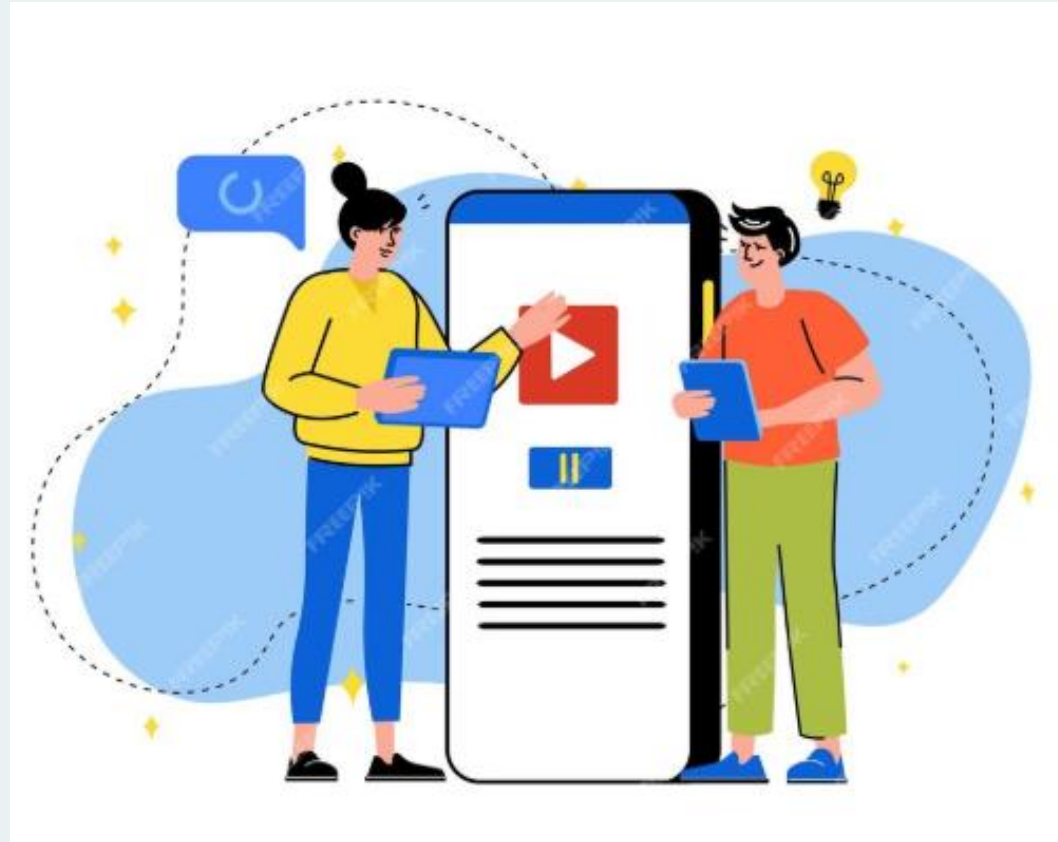


Mastering the Data Maze: Best Practice for Streamlined Research Data Management

Georgian Implementation Science Fogarty Training
(GIFT) Program
Ilia State University & Yale University

August 1, 2025

DATA IS YOUR FRIEND!



What We'll Cover Today

1. What is Data Management (DM)?
2. Goals & Objectives
3. Clinical-Trial Life-Cycle and Data Flow
4. Data Management's Role within a Trial
5. Case Report Forms (CRF/eCRF)
6. The Six Core Stages of the Data Management Process
7. Preparing Data for Analysis
8. Lock and Archival
9. Q&A

Defining Our Focus

- **Definition:** Data Management is a set of processes that plan, collect, clean, store, protect & deliver research study data in a form that is complete, accurate, and ready for statistical analysis, regulatory submission, and secondary use.
- **Core Function:** To ensure that data is accurate, complete, and reliable.
- **It's a Quality System:** It bridges the gap between research/clinical activities and statistical analysis, transforming raw observations into a trustworthy dataset.



Key Objectives:

- To collect all data specified in the study protocol.
- To support data monitoring activities.
- To minimize errors and missing data.
- To create a "locked" database that is clean and ready for analysis in a timely manner.

Goals of DM

Primary Goal: To guarantee the **integrity and quality** of the data.

This directly impacts:

Patient Safety:

Quickly identify adverse events and safety signals.



Study Validity:

Ensure conclusions drawn from the study are correct and defensible.

