U.S. Department of Energy Smart Grid Investment Grant Technical Advisory Group Guidance Document #11

Topic: CBS Data Reporting Process

September 5, 2012

U.S. Department of Energy Smart Grid Investment Grant

Technical Advisory Group Guidance Document #11[*]

Topic: CBS Data Reporting Process
September 5, 2012

OBJECTIVE

The Department of Energy (DOE) set the expectation early on in the SGIG Funding Opportunity Announcement (FOA) that recipients who undertake a rigorous consumer behavior study (CBS) would be obliged to collect and report highly granular customer-level consumption and demographic data (hereafter referred to as "Project data"). This guidance document provides information on the data that is to be reported to the SmartGrid.gov Data Hub. The submittal of this data is consistent with the reporting requirements of the Assistance Agreement, as provided within Attachment B, Federal Assistance Reporting Checklist and Instructions.

The first part of this document describes the process through which the data should be submitted into the database repository in the section titled "Process Overview for Submission and Acceptance of CBS Data". The second section of this document, entitled "Data Reporting Categories and Time Frame", lists the different categories of datasets that should be submitted and provides a submittal deadline for each category. The third section of this document, entitled "CBS Data Dictionary", provides a comprehensive description of each variable within every dataset that should be submitted.

"CBS Data Dictionary", provides a comprehensive description of each variable within every dataset that should be submitted.
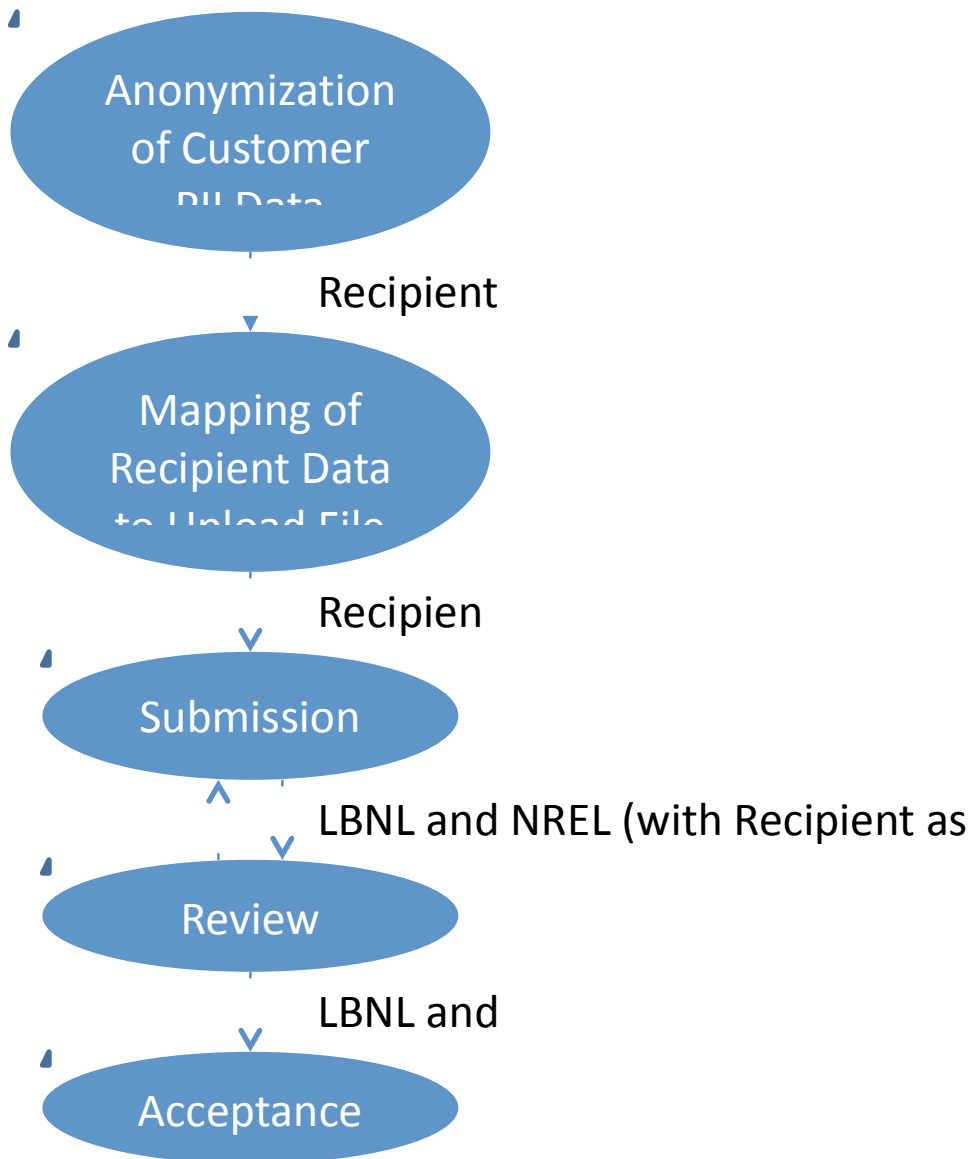
BACKGROUND

DOE provided a preliminary data dictionary in Appendix D of the Metrics and Benefits Guidebook, which expanded the data reporting requirement for utilities undertaking rigorous consumer behavior studies to include the highly granular customer-level Project data as well as more aggregate metrics produced by each recipient's own study evaluation effort. DOE then provided a guidance document that described the totality of data that should be *collected* by SGIG recipients undertaking consumer behavior studies, hereafter referred to as CBS data.[1]

This document describes the process by which this CBS data is to be reported, including a data submittal time frame, a description of the way in which the Project data is to be anonymized and uploaded to the National Renewable Energy Laboratory's Data Hub, and a CBS data dictionary.

PROCESS OVERVIEW FOR SUBMISSION AND ACCEPTANCE OF CBS DATA

The process for recipients submitting CBS data and having it accepted into the database repository is shown in the diagram below. The bubbles represent the various steps in the process and the names listed between the processes represent the party responsible for executing that process. For example, the SGIG recipient is responsible for anonymizing the customer Personally Identifiable Information (PII) data, whereas Lawrence Berkeley National Laboratory (LBNL) and National Renewable Energy Laboratory (NREL) are jointly responsible for reviewing and accepting the CBS data submitted by each SGIG recipient The steps of the process are described in more detail below, but are generally similar to the data submission and acceptance processes already used by recipients for other SGIG data, such as build metric data. The primary differences for CBS data, relative to build metric data, are the anonymization of PII step and the format of the data submission files.

---

[1] See Guidance Document #10: Consumer Behavior Study Data Collection Requirements for more details

Anonymization of Customer PII Data

↓ Recipient

Mapping of Recipient Data to Upload File

↓ Recipien

Submission

↕ LBNL and NREL (with Recipient as

Review

↓ LBNL and

Acceptance

Description of steps:

1. **Anonymization of Customer PII Data:** Recipients will anonymize PII through either their own methods or through a web service provided by NREL. In order to respect privacy concerns, no raw PII data will ever leave the recipient's system. Anonymization of customer PII will allow the recipient to submit data records without PII, but still allow for geographic and network topological analysis of the data at a customer level.

a. Recipient's own anonymization method: SGIG recipients with expertise in data anonymization may develop their own algorithms to remove any PII from the submitted data, subject to the proviso that the methods produce a data record that conforms to the CBS data reporting templates.

b. NREL's web service anonymization method: NREL has expertise in data anonymization and developing algorithms to establish differential privacy in databases. The same algorithms used to protect financial, medical and credit card data in commercial transactions and databases will be applied to this CBS data in order to safeguard individual customer privacy.

2. **Mapping of Recipient Data to Upload File:** SGIG recipients likely have stored the various elements of CBS data in a different format and in various locations across their IT systems. As such, an algorithm will have to be developed that translates the data from the format it is stored on the SGIG recipient's IT systems to the format associated with the CBS data reporting template. To facilitate this algorithm construction, a database dictionary is described below for the CBS data reporting template. The database dictionary details and describes the data to be submitted. NREL and LBNL will work with the recipients to provide tools and external database access to populate data fields that the recipient is not capable of populating.

3. **Submission:** The recipients will submit data in the CBS data reporting template file format, defined in this document and its appendices (including two excel spreadsheets). When the data submission is received by NREL, the recipient will receive an e-mail verifying submission of the data. Multiple secure methods of data submission (e.g. FTP, web service, physical mailing of hard drives, etc.) will be supported. Each recipient will work with NREL to determine the most applicable and acceptable method.

4. **Review:** Data checks will be performed by NREL and LBNL to validate the quality, continuity and completeness of the submitted CBS data. If deficiencies are identified in the submitted CBS data, the recipient will be notified by e-mail of the specific nature of the deficiency. NREL and LBNL will work with the recipient to resolve the deficiencies in a timely manner. Once these deficiencies are corrected, the recipient will need to re-submit corrected data files. Once the submitted data files are reviewed and no deficiencies are identified, the data files will move to the accepted state.

5. **Accepted:** When the submitted data files are accepted, the recipient will receive an acceptance e-mail. At this point the submitted data will be loaded into a secure database for subsequent analysis.

DATA REPORTING CATEGORIES AND TIME FRAME

There are 12 different categories of CBS data that will need to be reported by recipients, indicated by different data upload files listed in the table below and described in more detail in the following section. Each CBS data category may have up to three submittal deadlines as indicated in the table below: within six months of assigning customers to experimental cells; within two months of submitting an interim evaluation report to DOE (if applicable); and within two months of submitting a final evaluation report to DOE.[2] Each data category should be submitted in its totality by the required deadline even if there have been no changes to the data since the previous required deadline. Each recipient will be asked to submit the entirety of their required and applicable CBS data to the Data Hub in a form and media mutually agreed upon. For an extensive description of each dataset, see the next section, "CBS Data Dictionary".

