# Using Hierarchical Data to Manage Sample and Contact Attempts in Microsoft Access

**Brendan Day** 

University of Wisconsin Survey Center University of Wisconsin-Madison

IFD&TC 2017 May 22, 2017





## Contact attempts differ by mode

- CATI and CAPI surveys
  - Interviewers
    - Telephone calls
    - In-person visits
- Mail and Web surveys
  - Advance letters and postcards
  - Email invitations and reminders
  - Paper questionnaires



#### Mail and Web Surveys

- Self-administered surveys don't administer themselves
  - No interviewers
    - Who initiates contact?
      - Is sample member eligible to be contacted?
      - What address or email should be used?
    - Who records outcome of contact attempt?
      - Is sample member still eligible?
      - Is address or email still usable?
    - What tools do they use?
      - No CATI or CAPI front-end



#### Project Operations Database (POD)



- Microsoft Access database
  - Front-end
    - Case Management
    - Contact Management
    - Progress reports
  - Users
    - Mail Department
    - Tracking & Locating Department
    - Project Directors



#### Project Operations Database (POD)



- Where are data stored?
  - Microsoft Access Back-End
    - Single-mode mail surveys
  - MySQL Back-End
    - In-house web surveys
    - CAPI surveys
    - Central data repository
      - Multi-mode surveys
      - Longitudinal studies
- Open Database Connectivity (ODBC)
  - Link MySQL back-end and Access front-end



## Project Operations Database (POD)



How are data stored?



- How many peas are in a pod?
  - It depends on the variety
    - 2-3 snow peas
    - 6-7 green peas

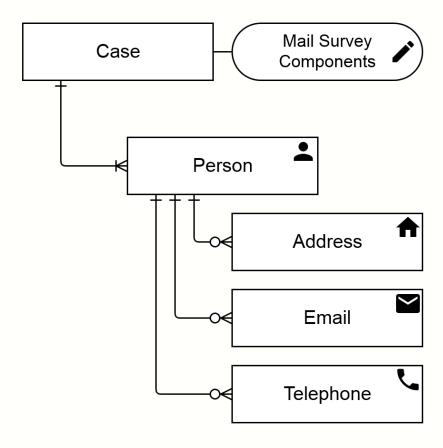


# Hierarchical Data



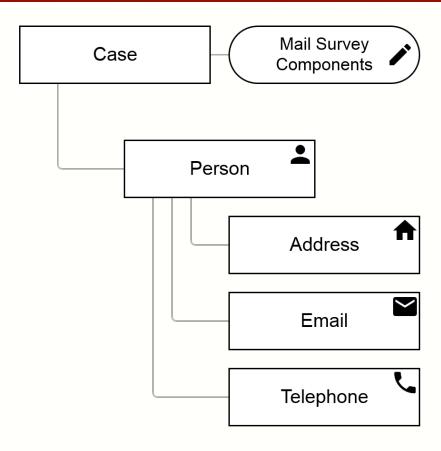


#### Hierarchical Database Structure





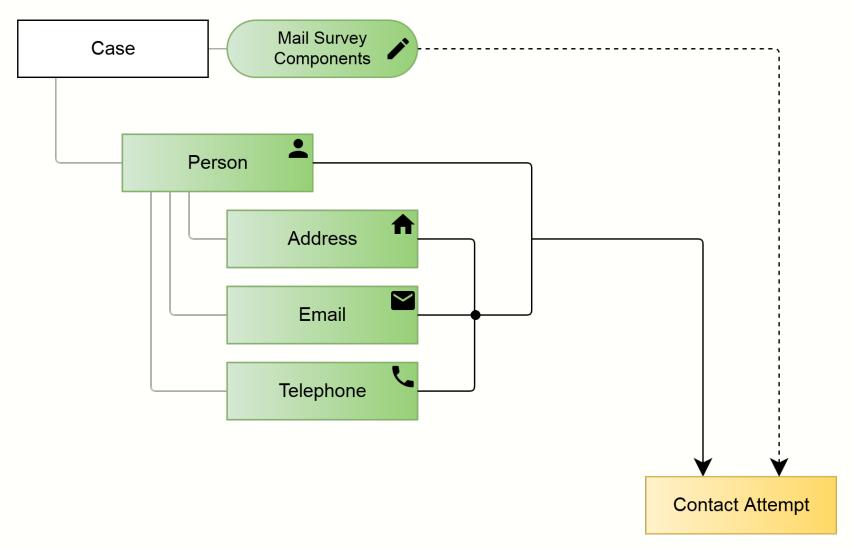
# **Data about Contact Attempts**



Contact Attempt

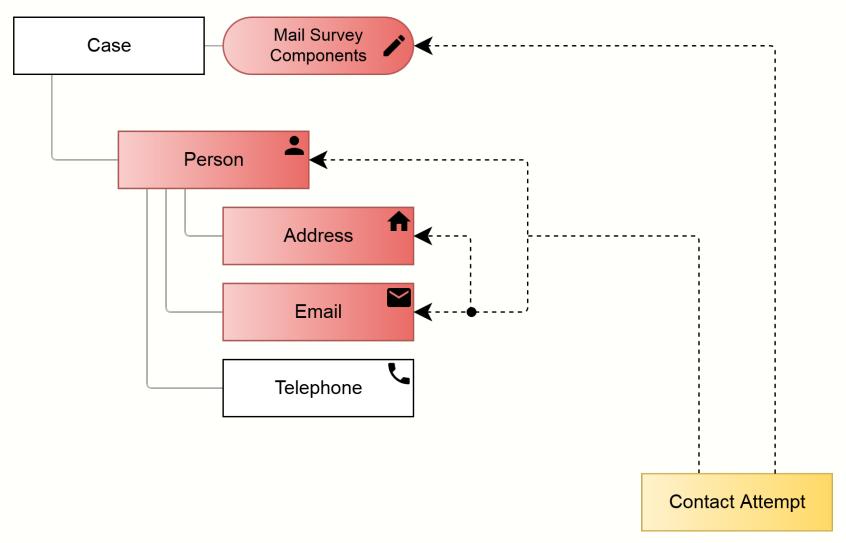


# Disposition Determines Eligibility for Contact





# Outcome of Contact May Affect Disposition





## Disposition and Outcome Code

- Common codes
  - Derived from AAPOR Final Disposition
  - Interim codes for Tracking & Locating
- Code Definitions Table
  - How does code affect disposition of...
    - ♣ People
    - ♠ Addresses

