# Data Management Practices in Health Research

### Part I

- Introduction
- Designing the Data Storage Systems

Brandon Guthrie, PhD
Departments of
Global Health & Epidemiology
University of Washington

### Relevance

What is data management?

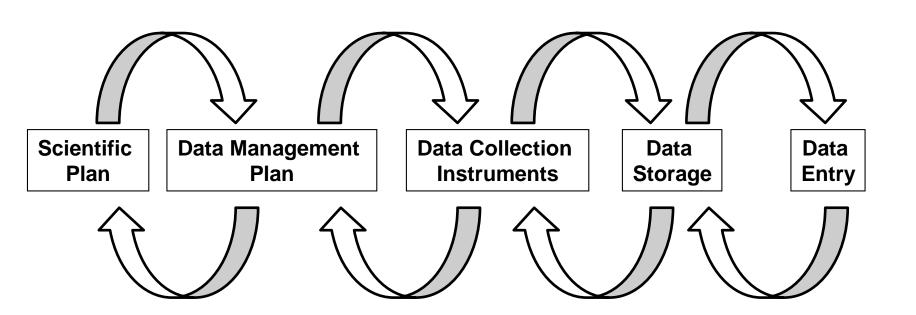
Why is good data management important?

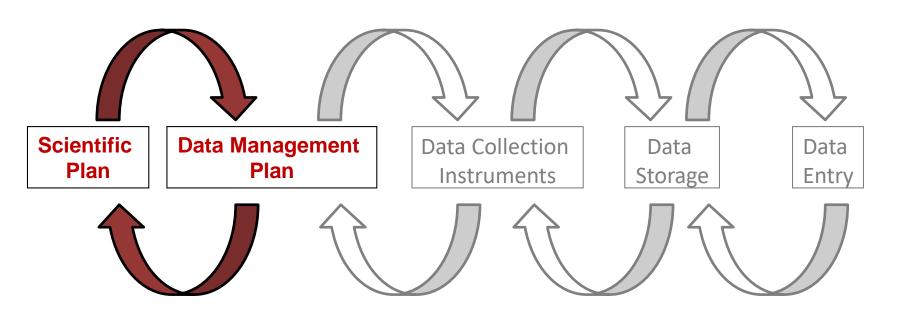
Who is responsible for data management?

How will progress be monitored

# **Ethical Obligations**

- Maintain confidentiality
  - Data is only shared with authorized personnel
  - De-link data from identifiable information to the greatest extent possible
  - Collect the minimal amount of identifiable information
- Ensure data integrity
- Unlink (anonymize) data after set period of time
- All uses of data must be approved by IRB/ERC
- When in doubt, ask your IRB/ERC





# Planning

- What are the aims of the study?
  - Data management must address all research aims
  - Anticipate future questions
- Descriptive, exploratory, or confirmatory
  - Scope of data collected
- How will the data be used?
  - Single analysis, regulatory approval, policy recommendations, etc.
- Who will interact with the data?

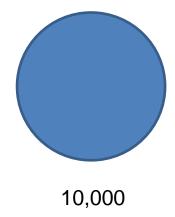
# Planning

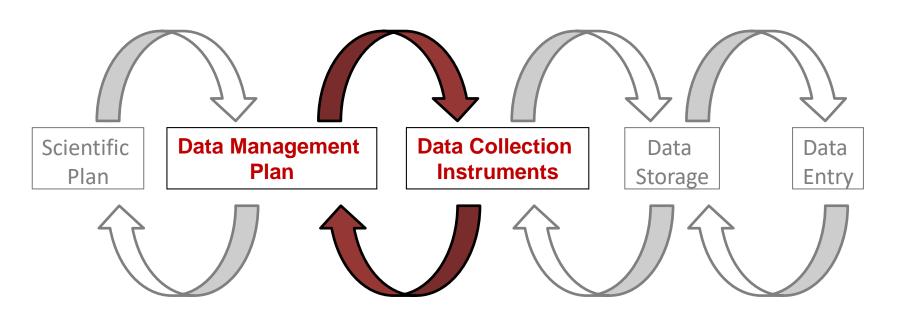
- Every study is different, but there are common themes:
  - Study design
    - Observational or experimental?
    - Cross-sectional?
    - Prospective or retrospective?
    - Longitudinal?
  - Sample size



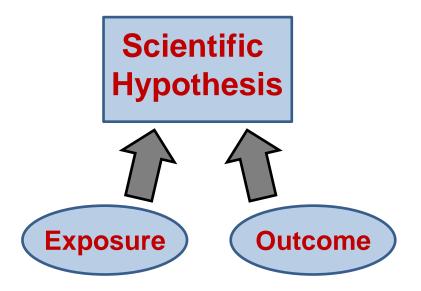


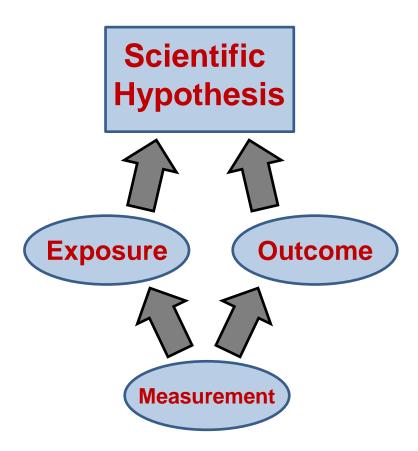


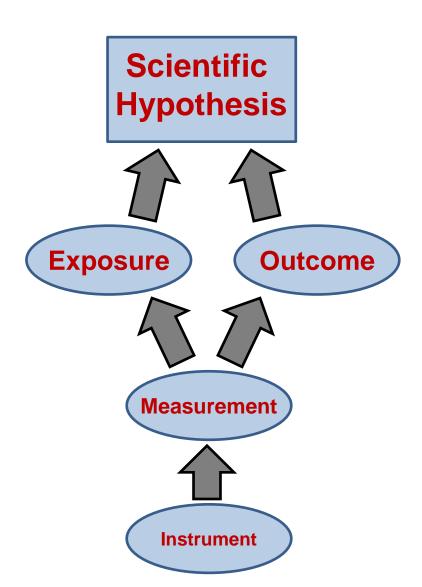




Scientific Hypothesis







### Example

A study is being designed to determine if HIV-specific immune responses in exposed, uninfected people provides some resistance to HIV infection.

Study population: HIV-discordant couples

Design: Observational study with longitudinal follow-up

Sample size: 600 couples

**Primary endpoint: HIV infection** 

### Considerations

#### What types of data will be collected

- Sociodemographic and behavioral
- Immunologic responses

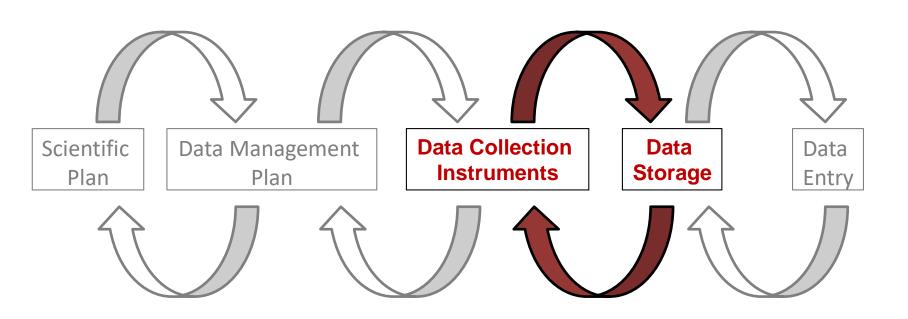
#### What instruments might be used

- Questionnaires
- Lab assays

#### What are some potential issues in managing these data?

- Data from different sources
- Longitudinal follow-up
- Multiple people conducting data entry

# Design the Data Storage Systems



### Motivation

 Put data into a format in which it can be manipulated, analyzed, and disseminated

- Requirements:
  - Efficient
  - Robust
  - Transparent and 'auditable'

## Data storage

Why do I need to know this?

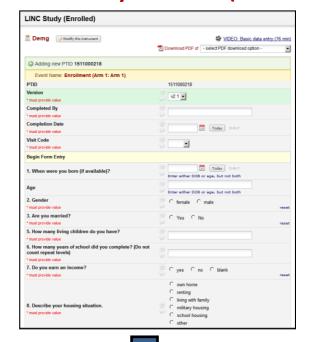
### **Data Flow**

#### Questionnaire

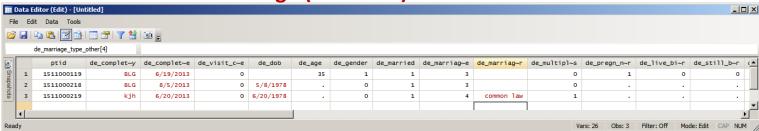




#### **Data Entry Interface (front end)**



**Database Storage (back end)** 



### Data storage: options

- Excel
- Epi Info
- Access
- REDCap
- Open Data Kit (ODK)
- Other applications
  - Open source systems (Open Clinica)
  - Custom built systems

## REDcap versus ODK

#### **REDcap Mobile App**

#### Pros

- Designed for epi study designs
- Strong project management
- Easy setup
- Integrates with standard entry
- HIPPA compliant

#### Cons

- Designed for epi study designs
- Limited error checking logic
- Poor user interface
- Limited control of user rights
- Some data synchronization problems

#### **ODK**

#### Pros

- High degree of flexibility
- Flexible error checking logic
- Excellent user interface
- Accommodates broad range of data types

#### Cons

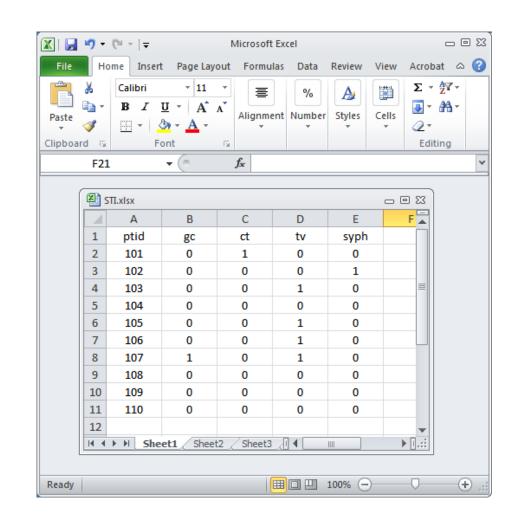
- High degree of flexibility
- HIPPA compliance not guaranteed
- Setup/administration more complicated
- More challenging data management
- Longitudinal records not possible

# Choice of Data Storage Systems

 Simple options may work for basic data storage needs

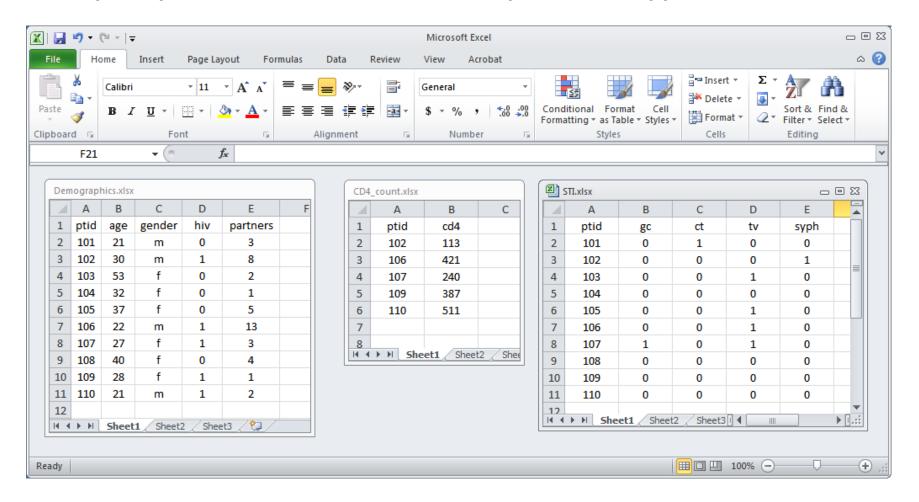
Easy to setup

 Compatible and transferable



# Choice of Data Storage Systems

As complexity of the data increases, more sophisticated approaches are needed

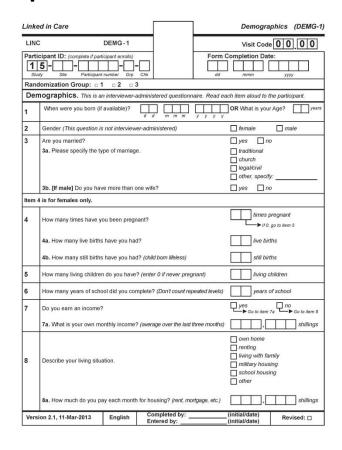


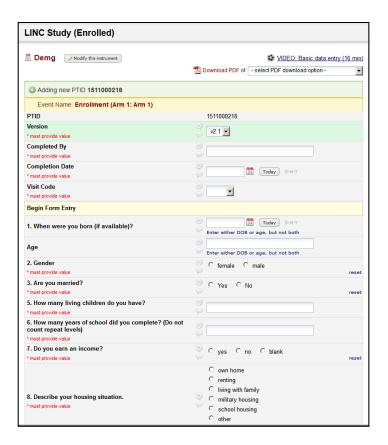
### Data Storage: General Principles

- The complexity of the system should be matched with the type of data being collected:
  - cross-sectional or longitudinal
  - number of forms
  - number of participants
  - number of study sites
  - number of people performing data entry

# Data Storage: General Principles

 Data entry "front end" should mirror the questionnaire/form



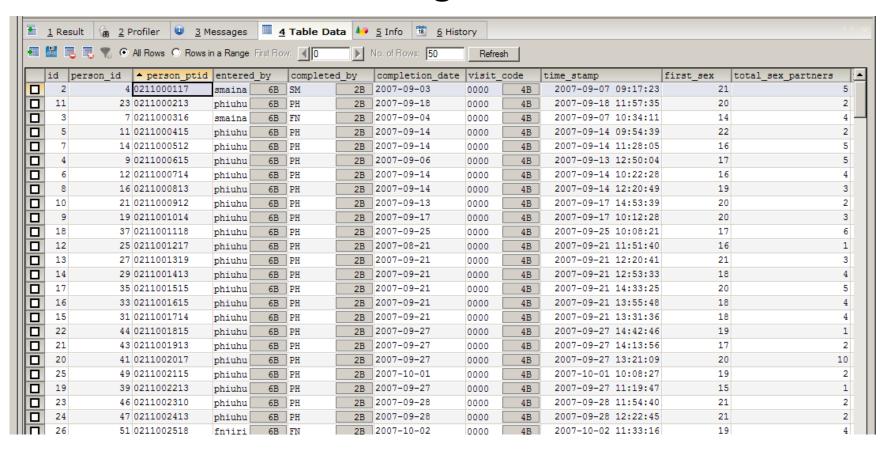


# Data Storage: General Principles

- Data should be stored in a database
  - ideally, the database should be accessible using SQL
  - it should be possible to easily export data in standard formats (tab or comma separated text file, excel, etc.)