Upload File Name (Data Category)	Data Level	Description	Submittal Deadline for Recipient to Provide Data		
			Within 6 months of customer assignment	Within 2 months of submitting interim evaluation	Within 2 months of submitting final evaluation
Exp Cell	Utility	For each experimental cell (i.e., each control and treatment group), a description of the rates, technology, and group requirements	●	●	●
Recruitment Data	Customer	Each recruitment contact made to each customer	●	●	●
Assignment Data	Customer	Each customer's assignment into each experimental cell (e.g., date of assignment, eligibility, etc.)	●	●	●

---

[2] Given the diversity of schedules that the different consumer behavior studies are on, some/all of these reporting deadlines may have already passed. In that case, the recipient should attempt to submit the data as soon as possible.

Survey Admin	Utility	A list of all of the different surveys administered	●	●	●
Survey Results	Customer	Each customer's response to each survey	●	●	●
Tariff Data	Utility	A list of all of the different tariff rates that occurred during and 12 months prior to the study	-	●	●
Customer Tariff Data	Customer	The tariff rates experienced by each customer during and 12 months prior to the study	-	●	●
Event Data	Utility	A list of all of the events that occurred during and 12 months prior to the study	-	●	●
Cust Event Note	Customer	The events experienced by each customer during and 12 months prior to the study	-	●	●
Billing Data	Customer	Monthly billing information for each customer during and 12 months prior to the study	-	●	●
Time Series	Customer	Hourly energy usage data for each customer during and 12 months prior to the study	-	●	●
Document Files	Utility	Copies of all applicable documents relevant to the study (e.g., tariff sheets, marketing collateral, etc.)	●	●	●
Evaluation Data	Utility	List of the recipient's evaluation efforts	-	●	●

CBS Data Dictionary

This section provides a comprehensive description of all of the data elements that are required to be reported by SGIG recipients undertaking consumer behavior studies (for an overview of all of the different datasets, see the previous section, "Data Reporting Categories and Time Frame"). The accompanying appendices, "Non Time Series Data File Formats.xlsx" and "Time Series Data File Formats.xlsx", provide examples of how the data should be organized to allow the CBS data to be loaded into the structured database.

The datasets were categorized into 13 different files in order to minimize the amount of repetitive data (e.g., each customer's treatment group is listed only once rather than being listed in conjunction with each hour of data provided for that customer). Many of the variables that are listed in the sections that follow were created to accommodate a wide variety of different types of study implementations. However, for each particular recipient, many of the data points may be identical to each other (e.g., the date that a technology is installed may be the same date that the technology is activated). NREL and LBNL will be available to discuss the data requirements with each recipient.

The data elements reported herein are all directly related to the current vision for the cross-project CBS evaluation effort LBNL and its subcontractors will be undertaking on behalf of DOE. For each table, the columns on the left indicate the reason that each variable is important to collect for the CBS evaluation effort. Ideally every variable would be provided by an SGIG recipient undertaking a rigorous consumer behavior study. However, we understand that it may be infeasible or impractical to report some of the data listed. At a minimum, variables listed as "Essential" and "Important" are required for data analysis purposes and should be reported by each SGIG recipient undertaking a rigorous consumer behavior study. If any difficulties arise, LBNL will work with each recipient on a case-by-case basis to determine how best to meet the SGIG reporting requirement of CBS data for that recipient's particular situation.

Analysis

These variables are needed to perform analysis at the utility level in order to answer research questions of interest for specific utilities

● Essential — Necessary in order to answer primary policy questions about the effectiveness of time based rates or enabling technologies (submission of these variables is **required**)

◉ Important — Necessary in order to answer secondary research questions about whether a technology or a rate had a larger effect for different selected customer demographic groups (submission of these variables is **required**)

○ Useful — Necessary to answer tertiary policy questions of interest (e.g., why some customers responded to rates while others did not, if some event notifications were more successful (e.g., phone notifications vs. email notifications) (submission of these variables is strongly encouraged)

Robustness

These variables are needed to perform robustness checks on the utility level analysis (i.e., make sure that the estimated results from the analysis are valid)

● Essential — Necessary for data cleaning and robustness checks (submission of these variables is **required**)

○ Useful — Nice to have for additional robustness checks (submission of these variables is strongly encouraged)

Meta Analysis

These variables are needed to group the utility level results together into a meta analysis in order to answer research questions of interest across all utilities

● Essential	Necessary to perform any cross cutting meta analysis across utilities (submission of these variables is **required**)
○ Useful	Nice to have for additional meta analyses (submission of these variables is strongly encouraged)

Exp Cell Data Upload File (Description of Control and Treatment Groups)

The purpose of the data in this file is to list and describe for the various important rate, technology and information elements that represent each experimental cell (i.e., each control group, treatment group, and group of customers not in the study).

Data for this file should be uploaded for each of the following:

- Each Experimental Cell
  - Each control group
  - Each treatment group
  - Each group of customers not in the study
- Time Periods for Tariff data
  - Historical Data (12 months prior to beginning of treatment)
  - Data throughout the study duration
- There should be one entry for each experimental cell (i.e., each control group, treatment group, and group of customers not in the study)
- Unique key fields for this file are denoted by an asterisk*

9

Analysis	Robust	Meta	Variable Name	Description	Type	Unit	Enumeration
●	●	●	ProjectID*	The SGIG project ID number	ENUM		
●	●	●	ExpCellID*	Name of each control group, treatment group, and group of customers not in the study (e.g., C1, C2, T1, Cntl1, A1, B2, N1)	ALPHANUMERIC		
–	○	–	ExpCellStartDate	The first date that any customer was exposed to the treatment specific to this experimental cell (the rates, technology, and/or information given below)	DATE		
–	○	–	ExpCellEndDate	The last date that any customer was exposed to the treatment specific to this experimental cell (the rates, technology, and/or information given below)	DATE		
●	–	–	TariffID1	The first tariff rate (as defined in the Customer Tariff Data Upload File) that customers in this group are placed on (e.g., their rate 12 months prior to the beginning of the study)	ALPHANUMERIC		

	Field	Description	Type
– –	TariffID2	The second tariff rate (as defined in the Customer Tariff Data Upload File) that customers in this group are placed on, if applicable	ALPHANUMERIC
– –	TariffID3	The third tariff rate (as defined in the Customer Tariff Data Upload File) that customers in this group are placed on, if applicable	ALPHANUMERIC
– –	TariffID4	The fourth tariff rate (as defined in the Customer Tariff Data Upload File) that customers in this group are placed on, if applicable	ALPHANUMERIC
– –	TariffID5	The fifth tariff rate (as defined in the Customer Tariff Data Upload File) that customers in this group are placed on, if applicable	ALPHANUMERIC
– ● ●	HasBillProtection	Bill protection is provided for this group at some point during the study	BOOLEAN **BOOLEAN Options:** YES NO

Field	Description	Type	Example
BillProtectionDesc	Description of bill protection given to this group	**ALPHANUMERIC**	
EducationID	ID number of the education given to this group. Each EducationID should be identical to the name, excluding the extension, of the corresponding education material file uploaded in the "Document Upload File" section.	**ALPHANUMERIC**	e.g. mailA_2012 mailB_2012 emailA_2013
EducationDescr	Description of education given to this group	**ALPHANUMERIC**	
InfoTechID	ID number of the informational technology given to this group. Each InfoTechID should be identical to the name, excluding the extension, of the corresponding informational technology description file uploaded in the "Document Upload File" section.	**ALPHANUMERIC**	e.g. IHD_1 IHD_2
InfoTechDescr	Description of informational technology given to this group	**ALPHANUMERIC**	
ControlTechID	ID number of the control technology given to this group. Each ControlTechID should be identical to the name,	**ALPHANUMERIC**	e.g. PCT_1

	Description	Type	Options
● -- -- HasUtilityControl	The control technology given to this group has utility control	**BOOLEAN**	BOOLEAN Options: YES NO
	excluding the extension, of the corresponding control technology description file uploaded in the "Document Upload File" section.		
● -- -- ControlTechDescr	Description of control technology given to this group	**ALPHANUMERIC**	
-- ○ ○ RequiresEligibleAMI	Assignment to this group requires AMI	**BOOLEAN**	BOOLEAN Options: YES NO
-- ○ ○ RequiresEligibleHistoricalUsage	Assignment to this group requires minimum amount of historical usage	**BOOLEAN**	BOOLEAN Options: YES NO

			BOOLEAN	BOOLEAN Options:
— ○	RequiresEligibleTariff	Assignment to this group requires customer to currently be on an eligible tariff rate		YES NO
— ○	RequiresEligibleInternet Access	Assignment to this group requires customer to currently have an eligible type of internet access	BOOLEAN	BOOLEAN Options: YES NO
— ●	RequiresNonParticipatio n	Assignment to this group requires non-participation in an existing DR program	BOOLEAN	BOOLEAN Options: YES NO
— ○	RequiresEligibleNonUtili tyEmployee	Assignment to this group requires customer to NOT be a utility employee to participate	BOOLEAN	BOOLEAN Options: YES NO
— ●	RequiresEligibleCentralA C	Assignment to this group requires customer to currently have central air conditioning	BOOLEAN	BOOLEAN Options: YES NO

– ○ ● RequiresEligibleWaterHeater	Assignment to this group requires customer to currently have an eligible water heater	BOOLEAN	BOOLEAN Options: YES NO
– ○ ● RequiresEligiblePoolPump	Assignment to this group requires customer to currently have an eligible pool pump	BOOLEAN	BOOLEAN Options: YES NO
– ○ ● RequiresEligibleYearRoundResident	Assignment to this group requires customer to currently be a year-round resident	BOOLEAN	BOOLEAN Options: YES NO
– ○ ○ RequiresEligibleStudyDuration	Assignment to this group requires customer to be a resident for the duration of the study	BOOLEAN	BOOLEAN Options: YES NO
– ○ ○ RequiresEligiblePaymentStanding	Assignment to this group requires customer to currently be in eligible payment standing	BOOLEAN	BOOLEAN Options: YES NO

☐ ○ ○	RequiresEligibleAddress Match	Assignment to this group requires customer to have their service and billing address match	**BOOLEAN**	BOOLEAN Options: YES NO
☐ ○ ●	RequiresEligibleDwelling Type	Assignment to this group requires customer to currently be in an eligible dwelling type	**BOOLEAN**	BOOLEAN Options: YES NO

Recruitment Data Upload File (Recruitment Contact for Each Customer)

The purpose of the data in this file is to list and describe each recruitment contact made to each customer. In order to better understand how customers accept AMI-enabled time-based rate programs and control/information technology, DOE would like each recipient to report information about their recruitment effort.