    - Telephone Numbers
    - Mail Survey Components



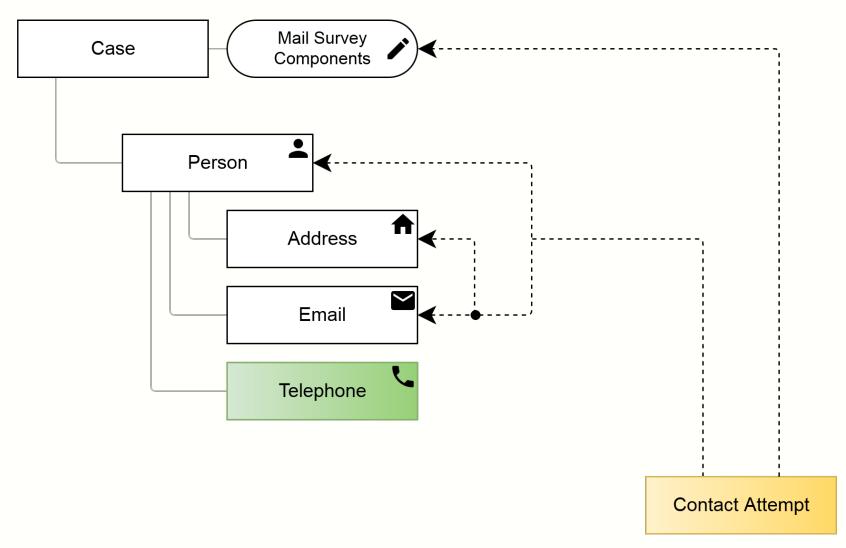
## Disposition



- More than just a code
- Contact Eligibility
  - Enabled
  - Disabled
  - Permanently Disabled (Final)
- Recency
  - When was contact last enabled?
- Tracking and Locating
  - Referred for Tracking & Locating
  - Referred for Review
  - Returned from Department

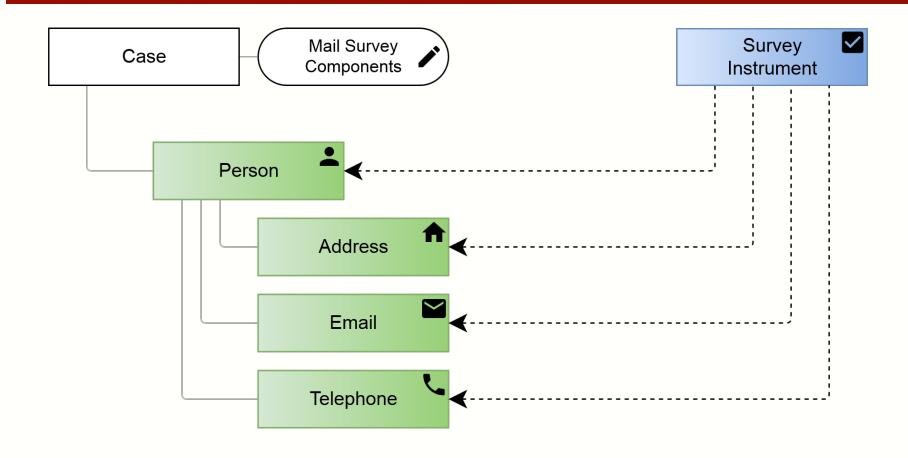


# Telephone as Mode of Contact



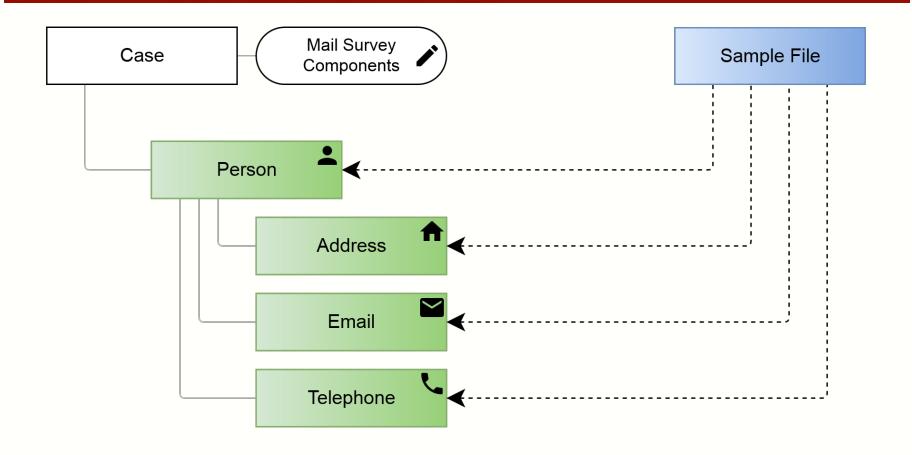


# Data Imported from Survey Instrument





# Data Imported from Sample File





## Why use hierarchical data?

- A case can any number of records
- Records can be of any type
- Each record has its own disposition
- Why not just use "current home address"?
  - When was it current?
    - Possible addresses, not yet confirmed
    - Past addresses, confirmed as incorrect



The Life of a Case
Importing Sample





#### Importing Sample

- Sample Field Definition Table
  - Crosswalk
- Example:
  - Fields in sample file:
    - haddr1, haddr2, hcity, hstate, hzip
  - Fields in database:
    - street1, street2, city, state, zip
  - Type of data: address
  - Type of record: home