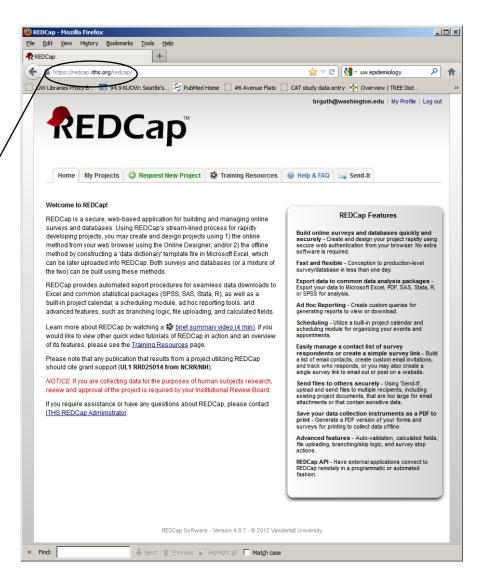
### Data Storage: Back-end Structure

 The "back end" of your data entry system should be a well designed database

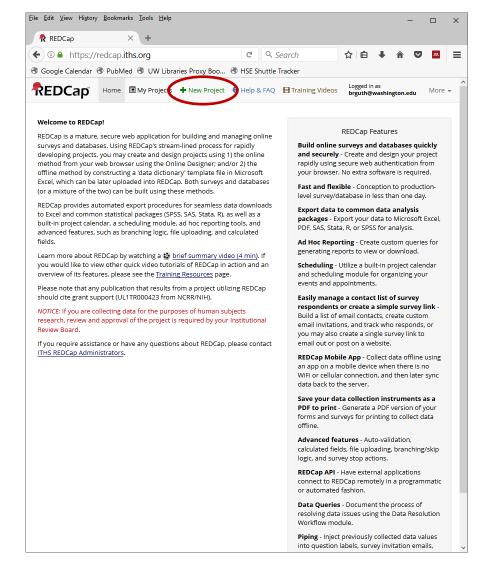


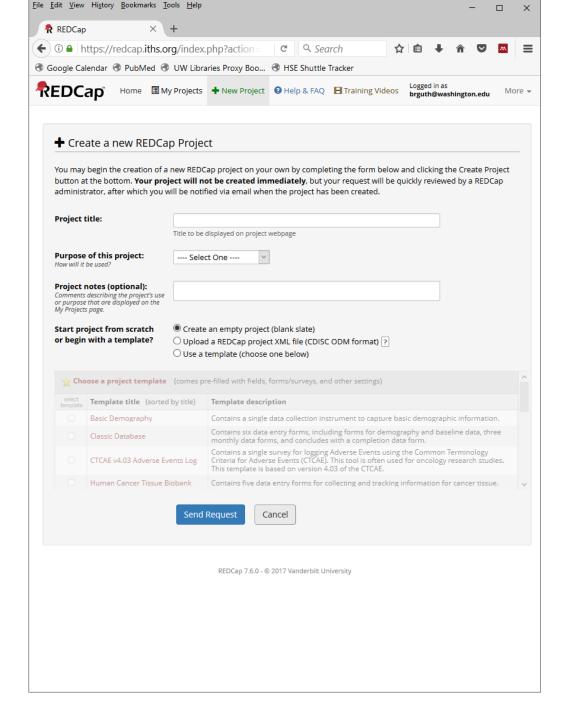
### **REDCap Demonstration**

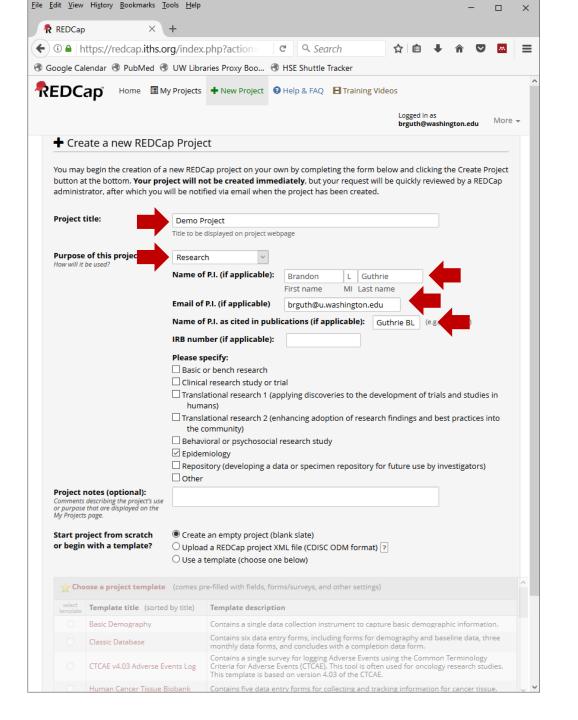
https://redcap.iths.org

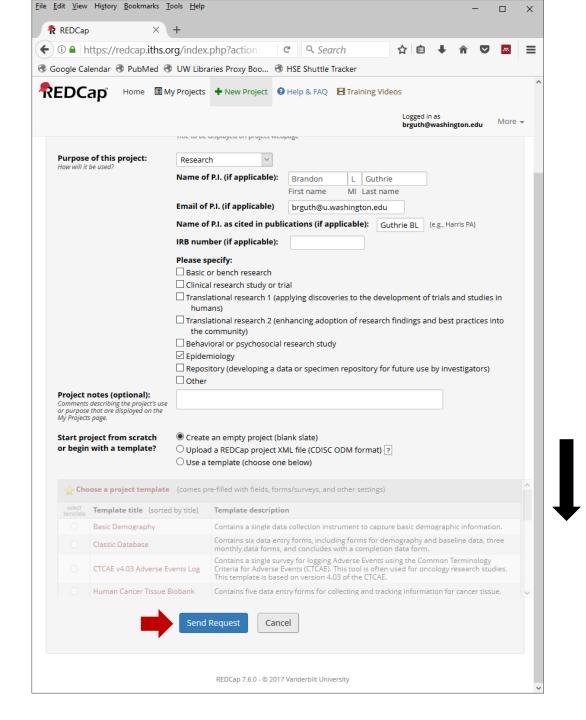


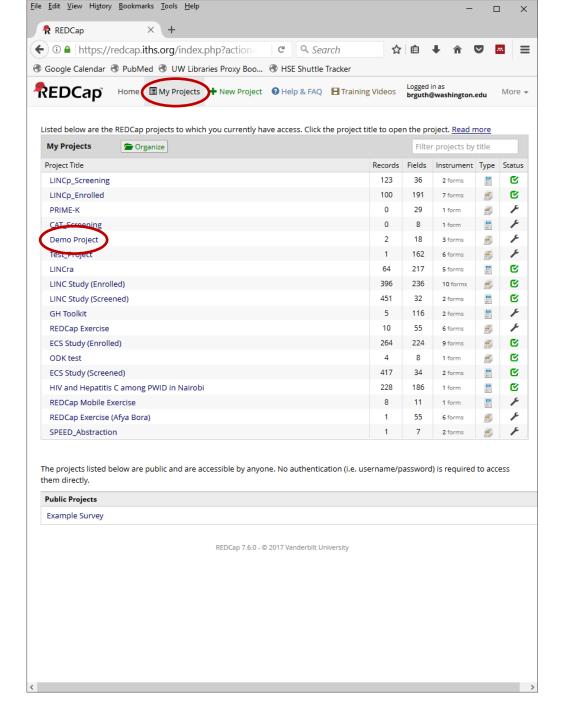
## **REDCap Demonstration**

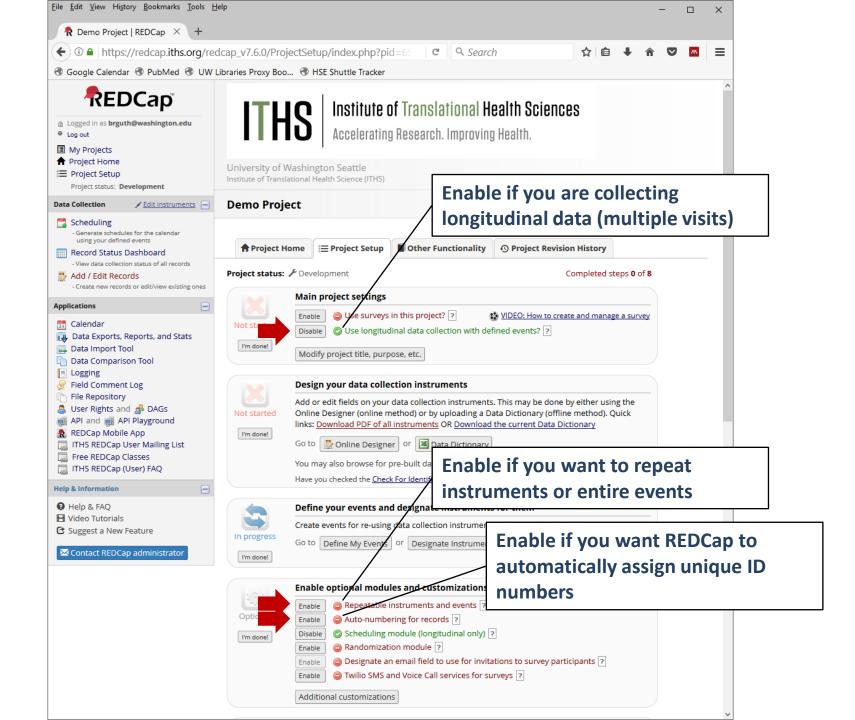


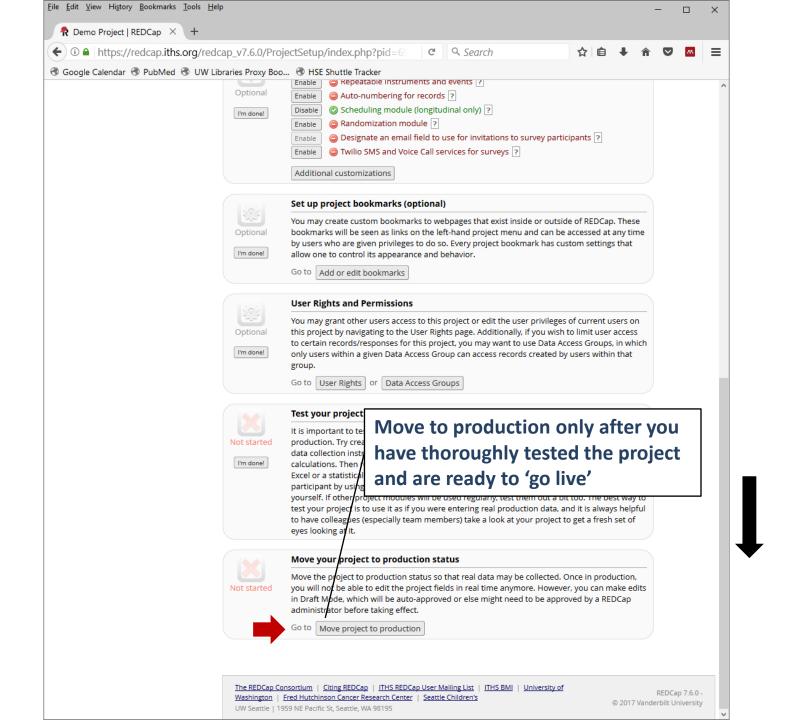






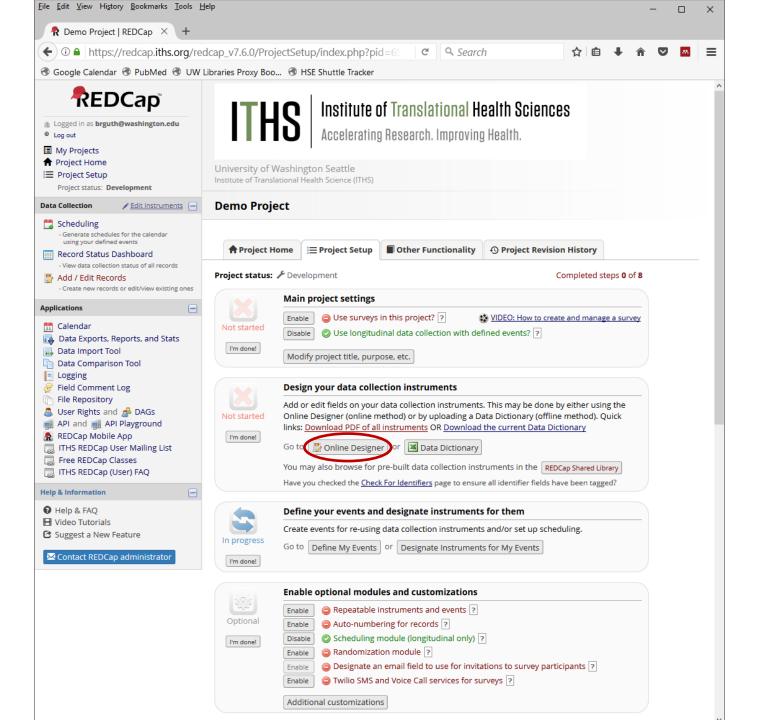


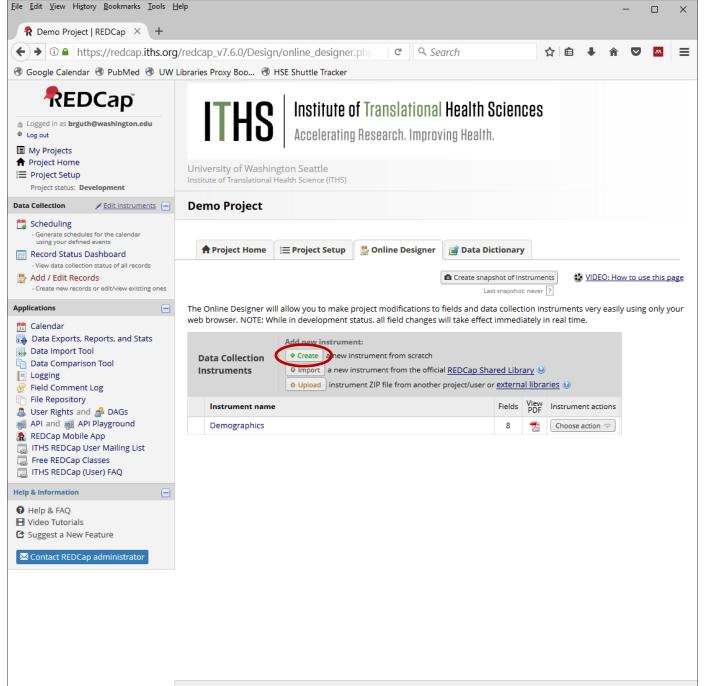




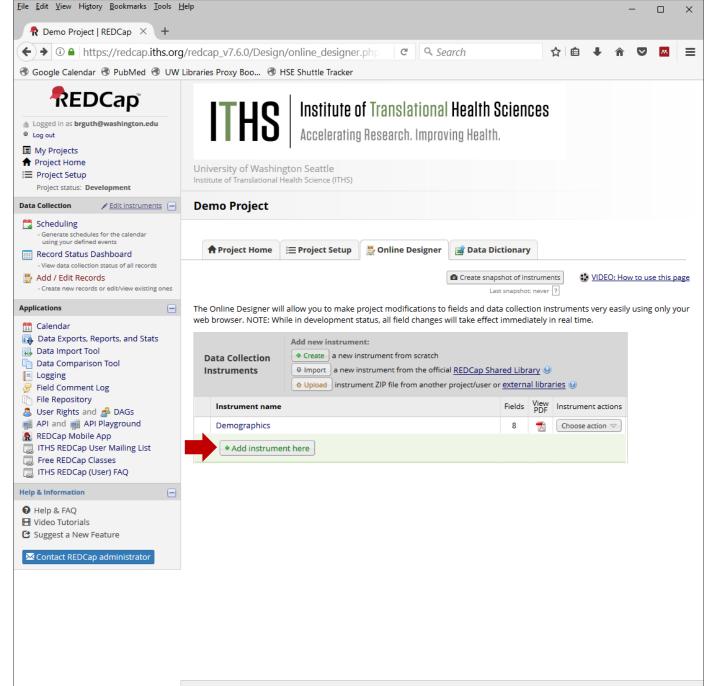
# Try it out!