This data will be used to determine whether customers respond differentially to different types of marketing and recruitment efforts. It is possible that even subtle differences in wording can lead to different response rates. By having access to all recipients' marketing and recruitment materials, DOE can examine differences in tactics across recipients, which will hopefully provide insights on which types of approaches are most effective.

In addition, a copy of all marketing collateral should be uploaded in the "Document Upload File" section. Each CollateralID used below should be identical to the name, excluding the extension, of the corresponding marketing collateral file uploaded in the "Document Upload File" section.

Data for this file should be uploaded for each of the following:

- Customers
  - Each treatment and control customer
  - A representative sample of those who were invited to participate in the study but declined the offer (each recipient should work with their TAG to determine the most appropriate size of the representative sample)
- Frequency
  - Each recruitment contact
- There should be one entry for each contact for each customer
- Unique key fields for this file are denoted by an asterisk*

	Variable Name	Description	Type	Unit	Enumeration
Analysis Robust Meta					
● ● ●	CustomerID*	Unique customer identifier	ALPHANUMERIC		
● ● ●	UsagePointID*	Geographic identifier for this customer's location	ALPHANUMERIC		
● ● ●	UsagePointType	The type of geographic identifier used for this customer's location. Census block data should be the level of geographic identifier reported. If census block data is not readily available, NREL can provide a web service that a recipient can	ENUM		**ENUM Options:** CENSUS BLOCK CENSUS BLOCK GROUP CENSUS TRACT ZIP CODE PLUS FOUR ZIP CODE

use on their internal system to convert an address into a census block level geographic identifier. If a recipient has concerns about reporting census block data, LBNL will work with that recipient to determine the most appropriate level of geographic identifier (e.g., census block group, zip code, etc.).

● ● ● ● ● ProjectID* Unique SGIG project ID number ENUM

● ● ● ContactID* Unique recruitment wave identifier. If a customer is contacted multiple times, there should be one entry for each contact. (E.g., if a customer is contacted two times on email and once on the phone, there should be three entries for that customer.) ALPHANUMERIC

─ ○ ○ ContactDate Date of this contact to this customer (date when mailer or periodical was sent out, date when phone call was made, date when email was sent, date when DATE

		billboard was erected, date when internet advertisement was posted, etc)	
— ○ ○	CollateralID	ID number of the marketing collateral given to this customer during this contact. Each CollateralID should be identical to the name, excluding the extension, of the corresponding marketing collateral file uploaded in the "Document Upload File" section.	ALPHAN UMERIC
— ○ ○	ContactType_DirectMail	During this contact, this customer was exposed to direct mail recruitment collateral	BOOLEA N	**BOOLEAN Options:** YES NO NULL
— ○ ○	ContactType_Periodical	During this contact, this customer was exposed to periodical recruitment collateral	BOOLEA N	**BOOLEAN Options:** YES NO NULL
— ○ ○	ContactType_Billboard	During this contact, this customer was exposed to billboard recruitment collateral	BOOLEA N	**BOOLEAN Options:** YES NO NULL

	Description	Type	Options
○ ○ ContactType_Internet	During this contact, this customer was exposed to internet recruitment collateral	BOOLEAN	**BOOLEAN Options:** YES NO NULL
○ ○ ContactType_Phone	During this contact, this customer was exposed to phone recruitment collateral	BOOLEAN	**BOOLEAN Options:** YES NO NULL
○ ○ ContactType_TV	During this contact, this customer was exposed to TV recruitment collateral	BOOLEAN	**BOOLEAN Options:** YES NO NULL
○ ○ ContactType_Radio	During this contact, this customer was exposed to radio recruitment collateral	BOOLEAN	**BOOLEAN Options:** YES NO NULL
○ ○ ContactType_Email	During this contact, this customer was exposed to email recruitment collateral	BOOLEAN	**BOOLEAN Options:** YES NO NULL
○ ○ ContactType_DirectSale	During this contact, this customer was exposed to direct sale recruitment collateral	BOOLEAN	**BOOLEAN Options:** YES NO NULL

				BOOLEAN Options:
☐ ☐ ContactType_SocialNetwork	During this contact, this customer was exposed to social network recruitment collateral	N	BOOLEA N	YES NO NULL

Assignment Data Upload File (Assignment of Each Customer into Experimental Cell)

The purpose of the data in this file is to list and describe the specific elements associated with enrollment into experimental cells for each customer, including types and dates of eligibility, technology and educational treatments, tariff rates, and other attributes. It is imperative that DOE understand the experimental cell that each participating customer was assigned to. This data will be used to examine whether differences in rates, information, and technologies both between and across recipients lead to differences in energy use. For example, two different educational programs run by two different recipients may have very different informational content. By collecting as much detail as possible about the customer experience, DOE can try to determine which elements have the most impact on energy use.

Some customers who chose to participate in these consumer behavior studies will be unable to complete the study due to a variety of reasons. This data will be used to gain a better understanding into what causes or helps explain attrition so that future time-based rate programs can better retain customers.

Data for this file should be uploaded for each of the following:

- Customers
  - Each treatment and control customer
  - A representative sample of those who were invited to participate in the study but declined the offer (each recipient should work with their TAG to determine the most appropriate size of the representative sample)
- Frequency
  - One per customer
- Unique key fields for this file are denoted by an asterisk*

Variable Name	Description	Type	Unit	Enumeration
Analysis Robust Meta				
● CustomerID*	Unique customer identifier	ALPHANU MERIC		
● UsagePointID*	Geographic identifier for this customer's location	ALPHANU MERIC		
● UsagePointType	The type of geographic identifier used for this customer's location. Census block data should be the level of geographic identifier reported. If census block data is not readily available, NREL can provide a web service that a recipient can use on their internal system to convert an address into a census block level geographic identifier. If a recipient has concerns about reporting census block data, LBNL will work with that recipient to determine the most appropriate level of geographic identifier (e.g., census block group, zip code, etc.).	ENUM		**ENUM Options:** CENSUS BLOCK CENSUS BLOCK GROUP CENSUS TRACT ZIP CODE PLUS FOUR ZIP CODE

Field	Description	Type	Options
ProjectID*	Unique SGIG project ID number	ENUM	
ExpCellID*	Name of the control group, treatment group, or group of customers not in the study that this customer was assigned to (e.g., C1, C2, T1, Cntl1, A1, B2, N1)	ALPHANUMERIC	
IsEnrolled	This customer decided to enroll (or was automatically enrolled or assigned) in their experimental cell	BOOLEAN	**BOOLEAN Options:** YES NO NULL
EnrollmentDate	Date that this customer was enrolled (or automatically assigned) in their experimental cell	DATE	
ExclusionCriteria	If this customer was excluded from the group they were assigned to after being enrolled, the reason for exclusion (e.g., technology communications problem)	ENUM	**ENUM Options:** INCOMPATIBILITY OF TECHNOLOGY WITH HOUSEHOLD EQUIPMENT TECHNOLOGY COMMUNICATIONS PROBLEM OTHER NULL

			Field	Description	Type
●	–	–	ExpCellStartDate	Date that this customer was assigned to their experimental cell (if customer is assigned to a group and then must decide to enroll, this date will be before the EnrollmentDate; if customer enrolls in the study and then is assigned to their group then this date will be after the EnrollmentDate; if customer is automatically assigned to a group then this date will be the same as the EnrollmentDate)	DATE
●	–	–	ExpCellEndDate	Date that this customer un-enrolled from their assigned group	DATE
–	○	–	ReasonForExit	Reason that this customer un-enrolled from their assigned group (e.g., opt-out, change of service address, end of study)	ENUM	**ENUM Options:** CHANGE OF SERVICE ADDRESS OPT OUT OTHER NULL
–	○	●	BillProtectionStartDate	Date that bill protection for this customer started (if never available for this customer leave blank)	DATE

BillProtectionEndDate	DATE	Date that bill protection for this customer ended (if never available for this customer leave blank)
EdTreatStartDate	DATE	Date that educational treatment for this customer started (if never available for this customer leave blank)
EdTreatEndDate	DATE	Date that educational treatment for this customer ended (if never available for this customer leave blank)
InfoTechInstallDate	DATE	Date that information technology for this customer was installed (if never available for this customer leave blank)
InfoTechUninstallDate	DATE	Date that information technology for this customer was uninstalled (if never available for this customer leave blank)

● -- -- InfoTechActiveDate	Date that information technology for this customer was activated (if never available for this customer leave blank)	DATE
● -- -- InfoTechDeactiveDate	Date that information technology for this customer was deactivated (if never available for this customer leave blank)	DATE
● -- -- InfoTechTreatStartDate	Date that information technology treatment for this customer began (if never available for this customer leave blank)	DATE
● -- -- InfoTechTreatEndDate	Date that information technology treatment for this customer ended (if never available for this customer leave blank)	DATE
● -- -- ControlTechInstallDate	Date that control technology for this customer was installed (if never available for this customer leave blank)	DATE
● -- -- ControlTechUninstallDate	Date that control technology for this customer was uninstalled (if	DATE