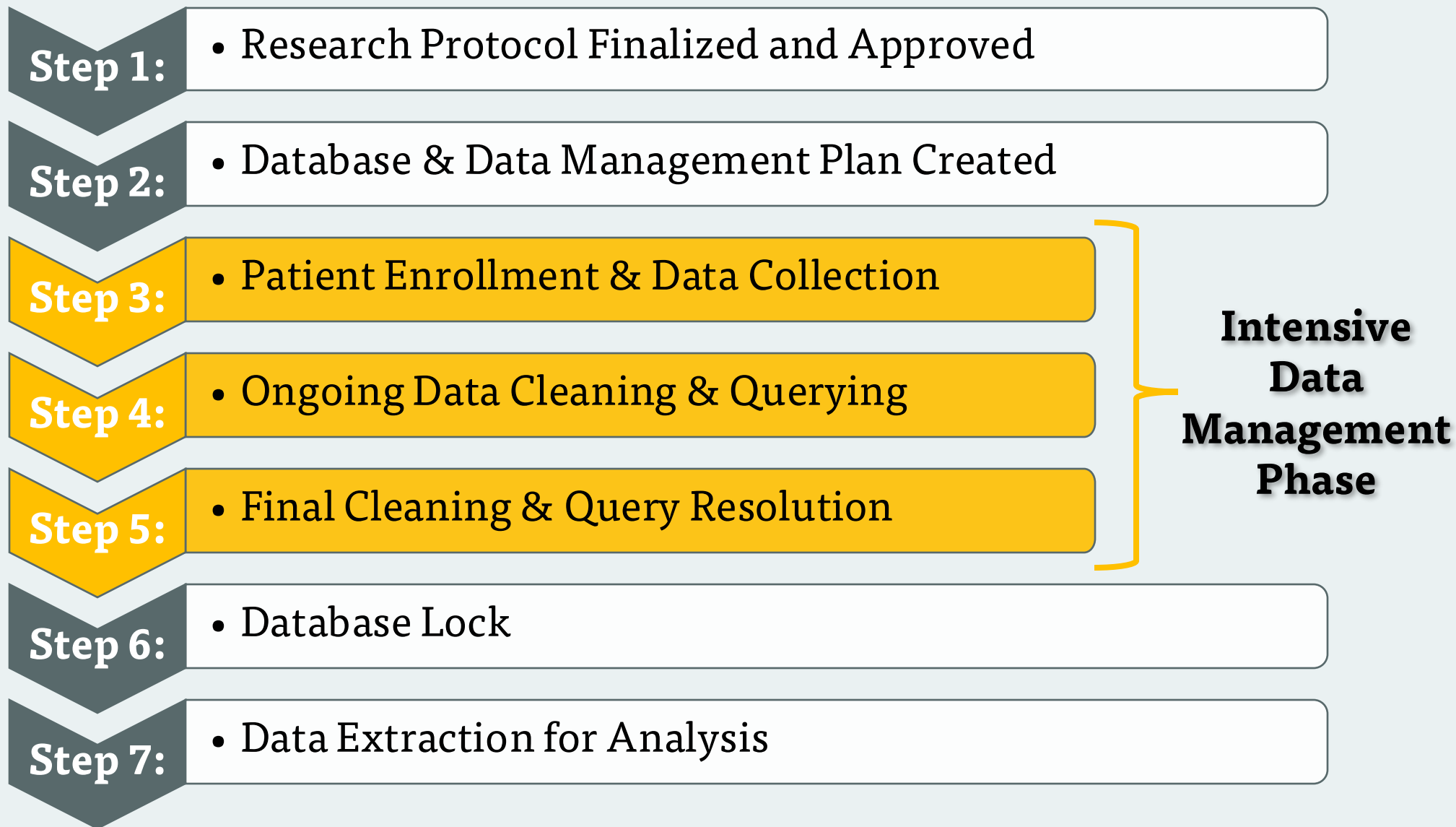


Regulatory Compliance:

Meet the rigorous standards of bodies like the FDA, GDPR, HIPAA.



The Data Flow Within a Trial



The Cornerstone: Case Report Form (CRF) Design

- **What is a Case Report Form?** A tool (paper or electronic) used to collect protocol-required data for each subject.
- **eCRF vs. Paper:** Most modern trials use electronic CRFs within an Electronic Data Capture (EDC) system.
- **Principles of Good CRF Design:**
 - **Clarity:** Unambiguous questions. Avoid jargon.
 - **Conciseness:** Collect only the data needed to answer the research questions (as defined in the protocol). "Nice to have" data adds noise and burden.
 - **Logical Flow:** Questions should follow the flow of a patient visit.
 - **Standardization:** Use standard instruments, when possible, to improve consistency and streamline analysis.

Electronic vs. Paper Case Report Form

- **eCRF Advantages**

- Real-time edit checks & query generation
- Remote monitoring
- Built-in compliant audit trails
- Save time on data entry (from paper into the database)

- **Paper CRF advantages**

- No need to rely on internet connection
- No need for multiple computers/tablets for data collection
- Training & change management are easier
- Ability to go back to paper records when querying data

If You Don't Ask for It Correctly, You Won't Get It

Medications Adherence Case Report Form

POOR DESIGN

Medication Status:

Problem:

Ambiguous. Does this mean "taking it," "stopped," "never took"? Leads to inconsistent data.

GOOD DESIGN

Was the study medication taken today?

Yes No Not Applicable

Benefit:

Clear, unambiguous, and produces clean, analyzable data



“I see you’re still fishing for that key piece of data.”

CRF Design: Purpose & Principles

- Translates protocol endpoints into collectable fields
- Good design reduces missing data & queries
- Protocol-driven
- Standardization & reuse
- Logical flow mirrors patient visit schedule
- Minimize free text; use controlled vocabularies (checkboxes, dropdowns)
- Built-in edit checks & skip logic
- User-friendly for site personnel

How many CRFs do you need?

Typical CRFs for a clinical trial or research study:

- Screening and/or Eligibility
- Demographics
- Randomization
- Medical History
- Physical Exam
- Vital Signs
- Concomitant medications/
therapies
- Adverse Event Form
- Lab forms
- Follow Up Visit Tracking
- Protocol Deviations
- Study Completion/ Early
Termination