- 1. Create a new project
- 2. Setup to project
  - Choose a longitudinal design
  - Decide whether you will assign ID numbers or whether you will let REDCap assign IDs (you can change this later)
  - Decide whether you will use repeating events/instruments (you can change this later)

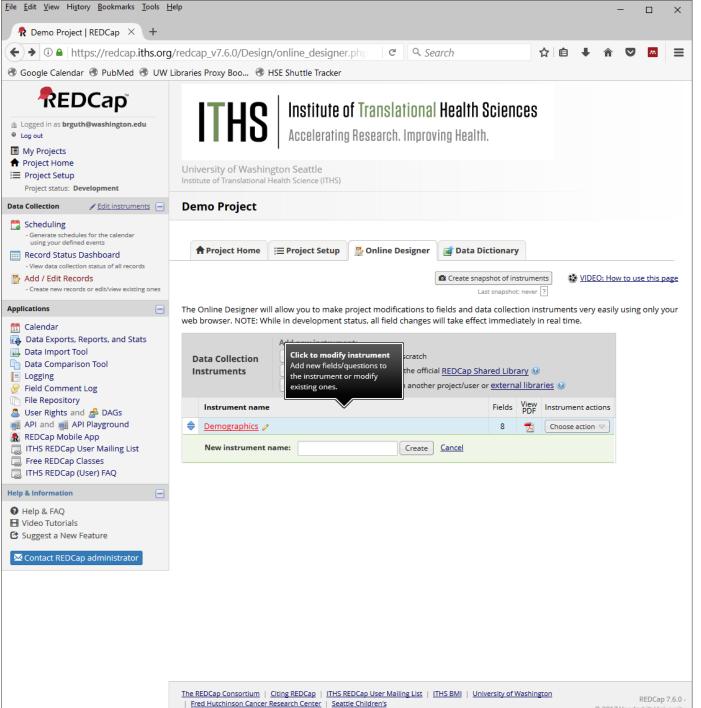




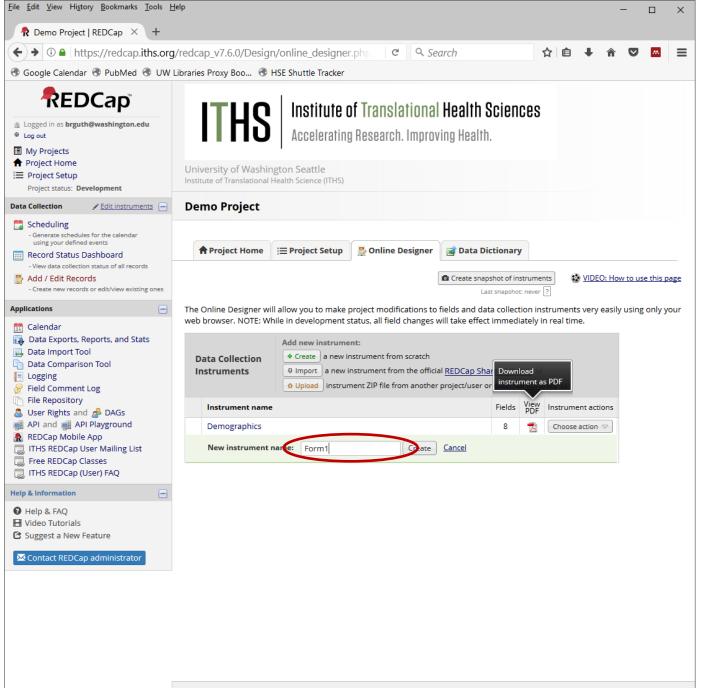
UW Seattle | 1959 NE Pacific St, Seattle, WA 98195



UW Seattle | 1959 NE Pacific St, Seattle, WA 98195



https://radgap.iths.org/radgap.y7.6.0/Dasign/opling.dasigner.php?pid=6008/paga=damagraphics



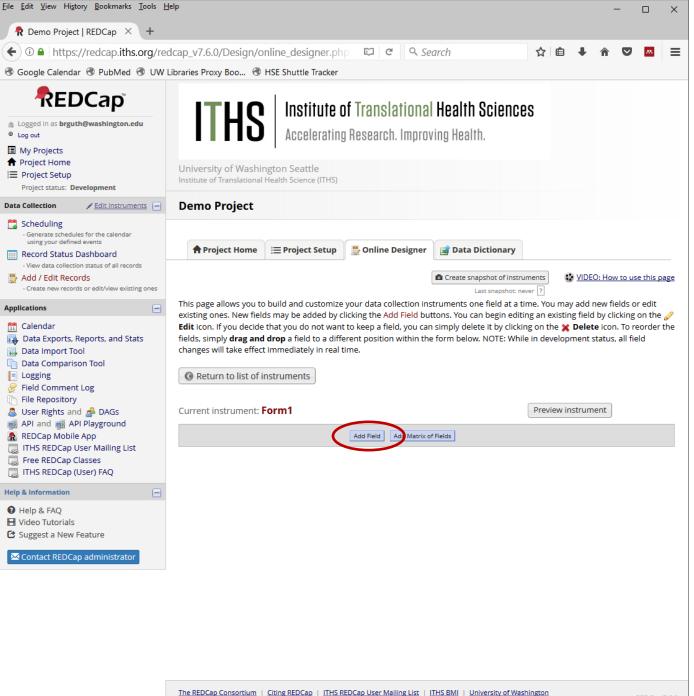
## Try it out!

## 1. Start your first form/instrument

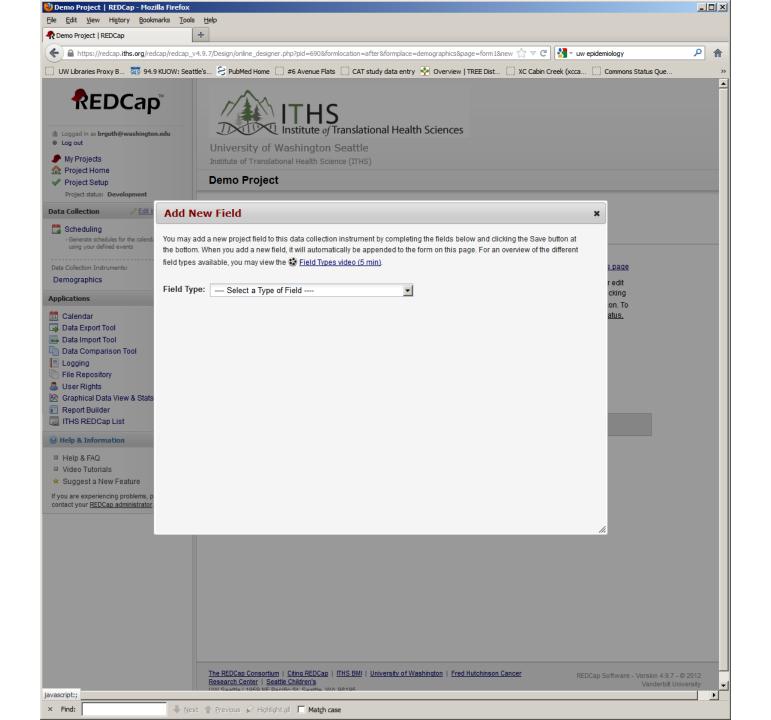
- Option 1: Use the default form (you will likely want to rename the form)
- Option 2: Create a new form
  - Create the form
  - Create the first variable and make it the variable you will use for the ID number
  - Move your new form to the first in the instrument order
  - Delete the default form

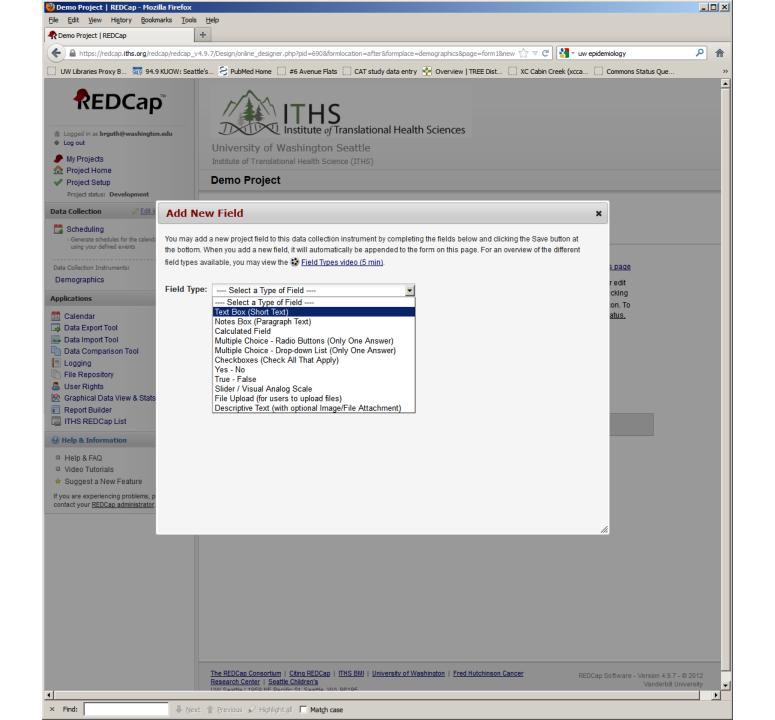
Linked in Care - pilot Main Form (MAIN-1)

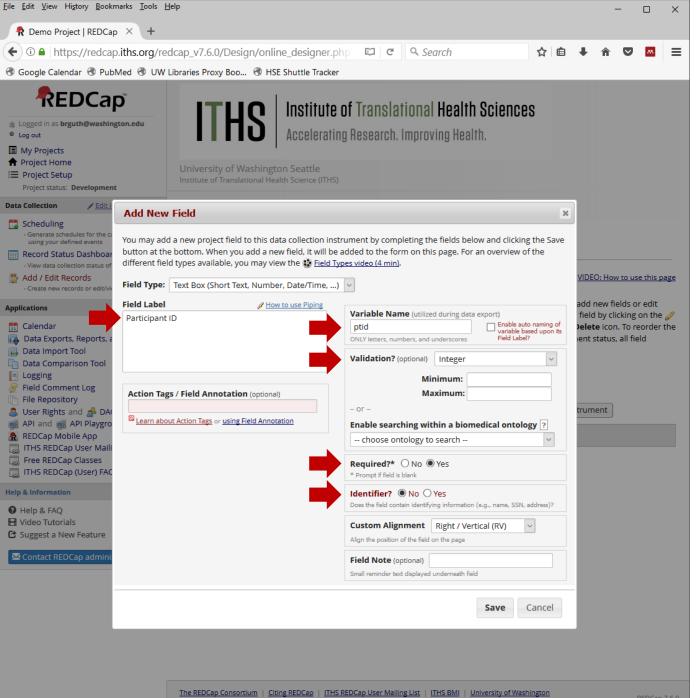
			z.	9	Visit Code	0 0 . 0 0				
(LINC-p) 14 (MAIN-1) 100		1898	Pandomi	1 2 3 4						
Participant ID:			- teal	Randomization Group:						
					T T T					
	<b>1 4 -</b> Study Site	Participant number Ci	The state of the s	dd	mmm	ууууу				
Ma	Main Form, page 1 This is an interviewer-administered form. Read each item aloud to the participant.									
1	When were you born (if	available)? dd mmm	уууу	OR	What is your age?	years				
2	(This question is not inte	erviewer-administered.) Gender		female	☐ male					
3	Are you married?			☐ yes	☐ no	If no go to itom 4				
	3a. Please specify the	e type of marriage.	traditional							
	3b. IF FEMALE: Do	you have more than one husband? -		yes	П по	not asked				
	IF MALE: Do yo	u have more than one wife?		☐ yes	☐ 110	not asked				
	ms 4 to 6 are for fen									
4	How many times have y	ou been pregnant?		t	imes pregnant					
5	How many live births ha	ve you had?			ive births					
6	How many still births ha	ve you had (child born lifeless)?		1	number of still birth	ıs				
7	How many living childre	n do you have?			number of living ch	ildren				
8	How many years of sch	ool did you complete? (Do not count r	epeat levels)	<u> </u>	/ears					
9	Do you earn an income	?		yes	по					
	9a. What is your own	monthly income? (average over the la	ast three months)		], 🗔	If no, go to item 10. shillings				
10	Describe your housing s		own home							
	10a. How much do you	mortgage, etc.)	, shillings							
11	How many rooms do yo		rooms in home							
12	How many people live in your household?			people in household						
13	Do you have indoor run		yes	по	not known					
14	Does your house have an indoor toilet?				no no	not known				
15	Is your toilet shared with other households?			yes	no no	not known				
16	(This question is not into formal or an informal se	erviewer-administered.) Does the part ttlement?	icipant live in a	formal	informa	I not sure				
Ver	sion 2.0, 04-09-12	Faultale	ed by:		(initials/date) (initials/date)	Revised:				
	rsion 2.0, 04-09-12 English Entered by: (initials/date) Revised:									



UW Seattle | 1959 NE Pacific St, Seattle, WA 98195







UW Seattle | 1959 NE Pacific St, Seattle, WA 98195

## Data Storage: General Principles

- Design database to fit the structure of the forms
- Use standardized variable names
  - easily understandable
  - consistent with other studies

## Data Storage: General Principles

- All records indexed by participant ID number (PTID) and visit code
- Time stamp for data entry should be used
- Follow consistent coding scheme
   no = 0 yes = 1 not known = 9999
- Ordinal categories should retain logical ordering less than primary education = 1 primary education = 2 secondary or higher education = 3

Think carefully how you will store the data

"What is your partner's HIV status"

ptid	partner_hiv		
1	Positive		
2	Positive		
3	р		
4	HIV+		
5	Neg		
6	n		
7	NA		
8	dont know		

 For analysis, your data should be numerically coded:

$$0 = no$$

$$1 = yes$$

 You will need to determine how data will be entered on the questionnaire

## **Option 1**

HIV results:	(0=Neg, 1=Pos)
--------------	----------------

### **Option 2**

HIV results:	Negative	Positive
·		

Think carefully how you will store the data

"What is the participant's gender?"

