MEMORANDUM

TO: Heidi Merchant, AIU

FROM: Riley Newbert, Bill Norton, Mary Sutter, Opinion Dynamics

DATE: May 8, 2009

RE: AIB Tracking Database Review

Task 3 of the AIU workplan indicates that we will verify the information in the program tracking database. Opinion Dynamics has conducted a review of the AIB Tracking database, exported on February 13, 2009. We outline below our findings with regards to outliers, missing values, and potentially missing variables. Generally speaking, the database seems to be well populated and contain the information that we need for our evaluation.

This memo is organized in five main sections:

- ➢ Program Ally Data
- ➢ Participant Data
- ➢ Project Data
- ➢ Measure-Specific Data
- ➢ General Database Recommendations

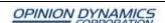We also have an appendix with the listing of all the variables reviewed.

The remainder of this memo discusses the data contained in the February 2009 extract and provides recommendations at the end of each section. While the data extract is now a few months old, we expect that the current database would have similar percentages input (unless SAIC has taken specific actions to clean the data) and reasonable to use for this review.

PROGRAM ALLY DATA

We reviewed all of the tables in the Access database that contained information pertaining to the program allies.

Availability of Ally Contact Information

Based on our review of the program ally data from a sample perspective, we have contact names and phone numbers available for nearly three-quarters (74%) of the 184 registered

allies. We have phone numbers for most (91%) of the 57 allies who are not registered with Ameren but we only have contact names for 21% of all non-registered allies. (See Table 1)

Table 1: Program Ally Contact Information

Registered Ally?	Total number of allies		Includes a phone number		Includes a contact name	
	#	% of total	#	% of ally type*	#	% of ally type*
Yes	184	76.3%	136	73.9%	136	73.9%
No	57	23.7%	52	91.2%	12	21.1%
Total	241	100.0%	188	78.0%	148	61.4%

* Percents are based on the total number of registered/non-registered allies (column 2).

Availability of Other Ally Information

Other information available for the trade allies includes area, expertise, services, customer type, and business information. The information that is most often populated about the program allies is customer type which is available for 78% of the program allies while business information is the least populated field available for only 57% of the program allies. This level of information for the registered allies is adequate for our purposes. While it may be interesting to see if there are systematic differences between registered and non-registered allies as seen in these areas, it probably is not worth the effort to attempt to obtain this as this group is obviously not providing it at this point in time.

Table 2: Availability of Other Information about the Program Allies

Information Type	Registered (n=184)	Not Registered (n=57)	Total (n=241)
Area	94%	7%	73%
Expertise	95%	7%	74%
Services	97%	7%	76%
Customer Type	99%	9%	78%
Business Information	74%	4%	57%

Recommendations

➢ **Include contact names and phone numbers for all of the allies:** In order to contact all of the trade allies as part of our process evaluation we would need a phone number and a contact name would help as well. Overall, 78 percent of phone numbers and 61 percent of contact names are populated.

➢ **Include 'date added' for all trade allies:** It would be helpful for the process evaluation to know when contractors are getting involved in the program process.

PARTICIPANT DATA

We reviewed all of the remaining tables in the Access database which pertained to program participants and actions taken.

Availability of Participant Contact Information

The February 13, 2009 data extract includes 377 projects representing 299 participants (based on ParticipantID) and 188 unique contacts. A contact name and phone number are available for every project, which is commendable as this is a key need for our evaluation. Email addresses are available for all except two of the 188 contacts.

It was not clear if the payee and customer were always the same entity. This may be an issue for the net analysis if we do not have the ability to talk with the decision makers (generally the payee) because we are uncertain who is the payee. Currently, the payee variable is input in 5% of the paid customer records.

Recommendations

➢ **Include payee information for all completed projects:** Only 5 out of the 95 completed projects have information on the payee. In order to complete a case study and talk to all of the decision makers we would like to have this information. If this person is the same as shown in the customer field, it would be useful for this variable to reflect that.

PROJECT DATA

Specific information around project data is needed for the impact evaluation in regards to sample design, ex ante estimates of savings, and the ability to extrapolate to the population if sampling occurs.

As shown in Table 3, one-quarter of the projects in the AIB Tracking database at the time of our extract had been completed, one-quarter were listed as pre-approved, one-quarter were listed as denied and the last quarter were a combination of projects on hold and under review. Obviously, this has changed since we last obtained the data.

Table 3: Project Summary

ProjectType	Denied	On Hold	Pre-Approved	Under Review	Check Cut	Total
Custom	42	8	20	3	9	82
HVAC	4	15	1	3	3	26
Lighting	23	31	70	24	54	202
Motor	0	1	0	0	2	3
Refrigeration	28	4	5	0	27	64
Total	97	59	96	30	95	377

The project summary information is not at issue for our review. However, other variables included in the project tables were studied with the following suggestions.

Recommendations

➢ **Report data in the 'Description' field in a consistent and clear manner:** The data reported in the 'Description' field in the *dbo_Projects* table contains a combination of descriptive values, such as New Construction & Replacement of Failed/Operating Equipment, and numbers. The program implementer indicated that the way that this information was recorded changed along the way. It would be helpful if the data could be fixed so that it is all recorded in the same way and it is clear what the values stand for if numbers are used.