Data Management Software

Data collection

Storage

Organization

Quality control

Security

Access control

- REDCap
- Qualtrics
- Medidata Rave
- Oracle Clinical

Data Analysis Software

Statistical analysis

Data mining

Data visualization

Reporting

Predictive modeling

- R
- SAS
- SPSS
- STATA

Examples of DM software/databases



- Qualtrics
- REDCap
- OpenClinica
- SurveyMonkey

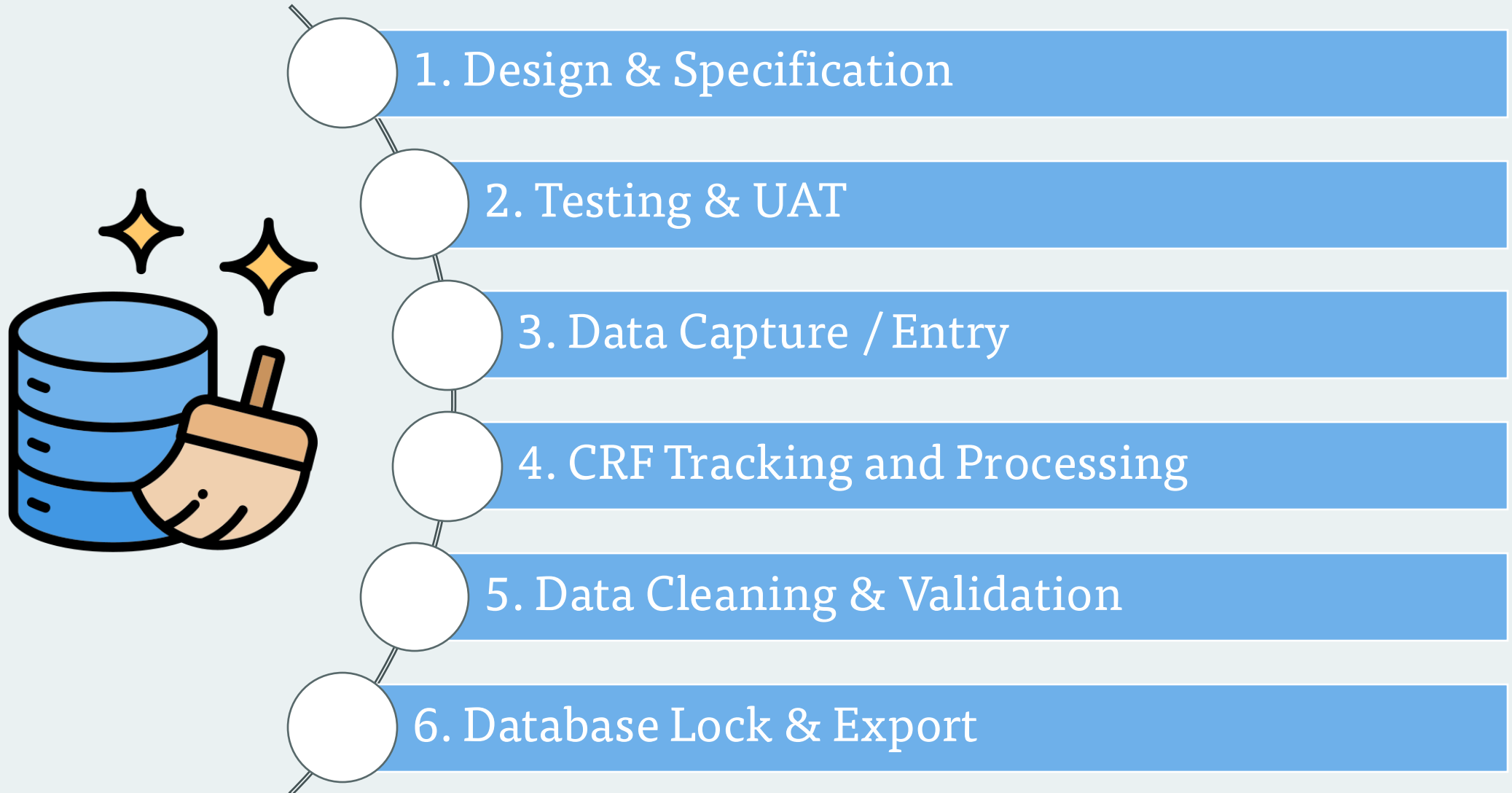


Excel is NOT a DM Tool

- Lacks the audit trail functionality
- No automated checks, logic rules, and real-time validation
- Limited security controls
- Accidental Deletion/Manipulation
- Limited Access Control
- Inefficient for Large Datasets
- Lack of Automation for querying data

Stages of Data Management

A Step-by-Step Journey to a Clean Database



Stage 1: Design & Specification

- **Database Design:** Database structure is determined by the study design and setup. The eCRFs are built in the system based on the protocol. The database is built in a TEST environment.
- **Validation Plan:** "Edit Checks" are programmed to automatically flag errors (e.g., impossible dates, out-of-range lab values).
- **Data Management Plan (DMP):** The master "how-to" guide for all data activities is written and approved.

Stage 2: Testing & User-Acceptance Testing (UAT)

“Measure Twice, Cut Once”

- **UAT:** Before the database goes “live” (PRODUCTION environment), a TEST version is rigorously evaluated by the data management team and then by the study team.
- **Purpose:**
 - To find bugs in the database build.
 - To ensure all CRFs are present and correct.
 - To test that the validation checks (edit checks) are firing correctly.
- **Process:** The team follows a test script, entering dummy data (both correct and incorrect) to see if the system behaves as expected.
- **Outcome:** A fully tested and validated database, ready for study launch.

Stage 3: Data Capture / Entry

- **Primary Method:** Direct data entry into an Electronic Data Capture (EDC) system by the site coordinator and/or by participant completing questionnaires on various devices.
- **Source:** Data is transcribed from original source documents (medical charts, lab reports, participant interviews & questionnaires).
 - Other examples of data source types: wearable devices, electronic health records
- **The Golden Rule:** Accuracy is paramount. The goal is a perfect transcription from source to the database.