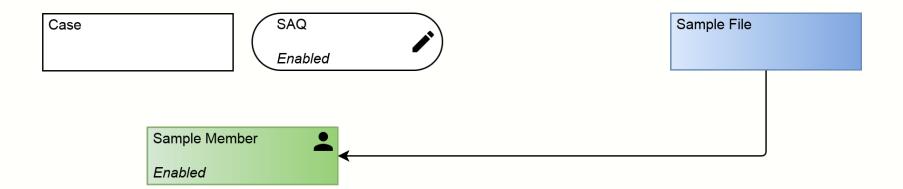


# Case-Level Data Imported into MySQL



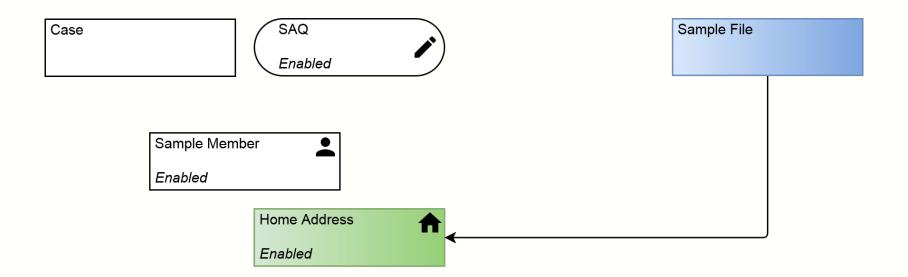


# Person Data Imported



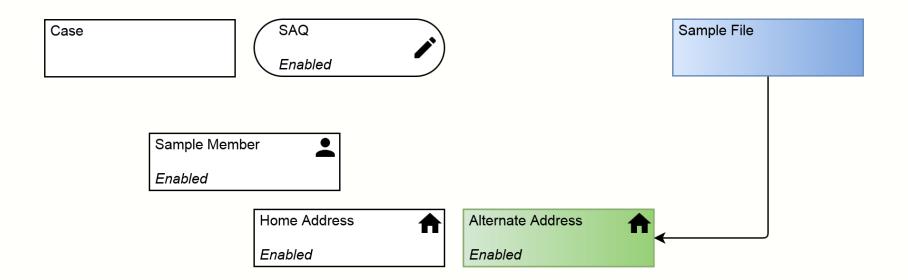


# Home Address Data Imported



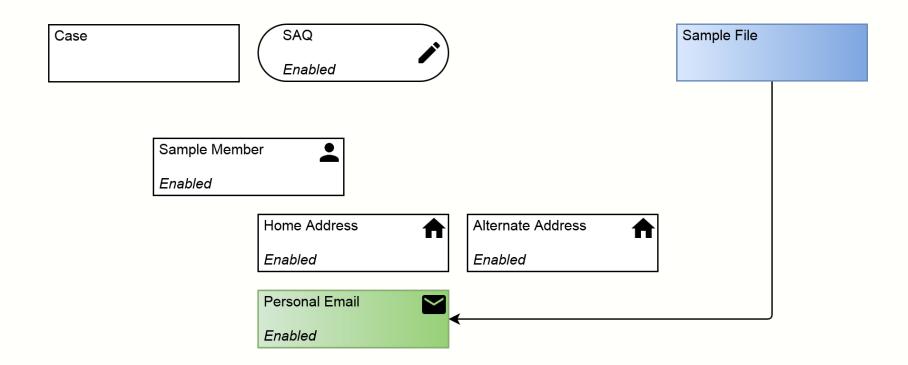


# Alternate Address Data Imported



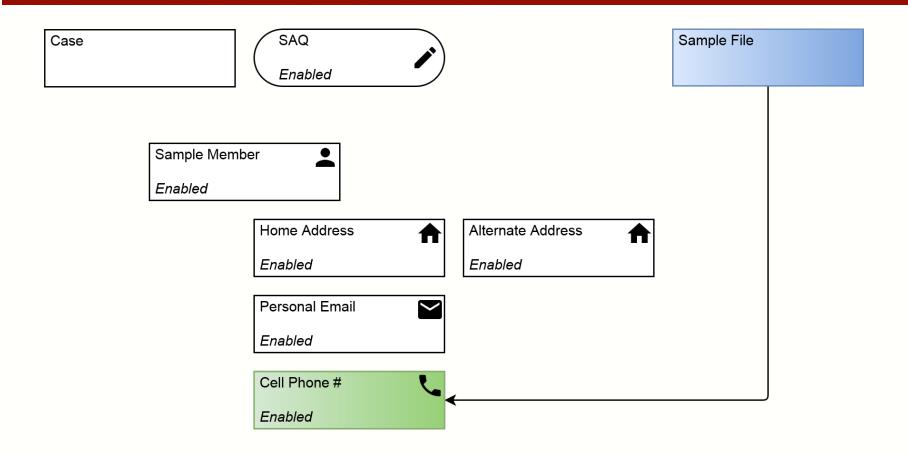


# **Email Data Imported**



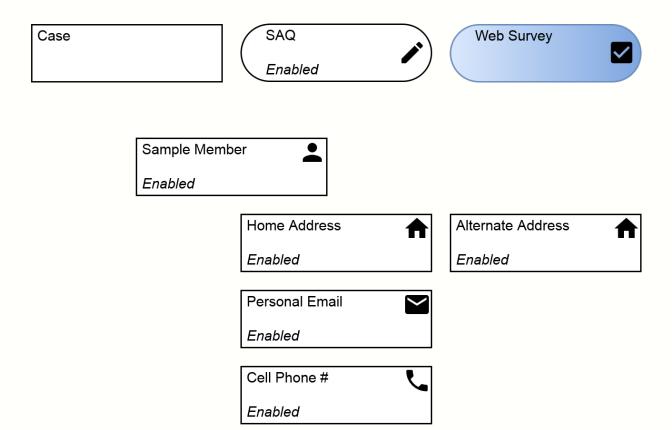


# Telephone Data Imported





# In-House Web Survey





#### **Record Selection**

- Which address, email, or telephone should be used?
- More than one usable address?
  - Prioritize by type
  - Which type of address should we try first?
- Addresses of the same type?
  - Prioritize by recency
  - Which address was most recently enabled?
- Addresses enabled at same time?
  - Prioritize by position
  - Which record was created first?



## Record Type Definitions Table

- Position of each unique record type in hierarchy
  - Which should be selected first?

#### Hierarchy of Address Types:

Home Mailing

Home

Seasonal Mailing

Seasonal

Alternate Mailing

**Alternate** 

Work



The Life of a Case

**Contact Attempts** 



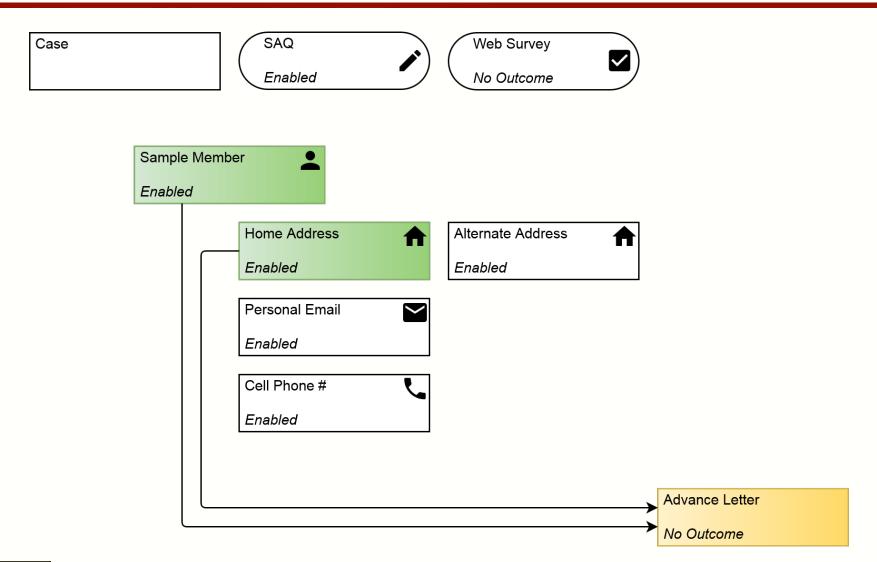


#### Contact Definition Table

- Defines each contact attempt
  - Who is the recipient?
  - What is the mode of contact?
  - Is a mails survey component attached?
  - Any conditions or pre-requisites?
- Other options
  - Resend materials returned undelivered
  - Attach incentives
  - Automate mail merges