 Where appropriate, provide flexibility in responses while ensuring consistency What concerns do you have about starting antiretroviral therapy in the future? (mark all that apply) cost of therapy side effects disclosure of HIV status challenge of adhering to therapy lack of support transportation to treatment center want to have children before starting interaction with other drugs other, specify: \_\_\_

 Anticipate data entry challenges

## Data Storage: General Principles

 Separate variables are needed for each option in "check all that apply items"

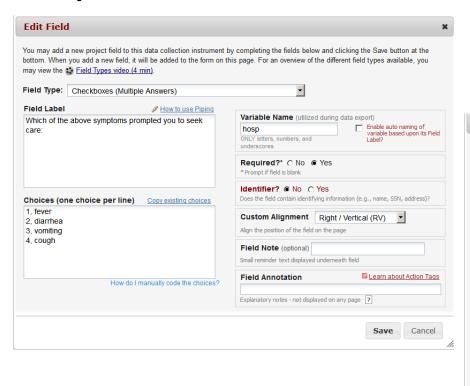
Which of the above symptoms prompted you to seek care:									
☐ fever ☐ diarrhea				□ vomiting	☐ cough				
							_		
hospital_:	fever	hospital_diarrh	nea	hospital_vomiting		hospital_cough			
no	•	no	-	no	•	yes	▼ 1		
no no	-	no	-	no	•	no	<b>-</b> !		
☐ yes	-	no	-	no	•	yes	<b>-</b> !		
☐ yes	-	no	-	no	•	yes	<b>-</b> 1		
no no	-	no	-	no	•	yes	<b>-</b> 1		
no no	-	no	-	no	•	no	<b>▼</b> 1		
no no	•	no	•	no	•	no	<b>-</b> 1		
☐ yes	•	no	-	no	•	no	<b>-</b> 1		
☐ yes	•	no	-	no	•	no	<b>▼</b> 1 <b>▼</b>		
1							<b>F</b>		
Database: cat_db_clinical_entry Table: enrollment_medical_history_partners									

# Check all that apply

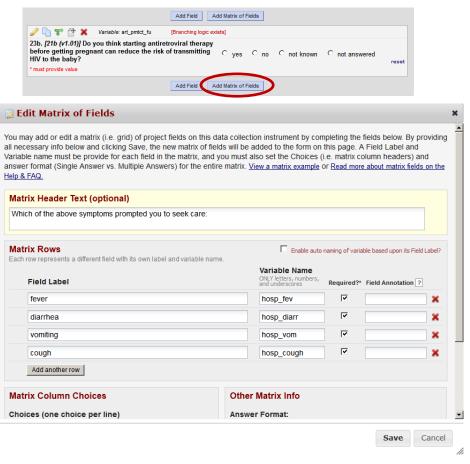
Which of the above symptoms prompted you to seek care:

☐ fever ☐ diarrhea ☐ vomiting ☐ cough

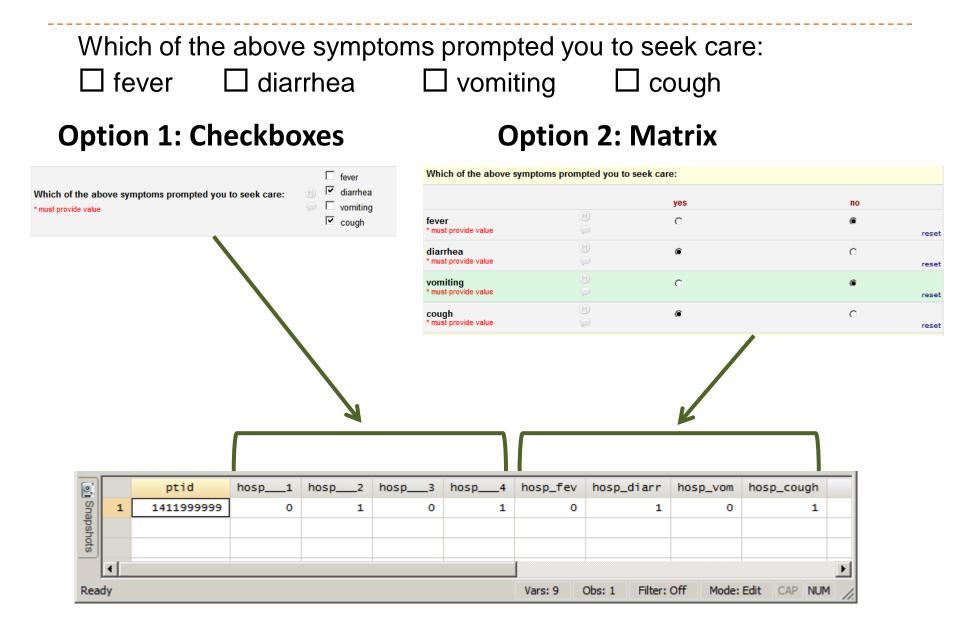
#### **Option 1: Checkboxes**



#### **Option 2: Matrix**



## Check all that apply



## Data Storage: Minimizing Errors

- Use drop-down menus and check boxes
- Hide fields that are not applicable
  - example: don't show oral contraceptive use for male participants
- Include error checking at point of entry to prevent common errors
  - missing data
  - data inconsistencies
  - impossible values

## Data Storage: Variable Names

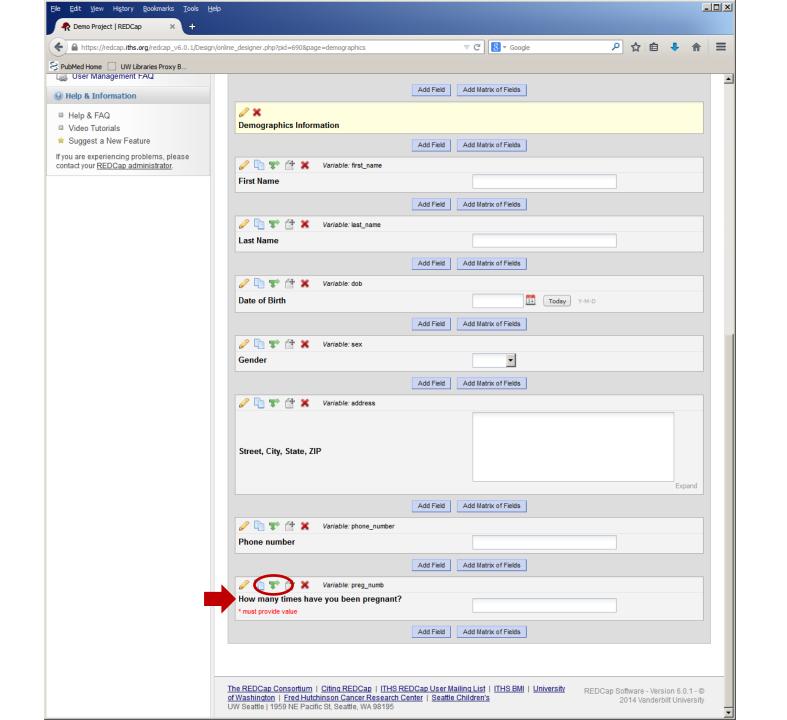
- Avoid excessively long variables.
  - Should be less than 26 characters
- Consider only using lower case letters.
- Avoid numbers at the beginning and end of variable name
- Use "\_" to improve readability
- Use consistent abbreviations:
  - yr (years)
  - oth (other)
  - sp (specify)
  - ptnr (partner)
- Follow consistent patterns:
  - barrier\_other
  - barrier\_other\_sp
- Consider consistent use of prefixes to group variables

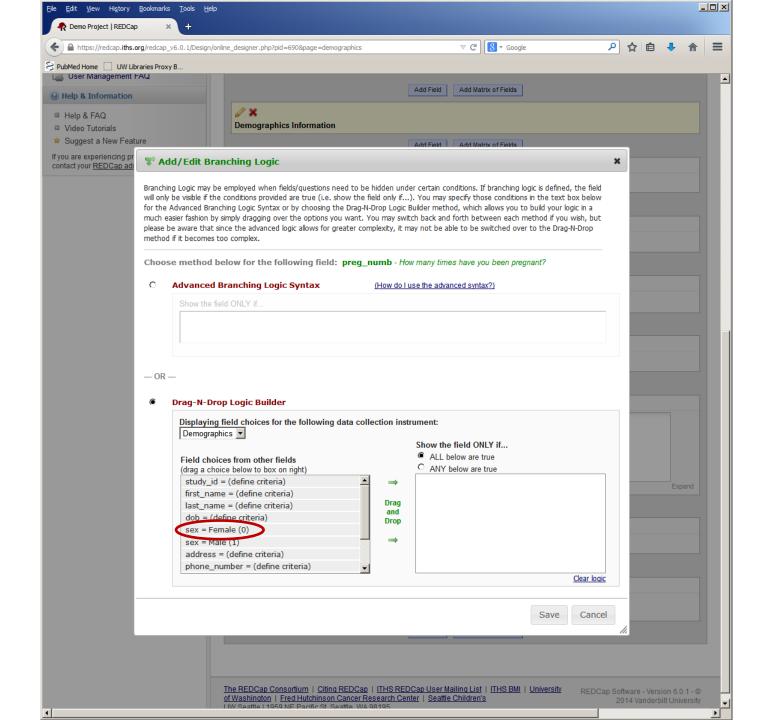
## Try it out: REDCap Exercise

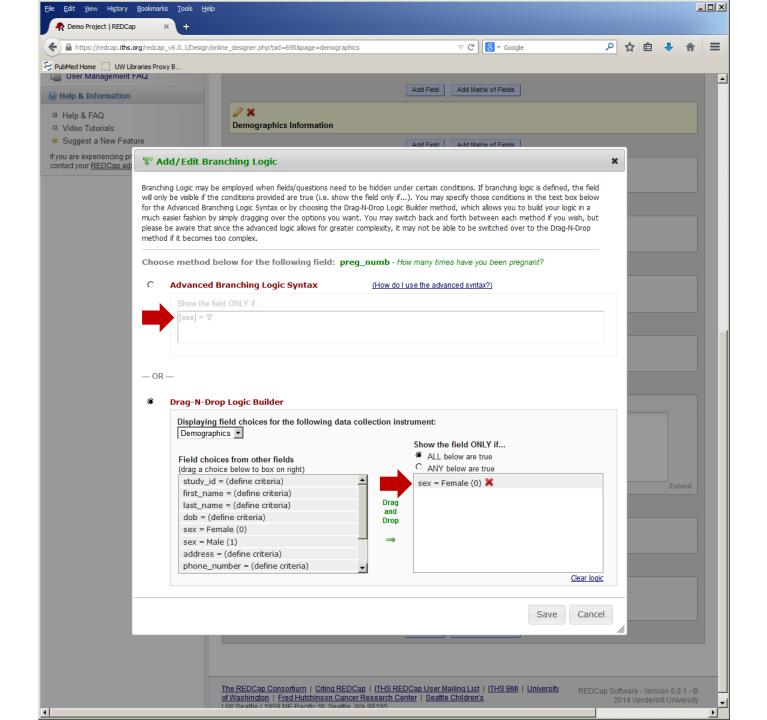
Category	Questionnaire Items			Follow-up
Demographics	Date of birth		Х	
	Gender		Х	
	What is your marital status?		Х	
	What is the legal nature of your marriage?		Х	
	Last month, did you earn an income?		Х	
	Last month, what was your income?			
	Have you had an income every month for the last	Х		
	Do you have regular employment?		Х	
	Classify your type of employment:	Х		
Care/Treatment Attitudes	How likely are you to go to a Comprehensive Care	e Center in the next 3 months?	Х	Х
	What do you think the benefits of antiretroviral t	herapy are?	х	Х
	select all that apply:			
	improve current health	improve job opportunities		
	prevent future health problems	reduce the risk of transmitting to a partner		
	live longer	other		
	look healthier			
Engagement in Care	Since enrolling in this study, did you go to a Comp	prehensive Care Center, or other clinic providing care for HIV.		Х
	When was your first visit to the Comprehensive C			Х
	Since you started this study, have you ever taken (if female, exclude ARVs to prevent mother-to-ch			Х
	When did you first start these antiretroviral medi			Х
	Are you currently taking any antiretroviral medica	ations?		Х
	What medications are you currently taking?			Х
	Mark all medications named:			
	Nevirapine (Viramune)	Stavudine (D4T)		
	Nelfinavir	Combination therapy (e.g., Triomune)		
	Zidovudine (AZT)	Indinavir		
	Tenofovir (TDF)	Lamivudine (3TC)		
	Efavirenz (Sustiva)	Other		
	Kaletra	Not Known		

Category	Laboratory Results Ba		Follow-up
CD4 Results	Results CD4 count (available 1 week after enrollment)		Х
Viral Load Results	Viral load (available after all follow-ups are completed)		Х

# Branching logic (aka skip patterns)





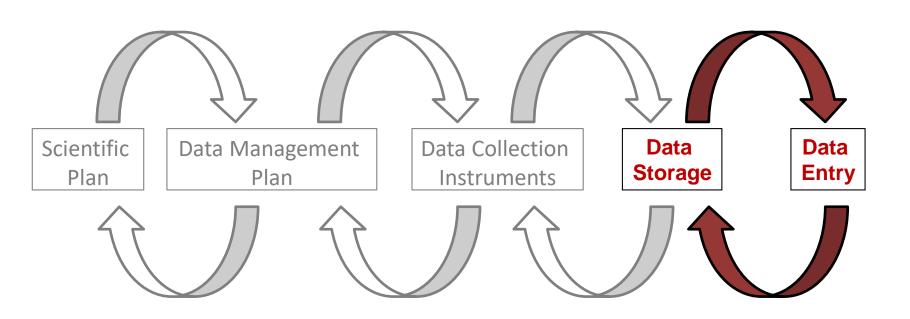


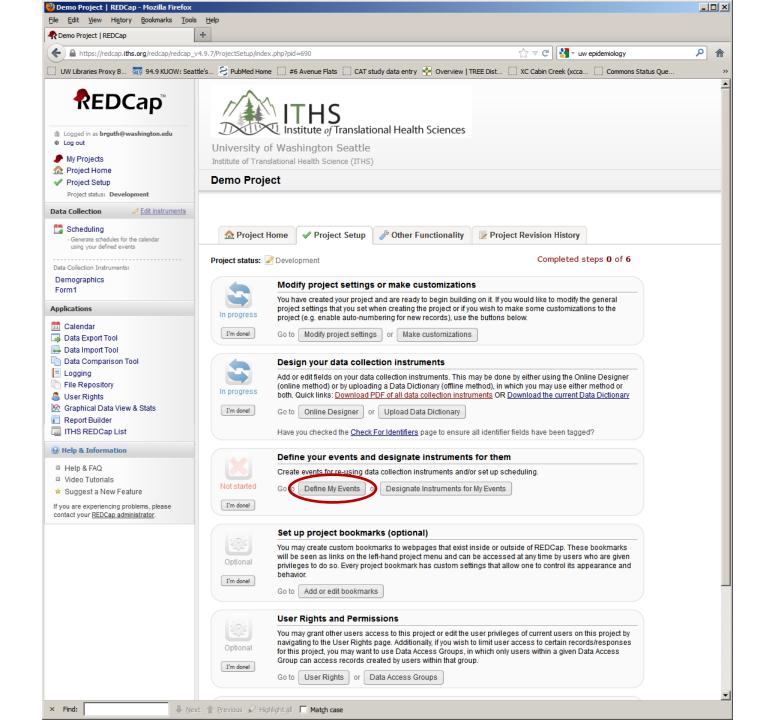
# Data Management Practices in Health Research