Stage 4: CRF Tracking & Processing

Knowing What We Have (and Don't Have)

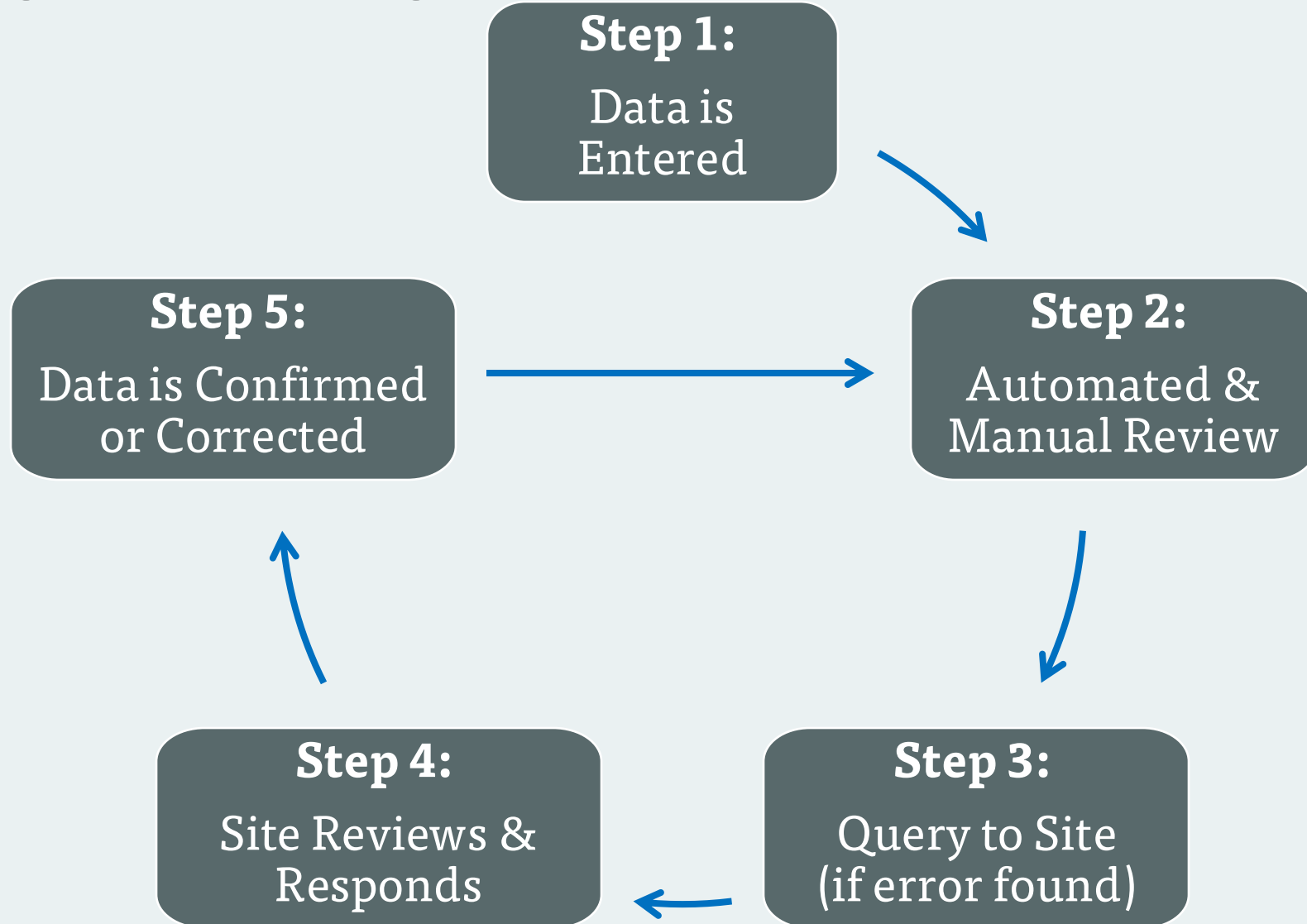
- **Goal:** Continuously track the status of all expected CRFs.
- **Key Questions Answered:**
 - Which patient visits have occurred?
 - Have the CRFs for those visits been entered into the database?
 - Are there overdue or missing pages?
- This ensures completeness and allows the team to proactively follow up with sites about missing data before it's too late.



**Perfect data is
an imaginary friend**

Stage 5: Data Cleaning & Validation

Finding and Fixing Errors



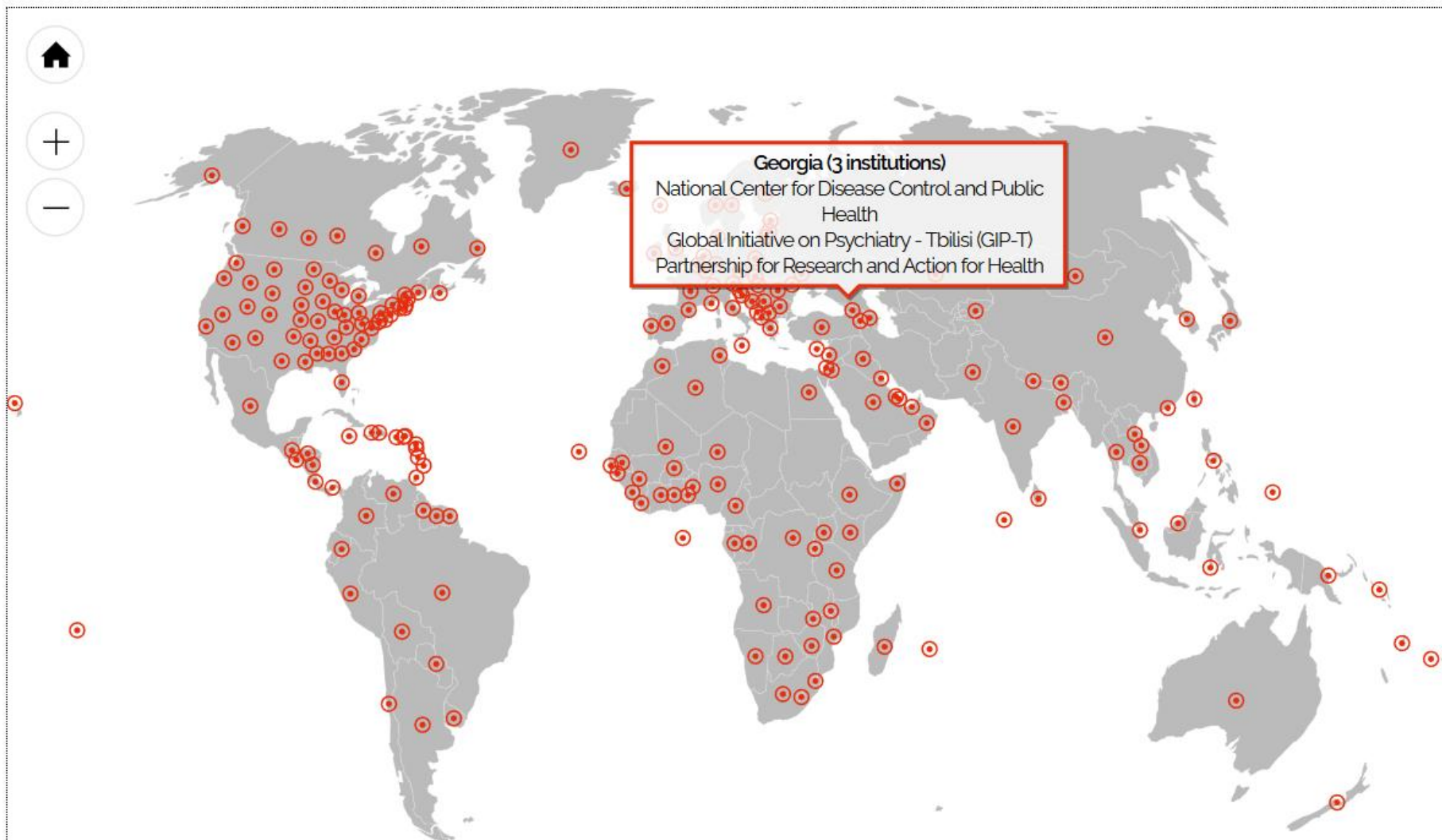
HE EXPECTS SO MUCH FROM ME BUT HIS DATA IS SO FLAWED