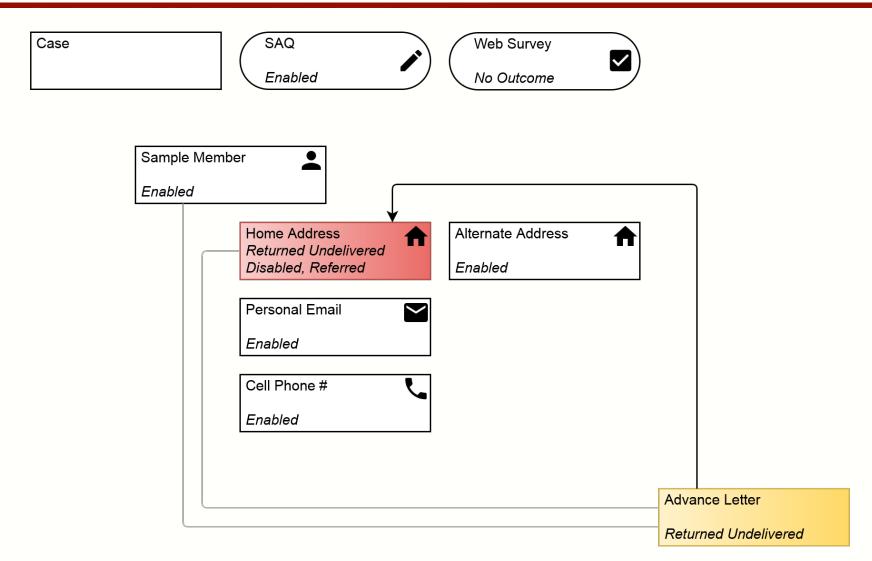


#### Advance Letter Created and Sent



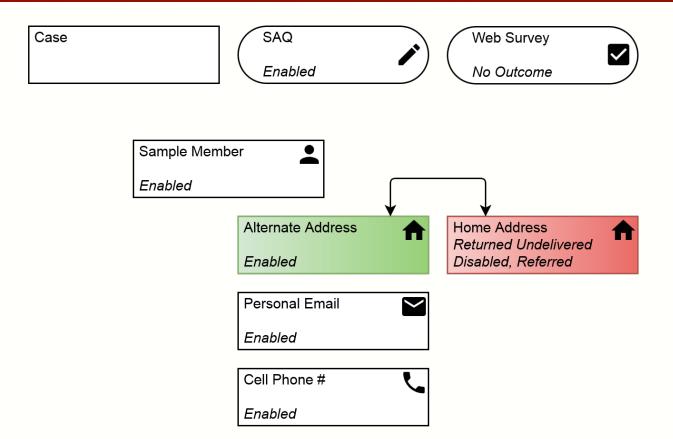


#### Advance Letter Returned Undelivered





#### Alternate Address Selected

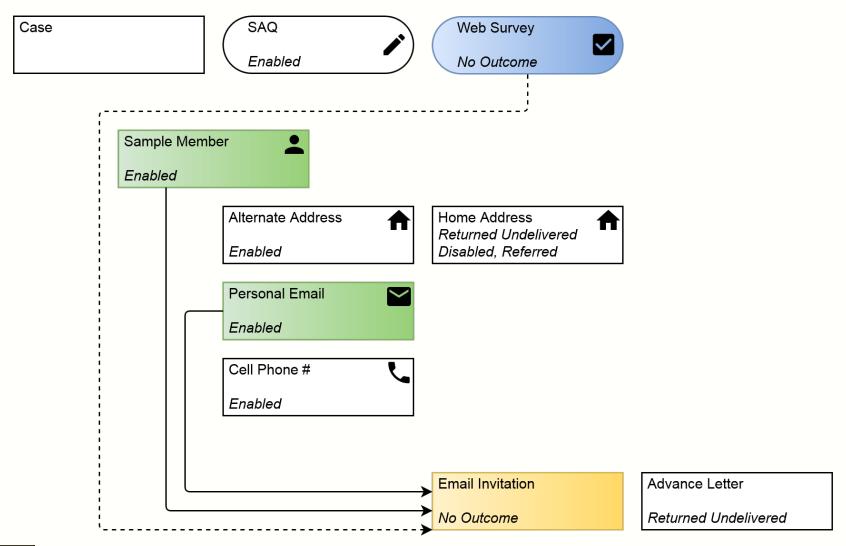


Advance Letter

Returned Undelivered

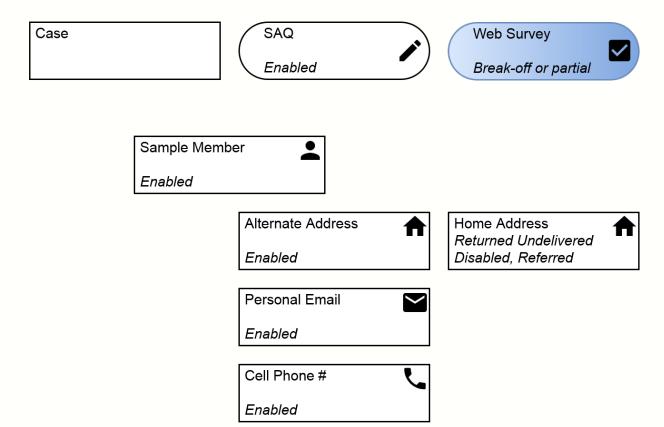


#### **Email Invitation Created and Sent**





# Web Survey Break-off



Email Invitation

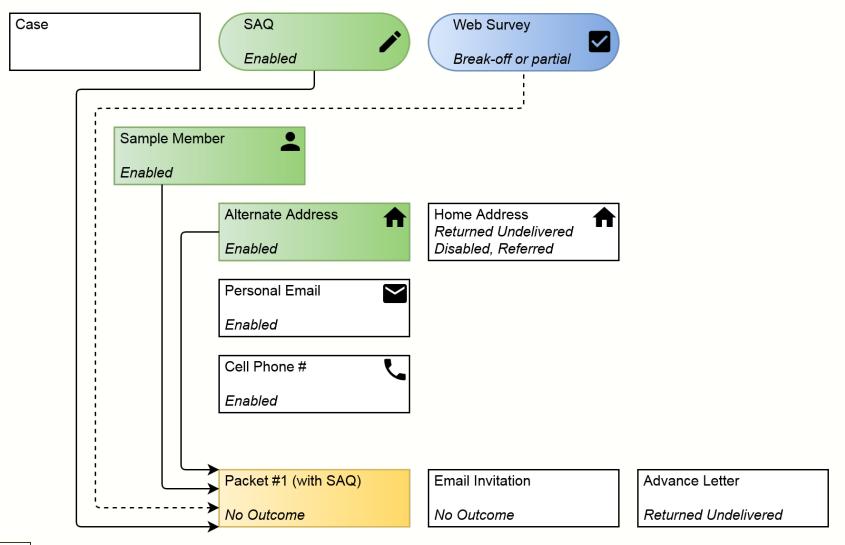
No Outcome

Advance Letter

Returned Undelivered



## SAQ Packet #1 Created





## Assembly and Delivery



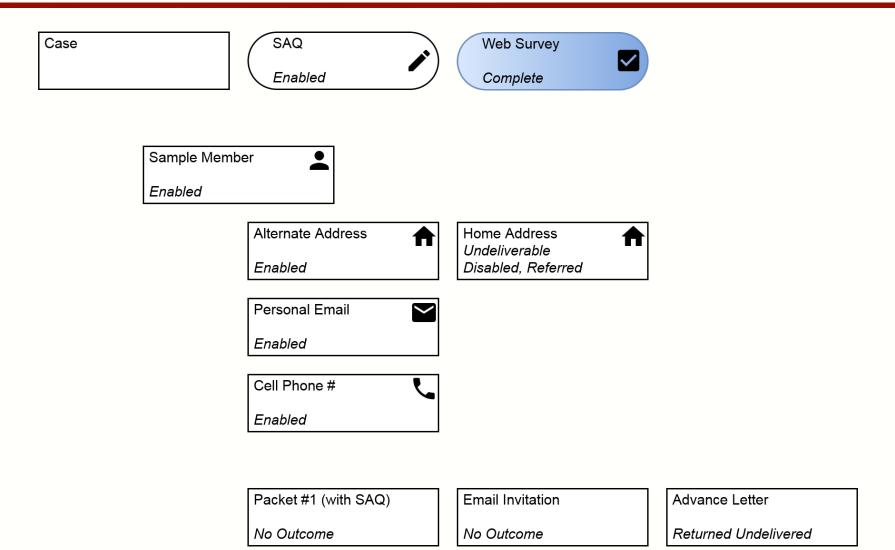
- Contact by mail is not instantaneous
  - **Step 1:** Create contact data records
    - Database evaluates eligibility
  - Step 2: Run mail merge
  - **Step 3:** Print and assemble materials
  - **Step 4:** Deliver materials
    - Recipients may no longer be eligible