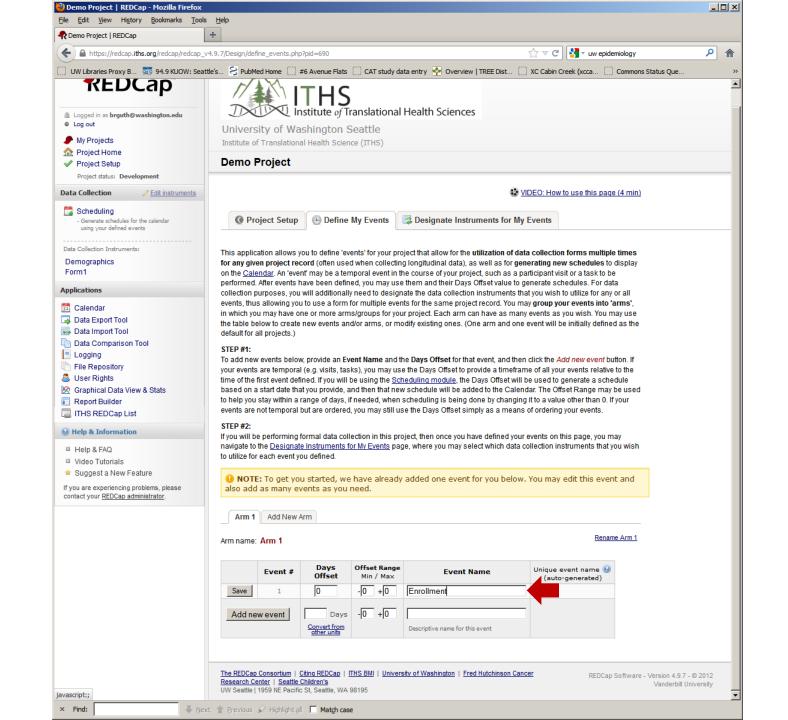
## Part II

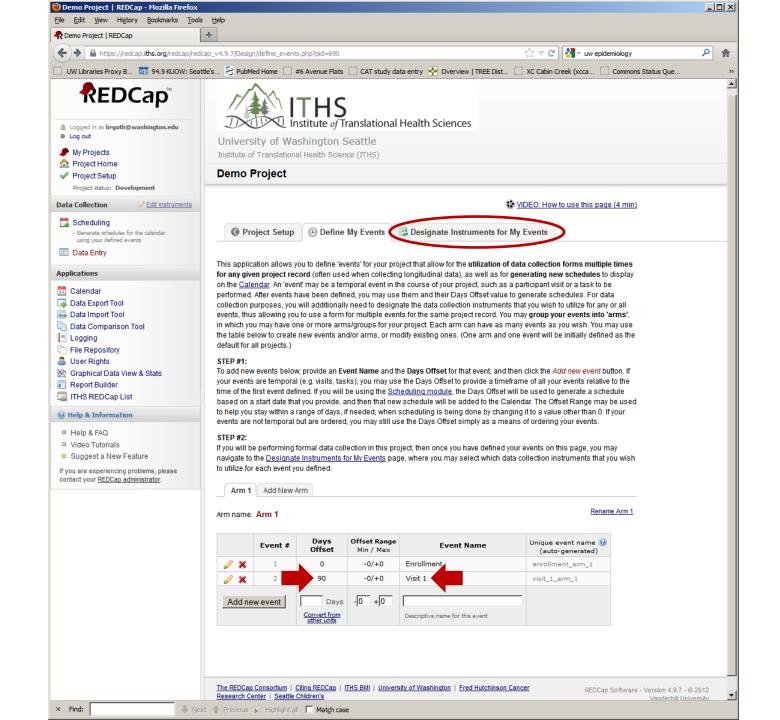
Managing data collection, storage, and dissemination