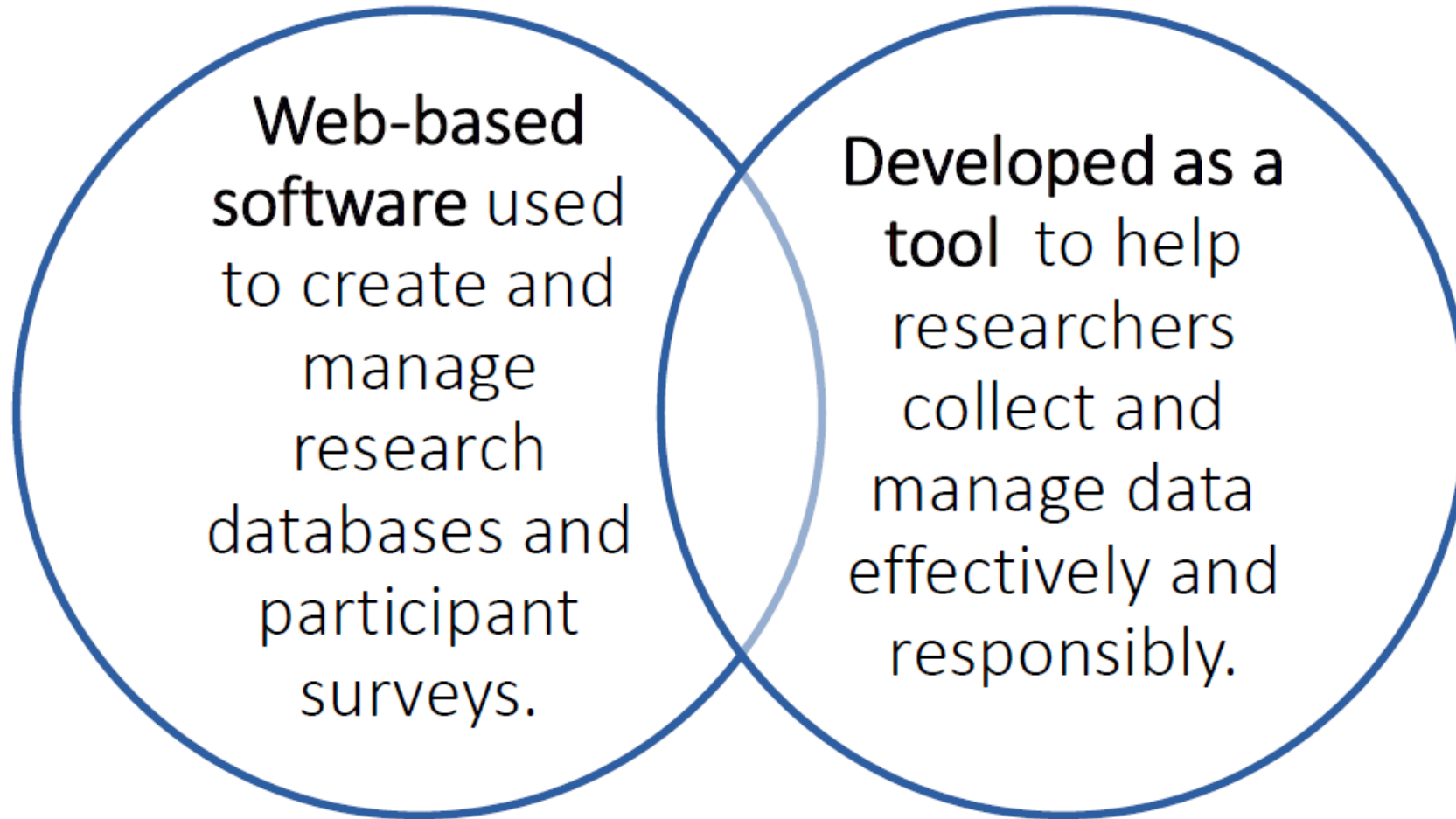
Stage 6: Database Lock & Export

- **Purpose:** To ensure no manipulation of study data during the final analysis.
- **Soft lock (or Database Freeze):** occurs after all data have been entered.
- **Final Database Lock:** occurs after all additional post-lock data cleaning, data validation and quality control activities have been finalized.
- **Export data** for statistical analysis





What is REDCap?