	Field	Type	Description	
			never available for this customer leave blank)	
●	ControlTechActiveDate	DATE	Date that control technology for this customer was activated (if never available for this customer leave blank)	
●	ControlTechDeactiveDate	DATE	Date that control technology for this customer was deactivated (if never available for this customer leave blank)	
●	ControlTechTreatStartDate	DATE	Date that control technology treatment for this customer began (if never available for this customer leave blank)	
●	ControlTechTreatEndDate	DATE	Date that control technology treatment for this customer ended (if never available for this customer leave blank)	
○ ○	HasRequiredAMI	BOOLEAN	If AMI is required based on Exp Cell Data Upload File, customer has required AMI	**BOOLEAN Options:** If this is not a requirement based on Exp

Field	Description	Type	Options
			Cell Data Upload File: N/A If this is a requirement based on Exp Cell Data Upload File: YES or NO
HasRequiredHistoricalUsage	If minimum amount of historical usage is required based on Exp Cell Data Upload File, customer has required minimum amount of historical usage	BOOLEAN	**BOOLEAN Options:** If this is not a requirement based on Exp Cell Data Upload File: N/A If this is a requirement based on Exp Cell Data Upload File: YES or NO
HasRequiredTariff	If eligible tariff rate is required based on Exp Cell Data Upload File, customer is currently on an eligible tariff rate	BOOLEAN	**BOOLEAN Options:** If this is not a requirement based on Exp Cell Data Upload File: N/A If this is a requirement based on Exp Cell Data Upload File: YES or NO
HasRequiredInternetAccess	If eligible type of internet access is required based on Exp Cell Data Upload File, customer currently has an eligible type of internet access	BOOLEAN	**BOOLEAN Options:** If this is not a requirement based on Exp Cell Data Upload File: N/A If this is a requirement based on Exp Cell Data Upload File: YES or NO

	Name	Description	Type	Options
○ ○	HasRequiredNonParticipation	If non-participation in an existing DR program is required based on Exp Cell Data Upload File, customer is not participating in an existing DR program	BOOLEAN	**BOOLEAN Options:** If this is not a requirement based on Exp Cell Data Upload File: N/A If this is a requirement based on Exp Cell Data Upload File: YES or NO
○ ○	HasRequiredNonUtilityEmployee	If not being a utility employee is required based on Exp Cell Data Upload File, customer is NOT a utility employee as required to participate	BOOLEAN	**BOOLEAN Options:** If this is not a requirement based on Exp Cell Data Upload File: N/A If this is a requirement based on Exp Cell Data Upload File: YES or NO
○ ○	HasRequiredCentralAC	If a central air conditioning system is required based on Exp Cell Data Upload File, customer currently has a required central air conditioning system	BOOLEAN	**BOOLEAN Options:** If this is not a requirement based on Exp Cell Data Upload File: N/A If this is a requirement based on Exp Cell Data Upload File: YES or NO
○ ○	HasRequiredWaterHeater	If an eligible water heater is required based on Exp Cell Data Upload File, customer currently has a required eligible water heater	BOOLEAN	**BOOLEAN Options:** If this is not a requirement based on Exp Cell Data Upload File: N/A If this is a requirement based on Exp Cell Data Upload File: YES or NO

○ ○ HasRequiredPoolPump	If an eligible pool pump is required based on Exp Cell Data Upload File, customer currently has a required eligible pool pump	BOOLEAN	**BOOLEAN Options:** If this is not a requirement based on Exp Cell Data Upload File: N/A If this is a requirement based on Exp Cell Data Upload File: YES or NO
○ ○ HasRequiredYearRoundResident	If year-round residency is required based on Exp Cell Data Upload File, customer currently is a year-round resident	BOOLEAN	**BOOLEAN Options:** If this is not a requirement based on Exp Cell Data Upload File: N/A If this is a requirement based on Exp Cell Data Upload File: YES or NO
○ ○ HasRequiredStudyDuration	If residency for the duration of the study is required based on Exp Cell Data Upload File, customer is a resident for the duration of the study	BOOLEAN	**BOOLEAN Options:** If this is not a requirement based on Exp Cell Data Upload File: N/A If this is a requirement based on Exp Cell Data Upload File: YES or NO

– ○ ○ HasRequiredPaymentStandi ng	If eligible payment standing is required based on Exp Cell Data Upload File, customer is currently in eligible payment standing	**BOOLEAN** **BOOLEAN Options:** If this is not a requirement based on Exp Cell Data Upload File: N/A If this is a requirement based on Exp Cell Data Upload File: YES or NO
– ○ ○ HasRequiredAddressMatch	If a service and billing address match is required based on Exp Cell Data Upload File, customer has the required service and billing address matching	**BOOLEAN** **BOOLEAN Options:** If this is not a requirement based on Exp Cell Data Upload File: N/A If this is a requirement based on Exp Cell Data Upload File: YES or NO
– ○ ○ HasRequiredDwellingType	If an eligible dwelling type is required based on Exp Cell Data Upload File, customer is in an eligible dwelling type	**BOOLEAN** **BOOLEAN Options:** If this is not a requirement based on Exp Cell Data Upload File: N/A If this is a requirement based on Exp Cell Data Upload File: YES or NO

Survey Admin Data Upload File (List of Recipient's Surveys)

The purpose of the data in this file is to list and describe each survey that is administered, in order to obtain some summary information on the survey efforts undertaken by each recipient to collect this customer characteristic and demographic data.

In addition, a copy of all survey instrument(s) should be uploaded in the "Document Upload File". Each SurveyID used below should be identical to the name, excluding the extension, of the corresponding survey instrument file uploaded in the "Document Upload File".

Data for this file should be uploaded for each of the following:

- Each Survey Administered
- There should be one entry for each survey that is administered
- Unique key fields for this file are denoted by an asterisk*

Variable Name	Description	Type	Unit	Enumeration
Analysis Robust Meta				
● ProjectID*	The SGIG project ID number	ENUM		e.g. 09-0111 09-0172
● ● ● PrimeRecipient	The name of the SGIG Recipient	ENUM		**ENUM Options:** City of Fort Collins Minnesota Power Lakeland Electric Marblehead Municipal Light Department FirstEnergy Service Ohio NV Energy , Inc. NV Energy , Inc. Oklahoma Gas and Electric Detroit Edison Company City of Auburn

● ● ● SurveyID*

Unique survey identifier. There should be one entry for each survey that is administered. Surveys should be considered different and have unique IDs if their wording or appearance is different or if any of the fields below differs. E.g., surveys administered with different start dates should have different IDs. Each SurveyID should be identical to the name, excluding the extension, of the corresponding survey instrument file uploaded in the "Document Upload File" section.

ALPHANU MERIC

-- ○ SurveyInstrumentDescription Briefly describe the survey effort associated with this survey ID

ALPHANU MERIC

-- ● ○ SurveyTargetAudience

Audience the survey was administered to (e.g., control group, MERIC treatment group, etc.)

ALPHANU

-- ● ● SurveyStartDate

Date survey was administered

DATE

	Field	Description	Type
-- ● ●	SurveyEndDate	Date survey was ended	DATE
-- ● ●	SurveyModalityType	Modality of survey (e.g., mail, phone call)	ENUM **ENUM Options:** MAIL PHONE INTERNET DOOR-TO-DOOR NULL
● ○ ●	SurveyType	Type of survey (i.e., enrollment, pre-treatment, event, post-treatment)	ENUM **ENUM Options:** PRE-ENROLLMENT ENROLLMENT PRE-TREATMENT EVENT POST-TREATMENT NULL
-- ● ●	NumSurveyed	Number of surveys solicited	INTEGER [COUNT]
-- ● ●	NumFullComplete	Number of surveys fully completed	INTEGER [COUNT]
-- ● ●	NumPartialComplete	Number of surveys partially completed	INTEGER [COUNT]
-- ○ ○	NumRefuse	Number of surveys returned unopened (mail) or refused (phone)	INTEGER [COUNT]

— ○ ○ ResponseRate3	Response Rate 3 (RR3, described in Appendix F on p.25 of Guidance Document #9)	DECIMAL [COUNT]
— ○ ○ ResponseRate4	Response Rate 4 (RR4, described in Appendix F on p.25 of Guidance Document #9)	DECIMAL

Survey Results Data Upload File (Each Customer's Survey Responses)

The purpose of the data in this file is to list and describe all responses to surveys for each participating customer that was chosen for surveying.

DOE believes that one of the major contributions of these studies is to provide a richer understanding of the segmentation of customer acceptance and response to AMI-enabled time-based rate programs and control/information technology. This segmentation is only possible, however, if demographic and other information is collected by recipients that more fully describes a customer's situation and

surroundings. Recipients are not asked to provide property value information because this data was deemed to be too risky in obtaining personally identifiable information (see the analysis described in the Appendix: Privacy Protection for SGIG Participants)

Data for this file should be uploaded for each of the following:

- Customers
  - Each treatment and control customer
  - A representative sample of those who were invited to participate in the study but declined the offer (each recipient should work with their TAG to determine the most appropriate size of the representative sample)
- Frequency
  - Each survey administered
- There should be one entry for each survey administered for each customer
- Unique key fields for this file are denoted by an asterisk*

Variable Name	Description	Type	Unit	Enumeration
Analysis Robust Meta				
● ● CustomerID*	Unique customer identifier	ALPHANU MERIC		
● ● ● UsagePointID*	Geographic identifier for this customer's location	ALPHANU MERIC		

• • • UsagePointType	The type of geographic identifier used for this customer's location. Census block data should be the level of geographic identifier reported. If census block data is not readily available, NREL can provide a web service that a recipient can use on their internal system to convert an address into a census block level geographic identifier. If a recipient has concerns about reporting census block data, LBNL will work with that recipient to determine the most appropriate level of geographic identifier (e.g., census block group, zip code, etc.).	ENUM	**ENUM Options:** CENSUS BLOCK CENSUS BLOCK GROUP CENSUS TRACT ZIP CODE PLUS FOUR ZIP CODE
• • • ProjectID*	The SGIG project ID number	ENUM	
• • • SurveyId*	Unique survey identifier	ALPHANUMERIC	