#### Solution?

- Database re-evaluates eligibility
- Mail Department removes materials

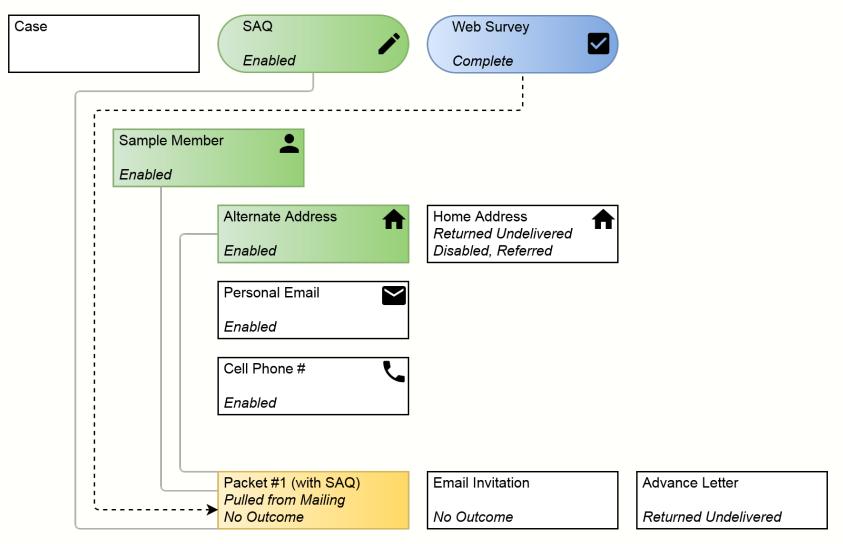


## Web Survey Completed During Assembly



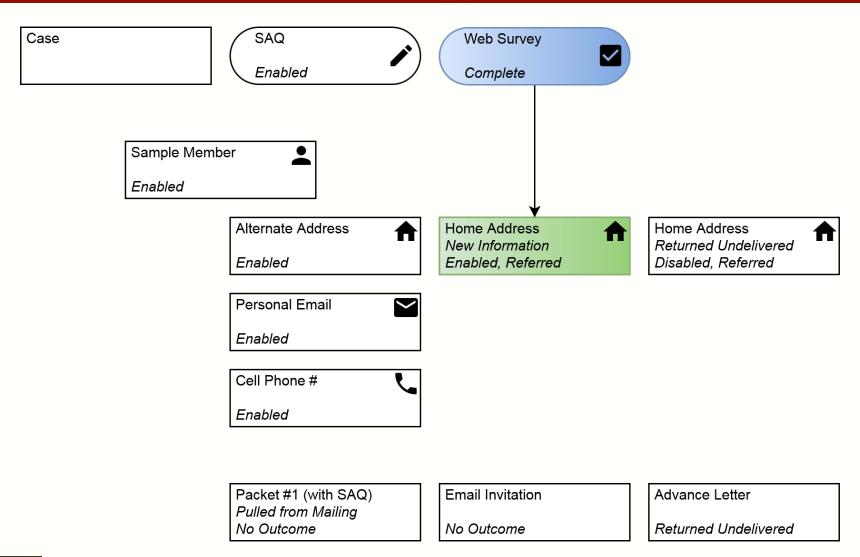


## SAQ Packet #1 Pulled from Mailing



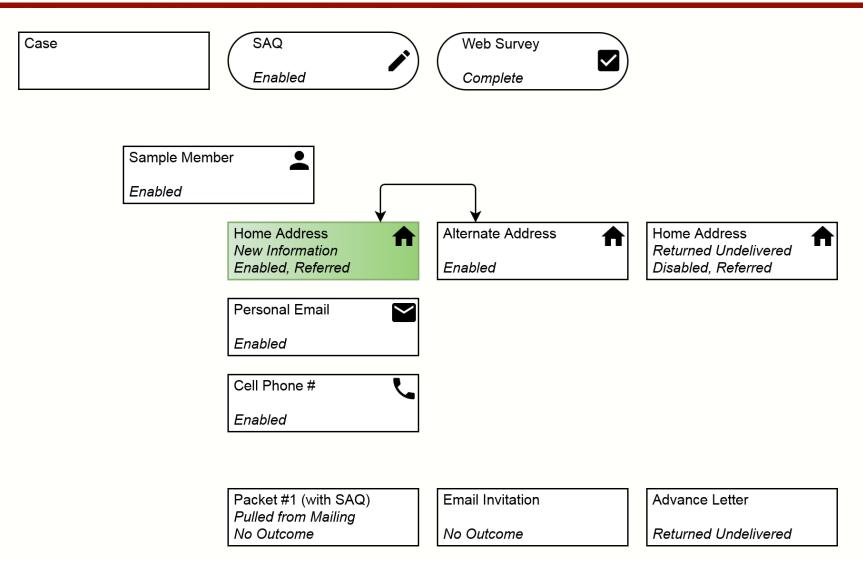


#### New Home Address Imported From Web Survey



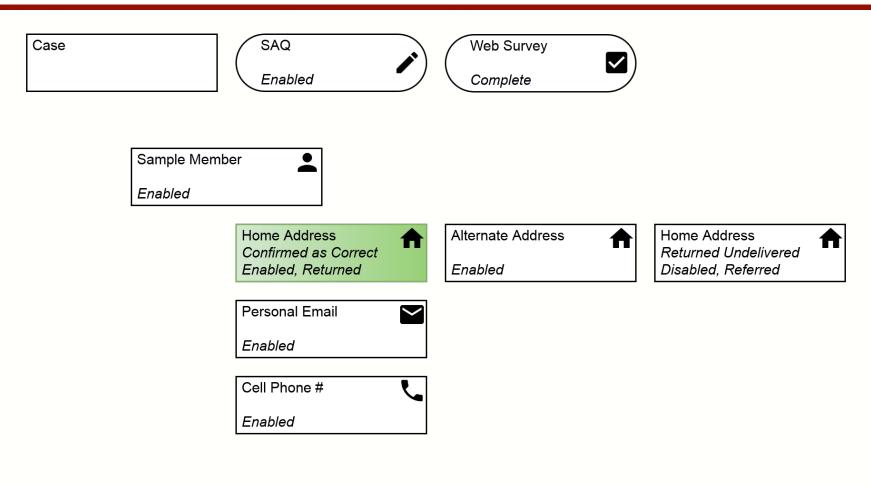


#### New Home Address Selected





#### New Home Address Reviewed



Packet #1 (with SAQ)
Pulled from Mailing
No Outcome

Email Invitation

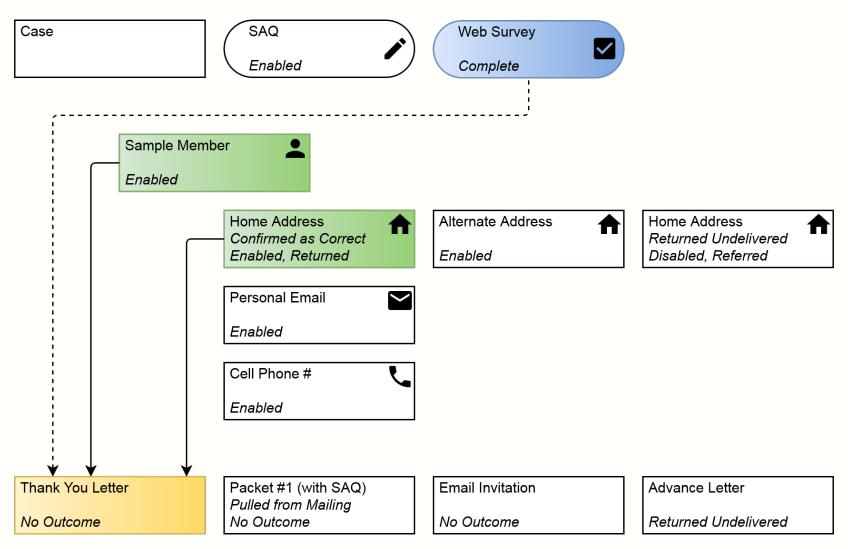
No Outcome

Advance Letter

Returned Undelivered



#### Thank You Letter Created and Sent





# **Data Duplication**



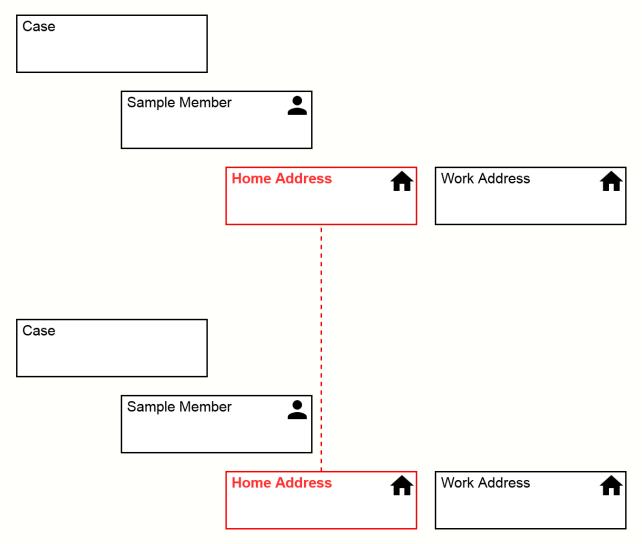


#### What makes the data "hierarchical"

- Each case is self-contained
  - Sample members who live together cannot share the same data
    - Data stored in separate cases
  - What if sample members live together?
    - May share the same address
    - But cannot share the same address record

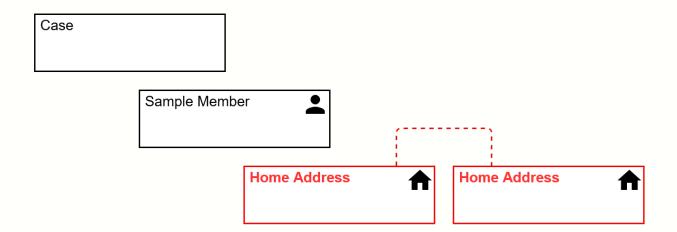


## Address Duplicated Across Cases





## Address Duplicated Within Case





## Why is Data Duplication Problematic?

- Each record has own disposition
  - Ambiguous if recorded twice
    - Which record should we use?
    - Current address or a past address?
    - Confirmed correct or incorrect?
- Loss of functionality
  - Is the address eligible for selection or not?
- Increased complexity for users
  - Which record should be updated?
  - Identifying and disabling duplicates takes time