REDCap Features

- **Build online surveys and databases quickly and securely in your browser** - Create and design your project using a secure login from any device. No extra software required. Access from anywhere, at any time.
- **Fast and flexible** - Go from project creation to starting data collection in less than one day. Customizations and changes are possible any time, even after data collection has begun.
- **Advanced instrument design features** - Auto-validation, calculated fields, file uploading, branching/skip logic, and survey stop actions.
- **Data quality** - Use field validation, branching/skip logic, and Missing Data Codes to improve and protect data quality during data entry. Open data queries to automatically identify and resolve discrepancies and other issues real-time.

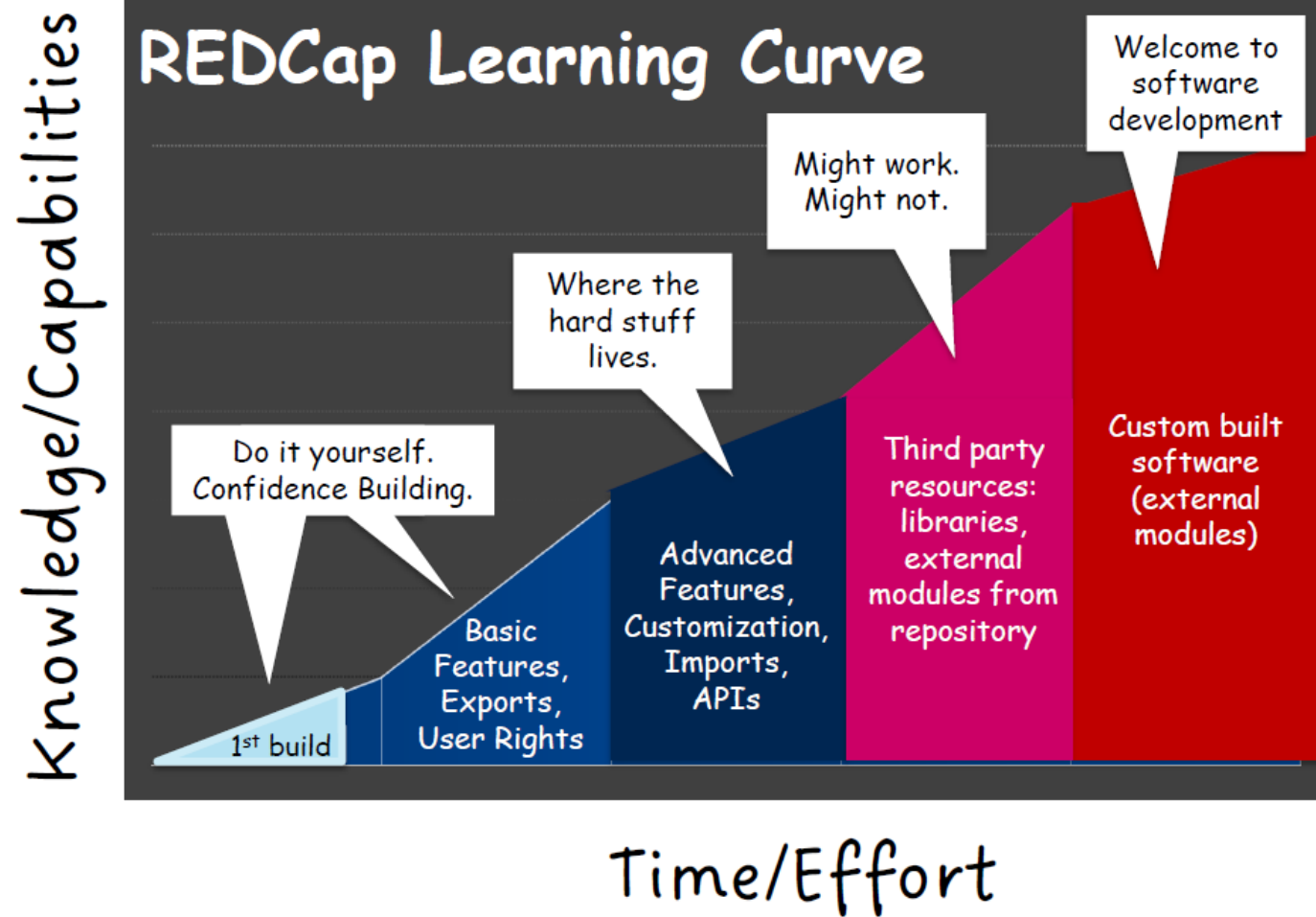
More REDCap Features

- **Diverse and flexible survey distribution options** - Use a list of email addresses or phone numbers for your survey respondents and automatically contact them with personalized messages, and track who has responded. Or create a simple link for an anonymous survey for mass email mailings, to post on a website, or print on a flyer.
- **Custom reporting** - Create custom searches for generating reports to view aggregate data. Identify trends with built-in basic statistics and charts.
- **Export data to common analysis packages** - Export your data as a PDF or as CSV data for easy analysis in SAS, Stata, R, SPSS, or Microsoft Excel.

Even more REDCap Features!

- **Secure file storage and sharing** - Upload and share any type of file with anyone in the world through the File Repository feature or Send-It tool. Also works with exports and other built-in file uploading features.
- **Data-based triggers and alerts** - Send real-time alerts and notifications to your team or other stakeholders via email, text, or phone based on certain data being entered or specific questions having a particular answer.
- **Connect to other resources** - Use built-in features (API) to move data to/from your project. Build your own custom software development features to connect your project to other systems.