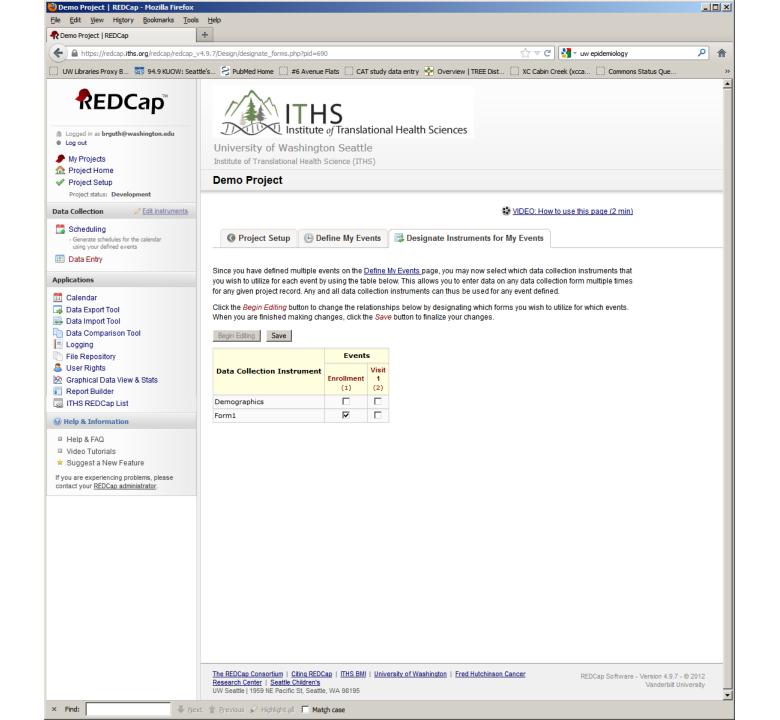
# Managing the Operation

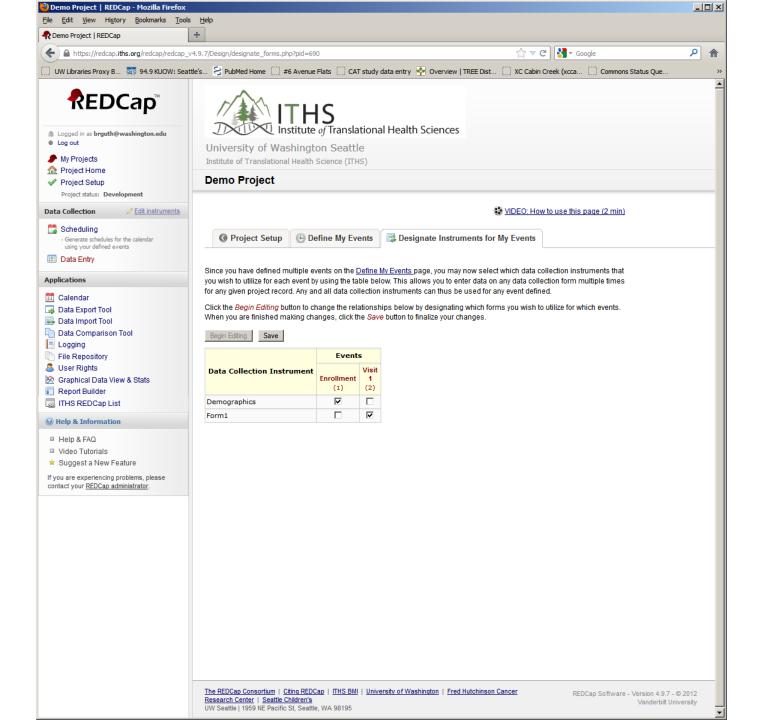


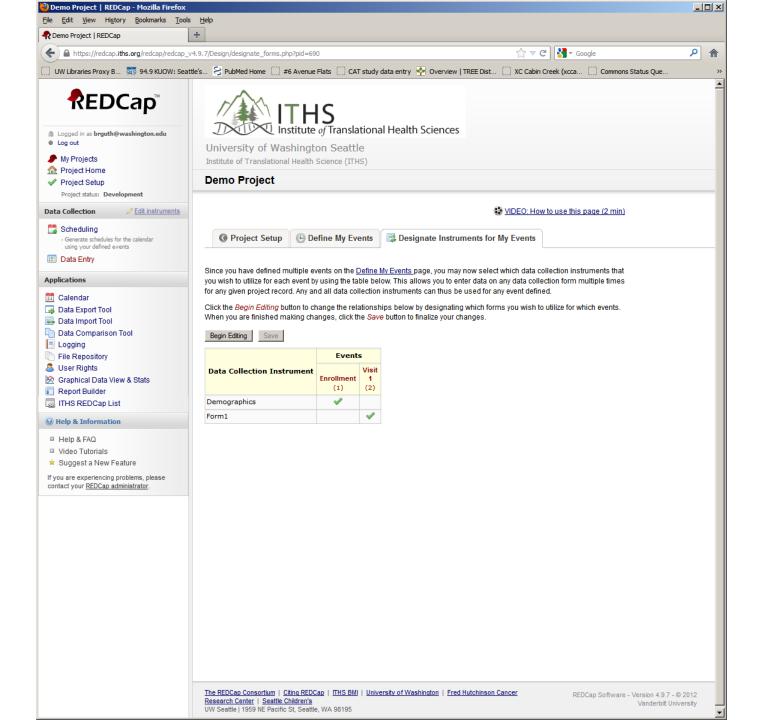


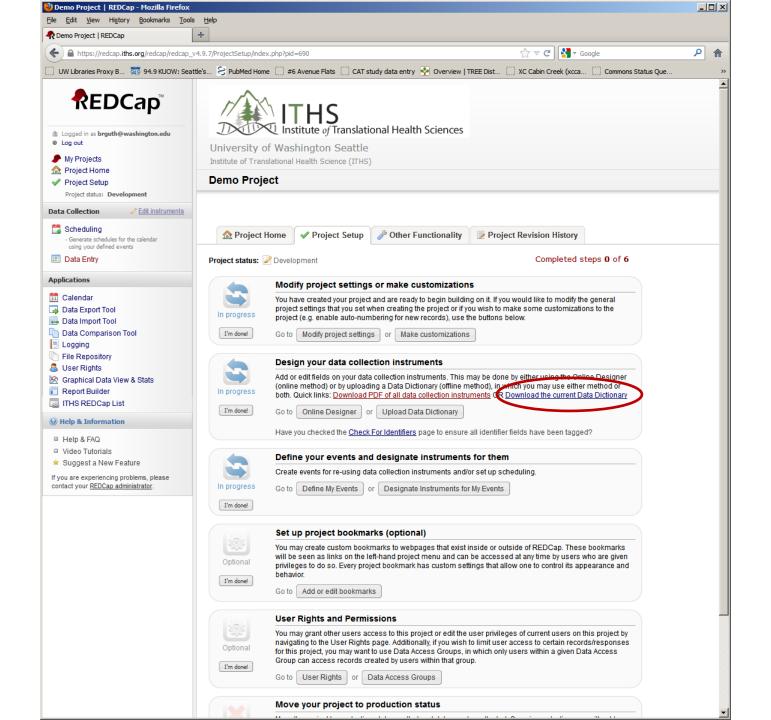


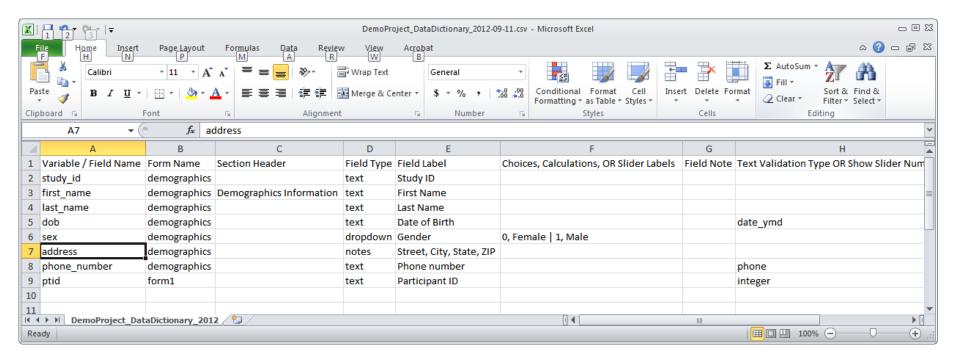


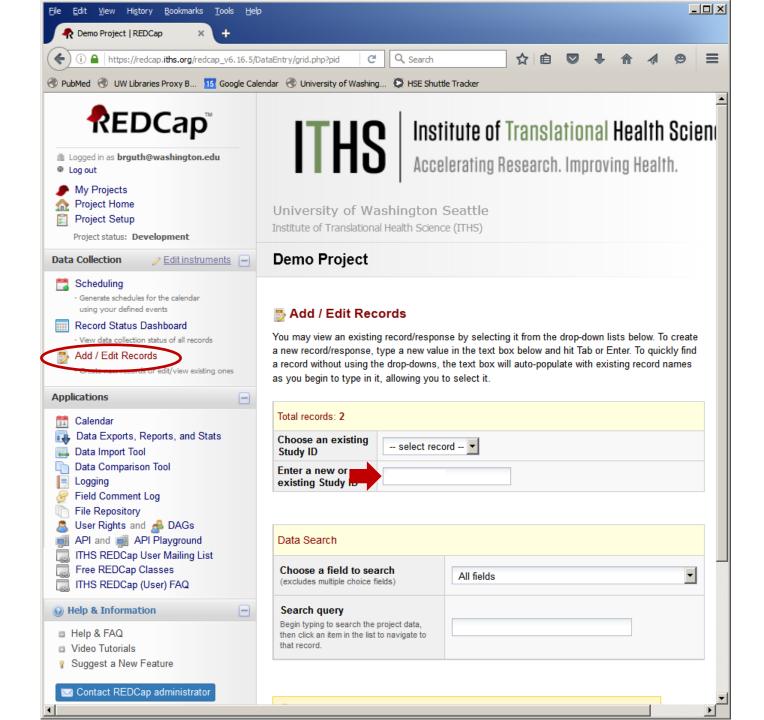


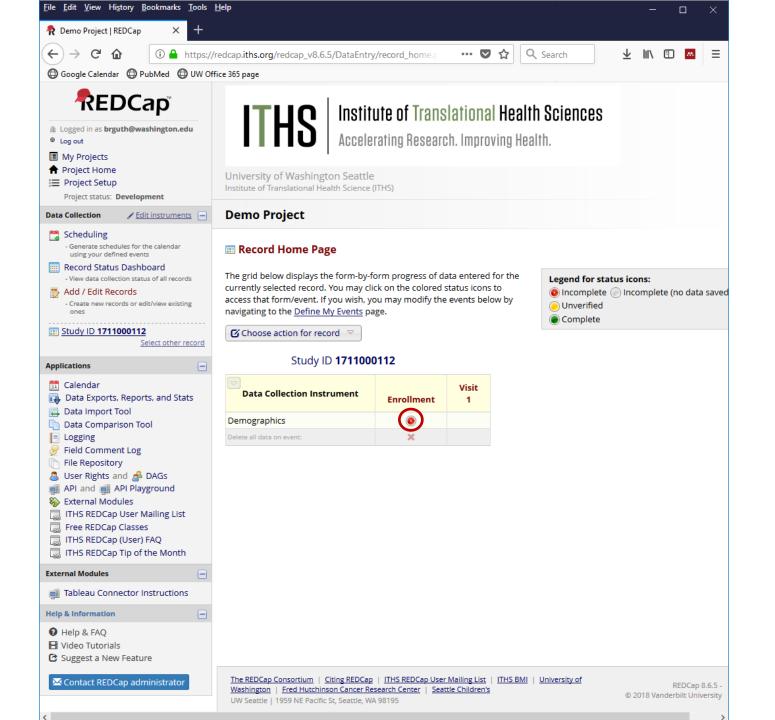


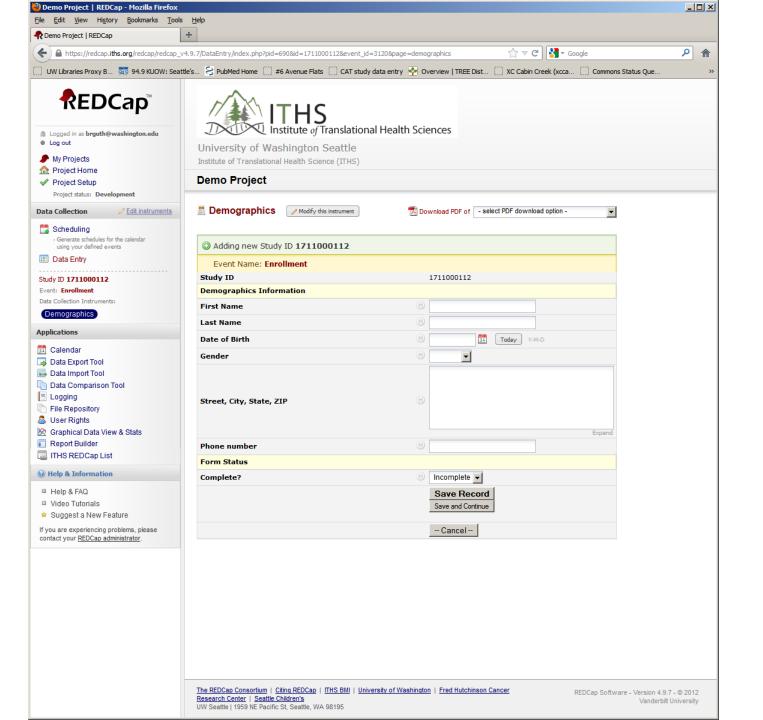


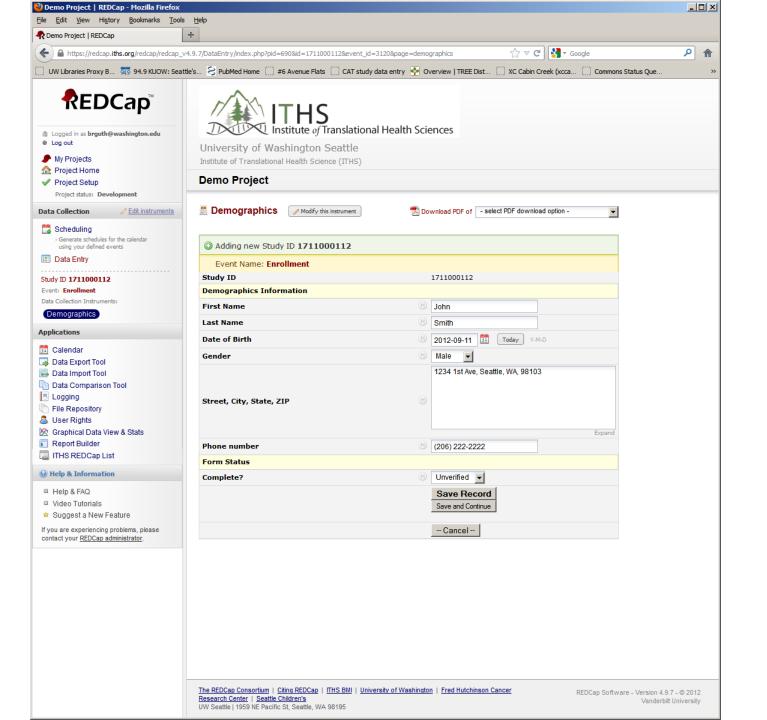


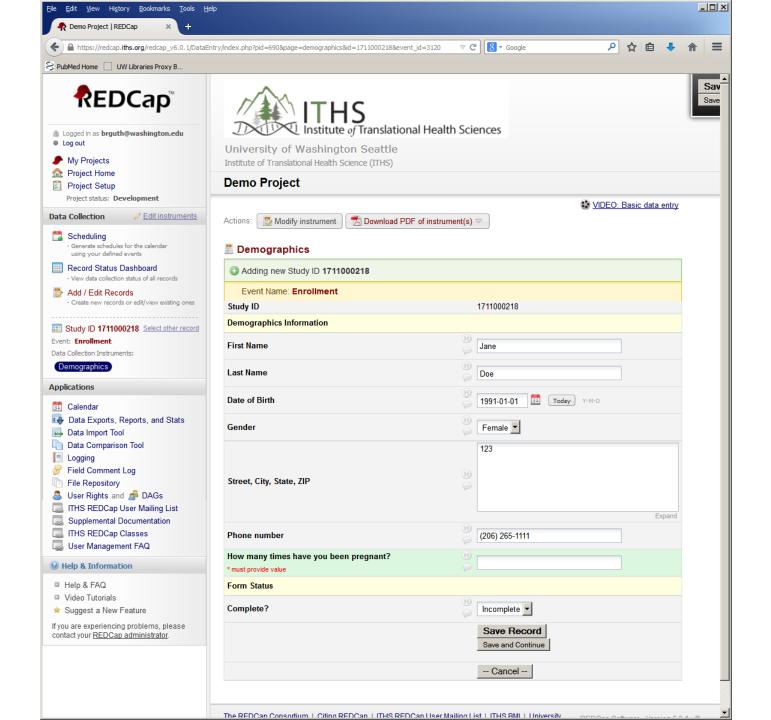


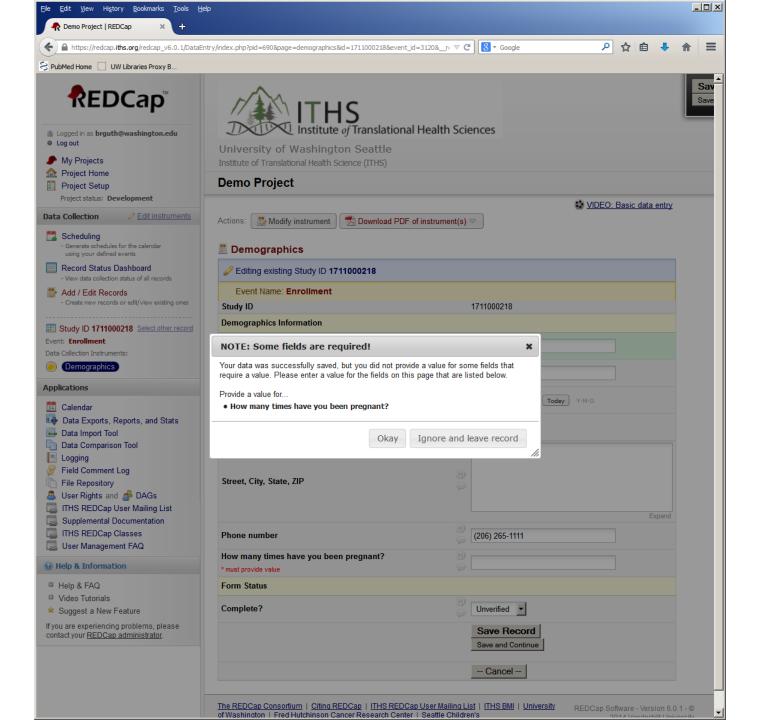


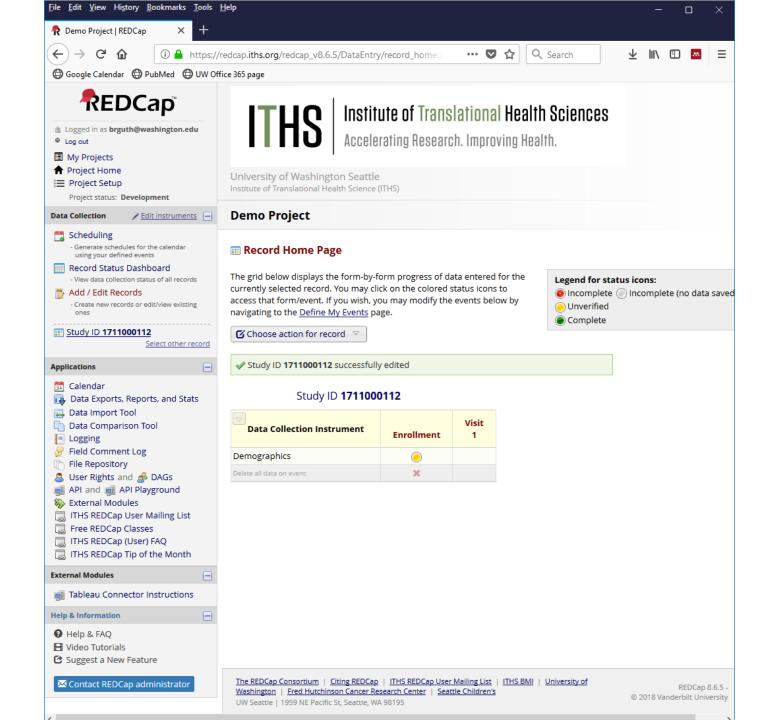


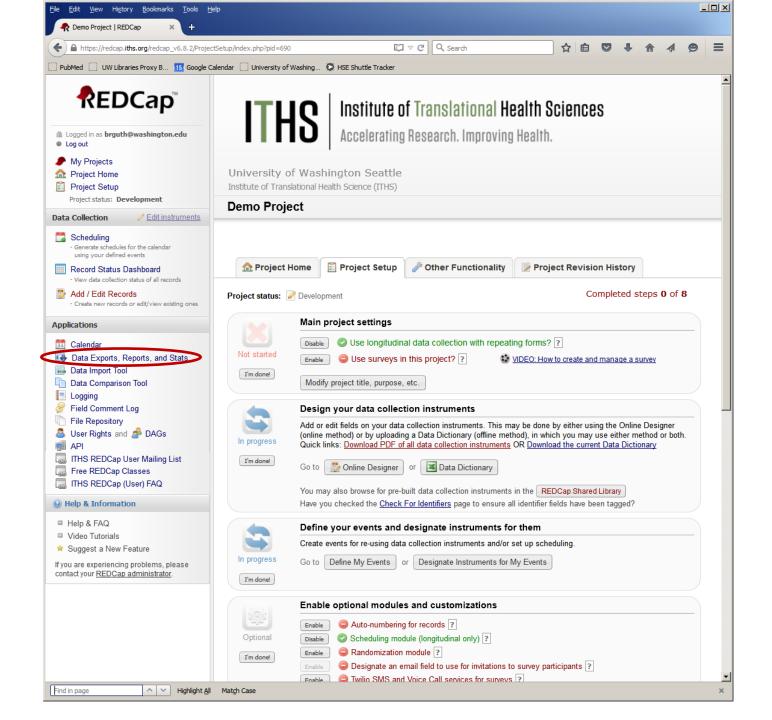


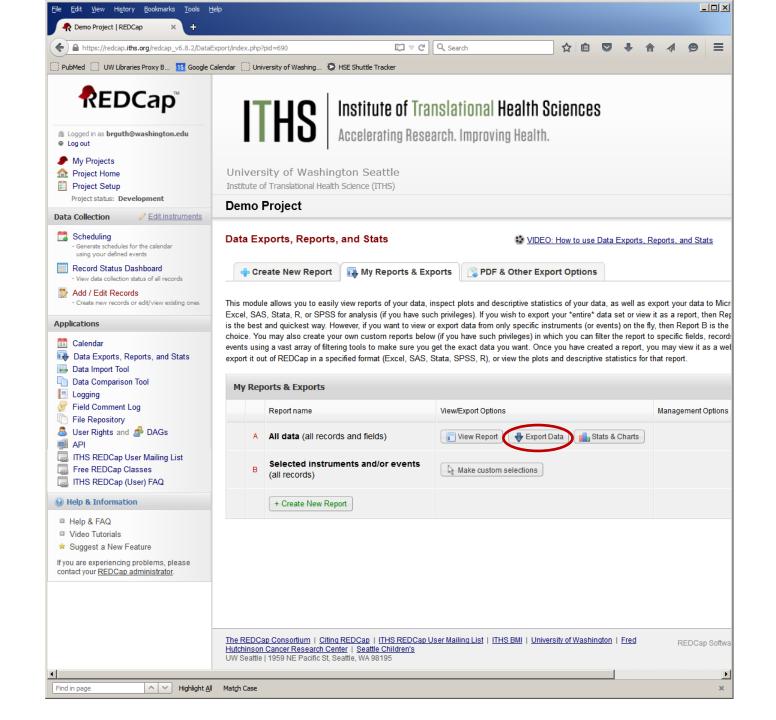


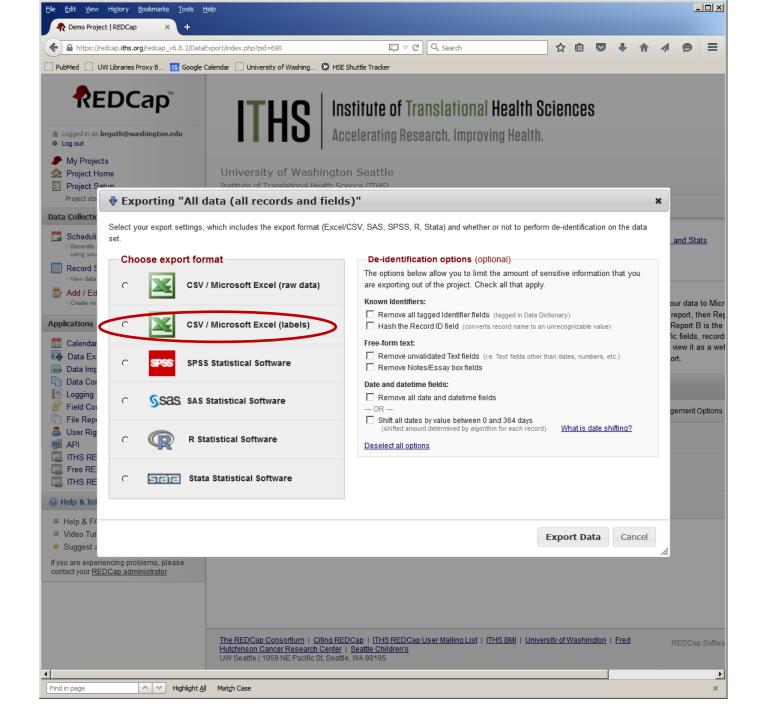


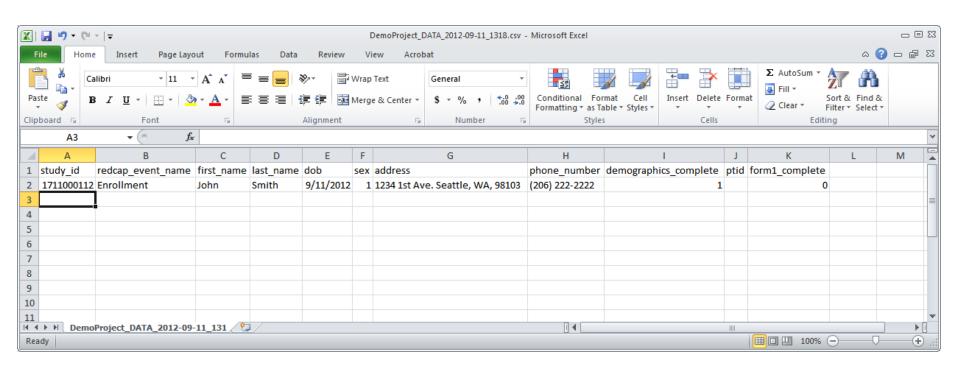


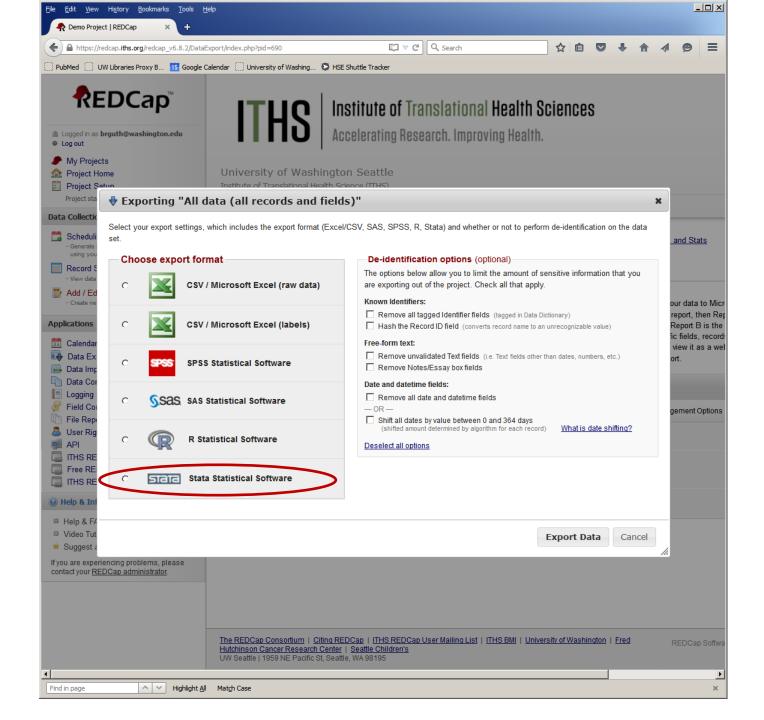








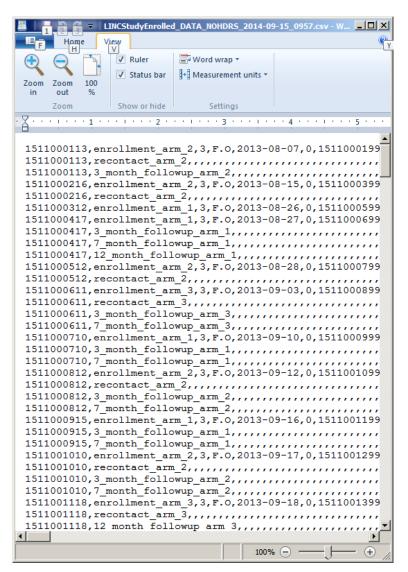


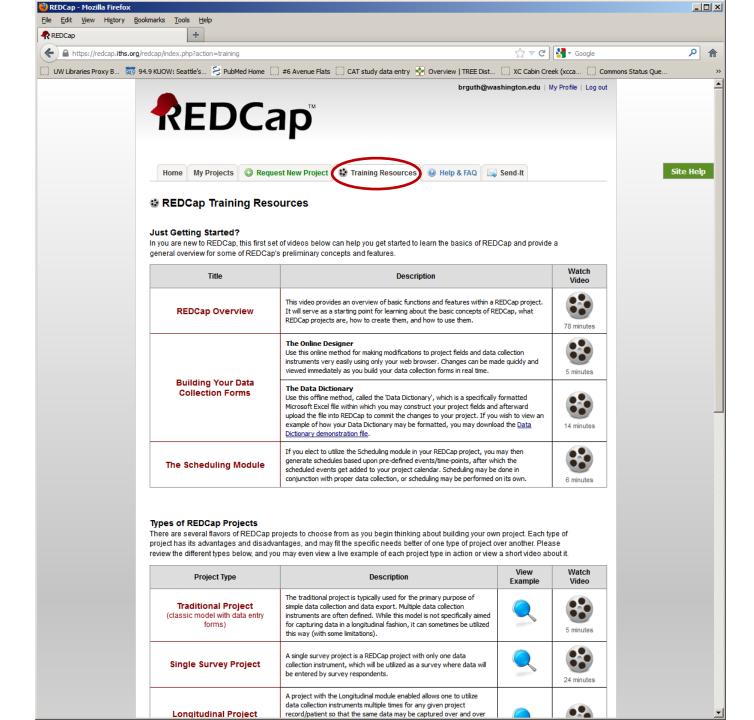


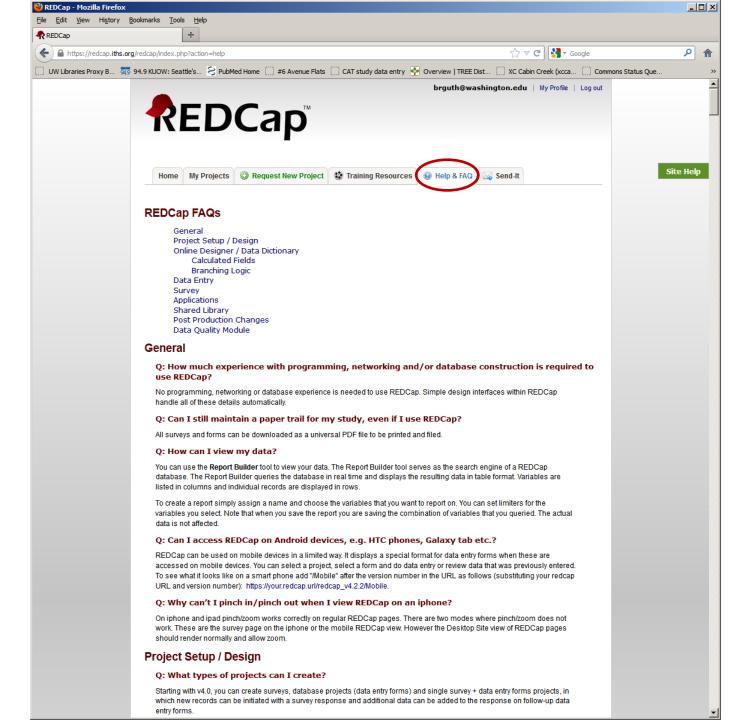
### Stata .do file

### **Data file**

```
🗹 Do-file Editor - LINCStudyEnrolled STATA 2014-09-15 0957.do
File Edit Tools View
D 🚅 🖫 🖶 A X 📭 🖺 19 (*) 1-至 🐨 🖢 🖓 🖟 🖟
  LINCStudyEnrolled_STATA_2014-0... Untitled.do
       clear
       insheet ptid redcap event name en version en completed by en completion date en visit
       label data "LINCStudyEnrolled DATA NOHDRS 2014-09-15 0957.csv"
       label define en version 3 "v2.1"
       label define en visit code 0 "00.00" 1 "00.01" 100 "01.00" 200 "02.00" 300 "03.00"
       label define en eligible 1 "Yes" 0 "No"
 10
 11
       label define en consent given 1 "Yes" 0 "No"
 12
       label define en consent recontact given 1 "Yes" 0 "No"
 13
       label define en exclude 1 "Yes" 0 "No"
 14
       label define en randomization group 1 "1" 2 "2" 3 "3"
       label define enrm complete 0 "Incomplete" 1 "Unverified" 2 "Complete"
16
       label define de version 3 "v2.1"
       label define de visit code 0 "00.00" 1 "00.01" 100 "01.00" 200 "02.00" 300 "03.00"
 17
18
       label define de gender 1 "female" 0 "male"
19
       label define de married 1 "Yes" 0 "No"
       label define de marriage type 1 "traditional" 2 "church" 3 "legal/civil" 4 "other" 99
 21
       label define de multiple spouses 1 "yes" 0 "no" -1 "not asked"
       label define de earn income 1 "ves" 0 "no" -1 "blank"
22
       label define de housing type 1 "own home" 2 "renting" 3 "living with family" 4 "milit
24
       label define de home water 1 "yes" 0 "no" 99 "not known" -1 "blank"
       label define de home indoor toilet 1 "yes" 0 "no" 99 "not known" -1 "blank"
       label define de home toilet shared 1 "yes" 0 "no" 99 "not known" -1 "blank"
 26
       label define de residence 2 "formal" 1 "informal" 3 "not sure"
 27
       label define demg complete 0 "Incomplete" 1 "Unverified" 2 "Complete"
 29
       label define bs version 3 "v2.1"
       label define bs visit code 0 "00.00" 1 "00.01" 100 "01.00" 200 "02.00" 300 "03.00"
       label define bs ever arv 1 "ves" 0 "no" 2 "not known" -1 "blank"
 31
       label define bs current arv 1 "yes" 0 "no" 2 "not known" -1 "blank"
       label define bs current arv drug 1 0 "Unchecked" 1 "Checked"
 34
       label define bs current arv drug 2 0 "Unchecked" 1 "Checked"
 35
       label define bs current arv drug 3 0 "Unchecked" 1 "Checked"
       label define bs current arv drug 4 0 "Unchecked" 1 "Checked"
 37
       label define bs current arv drug 5 0 "Unchecked" 1 "Checked"