#### Standardizing Address Data

#### Address #1:

10 West Sunshine Ave AL 12345

#### Address #2:

10 W. Sunshine Avenue Alabama 12345

# Standardized Value:

10WSUNSHINEAVEAL12345

Standardized Value	Original Value
AL	ALABAMA
AK	ALASKA
AZ	ARIZONA
AR	ARKANSAS
ALY	ALLEY
ANX	ANNEX
APT	APARTMENT
ARC	ARCADE
AVE	AVENUE
BSMT	BASEMENT
BYU	BAYOU
ВСН	BEACH
BND	BEND
BLF	BLUFF
ВТМ	воттом
BLVD	BOULEVARD
BR	BRANCH
BRG	BRIDGE
BRK	BROOK
BLDG	BUILDING
BG	BURG
ВҮР	BYPASS
W	WEST



# Hierarchical Data Advantages and Disadvantages





#### Advantages of Hierarchical Data

- Scalable
  - Any number of addresses
  - Projects of varying size and complexity
- Adaptable
  - Different types of addresses
  - Complex or unpredictable sample files
- Comprehensive
  - Old addresses are not overwritten with new



#### Advantages of Hierarchical Data

- Central Data Repository
  - Multiple Departments and Users
    - Mail Department
    - Tracking & Locating Department
    - Project Directors
  - Multiple Applications
    - Email delivery scripts
    - CAPI survey instruments
  - Multi-mode projects
  - Longitudinal studies



## Disadvantages of Hierarchical Data

- Increased complexity for users
  - Each record has its own disposition
  - What is the overall disposition of the case?
    - Multi-mode projects
      - Sample member
      - Mail survey
      - Web survey
- Data duplication
  - Coding and cleaning
  - More prevalent if multiple people per case



#### Thank You!

For copies of this presentation or more information, contact:

Brendan Day brendan.day@wisc.edu

Please visit us at: www.uwsc.wisc.edu









## **Access Database**

## Forms and Tables



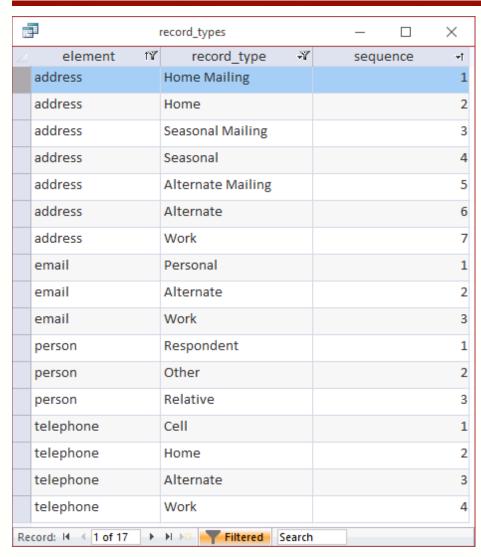


## Sample Field Definitions

l			sample_field_definition	S		- □ ×
sequence 🕣	element 🔻	field_name ⊸7	record_type -	record_source -	sample_field_name -	default_value
100	survey	caseid	Production	sample	caseid	
200	person	person_caseid	Respondent	sample	caseid	
201	person	name_last	Respondent	sample	name_last	
202	person	name_first	Respondent	sample	name_first	
203	person	relationship	Respondent	sample		Respondent
301	address	street1	Home	sample	home_street1	
302	address	street2	Home	sample	home_street2	
303	address	city	Home	sample	home_city	
304	address	state	Home	sample	home_state	
305	address	zip	Alternate	sample	home_zip	
311	address	street1	Alternate	sample	alternate_address	
313	address	city	Alternate	sample	alternate_city	
314	address	state	Alternate	sample	alternate_state	
315	address	zip	Alternate	sample	alternate_zip	
401	telephone	telephone	Cell	sample	cell_phone	
501	email	email	Personal	sample (email1)	email1	
502	email	email	Personal	sample (email2)	email2	
	survey					
	person address					
ord: I4	email	Filtered Search				

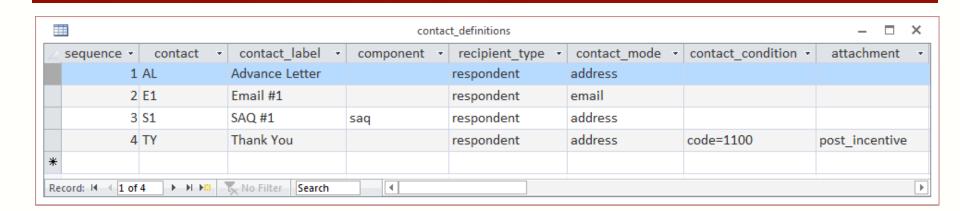


## **Record Type Definitions**





#### **Contact Definitions**





#### **Contact Definition**

Identifier: AL

Label: Advance Letter

Recipient: Sample Member

Mode of contact: Address

Mail Survey Component? No

Conditions or pre-requisites? No \*

\* Default rule for in-house web surveys:

Ineligible if web survey complete (code=1100)

unless overridden



#### **Contact Definition**

Identifier: TY

Label: Thank You

Recipient: Sample Member

Mode of contact: Address

Mail Survey Component? No

Incentive: \$10

Conditions or pre-requisites? code=1100 \*

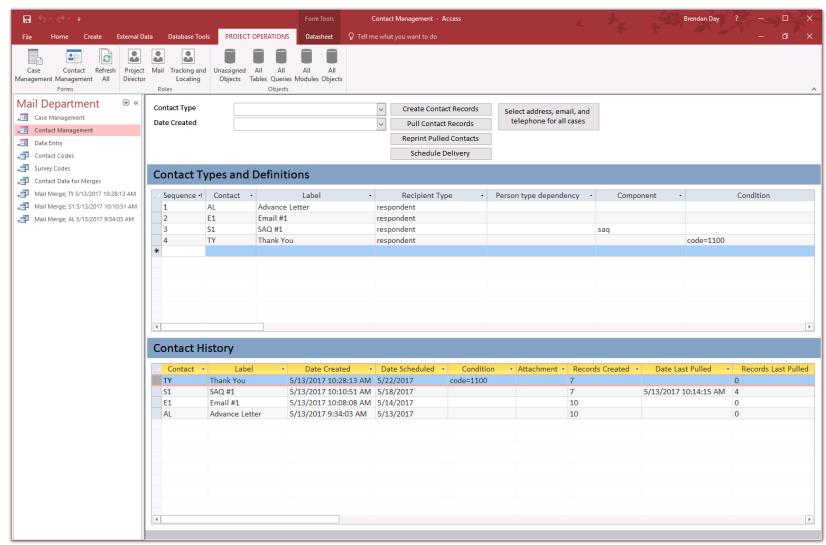
\* Default rule for in-house web surveys:

Ineligible if web survey complete (code=1100)