Advantages: User Friendly



Data Collection

- Record Status Dashboard
- Add / Edit Records

Applications

- Calendar
- Data Exports, Reports, and Stats
- Data Import Tool
- Data Comparison Tool
- Logging
- Field Comment Log
- File Repository
- User Rights and DAGs
- Record Locking Customization
- E-signature and Locking Mgmt
- Data Quality
- API and API Playground
- REDCap Mobile App

Reports

- 1) Enrolled
- 2) Discontinued subjects
- 3) Climb Flight of Stairs
- 4) Intervention or Control
- 5) Step Counts
- 6) Outpatient Visits

Help & Information

- Help & FAQ
- Video Tutorials
- Suggest a New Feature

Contact REDCap administrator

Empower

How does REDCap work?
Direct data entry (i.e., form), survey, or data import
Exports to Excel, CSV, SPSS, SAS, Stata and R, and CDISC ODM.
Anyone can learn to use it

Record ID	Screening	Baseline	Gait Speed 3 Meters	Physical Activity	ADL	Fall Efficacy	Functional Assessment	SF12	Mini COG	PHQ	Most Sedentary Behavior	UAB Life Space	Kinesiophobia	SPPB	End Of Interview	Intervi Observ
9																
12																
30																
37																
46																
54																
63																
70																
86																
87																
91																
114																
122																
123																
134																
140																
151																
155																

UAB Life Space

Editing existing Record ID 12

Event Name: **Baseline**

Record ID: 12

UAB Study of Aging Life - Space Assessment

1a. Past 4 wks been to other rooms besides room where you sleep? Yes No reset

1b. How often get to other rooms? reset

1c. Did you use aids or equipment? Did you need help from another person? (other rooms) reset

2a. Past 4 wks been to an area outside your home such as your porch, deck or patio, hallway, etc? Yes No reset

2b. How often get to areas outside your home? reset

2c. Did you use aids or equipment? Did you need help from another person? (outside areas) reset

REDCap: Project Setup

The screenshot displays the REDCap Project Setup interface for a project named "Test project Neuro QOL" (PID: 2362). The interface is divided into a left-hand navigation menu and a main content area. The main content area is highlighted with a red border and contains several sections for configuring the project.

Navigation Menu (Left):

- Logged in as bt24 | Log out
- My Projects | Control Center
- REDCap Messenger
- Contact REDCap administrator
- View project as user: -- select a user --
- Enter PID to go to project
- Project Home and Design
 - Project Home
 - Project Setup
 - Designer
 - Dictionary
 - Codebook
 - Project status: Development
- Data Collection
 - Survey Distribution Tools
 - Record Status Dashboard
 - Add / Edit Records
 - Show data collection instruments
- Applications
 - Project Dashboards
 - Alerts & Notifications
 - Multi-Language Management
 - Calendar
 - Data Exports, Reports, and Stats
 - Data Import Tool
 - Data Comparison Tool
 - Logging and Email Logging
 - Field Comment Log
 - File Repository
 - User Rights and DAGs
 - Customize & Manage Locking/E-signatures
 - Data Quality
 - API and API Playground
 - REDCap Mobile App
- Project Dashboards (Add/Edit)
- External Modules (Manage | View Logs)
- Help & Information
 - Help & FAQ
 - Video Tutorials
 - Suggest a New Feature
 - Contact REDCap administrator

Main Content Area (Right):

Test project Neuro QOL PID: 2362

Project status: Development Completed steps 0 of 7

Main project settings (Not started)

- Disable Use surveys in this project? [?]
- Enable Use longitudinal data collection with defined events? [?]
- Enable Use the MyCap participant-facing mobile app? [?]
- Modify project title, purpose, etc.

Design your data collection instruments & enable your surveys (Not started)

Add or edit fields on your data collection instruments (survey and forms). This may be done by either using the Online Designer (online method) or by uploading a Data Dictionary (offline method). You may then enable your instruments to be used as surveys in the Online Designer. Quick links: [Download PDF of all Instruments](#) OR [Download the current Data Dictionary](#)

Go to [Online Designer](#) or [Data Dictionary](#) Explore the [REDCap Instrument Library](#)

Have you checked the [Check For Identifiers](#) page to ensure all identifier fields have been tagged?

Learn how to use: [Smart Variables](#) [Piping](#) [Action Tags](#) [Field Embedding](#) [Special Functions](#)

Enable optional modules and customizations (Optional)

- Enable Repeating instruments [?]
- Disable Auto-numbering for records [?]
- Enable Scheduling module (longitudinal only) [?]
- Enable Randomization module [?]
- Enable Designate an email field for communications (including survey invitations and alerts) [?]
- Additional customizations

Settings displayed to Administrators only:

- Enable Twilio SMS and Voice Call services for surveys and alerts [?]
- Enable Mosio SMS services for surveys and alerts [?]
- Enable SendGrid Template email services for Alerts & Notifications [?]

Set up project bookmarks (optional) (Optional)

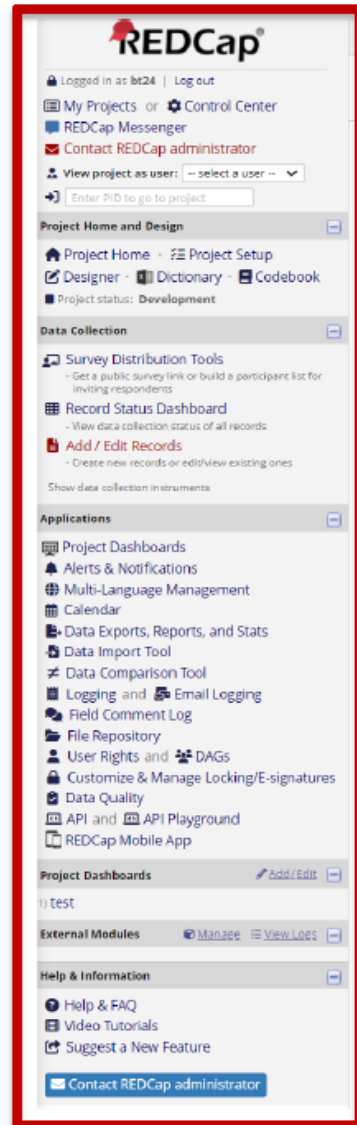
You may create custom bookmarks to webpages that exist inside or outside of REDCap. These bookmarks will be seen as links on the left-hand project menu and can be accessed at any time by users who are given privileges to do so. Every project bookmark has custom settings that allow one to control its appearance and behavior.