       label define bs current arv drug 6 0 "Unchecked" 1 "Checked"
       label define bs_current_arv_drug 7 0 "Unchecked" 1 "Checked"
       label define bs current arv drug 8 0 "Unchecked" 1 "Checked"
                                                                        Line: 1, Col: 0 CAP NUM OVR
```







## Data entry

- Develop a protocol and follow it
- Storage of files
  - secure, organized, accessible to those who need access
- Quality control
  - reviewed by person filling out form
  - reviewed by second person in clinic
  - reviewed by data entry personnel before entry
- Forms should be corrected only by the person who completed the form, or a designated person
  - do not erase original entry; clearly record the correction and initial/date the entry

## Data entry

- Movement of files
  - files should only be moved by designated personnel in a structured manner.
  - Movement of files should be logged (sending and receiving)
- Data entry should occur in a timely manner to avoid backlogs
- Duplicate entry is preferred
  - duplicate entry of a subset to validate data entry and identify problems
  - duplicate entry of every form can ensure accuracy, but can be time consuming



What is a barcode?

What can you do with a barcode?





**Symbology** 

Content: 1411000117

**Code 128** 

**PDF 417** 



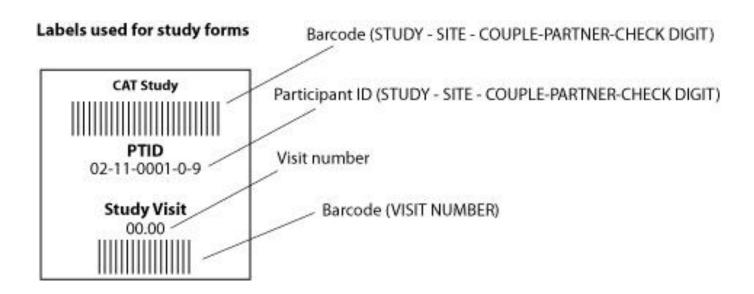
**QR Code** 



**Data Matrix 2D** 



### **Barcodes and Labels**



### Example specimen barcode:

Speed up data entry

Reduce transcription errors

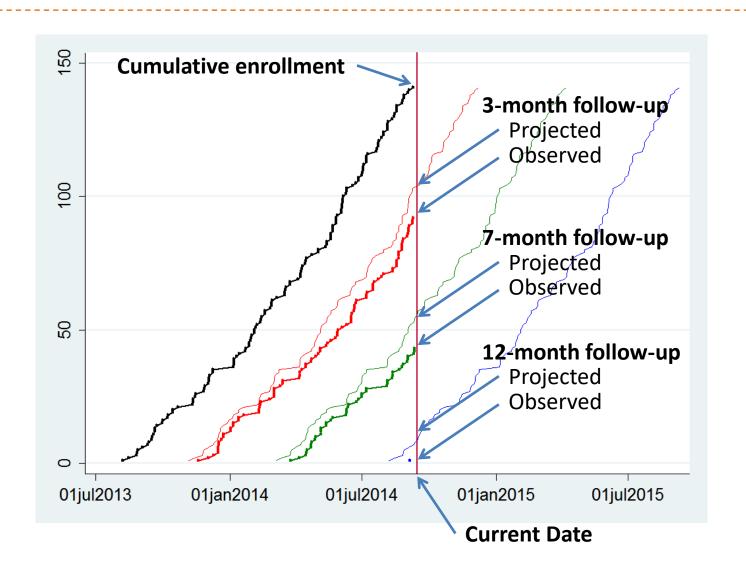
Improve storage of specimens

## **Monitor Progress**

- Monitor study progress in real-time
  - Missing forms / visits
  - Identify data entry errors
  - Identify enrollment issues
- Automate as much as possible
- Disseminate to relevant personnel

```
Missing_forms_2014_9_15.txt - Notepad
File Edit Format View Help
SUMMARY
Enrolled: 141
3 Month: 88
  Month: 39
List participants with missing LINT form (3 month follow-up)
     1511000216
     1511003315
     1511004519
List participants with missing LINT form (7 month follow-up)
     1511000113
Arm 2: 21 (81%) of 26 [5 missed]
Arm 3: 35 (88%) of 40 [5 missed]
Arm 1: 18 (90%) of 20 [2 missed]
Arm 2: 10 (67%) of 15 [5 missed]
Arm 3: 15 (75%) of 20 [5 missed]
12 Month:
Arm 1: 1 (97%) of 3 [2 missed]
Arm 2: 0 (81%) of 4 [4 missed]
```

## **Monitor Progress**



## Data Storage: Back-up

- Have a plan
- Make a schedule
- Test back-ups
- Store off-site
- Use multiple types of media
- "Outsource" your back-ups

## Data cleaning

- Garbage in = garbage out
  - the quality of your data depends on thorough data cleaning
- Develop a plan
- Don't fall behind
- Main categories
  - missing data
  - impossible values
  - inconsistencies
  - transcription errors

## Data cleaning

- Changes should NOT be made to the raw data set
- All changes should be documented
- Using "code" is preferred

### SQL:

```
UPDATE followup_medical_history_indices SET malaria =
   'no' WHERE person_ptid = '0211004818' AND visit_code
   = '0701';
```

#### Stata:

```
replace cd4 count = 437 if ptid == 211014613
```

### Documentation

- Study protocols
- Data dictionaries
- Changes to questionnaires and data collection instruments
- Systematic problems identified during the study
- Changes to the database structure
- Data cleaning

You probably won't remember what you did... so WRITE IT DOWN!