unless overridden

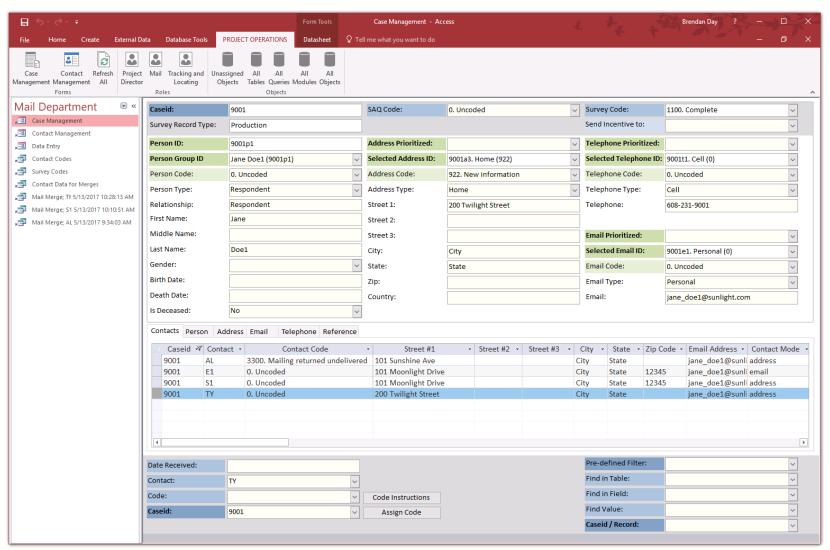


## **Contact Management Form**



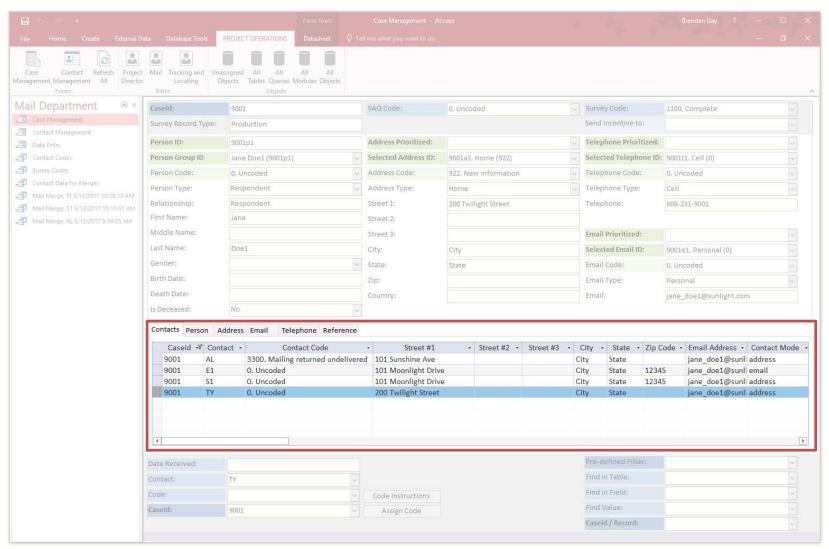


## Case Management Form



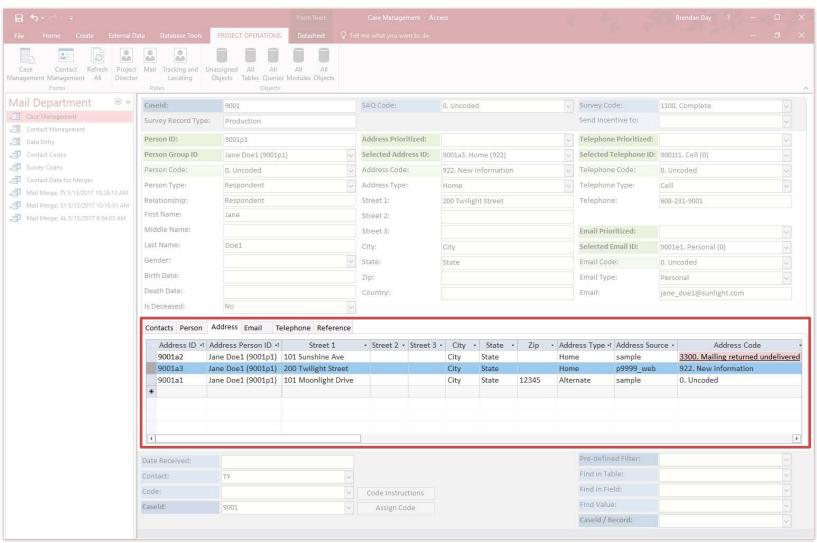


#### Case Management Form (Contact Attempts)





#### Case Management Form (Addresses)





## **Code Definitions**





## **Code Definitions**

<b>.</b> ♠ <b></b>													
Code	Label	Contact	Person		Address			Email			Component		
0	Uncoded	Yes	Yes			Yes			Yes			Yes	
1100	Complete	Yes	No			No			No			Yes	Finalized
2110	Explicit refusal	Yes	Yes	Finalized		No			No			Yes	Finalized
2111	Other person refusal	Yes	Yes	Finalized		No			No			Yes	Finalized
2112	Known respondent-level refusal	Yes	Yes	Finalized		No			No			Yes	Finalized
2113	Blank questionnaire mailed back, implicit refusal	Yes	No			No			No			Yes	Disabled
2260	Respondent was unavailable during field period	Yes	Yes	Finalized		No			No			Yes	Finalized
2270	Completed, but not returned during field period	Yes	No			No			No			Yes	Finalized
2310	Deceased	Yes	Yes	Finalized		No			No			Yes	Finalized
2320	Physically or mentally unable or incompetent	Yes	Yes	Finalized		No			No			Yes	Finalized
2330	Language barrier	Yes	Yes	Finalized		No			No			Yes	Finalized
2400	Mailing forwarded by USPS with address update	Yes	No			Yes	Disabled		No			No	
3190	Nothing ever returned	Yes	No			No			No			Yes	
3191	Email returned with automatic response	Yes	No			No			Yes			No	
3300	Mailing returned undelivered	Yes	No			Yes	Disabled	Refer	Yes	Disabled	Refer	Yes	
3313	No such address	Yes	No			Yes	Disabled	Refer	No			Yes	
3314	Vacant	Yes	No			Yes	Disabled	Refer	No			Yes	
3400	Mailing returned with forwarding information	Yes	No			Yes	Disabled		Yes	Disabled		Yes	
4700	No eligible respondent	Yes	Yes	Finalized		No			No			Yes	Finalized
4910	Duplicate complete (survey already complete)	Yes	No			No			No			Yes	Finalized
4911	Duplicate complete (mail survey already complete)	Yes	No			No			No			No	
8000	Request resend or reprint	Yes	No			No			No			No	



## **Code Definitions**

<b>.</b>													
Code	Label	Contact	Person		Address			Email			Component		
910	Refer to tracking and locating	No	Yes		Refer	Yes		Refer	Yes		Refer	No	
911	Refer for tracing review	No	Yes		Review	Yes		Review	Yes		Review	No	
920	No new information	No	Yes		Return	Yes		Return	Yes		Return	No	
921	Confirmed as correct	No	Yes	Enabled	Return	Yes	Enabled	Return	Yes	Enabled	Return	No	
922	New information	No	Yes	Enabled	Return	Yes	Enabled	Return	Yes	Enabled	Return	No	
931	Confirmed as incorrect	No	Yes	Disabled	Return	Yes	Disabled	Return	Yes	Disabled	Return	No	
942	New information, Review	No	Yes	Enabled	Review	Yes	Enabled	Review	Yes	Enabled	Review	No	
951	Confirmed as incorrect, Refer	No	Yes	Disabled	Refer	Yes	Disabled	Refer	Yes	Disabled	Refer	No	
1200	Partial or break-off with sufficient information	No	No			No			No			Yes	Disabled
2100	Break-off or partial with insufficient information	No	No			No			No			No	
2121	Logged on to survey, did not complete any items	No	No			No			No			No	
4100	Selected respondent screened out of sample	No	Yes	Finalized		No			No			Yes	Finalized
4900	Duplicate listing	No	Yes	Finalized	Return	Yes	Finalized	Return	Yes	Finalized	Return	No	
9001	Enable contact	No	Yes	Enabled		Yes	Enabled		Yes	Enabled		Yes	Enabled
9002	Disable contact	No	Yes	Disabled		Yes	Disabled		Yes	Disabled		Yes	Disabled



#### Code 3300: Returned Undelivered

#### Contact Outcome? Yes

- Applicable to Person? No
- Applicable to Address? Yes

  Effect on Address: Disable, Refer for Tracking & Locating
- Applicable to Telephone? No



#### Code 920: No New Information

- Applicable to Person? Yes
  Effect on Person: Return from Tracking & Locating
- ♠ Applicable to Address? Yes
  Effect on Address: Return from Tracking & Locating
- Applicable to Telephone? Yes
  Effect on Telephone: Return from Tracking & Locating



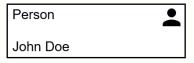
# **Entity Association Records**





#### Non-Hierarchical Data







Address	A
123 Sunshine Drive	

Person \_\_\_\_\_
Jane Doe





## **Entity Association Records**