● – – CompletionStatus	Describes the completion status of this survey with this customer	ENUM	**ENUM Options:** COMPLETED PARTIALLY COMPLETED REFUSED NOT RETURNED NULL
● ○ ○ ResOwnership	Customer's response to housing ownership	ENUM	**ENUM Options:** RENT OWN DON'T KNOW REFUSED OTHER NULL
● ○ ○ ResType	Customer's response to residence type	ENUM	**ENUM Options:** SINGLE FAMILY DUPLEX APARTMENT 2-4 UNIT APARTMENT > 4 UNIT TOWNHOUSE MOBILE HOME DON'T KNOW REFUSED OTHER NULL
● ○ ○ ResCAC	Customer's response to whether	BOOLEAN	**BOOLEAN**

	there is central air conditioning		**OPTIONS:** YES NO NULL
● ○ ○ ResRoomAC	Customer's response to whether there is room A/C	BOOLEAN	**BOOLEAN OPTIONS:** YES NO NULL
● ○ ○ ResNumRoomAC	Customer's response to number of room A/C units in the house	INTEGER [COUNT]	
○ ○ ○ ResPT	Customer's response to programmable thermostat	BOOLEAN	**BOOLEAN OPTIONS:** YES NO NULL
○ ○ ○ ResPTSet	Customer's response to programmable thermostat set to change temperature during the day	BOOLEAN	**BOOLEAN OPTIONS:** YES NO NULL
○ ○ ○ ResElectricDryer	Customer's response to whether there is an electric dryer	BOOLEAN	**BOOLEAN OPTIONS:** YES NO NULL

○ ○ ○ ResNumAdults	Customer's response to number of adults (18 or older) living in the residence	INTEGER	[COUNT]
● ○ ○ ResNumElderly	Customer's response to number of elderly (65 or older) living in the house	INTEGER	[COUNT]
○ ○ ○ ResNumKids	Customer's response to number of children (under 18) living in the house	INTEGER	[COUNT]
● ○ ○ ResChronicIllness	Customer's response to chronic illness	BOOLEAN	**BOOLEAN OPTIONS:** YES NO NULL
○ ○ ○ ResHome1To5	Customer's response to someone home M-F between hours of 1 and 5 PM	BOOLEAN	**BOOLEAN OPTIONS:** YES NO NULL
○ ○ ○ ResWorkFulltime	Customer's response to someone in the home works full-time for pay	BOOLEAN	**BOOLEAN OPTIONS:** YES NO NULL

○ ○ ○ ResWorkHome	Customer's response to someone in the home works from home at least once a week	BOOLEAN	**BOOLEAN OPTIONS:** YES NO NULL
○ ○ ○ ResReceiptInfo	Customer's response to received information about the pilot	BOOLEAN	**BOOLEAN OPTIONS:** YES NO NULL
○ ○ ○ ResInfoHelpful	Customer's response to found this information to be useful	BOOLEAN	**BOOLEAN OPTIONS:** YES NO NULL
○ ○ ○ ResPrimaryLanguage	Customer's response to primary language	ENUM	**ENUM Options:** ENGLISH SPANISH CHINESE KOREAN VIETNAMESE RUSSIAN REFUSED OTHER NULL

ResIncomeLevel — Customer's response to income level — ENUM

ENUM Options:
LESS THAN $10,000
$10,000 TO LESS THAN $20,000
$20,000 TO LESS THAN $30,000
$30,000 TO LESS THAN $40,000
$40,000 TO LESS THAN $75,000
$75,000 TO LESS THAN $90,000
$90,000 TO LESS THAN $100,000
$100,000 TO LESS THAN $150,000
$150,000
$150,000 OR MORE
DON'T KNOW
REFUSED
NULL

ResEducationLevel — Customer's response to education level — ENUM

ENUM Options:
NONE OR GRADE 1-8
GRADE 9-11
HIGH SCHOOL INCOMPLETE
HIGH SCHOOL GRADUATE
GRADE 12 OR GED
TRADE OR VOCATIONAL SCHOOL
AFTER HIGH SCHOOL
SOME COLLEGE INCLUDING
ASSOCIATE DEGREES
COLLEGE GRADUATE
POST GRADUATE OR PROFESSIONAL

SCHOOL AFTER COLLEGE
DON'T KNOW
REFUSED
NULL

⦿ ○ ○	BusinessType	Customer's response to business type	ENUM	**ENUM Options:** AGRICULTURE/AGRICULTURAL PROCESSING ASSEMBLY/LIGHT INDUSTRY CHEMICALS/PAPER/REFINING FOOD PROCESSING GROCERY STORE/RESTAURANT LODGING HIGH TECH LUMBER/MINING/PLASTICS OFFICE OIL/GAS EXTRACTION RETAIL STONE/CLASS/CLAY/CEMENT TRANSPORTATION UTILITY OTHER NULL
○ ○ ○	NonResSqFt	Non-residential customer's response to square feet	INTEGER [SQFT]	

○ ○ ○ NonResFTEmployees	Non-residential customer's response to full time employees	INTEGER [COUNT]
○ ○ ○ NonResBusinessValue	Non-residential customer's response to business value	INTEGER [$/YEAR]
○ ○ ○ NonResElectricityShareofBudget	Non-residential customer's response to electricity's share of the budget	DECIMAL [ANNUAL PERCENTAGE] **Scale Range:** 0 to 1

Tariff Data Upload File (List of Tariff Rates)

The purpose of the data in this file is to list and describe each tariff rate that occurred before and during the study. If one tariff rate, indicated by TariffID, has rates that change over time, then a separate entry should be listed for each one indicated by different start dates in TariffStartDate.

In addition, a copy of all applicable tariff sheets should be uploaded in the "Document Upload File". Each TariffID and TariffStartDate used below should be identical to the name, excluding the extension, of the corresponding uploaded tariff sheet file, where TariffID and TariffStartDate are separated by an underscore " _ ", and the date in the file name is in mmddyyyy format (e.g., for TariffID="RES1" and TariffStartDate="6/1/2011 ", the corresponding tariff sheet uploaded in the "Document Upload File" should be named "RES1_06012011" (see examples below).

Data for this file should be uploaded for each of the following:

- Each tariff rate that occurred for every customer in the treatment group, control group, and group of customers not in the study.
- Time Periods for tariff data
  - Historical (12 months prior to beginning of treatment)
  - Throughout the study duration
- There should be one entry for each tariff rate.
- Unique key fields for this file are denoted by an asterisk*

Examples

Entries in the Tariff Data Upload File

	ProjectID	TariffID	TariffStartDate	BaseRateType	OverlayRateType	TariffIDDesc	ReasonForTariffChange	Name of the uploaded file	Corresponding tariff sheet uploaded in the "Document Upload File"
1	FFLP	RES1	6/1/2011	FLAT	CPR	Flat rate of $0.089/kWh with a CPR overlay			"RES1_06012011.pdf"
2	"	RES1	6/1/2012	FLAT	CPP	Flat rate of $0.089/kWh with a CPP overlay	Part of the CBS study		"RES1_06012012.pdf"

Example 1: A utility with ProjectID="FFLP" has a tariff rate "RES1" which a certain group of customers in the CBS will be exposed to. Starting on 6/1/2011, tariff rate "RES1" is defined as a flat rate of 8.9 cents/kWh with CPR overlay but then changes on 6/1/2012 to a flat rate with CPP overlay. This utility should submit two records in this data file for "RES1": one that describes the flat rate with CPR overlay with a TariffID="RES1" and TariffStartDate=6/1/2011, and one that describes the flat rate with CPP overlay with a TariffID="RES1" and TariffStartDate=6/1/2012. Two tariff sheets should also be uploaded in the "Document Upload File", one that is named "RES1_06012011.pdf" and one that is named "RES1_06012012.pdf".

Example 2: This same utility also has a tariff rate "R2" which a certain group of customers in the CBS will be exposed to. Tariff rate "R2" is defined as a flat rate. Since 5/15/2009, tariff rate "R2" has been set at 9.2 cents/kWh but then increased on 11/23/2011 to a flat rate of 11.3 cents/kWh. This utility should submit two records in this data file for "R2": one that describes the flat rate for "R2" with a TariffID="R2" and TariffStartDate=5/15/2009, and one that describes the flat rate at 11.3 cents/kWh with a TariffID="R2" and TariffStartDate=11/15/2011. Two tariff sheets should also be uploaded in the "Document Upload File", one that is named "R1_05152009.pdf" (assuming the tariff sheet is in PDF format), and one that is named "R1_11232011.pdf"

	Variable Name	Description	Type	Unit	Enumeration
Analysis Robust Meta ● ● ●	● ProjectID*	The SGIG project ID number	ENUM		
	● TariffID*	Tariff rate identifier	ALPHAN		

● ● ● TariffStartDate*		UMERIC
	For this tariff, the date that the rate descriptions below took effect.	DATE
● — — BaseRateType	The base tariff rate type for this tariffID and TariffStartDate	ENUM — **ENUM Options:** FLAT / INCLINING BLOCK / DECLINING BLOCK / TOU / VPP / RTP / DEMAND ENERGY / NULL
● — — OverlayRateType	The overlay rate type for this tariffID and TariffStartDate	ENUM — **ENUM Options:** CPP / CPR / NONE / NULL
● — — TariffIDDesc	Description of this tariff ID and TariffStartDate	ALPHANUMERIC
● — — ReasonForTariffChange	Description of the reason that this TariffID changed over time indicated by different TariffStartDates	ALPHANUMERIC

Customer Tariff Data Upload File (List of Rate Experienced by Each Customer)

The purpose of the data in this file is to list and describe the period in which each customer experienced each tariff rate (where the tariff rates are defined in the Tariff Data Upload File).