Go to [Add or edit bookmarks](#)

User Rights and Permissions (Optional)

You may grant other users access to this project or edit the user privileges of current users on this project by navigating to the User Rights page. Additionally, if you wish to limit user access to certain records/responses for


REDCap: Project Navigation Bar



Access to most parts of REDCap:

- Project design pages
- Data Collection
- Applications
- Reports and Dashboards
- Help and Training

Your Navigation Bar will vary based on your user rights

 Upload or download arms/events ▾















Arm 1: U19

Arm 2: IMAGIN

+Add New Arm



Arm name: U19

[Rename Arm 1](#) | [Delete Arm 1](#)


	Event #	Days Offset	Offset Range Min / Max	Event Name	Custom Event Label  (optional)	Unique event name  (auto-generated)
 	1	0	-0/+0	Flu Clinic		flu_clinic_arm_1
 	2	2	-0/+2	Day 2-4 Blood Draw		day_24_blood_draw_arm_1
 	3	7	-0/+0	Day 7 Blood Draw		day_7_blood_draw_arm_1
 	4	28	-0/+0	Day 28 Blood Draw		day_28_blood_draw_arm_1
 	5	70	-0/+0	Day 70 Blood Draw		day_70_blood_draw_arm_1
 	6	100	-0/+0	Non-Flu Blood Draw		nonflu_blood_draw_arm_1
<input type="button" value="Add new event"/>	<input type="text" value=""/> Days	- <input type="text" value="0"/> + <input type="text" value="0"/>		<input type="text"/>	<input type="text"/>	
	Convert from other units			Descriptive name for this event	Custom Event Label (optional) Example: [visit_date], [weight] kg	

Building Forms: Field Embedding

Without field embedding:

Test Date 1	<input type="text"/>  Today M-D-Y	↓ Field Embedding
Test Result 1A	<input type="text"/>	
Test Result 1B	<input type="text"/>	
Test Date 2	<input type="text"/>  Today M-D-Y	
Test Result 2A	<input type="text"/>	
Test Result 2B	<input type="text"/>	

With field embedding:

Date	Test Result A	Test Result B
<input type="text"/>  Today M-D-Y	<input type="text"/>	<input type="text"/>
<input type="text"/>  Today M-D-Y	<input type="text"/>	<input type="text"/>

Surveys

Not enabled as a survey:

The screenshot displays the REDCap interface for a survey titled "Neuro-QOL SF v1.0 - Fatigue". The left sidebar contains navigation options such as "My Projects or Control Center", "REDCap Messenger", "Contact REDCap administrator", and "View project as user". The main content area shows the survey form for Record ID 3. The form includes three questions, each with five radio button options: "none of the time", "a little bit of time", "some of the time", "most of the time", and "all of the time". The first question is "In the past 7 days I felt tired", the second is "In the past 7 days I had trouble starting things because I was too tired", and the third is "In the past 7 days I had trouble finishing things because I was too tired". Each question has a "reset" link to its right. The top right of the form area shows "Invitation status: [envelope icon]" and a "Survey options" dropdown menu.

My Projects or Control Center
REDCap Messenger
Contact REDCap administrator
View project as user: -- select a user --
Enter PID to go to project

Project Home and Design
Project Home · Project Setup
Designer · Dictionary · Codebook
Project status: Development

Data Collection
Survey Distribution Tools
- Get a public survey link or build a participant list for inviting respondents
Record Status Dashboard
- View data collection status of all records
Add / Edit Records
- Create new records or edit/view existing ones

Record ID 3 [Select other record](#)

Data Collection Instruments:
Form 1
Consent
Neuro-QOL SF v1.0 - Fatigue
Tracking Form

Neuro-QOL SF v1.0 - Fatigue Invitation status: [envelope icon] Survey options

Editing existing Record ID 3.

Record ID	3
In the past 7 days I felt tired	<input type="radio"/> none of the time <input type="radio"/> a little bit of time <input type="radio"/> some of the time <input type="radio"/> most of the time <input type="radio"/> all of the time reset
In the past 7 days I had trouble starting things because I was too tired	<input type="radio"/> none of the time <input type="radio"/> a little bit of time <input type="radio"/> some of the time <input type="radio"/> most of the time <input type="radio"/> all of the time reset
In the past 7 days I had trouble finishing things because I was too tired	<input type="radio"/> none of the time <input type="radio"/> a little bit of time <input type="radio"/> some of the time <input type="radio"/> most of the time <input type="radio"/> all of the time reset

Enabled as a survey

Neuro-QOL SF v1.0 - Fatigue

AAA

 [English](#)  

Please complete the survey below.

Thank you!

**In the past 7 days
I felt tired**

- none of the time
- a little bit of time
- some of the time
- most of the time
- all of the time

[reset](#)

**In the past 7 days
I had trouble starting things because I was too tired**

- none of the time
- a little bit of time
- some of the time
- most of the time
- all of the time

[reset](#)

**In the past 7 days
I had trouble finishing things because I was too tired**

- none of the time
- a little bit of time
- some of the time
- most of the time
- all of the time

[reset](#)

Why Data Management Matters

- Data Management is a foundational pillar of successful clinical research.
- It is a **proactive system**, not a reactive cleanup job.
- Its mission is to ensure the final dataset is a **true and accurate representation** of the trial.
- Technology is evolving, but principles of accuracy, integrity, and traceability remain constant.



"Without data, all we have is an opinion."

W. Edwards Deming

Q&A

Thank
you

Lily Katsovich

Lily.Katsovich@yale.edu