ICU Day 0 (Arm 1: Arm 1)</th>
<th>ICU Assess ICU (Arm 1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0000</td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
</tr>
<tr>
<td>1</td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
</tr>
<tr>
<td>2</td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
</tr>
<tr>
<td>3</td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
</tr>
<tr>
<td>4</td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
</tr>
<tr>
<td>5</td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
</tr>
<tr>
<td>23</td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
</tr>
<tr>
<td>0000000001</td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
<td><img src="image" alt="Status Icon" /></td>
</tr>
</tbody>
</table>
## EDC SYSTEMS

### Advantages
- Basic setup requires little expertise
- Web-based systems on secure servers
- Easy multi-user
- Range of solutions and costs
- Fairly rigorous of designed correctly

### Disadvantages
- Some solutions expensive
- Difficult to customize
- Limited built-in logic
- Limited structure
- Some solutions have limited output
CLINICAL TRIAL MANAGEMENT SYSTEMS
Physical Exam English

There are issue(s) with your submission. The data has NOT been saved. See below for details.

- [Out of expected range. Diastolic expected range of 70-90.]

Visit Information:
- Date: 27-Oct-2015
- Time: 14:23

Physical Exam Information:
- Height: 69 in
- Weight: 175 lb
- Temperature: 98.2 F
- Pulse Rate: 63 per min
- Respiration Rate: 
- Blood Pressure:
  - Systolic: 110 mm
  - Diastolic: 120 mm

Return to top
CLINICAL TRIAL MANAGEMENT SYSTEMS

OpenClinica Home page and menu of tasks for a Study’s Data Manager:

Welcome to Docetaxel in Patients With Con...

Subject Enrollment By Site:

- Cambridge Cancer for Surgical Oncology: 8/20 (40%)
- Center for Cancer Research at Cambridge: 3/20 (15%)
- Somerville Cancer Research Consortium: 12/20 (60%)
- Somerville Medical Center: 3/20 (15%)

Study Progress:

- scheduled: 53 (53%)
- data entry started: 13 (15%)
- completed: 14 (17%)
- signed: 1 (1%)
- locked: 0 (0%)
- skipped: 1 (1%)
- stopped: 2 (2%)

Subject Status Count:

- Study Subject Status: available: 26 (100%)
- Study Subject Status: signed: 0 (0%)
- Study Subject Status: removed: 0 (0%)

© 2004-2011 Akaza Research LLC and collaborators. The Program is provided AS IS, without warranty. Licensed under LGPLv2.1. The program is free software; you can redistribute it and/or modify it under the terms of the GNU Lesser General Public License version 2.1 as published by the Free Software Foundation.
# CLINICAL TRIAL MANAGEMENT SYSTEMS

## Advantages
- Powerful, customized for trials
- Full auditing, review and monitoring
- Complex logic built in

## Disadvantages
- Can be expensive or difficult to install/configure
- Steep learning curve for setup
- Steeper learning curve for customization
### CUSTOM APPLICATIONS

**Statistical Analysis Center**

- **Data Management**
- **Informatics Support Center**

**Breast Cancer Family Registries**

**Columbia University, New York - Breast**

**Uploaded Date**

Click on a node to download the corresponding list of errors. File type: Report (pdf) Spreadsheet (xlsx)

**View by:**
- Batch
- Table

**Suppressed errors:**
- High
- View

**Warnings:**
- High
- View

- **Center 2 (465 errors)**
  - **Batch 11/19/2010 09:23:54 (114 errors)**
    - **BREASTCONFIRMATION** (1079 rows imported, 0 errors)
    - **BREAST_EPI** (4591 rows imported, 2 errors)
    - **BREAST_EFLU** (3197 rows imported, 9 errors)
    - **BREAST_EFLU2** (2694 rows imported, 0 errors)
    - **BREAST_EFLU3** (1145 rows imported, 0 errors)
    - **BREAST_EFLU_6** (585 rows imported, 0 errors)
    - **BREAST_INVASIVE** (841 rows imported, 0 errors)
    - **BREAST_MUTATION** (215 rows imported, 0 errors)
    - **BREAST_MUTATION_DETAIL** (4881 rows imported, 0 errors)
    - **BREAST_NONINVASIVE** (838 rows imported, 0 errors)
    - **BREAST_RX** (882 rows imported, 0 errors)
    - **BREAST_DX_FU** (3964 rows imported, 0 errors)
    - **BREAST_SURGERY** (144 rows imported, 1 error)
    - **CANCER** (12676 rows imported, 2 errors)
      - *CANCER*: Duplicate tumor - another tumor exists with the same features and an identical AGE_DX or a DDX_DATE within 30 days (2 errors)
    - **CANCER TISSUE** (1350 rows imported, 0 errors)
    - **FAMILY** (1342 rows imported, 0 errors)
    - **FAMILY_MEMBERSHIP** (39,995 rows imported, 0 errors)
    - **INDIVIDUAL** (39,995 rows imported, 0 errors)
    - **OVARIAN_COMPARISON** (172 rows imported, 0 errors)
    - **OVARIAN_PATH** (172 rows imported, 0 errors)
    - **PREGNANCY** (7948 rows imported, 0 errors)
    - **SPECIMENS** (6624 rows imported, 0 errors)
    - **Batch 21/12/2016 17:03:52 (451 errors)**
CUSTOM APPLICATIONS
CUSTOM APPLICATIONS

Advantages
- Completely flexible
- Full range of interfaces/outputs
- Can interact with other systems
- Can accommodate complex logic, formatting, etc.
- Can accommodate many users, roles, etc.
- Flexible pricing

Disadvantages
- Can be expensive
- Cannot be modified by users
- Requires planning and development time
- Requires debugging/validation
DATA MANAGEMENT BEST PRACTICES

- Define domain
  - Analysis data vs. logistical data
  - “Bind” users to data system

- Object modelling
  - Entities
  - Relationships
  - Behavior
  - Workflow
  - Interfaces/interaction
DATA MANAGEMENT BEST PRACTICES

- Design from object model up
  - Entities
  - Data tables
  - Interfaces
  - Behavior
  - Output
  - Tools
Proximity to data

- Limit steps
- Proximal in time as well as process
- Early feedback
- Empower users
DATA MANAGEMENT BEST PRACTICES

- **Security**
  - Design for security
  - Role-based access
  - Principle of least privilege
  - Server vs. file

- **Availability**
  - Appropriate access
  - Keep current
  - Back up!!
### Data management plan

- **Data acquisition**
  - What?
  - Who?
  - How?
  - Where?

- **Data access**
  - What?
  - Who?
  - How?

- **Final disposition**
- Needs assessment prior to funding
- Request funding
- Irving Institute DBR consultation
RESOURCES

- Needs assessment prior to funding
- Request funding
- Irving Institute DBR consultation
DATA MANAGEMENT OPTIONS

- DCC: Howard Andrews
- SAC: Seamus Thompson/Richard Buchsbaum
- External consultancy
- Dedicated hire
NEED/BENEFIT VS COST/EFFORT
“But I need more funding to fill in missing data!”

“Good thing I budgeted for data management up front. Here’s my carry-over.”