Data for this file should be uploaded for each of the following:

- Customers
  - Each treatment and control customer
  - A representative sample of those who were invited to participate in the study but declined the offer (each recipient should work with their TAG to determine the most appropriate size of the representative sample)
- Time Periods for tariff data
  - Historical Data (12 months prior to beginning of treatment)
  - Data throughout the study duration
- There should be one entry for each tariff rate for each customer.
- Unique key fields for this file are denoted by an asterisk*

Examples

	ProjectID	CustomerID	TariffID	TariffStartDate	TariffStartDateCust	TariffEndDateCust
1	FFLP	1001	R2	5/15/2009	8/6/2009	12/5/2011
2	"	"	R2	11/23/2011	12/6/2011	1/1/2013

Entries in the Customer Tariff Data Upload File

3	1002		R2	5/15/2009	12/9/2010	8/8/2011
4	"	"	RES1	6/1/2011	8/9/2011	6/5/2012
5	"	"	RES1	6/1/2012	6/9/2012	1/1/2013

Example 3: A utility (ProjectID="FFLP") has the tariff rates described above in Examples 1 and 2, rates "R2" and "RES1". The TariffID = "R2" has two different rates associated with it: starting on TariffStartDate = "5/15/2009", "R2" is a flat rate set at 9.2 cents/kWh, but then on TariffStartDate = "11/23/2011", "R2" increased to a flat rate set at 11.3 cents/kWh. The TariffID = "RES1" also has two different rates associated with it: starting on TariffStartDate = "6/1/2011", "RES1" is a flat rate with a CPR overlay, but on TariffStartDate = "6/1/2012", "RES1" changes to a CPP overlay.

Customer 1001 started to receive rate "R2" on 8/6/2009, and was kept on "R2" for the duration of the study until 1/1/2013. She always receives her bill on the 6th day of each month, and so if her rate changes, the change will occur for her on the 6th day of the month. Therefore for this customer, because "R2" changes from 9.2 to 11.3 cents/kWh, there should be two entries for CustomerID = "1001": first, starting on TariffStartDateCust = "8/6/2009", she experiences the rate defined by TariffID = "R2" and TariffStartDate = "5/15/2009"; and second, starting on TariffStartDateCust = "12/6/2011", she experiences the rate defined by TariffID = "R2" and TariffStartDate = "11/23/2011". (In this example, even though the tariff rate changed on TariffStartDate = "11/23/2011", she didn't experience the change until TariffStartDateCust = "12/6/2011".)

Example 4: Using the same utility and the same rates as in Examples 1-3, customer 1002 started to receive rate "R2" on 12/9/2010, and then as part of the study was transitioned to "RES1" on 8/9/2011 (rate changes always occur for her on the 9th day of the month), and remained on "RES1" for the duration of the study until 1/1/2013. Because "RES1" changed from a CPR to a CPP rate on 6/1/2012, this customer's rate will change from a CPR to a CPP on 6/9/2012. Therefore for this customer, there should be three entries for CustomerID = "1002": first, starting on TariffStartDateCust = "12/9/2010", she experiences the rate defined by TariffID = "R2" and TariffStartDate = "5/15/2009"; second, starting on TariffStartDateCust = "8/9/2011", she experiences the rate defined by TariffID =

"RES1" and TariffStartDate = "6/1/2011"; third, starting on TariffStartDateCust = "6/9/2012", she experiences the rate defined by TariffID = "RES1" and TariffStartDate = "6/1/2012".

Variable Name	Description	Type	Unit	Enumeration
Analysis Robust Meta				
● CustomerID*	Unique customer identifier	ALPHANUMERIC		
● UsagePointID*	Geographic identifier for this customer's location	ALPHANUMERIC		
● UsagePointType	The type of geographic identifier used for this customer's location. Census block data should be the level of geographic identifier reported. If census block data is not readily available, NREL can provide a web service that a recipient can use on their internal system to convert an address into a census block level geographic identifier. If a recipient has concerns about reporting census block data, LBNL will work with that recipient to	ENUM		**ENUM Options:** CENSUS BLOCK CENSUS BLOCK GROUP CENSUS TRACT ZIP CODE PLUS FOUR ZIP CODE

Field	Description	Type
	determine the most appropriate level of geographic identifier (e.g., census block group, zip code, etc.).	
●●● ProjectID*	The SGIG project ID number	ENUM
●●● TariffID*	Tariff rate identifier	ALPHANUMERIC
●●● TariffStartDate*	Start date of the TariffID (as defined in the Tariff Data Upload File)	DATE
● -- -- TariffStartDateCust	Date that this customer began taking service under the rate defined by the above TariffID and TariffStartDate due to their particular bill cycle	DATE
● -- -- TariffEndDateCust	Date that this customer stopped taking service under this TariffID rate and accompanying TariffStartDate due to their particular bill cycle	DATE

Event Data Upload File (List of All Events)

The purpose of the data in this file is to list and describe each event that occurred before and during the study.

Data for this file should be uploaded for each of the following:

- Frequency
  - Each event
- Time Periods for event data
  - Historical Data (12 months prior to beginning of treatment)
  - Data throughout the study duration
- There should be one version for each event
- Unique key fields for this file are denoted by an asterisk*

Variable Name	Description	Type	Unit	Enumeration
Analysis Robust Meta				
● ● ProjectID*	The Unique SGIG project ID number	ENUM		
● ● ● EventID*	Unique identification number of the event (there should be one entry for each event).	ALPHAN UMERIC		
● ‒ ‒ EventStartTimeUtility	Day and time that the event started	MM/DD/ YYYY HH:MM: SS		

● EventDurationUtility	Number of seconds that the event lasted	NUMERIC	[SECONDS]
○ ○ EventCriteriaCode	The reason that this event occurred	ENUM	**ENUM Options:** ECONOMICS SHORT TERM RELIABILITY LONG TERM RELIABILITY
○ ○ EventCriteriaDesc	A description of the criteria for this event occurring (e.g., there is a temperature or a price threshold that triggers an event)	ALPHANUMERIC	

Cust Event Note Upload File (Events Experienced by Each Customer)

The purpose of the data in this file is to list and describe each event that is experienced by each customer. It catalogs attempts to notify each appropriate participating customer of an impending curtailment event. During the consumer behavior study, data will be collected from each participating treatment and control customer in order to better understand how being exposed to treatment affected electricity consumption behavior.

Data for this file should be uploaded for each of the following:

- Customers
  - Each treatment and control customer that experiences or is notified about an event
- Frequency
  - Each event for each customer

- Time Periods for event data
  - The duration of time that the customer experienced or was notified about events (e.g., throughout the duration that the customer is enrolled in the study and prior to the study if the customer experienced or was notified of events in the 12 months before the study began).
- There should be one entry for each event for each customer
- Unique key fields for this file are denoted by an asterisk*

Variable Name	Description	Type	Unit	Enumeration
Analysis Robust Meta				
● ● ● CustomerID*	Unique customer identifier	ALPHANUMERIC		
● ● ● UsagePointID*	Geographic identifier for this customer's location	ALPHANUMERIC		
● ● ● UsagePointType	The type of geographic identifier used for this customer's location. Census block data should be the level of geographic identifier reported. If census block data is not readily available, NREL can provide a web service that a recipient can	ENUM		**ENUM Options:** CENSUS BLOCK CENSUS BLOCK GROUP CENSUS TRACT ZIP CODE PLUS FOUR ZIP CODE

use on their internal system to convert an address into a census block level geographic identifier. If a recipient has concerns about reporting census block data, LBNL will work with that recipient to determine the most appropriate level of geographic identifier (e.g., census block group, zip code, etc.).

- ProjectID* — The Unique SGIG project ID number — ENUM

- EventID* — Unique identification number of the event as defined in the Event Data Upload File. There should be one entry for each event per customer (e.g., if one customer experiences 8 events, there should be 8 entries for that customer). — ALPHANUMERIC

- EventStartTiimeCust — Day and time that this event started for this customer — MM/DD/YYYY HH:MM:SS

	Field	Description	Type	Options
● -- --	EventDurationCust	Number of seconds that this event lasted for this customer	NUMERIC [SECONDS³]	
● -- --	PCTUtilitySetDuringEvent	This customer's PCT was set by the utility during this event	BOOLEAN	**BOOLEAN Options:** YES NO
-- ○ ●	NotificationType	Type of notification provided to this customer for this event (e.g., email, phone, text, pager)	ENUM	**ENUM Options:** EMAIL PHONE TEXT MESSAGE PAGER NULL
-- ○ ●	AdvanceNotificationHours	Number of hours ahead of time notification is provided to this customer for this event	DECIMAL [HOURS]	
-- ○ ○	NotificationSuccessful	The notification to this customer for this event was successfully delivered	BOOLEAN	**BOOLEAN Options:** YES NO

---

3 The industry standard format for reporting utility event duration is in seconds (see the Standards Reference NAESB REQ 18/WEQ19).

-- ○ ○ FailureReason	If notification delivery failed, reason (email bounced, phone - no answer)	ENUM	**BOOLEAN Options:** EMAIL BOUNCED PHONE NOT ANSWERED OTHER NULL

Billing Data Upload File

The purpose of the data in this file is to list and describe key monthly billing information for each customer during pre-treatment and post-treatment periods.