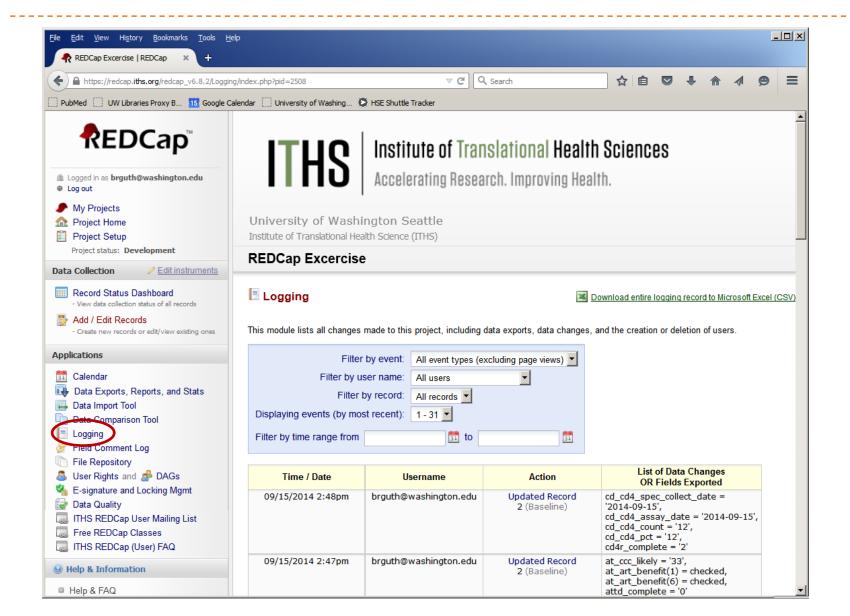
## Documentation: data code books

Form	Number	Question	Variable	Variable Type	Coding
SDEM	1	When were you born (if available)?	dob	DATE	
SDEM	1	OR what is your age?	age	SMALLINT	
SDEM	2	What is your gender?	gender	TINYTEXT	0 = MALE 1 = FEMALE 2 = Not listed
SDEM	3	Are you married?	married	TINYTEXT	0 = NO 1 = YES
SDEM	3a	Please specify the type of marriage	marriage_type	TINYTEXT	1 = TRADITIONAL 2 = CHURCH 3 = LEGAL 4 = OTHER
SDEM	3a	specify "other" marriage type	marriage_type_other	TINYTEXT	free entry
SDEM	3b	Do you have more than one wife/husband?	multiple_spouses	TINYTEXT	0 = NO 1 = YES
SDEM	4	How many times have you been pregnant?	pregn_number	TINYINT UNSIGNED	
SDEM	5	How many live births have you had?	live_birth_number	TINYINT UNSIGNED	
SDEM	6	How many still births have you had (child born lifeless)?	still_birth_number	TINYINT UNSIGNED	
SDEM	7	How many living children do you have?	children_living	TINYINT UNSIGNED	
SDEM	8	What is your ethnic group or tribe?	ethnic	TINYTEXT	free entry
SDEM	9	How many years of school did you complete?	school_years	TINYINT UNSIGNED	
SDEM	10	Do you own your own home?	own_home	ENUM	0 = NO 1 = YES
SDEM	10	What is your monthly rent?	rent	DECIMAL(8,2) UNSIGNED	
SDEM	11	Do you earn an income?	earn_income	TINYTEXT	0 = NO 1 = YES
SDEM	11a	What is your own monthly income in Kenyan shillings?	income_shill	DECIMAL(8,2) UNSIGNED	
SDEM	12	Does the participant live in a formal or an informal settlement?	residence	TINYTEXT	1 = INFORMAL 2 = FORMAL 3 = NOT SURE
SDEM	13a	Are you and your study partner married?	married_to_partner	TINYTEXT	0 = NO 1 = YES
SDEM	13b	Do you and your study partner live together?	live_with_partner	TINYTEXT NULL	0 = NO 1 = YES
SDEM	13b1	How long have you been living together? (years)	live_with_partner_years	SMALLINT UNSIGNED	
SDEM	13b1	How long have you been living together? (months)	live_with_partner_months	SMALLINT UNSIGNED	
SDEM	13c	How many living children do you have with the study partner?	children_with_partner	TINYINT UNSIGNED	

## Documentation: data revisions

	А	В	С	D	Е	F	G	Н	1	J
1		Requestor	PTID	Visit	Form	Change needed	Variable name		Corrected by	Comments
	21-Oct-09	Anne/Rob	0211-0131-03	555	???	Issue came up during our conversation on how to fix the	NO CHANGE MADE	22-Oct-09	BLG	male_circumcised already set to
						switched couples: this man is circumcised but the CRF				
9						says he is not circumcised.				
	13-Oct-09	Priscilla	0211-0301-0-4	S400	Visit	This visit is not yet due was erroneously created ,kindly	record deleted	22-Oct-09	BLG	
10						destory the record .				
11	15-Oct-09		0211-0340-1-1		IFPE	Confirm that the form was created.	NO CHANGE MADE	22-Oct-09		Form was entered on 15 Oct 200
12	15-Oct-09		0211-0083-0-5		visit	Change the completion date to 30th September 2009	completion_date	22-Oct-09		
13	16-Oct-09		0211-0344-1-2		visit	Change the completion date to 16th July 2009	completion_date	22-Oct-09	BLG	
	16-Oct-09	Priscilla	0211-0344-1-2	Visit 0100/0	IFSX, IFPE, FSHM	The information in the CRFs appearing as Visit 0100 is				
14						actually for visit 0200 and vice versa				
15	16-Oct-09	Priscilla	0211-0126-0-8	Visit 0700	visit	Change the completion date to 8th October 2009	NO CHANGE MADE	22-Oct-09	BLG	completion_date already set to '
	16-Oct-09	Priscilla	0211-0126-0-8	Visit 0800	visit	This visit is not yet due was erroneously created ,kindly	record deleted	22-Oct-09	BLG	
16						destory the record .				
17	16-Oct-09	Priscilla	0211-0261-1-9	Visit 0300	visit	Change the completion date to 25th June 2009	completion_date	22-Oct-09	BLG	
18	16-Oct-09	Priscilla	0211-0233-1-0	Visit 0300	visit	Change the completion date to 6thApril 2009	NO CHANGE MADE	22-Oct-09	BLG	completion_date already set to '
19	16-Oct-09	Priscilla	0211-0160-0-8	Visit 0600	visit	Change the completion date to 1st September 2009	completion_date	22-Oct-09	BLG	
20	16-Oct-09	Priscilla	0211-0015-0-4	Visit 0701	visit	Change the completion date to 19th August 2009	completion_date	22-Oct-09	BLG	
21	16-Oct-09	Priscilla	0211-0388-1-2	Visit 0100	visit	Change the completion date to 10th September 2009	completion_date	22-Oct-09	BLG	
22	16-Oct-09	Priscilla	0211-0326-0-8	Visit 0200	visit	Change the completion date to 22nd September 2009	completion_date	22-Oct-09	BLG	PTID should be 0211-0326-0-0
23	16-Oct-09	Priscilla	425	Visit 0000	visit	Change the completion date to 10th August 2009	NO CHANGE MADE	22-Oct-09	BLG	completion_date already set to '
24	16-Oct-09	Priscilla	0211-0014-1-3	Visit 0701	visit	Change the completion date to 24th August 2009	completion_date	22-Oct-09	BLG	
25	16-Oct-09	Priscilla	0211-0399-1-0	Visit 0000	visit	Change the completion date to 30 June 2009	completion_date	22-Oct-09	BLG	
	16-Oct-09	Priscilla	0211-0048-1-8	Visit 0701	IFMH-1	Q4-6 should be 'NO	malaria	26-Oct-09	BLG	only malaria, malaria_diagnose
							malaria_diagnosed			and malaria_treated changed
26							malaria_treated	<u> </u>		
	16-Oct-09	Priscilla	0211-0048-1-8	Visit 0701	IFMH	Q15 should be 'NO'	septrin	26-Oct-09	BLG	
27							septrin_start			
28	15-Oct-09	Priscilla	0211-0249-1-3	Visit 0400	IFPE	Confirm that the form was created.	NO CHANGE MADE	26-Oct-09	BLG	Form was entered on 10 Sept 200
29	16-Oct-09	Priscilla	0211-0325-1-0	Visit 0200	IFPE	Q15 should be 'YES'	NO CHANGE MADE	26-Oct-09	BLG	pap_collected already set to 'yes
30	16-Oct-09	Priscilla	0211-0018-1-5	Visit 0800	IFPE	Q15 should be 'YES'	pap_collected	26-Oct-09	BLG	
	16-Oct-09	Priscilla	0211-0104-0-2	Visit 0600	PFPE	Q13-15 should be 'NOT EXAMINED'				PLEASE CONFIRM PTID AND FORM
										TO CHANGE. THE PTID GIVEN IS FO
										A PARTNER, BUT THE FORM TO
										CHANGE IS FOR INDEX
31										PARTICIPANTS.
	16-Oct-09	Priscilla	0211-0394-0-4	Visit 0100	PFPE	Q13-14 should be 'NO' and Q15 Should be 'YES'				PLEASE CONFIRM PTID AND FORM
										TO CHANGE. THE PTID GIVEN IS FO
										A PARTNER, BUT THE FORM TO
										CHANGE IS FOR INDEX
32										PARTICIPANTS.
33	16-Oct-09	Priscilla	0211-1057-99	visit 9900	SCRN	Change refering VCT to BARAKA				V
14 4	K KI PALLALI:		D-4- :4		/ D-1	· / 🔄 /				

### Documentation: data revisions



## Tips and Tricks

### Prepare for some detective work

ptid	visit_date	visit_code	time_stamp		
02-11-0003-1-7	05-01-2012	0	05-05-2012-10:13:17		
02-11-0003-1-7	08-05-2012	1	08-07-2012-09:05:19	96	
02-11-0003-1-7	11-02-2011	2	11-04-2012-14:21:41	-277	<b>—</b>
02-11-0003-1-7	05-01-2013	3	05-07-2013-12:02:22	546	<b>—</b>
02-11-0003-1-7	02-03-2013	4	02-06-2013-15:33:58	-87	
02-11-0030-1-5	10-03-2012	0	10-06-2013-14:14:42		
02-11-0030-1-5	01-04-2013	1	01-05-2012-13:21:18	93	
02-11-0030-1-5	04-02-2013	2	04-05-2013-16:22:38	88	
02-11-0030-1-5	07-01-2013	3	07-03-2013-10:15:47	90	
02-11-0030-1-7	08-02-2013	5	08-04-2013-11:04:31	32	<b>4</b>

## Tips and Tricks

Prepare for some detective work

Incorporate redundancy

Anticipate reports

Expect the unexpected

## Other Forms of Data Entry

- Electronic data capture
- Mapping

## Electronic Data Capture



Handheld computers
Tablets
Smart Phones



Audio Computer Assisted Self-Interview (ACASI)

## REDCap, Open Data Kit, etc.





Open source (free) platform for electronic data capture http://opendatakit.org/





Source: www.kiva.org/updates/fellows/2012/09/12/technology-in-the-field-the-future-of-microfinance

## REDcap versus ODK

#### **REDcap Mobile App**

#### **Pros**

- Designed for epi study designs
- Strong project management
- Easy setup
- Integrates with standard entry
- HIPPA compliant

#### Cons

- Designed for epi study designs
- Limited error checking logic
- Poor user interface
- Limited control of user rights
- Some data synchronization problems

#### **ODK**

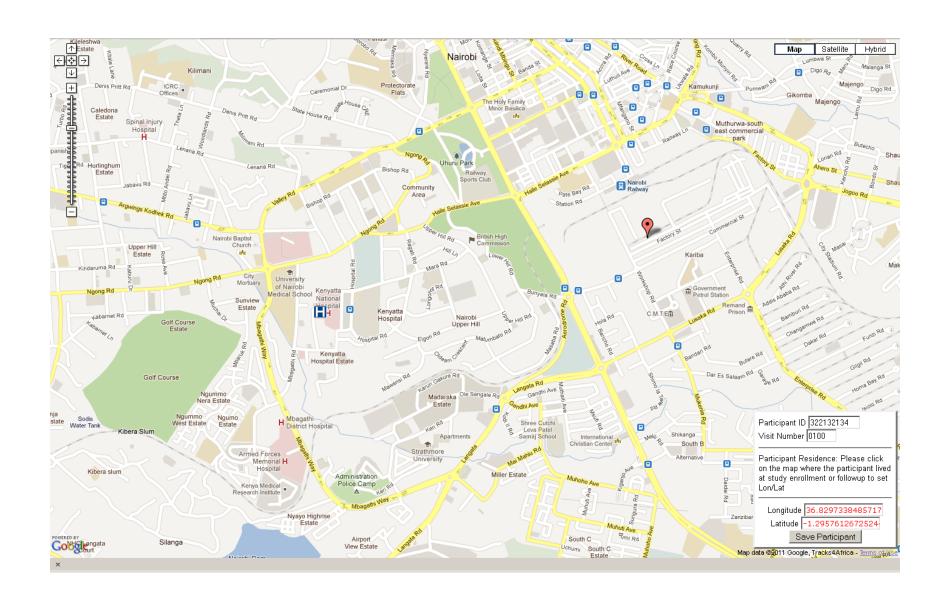
#### Pros

- High degree of flexibility
- Flexible error checking logic
- Excellent user interface
- Accommodates broad range of data types

#### Cons

- High degree of flexibility
- HIPPA compliance not guaranteed
- Setup/administration more complicated
- More challenging data management
- Longitudinal records not possible

# Capturing Geographic Data



## Summary

- Plan your data management system ahead of time
- Tailor data management to the study
- Data entry should be timely, efficient, and accurate
- Data cleaning is critical
- Think carefully about how data will be distributed