Data for this file should be uploaded for each of the following:

- Customers
  - ○ Each treatment and control customer
  - ○ A representative sample of those who were invited to participate in the study but declined the offer (each recipient should work with their TAG to determine the most appropriate size of the representative sample)
- Frequency
  - ○ Each Bill (e.g., monthly)
- Time Periods for billing data
  - ○ Historical Data (12 months prior to beginning of treatment)
  - ○ Data throughout the study duration
- There should be one entry for each bill cycle for each customer
- Unique key fields for this file are denoted by an asterisk*

Variable Name	Description	Type	Unit	Enumeration
Analysis Robust Meta				
● CustomerID*	Unique customer identifier	ALPHANUMERIC		
● UsagePointID*	Geographic identifier for this customer's location	ALPHANUMERIC		
● UsagePointType	The type of geographic identifier used for this customer's location. Census block data should be the level of geographic identifier reported. If census block data is not readily available, NREL can provide a web service that a recipient can use on their internal system to convert an address into a census block level geographic identifier. If a recipient has concerns about reporting census block data, LBNL will work with that recipient to determine the most appropriate level of geographic identifier (e.g., census block group, zip code, etc.).	ENUM		**ENUM Options:** CENSUS BLOCK CENSUS BLOCK GROUP CENSUS TRACT ZIP CODE PLUS FOUR ZIP CODE

Field	Description	Data Type
● ● ● ProjectID*	The SGIG project ID number	ENUM
● ● ● BillCycleID*	Unique bill cycle identifier. There should be one entry per customer per month.	ALPHANUMERIC
● ● ● BillCycleStartDate*	Date that this bill cycle started for this customer	DATE
● -- -- BillCycleEndDate	Date that this bill cycle ended for this customer	DATE
● -- -- BillCycleAmount	Monthly electricity bill amount in $	DECIMAL [$/BILLCYCLE]
BillCycleAllInRate	The all-in volumetric retail rate, in $/kWh, that is applicable for this customer during this billing cycle. This includes electricity commodity charges (e.g., fuel costs), transmission and distribution charges, and additional cost trackers and/or riders, but excluding monthly customer charges.	DECIMAL [$/kWh]

● -- -- BillCyclekWh	Monthly electricity amount in kWh	DECIMAL [kWh/BILLCYCLE]
● -- -- BillCyclekW	Monthly demand in kW, if applicable	DECIMAL [kW/BILLCYCLE]
● -- -- ShadowBillAmount	Monthly shadow bill amount in $, if applicable	DECIMAL [$/BILLCYCLE]
● -- -- BillProtectionAmount	Bill protection adjustment amount in $, if applicable	DECIMAL [$/BILLCYCLE]

Time Series Upload File Format (Hourly Energy Use Data)

The purpose of the data in this file is to list and describe the relevant hourly data (e.g., energy, weather, retail rate) for each customer during pre-treatment and post-treatment periods.

Data for this file should be uploaded for each of the following:

- Customers
  - Each treatment and control customer
  - A representative sample of those who were invited to participate in the study but declined the offer (each recipient should work with their TAG to determine the most appropriate size of the representative sample)
- Frequency
  - Hourly
- Time Periods for usage data

- o Historical Data (12 months prior to beginning of treatment)
- o Data throughout the study duration
- There should be one entry for each hour for each customer
- Unique key fields for this file are denoted by an asterisk*

Analysis Robust Meta	Variable Name	Description	Type	Unit	Enumeration
● ● ●	ProjectID*	The SGIG project ID number	ENUM		
● ● ●	CustomerID*	Unique customer identifier	ALPHANUMERIC		
● ● ●	DateTimeStamp*	Time and date for this energy usage	DATETIME	[MM/DD/YYYY HH:MM:SS]	
● ● ●	UsagePointID*	Geographic identifier for this customer's location	ALPHANUMERIC		
● ● ●	UsagePointType	The type of geographic identifier used for this customer's location. Census block data should be the level of geographic identifier	ENUM		**ENUM Options:** CENSUS BLOCK CENSUS BLOCK GROUP CENSUS TRACT

	reported. If census block data is not readily available, NREL can provide a web service that a recipient can use on their internal system to convert an address into a census block level geographic identifier. If a recipient has concerns about reporting census block data, LBNL will work with that recipient to determine the most appropriate level of geographic identifier (e.g., census block group, zip code, etc.).		ZIP CODE PLUS FOUR ZIP CODE
● – – Energy	Energy in kWh for this customer at this time	NUMERIC	[kWh]
● – – EnergyAccumulationType	The way in which this energy data is accumulated (ENUM options are from the Standards Reference NAESB REQ 18/WEQ19; see footnote at the end of the table for definitions)	ENUM	**ENUM Options*:** NULL Bulk Quantity Cumulative Delta Data Indicating Summation Instantaneous

Field	Description	Type	Units / ENUM Options
○ – – Power (Not Required)	Power in kW for this customer at this time	NUMERIC	[kW]
○ – – PowerAccumulationType (Not Required)	The way in which this power data is accumulated	ENUM	**ENUM Options*:** NULL, Bulk Quantity, Cumulative, Delta Data, Indicating, Summation, Instantaneous
○ – – PowerFactor (Not Required)	The Power Factor for this customer at this time	NUMERIC	
○ – – PowerFactorAccumulationType (Not Required)	The way in which this power factor data is accumulated	ENUM	**ENUM Options*:** NULL, Bulk Quantity, Cumulative, Delta Data, Indicating, Summation, Instantaneous

	Field	Description	Type	ENUM Options
O --	Voltage (Not Required)	The voltage in Volts for this customer at this time	NUMERIC [Volts]	
O --	VoltageAccumulationType (Not Required)	The way in which this voltage data is accumulated	ENUM	**ENUM Options*:** NULL, Bulk Quantity, Cumulative, Delta Data, Indicating, Summation, Instantaneous
● --	WeatherStationID	A unique weather station identifier	ALPHANUMERIC	
● --	WeatherAccumulationType	The way in which the below weather data is accumulated	ENUM	**ENUM Options*:** NULL, Bulk Quantity, Cumulative, Delta Data, Indicating, Summation, Instantaneous

● -- -- DryBulbTemp	Dry bulb temperature in degrees Fahrenheit from the nearest weather station to this customer during the time indicated by the DateTimeStamp	NUMERIC [C]
● -- -- WetBulbTemp	Wet bulb temperature in degrees Fahrenheit from the nearest weather station to this customer during the time indicated by the DateTimeStamp	NUMERIC [C]
● -- -- DewPointTemp	Dew point temperature in degrees Fahrenheit from the nearest weather station to this customer during the time indicated by the DateTimeStamp	NUMERIC [C]
● -- -- TariffRate	The tariff rate stated on the utility tariff sheets, in $/kWh, that is applicable for this customer during the time indicated by the DateTimeStamp (i.e., excluding monthly customer charges and excluding all trackers and/or riders).	NUMERIC [$/kWh]

	Field	Description	Type
O	PageViews	Number of page views and click-through for data on the website for this customer during the time indicated by the DateTimeStamp	INTEGER [Count/hour]
O	PCTSetPoint	The temperature set by the customer on their PCT during the time indicated by the DateTimeStamp (leave blank if the customer's PCT was not set by the customer at this time and/or the set point was not collected by the utility).	NUMERIC [C]
O	UtilityPCTSetPoint	The temperature set by the utility on this customer's PCT during the time indicated by the DateTimeStamp (leave blank if the customer's PCT was not set by the utility at this time and/or the set point was not collected by the utility).	NUMERIC [C]
●	CustPCTOverride	This customer overrode the utility's setting of their PCT during the time	BOOLEAN **BOOLEAN Options:** YES

	indicated by the DateTimeStamp	NO NULL
O -- - -	IHDBacklightCount	Number of times this customer activated the IHD's backlight during the time indicated by the DateTimeStamp	INTEGER [Count/hour]
O -- - -	HelpCenterAccessCount	Number of times this customer accessed the Help Center during the time indicated by the DateTimeStamp	INTEGER [Count/hour]

*The definition of AccumulationType field ENUM options are from the Standards Reference NAESB REQ 18/WEQ19:

"Accumulation behavior of a reading over time, usually 'measuringPeriod', to be used with individual endpoints (as opposed to 'macroPeriod' and 'aggregate' that are used to describe aggregations of data from individual endpoints).

1. *NULL: Initial or default value for data element when the database structure is created.*

2. *Bulk Quantity: A value from a register which represents the bulk quantity of a commodity. This quantity is computed as the integral of the commodity usage rate. This value is typically used as the basis for the dial reading at the meter, and as a result, will roll over upon reaching a maximum dial value.*

3. *Cumulative: The sum of the previous period values. Note: "Cumulative" is commonly used in conjunction with "demand." Each demand reset causes the maximum demand value for the present billing period (since the last demand reset) to accumulate as an accumulative total of all maximum demands. So instead of "zeroing" the demand register, a demand reset has the affect of adding the present maximum demand to this accumulating total.*

4. *Delta Data: The difference between the value at the end of the prescribed interval and the beginning of the interval. This is used for incremental interval data. Note: One common application would be for load profile data, another use might be to*

report the number of events within an interval (such as the number of equipment energizations within the specified period of time.)

5. *Indicating: As if a needle is swung out on the meter face to a value to indicate the current value. (Note: An "indicating" value is typically measured over hundreds of milliseconds or greater, or may imply a "pusher" mechanism to capture a value. Compare this to "instantaneous" which is measured over a shorter period of time.)*

6. *Summation: A form of accumulation which is selective with respect to time. Note : "Summation" could be considered a specialization of "Bulk Quantity" according to the rules of inheritance where "Summation" selectively accumulates pulses over a timing pattern, and "BulkQuantity" accumulates pulses all of the time.*

7. *Instantaneous: Typically measured over the fastest period of time allowed by the definition of the metric (usually milliseconds or tens of milliseconds.) Basically the value of the reading at the moment of the time stamp. (Note: "Instantaneous" was moved to attribute #3 in 61968-9Ed2 from attribute #1 in 61968-9Ed1.)*

Document Upload File (A Copy of All Applicable Files)

The purpose of this file is to upload copies of any documents that are relevant to the study.

- File Format
  - The uploaded files can be in Word, PDF, Excel, JPEG, or another type of file format.
- File Name
  - The file name of each document must be unique and not contain any special characters. Valid characters for this field are 0 through 9, A through Z, single space " ", and underscore " _ ". Don't use other characters (e.g. #,/, ',&,~, double space, etc.) in the filename as these are special characters that will need to be removed prior to accepting the collateral file and this data table.
  - Each file name should correspond with an ID entry used in another file (as described in detail under each type of file to be uploaded below).
- Types of files to be uploaded:
  - Survey Instruments

- Include a copy of the survey instrument(s) or list of questions for each survey administered to customers for the study.
- The name of each uploaded file (excluding the extension) should have an identical SurveyID entry used in the "Survey Admin Data Upload File".

○ Marketing Collateral
- Include a copy of any marketing materials used to recruit customers for the study (e.g., mailers, advertisements, or phone contact description).
- The name of each uploaded file (excluding the extension) should have an identical CollateralID entry used in the "Recruitment Data Upload File".

○ Tariff Sheets
- Include a copy of tariff sheets for all tariff rates that occurred for at least 12 months before the study began, and throughout the study duration, for any customer in the treatment group, control group, and group of customers not in the study.
- Each TariffID and TariffStartDate used in the "Tariff Data Upload File" should be identical to the name, excluding the extension, of the corresponding tariff sheet file uploaded here, where TariffID and TariffStartDate are separated by an underscore "_", and the date in the file name is in mmddyyyy format (e.g., for TariffID="RES1" and TariffStartDate="6/1/2011", the corresponding tariff sheet uploaded here should be named "RES1_06012011".

○ Education Material
- Include a copy of any educational materials (e.g., letters, emails, or web-based tips or other education) given to customers in the study.
- The name of each uploaded file (excluding the extension) should have an identical EducationID entry used in the "Exp Cell Data Upload File".

○ Informational Technology Description
- Include a picture of any information technology devices (e.g., IHDs, web-based feedback displays) used in the study and a description of its operation.

- The name of each uploaded file (excluding the extension) should have an identical InfoTechID entry used in the "Exp Cell Data Upload File".

  o Control Technology Description
    - Include a picture of any control technology devices (e.g., PCTs, DLCs) used in the study and a description of its operation.
    - The name of each uploaded file (excluding the extension) should have an identical ControlTechID entry used in the "Exp Cell Data Upload File".

Evaluation Data Upload File (List of Evaluation Efforts)

The purpose of the data in this file is to list and describe the key metrics that come out of the recipient's own evaluation effort. Each recipient is responsible for evaluating their SGIG consumer behavior study. The results of those evaluations are of keen interest to DOE as this information will be used, in part, to summarize experiences, results and lessons learned across the different studies. See the updated Guidance Document 6 for a description of what should be included in the recipient's evaluation reports.

The data used to construct any table or figure included in the recipient's interim and/or final evaluation report(s) that lists treatment impacts (e.g., econometric estimates of energy savings) should be submitted in an electronic format (e.g., Microsoft Excel tables, SAS file; but not PDF or JPEG formats). The LBNL research team will work with each of the recipients on an as needed basis if there are concerns about which tables/figures to include, what data from these tables/figures to include, and what format to provide this data in.[4]

_____

[4] It is understood that not every CBS recipient will be producing all of the load impacts and elasticity values originally included in the SGIG Metrics and Benefits Guidebook. As such, we have tried to make this file generic enough to accommodate the various results that will come out of all the different CBS recipient's evaluation efforts. Developing the content for this file, more than any other, will likely require direct interaction between each recipient and the LBNL research team.

APPENDIX: PRIVACY PROTECTION FOR SGIG PARTICIPANTS

MEMO

Date: April 22, 2011

To: Peter Cappers

From: Mike Sullivan, FSC GROUP

Re: *Privacy Protection for SGIG Participants*

This memo documents an analysis done by FSC to determine whether the identities of SGIG participants can be discovered by matching Smart Grid Investment Grant (SGIG) survey responses information with other publically available information about households (e.g., assessor records and market research data from third parties). To assess the extent of this problem, FSC randomly chose 5 targets from a survey it conducted during 2010 that contains information similar to what would be available from a SGIG survey (this will be referred to below as the pseudo-SGIG survey). An analyst was then given access to the survey responses of the targets, along with demographic profiles from a credit rating agency for all households in the targets' census block groups (CBGs) along with publicly available property values from all the households in the targets' CBGs. The analyst was given increasing amounts of information about the targets' property values, and each time, was asked to make 10 or fewer guesses for each target, and in some cases, to rank the guesses. The results are summarized in the table below.

Table 1
Results of Identification Experiment

Amount of Property Value Information Given to Analyst	Result
No property value information	One of five lists included a target household
Value rounded to $50k, Top-Coded for all above 90th percentile in CBG	Four of five lists included target households (no ranking of guesses)
Value rounded to $25k, Top-Coded for all above 90th percentile in CBG	Four of five lists included target households, target households were never the top choice
Decile of value within each CBG	Four of five lists included target households, target household top choice in one exercise

The experiment was meant to simulate the information available to a person trying to use the SGIG data to identify particular individuals. The SGIG surveys are designed to collect survey responses to demographic questions; and to that utilities are requested to add assessed property value for the address of the participant and the Census Block Group (CBG) number associated with the area in which the participant lives. Using the CBG number it is possible to obtain the addresses for all of the properties in the geographic area, demographic profiles for those addresses and property values from assessors' records. From the point of view of discovering participants' identities, the most worrisome data element is the assessed property value.

During the summer of 2010, FSC surveyed a representative sample of program participants and non-participants in an AC load control program in which customers were asked a number of questions including some of the key questions in the recommended Enrollment and Pre-treatment surveys. The information potentially of use to someone trying to identify individuals from that survey are: single versus multifamily home, square footage, the number of people in the household, age of respondent, education, income, gender and CBG number.

The CBG was known for each survey respondent – which accurately reflects the information available to anyone with access to the SGIG response data. From this information it is possible to ascertain a list of all the street addresses in the Census Block Group and to match these addresses to assessor records to obtain publically available information on property value and to demographic profiles available from companies such as Experian or Claritas (purveyors of information about households from credit records, magazine subscriptions and other sources). In this experiment, the demographic information obtained from Experian consisted of the name of one member of the household, address, city, state, zip, gender, age, and income and education levels.

Each CBG in the study contained between 314 and 499 households. Given the sample size of the survey (about 600) they were spread fairly thinly across a number of CBGs. No more than one or two households were found in each CBG. The task of matching the pseudo-SGIG survey data with the information available in the public record amounted to finding 1 or 2 target "needles" in haystacks containing 300 to 500 pieces of straw.

The cost of assembling the matching information from the public records was found to be significant. Each matching assessor record within a CBG costs $1 and the matching Experian records cost about $0.1 each. The effort to match the survey data to the information in the public record will be expensive – around $300 to $600 *per target observation*. The possible benefits obtained from carrying out the exercise (regardless of the success rate) would have to be quite large indeed to support the exercise. It would not be something one would try, for example, to support a home burglary or some other such low return activity. On the other hand it is impossible to predict the motivations that might drive someone to undertake such a matching exercise.

Given the sizes of the populations in the CBGs, the needed records were assembled for 5 targets, each from a different CBG. To assess the ability of some outside party to find the

needle in the haystack, two analysts[5] were assigned the task of identifying the target individuals in each CBG using only the survey data and varying levels of property value data. One of the analysts analyzed the degree to which targets could be positively identified if housing values were provided directly. This part of the exercise was performed to learn whether information about property value and CBG can be used to directly match pseudo-SGIG participants without other information. The conclusion from this part of the exercise was that it is relatively easy, with fresh assessor record information, to create a near perfect match between pseudo-SGIG participant addresses and the information in the public record. So, providing public access to unaltered assessor data and CBG number along with the SGIG data is definitely not recommended.

The next step in the study was to determine whether it is possible to mask the assessed value of the property in such a way as to eliminate the possibility of matching the pseudo-SGIG respondent addresses with information in the public record.   It was also found that at the tails of the property value distribution, that the relatively few high value assessments cannot be blurred by simple rounding. Some form of top coding will have to be done for these values (i.e., setting everything above a certain point in the distribution – say the 90th percentile to a maximum value).

The purpose of the above described step was to remove all the matches that are more or less *certain* to be found and must be corrected by adjusting the data before it is released. In the next step the pseudo-SGIG data without property values for each target was given to another analyst to try to match with the publically available data on households within the CBGs. This analyst was blind to the identity of the pseudo-SGIG participants and was instructed to list up to 10 households in each CBG that he thought were likely to match the target pseudo-SGIG participants using whatever means he could think of. This analyst completed this exercise for all five target pseudo-SGIG participants. In this case, one of the lists of guesses contained the target household.

The analyst was not told his success rate and was then asked to repeat the exercise, but with the addition of property values rounded to $50,000 and with the top decile of values within each CBG top coded (i.e.. all values in the 90th percentile or above were coded at the value of the 90th percentile. Again the analyst made a list of guesses for each target. In this case, four out of five of the lists of guesses contained the target household.

Again, the analyst was not told his success rate and was then asked to repeat the exercise, but with the addition of property values rounded to $25,000 and with the top decile of values within each CBG top coded. Again the analyst made a list of guesses for each target, and this time the lists were ranked according to the analyst's believed likelihood that each guess was the target. In this case, four of the lists of guesses contained the target household, but none had the target as one of the top two guesses.

---

[5] Both analysts are Ph.D. economists – one an econometrician the other a game theorist.

Again, the analyst was not told his success rate and was then asked to repeat the exercise, but with property values expressed only as deciles within each CBG. Again the analyst made a list of guesses for each target, and again the lists were ranked according to the analyst's believed likelihood that each guess was the target. In this case, four of the lists of guesses contained the target, and one list had the target household as the top guess. None of the other lists had the target as one of the top two guesses.

In conclusion, the above described simple test indicates that releasing the CBG number *and* assessed property value collected from the SGIG data to the public through a FOIA request risks disclosing the identity of the study participants.