➢ **Clarify what the contents are in the 'Approval Type' field:** It was not obvious upon review how the 'Approval Type' column in the *dbo_Projects* table was used because this field did not seem to correspond to the status column (i.e., many projects are listed in 'Approval Type' column as pre-approval but 'Check Cut' in the 'Status' column). During a conversation with the program implementer, it was explained this field was created because the Standard Application form (not Standard Revised) included an option for the project to go through a pre-approval process or straight to the final approval process. It would be helpful if the "Projects Table Desciption" worksheet in *the AIB Table Descriptions 2-12-2009.xls* file included some of this information.

➢ **Clarify the use of the 'GasAccountNo' field:** It was not obvious upon review of the database how the 'GasAccountNo' field was being used because it was not populated for any of the records in the February 2009 extract data. The program implementer explained that this field only needs to be filled in for participants in the Small Business HVAC Pilot Program who are getting a rebate for gas measures. Again, it would be helpful if the "Projects Table Desciption" worksheet in *the AIB Table Descriptions 2-12-2009.xls* file included some of this information.

OPINION DYNAMICS
CORPORATION

MEASURE-SPECIFIC INFORMATION

This next section focuses specifically on the measures associated with the 95 completed projects which are listed with a status of "Check Cut". While this review is only for part of the projects currently with checks cut, we expect that issues seen here are ones that are systematic within the database and are applicable even now.

Custom Projects
A total of nine custom measures had been completed as of the 2/13/2009 database extract. We have complete information for seven of the nine projects. One project does not include a description and one retrofit project is missing baseline and proposed information. This type of information is especially crucial for custom projects to allow an appropriate choice for sampling as well as the evaluation of impacts

HVAC Projects
A total of three HVAC measures had been completed as of 2/13/2009. All desired data for this end use was present.

Lighting Projects
A total of 120 lighting measures had been completed as of 2/13/2009. KW savings estimates are missing for six of the lighting measures and incentives are zero for eight of the lighting measures. Measure descriptions are missing for four of the lighting measures and a measure code is also missing for one of those projects. Of these missing data, the one that is most crucial is the measure code information. The other data can have a small level of missing data and not cause difficulties in our analysis.

Motor Projects
A total of 2 motor measures had been completed as of 2/13/2009. The information that we would want in the database is complete for both of these measures.

Refrigeration Projects
A total of 74 refrigeration measures had been completed as of 2/13/2009. The information that we would want in the database is complete for both of these measures.

Recommendations

➢ **Include measure codes and descriptions for all completed projects:** Based on a conversation with the program implementer, for the "Standard Revised" projects, only the information that was needed to do the savings estimation was entered. We recommend including a measure code and measure description for all projects.

➢ **Fill-in any missing measure-specific information on completed projects:** For the most part the completed projects include full detail on the type of measure installed and savings associated with that change. However, there are some cases, where there is some missing information. The incompletely input variables that would cause the most difficulty for the impact evaluation are: kWh, kW, and measure code.

OPINION DYNAMICS
CORPORATION

GENERAL DATABASE RECOMMENDATIONS

➢ **Create a Users Guide for database extract users:** The *AIB Table Descriptions 2-12-2009.xls* file and the diagrams of the database relationships provide a good starting point from which one can use the data in the data extract. However, it would be helpful to have a more developed User Guide to help understand the data. Examples of information that we would like to be included in that user guide are:

- Table and data field descriptions,

- Description of missing data values such as '9999' in the AllyID field (when applicable) and '9999999999' in the Account Number field,

- Table relationships, and

- Data warnings such as:

  – There is some data that is used for test purposes only (ProjectID 100472 and measure-specific data with ProjectID 12345). This data should not be included in any database analysis. During a conversation with the program implementer, it was explained that these records are included so that when changes are made they can be tested without changing the data in an active project record. This makes a lot of sense but should probably be documented somewhere.

  – Date fields populated with 1/1/1900 and 1/1/1753 (default minimum dates) have been pre-populated in the database. Users should be made aware that these dates should not be used in any analysis.

OPINION DYNAMICS CORPORATION

Appendix A. REVIEW OF ALL VARIABLES

This appendix provides the listing of all the variables in the tables we reviewed. The variables are separated into groups based on our evaluation need for the information. As shown in Table 4, there are 19% of the variables that are not 100% populated, but we feel should be for evaluation purposes. There are another 7% that are not fully populated, that the evaluation team would like, but if the data are not present, we are OK with it. The remaining data is either all present or not needed by the evaluation team.

Table 4: Evaluation Needs of Database Variables

Evaluation Need	Percent of Variables (n=241)
Not fully populated, but need this information	19%
Not fully populated and would like this information	7%
Population did not matter as we do not need this information	22%
All data present, not an issue	52%

Next is a listing of each variable along with the percent populated and any noted data quality issues for that variable. You will note that many of the variables which are not fully populated are actually populated to a high degree. We recommend that the variables in this first group be looked at carefully and updated. If we have misunderstood anything here, though, please let us know. For Table 6, we have no expectations that SAIC would go back and populate these variables, but if it is possible to do so without much problem, it would be useful.

Those variables which are not fully populated are shown next, followed by those we would like. We do not present any of the remaining variables.

Table 5: Listing of Variables that are Not Fully Populated, but needed by Evaluation Team

N	Table	Field Name	Pop N	# Populated	% Populated	Data quality issues
1	dbo_Allies	ContactName	241	148	61%	
2	dbo_Allies	DateAdded	241	72	30%	
3	dbo_Allies	Phone	241	188	78%	One phone number punched as "0000000000" otherwise numbers seem to be valid and not duplicated
4	dbo_CustomBaselineSummary	BaselineLaborCost	5	4	80%	
5	dbo_CustomBaselineSummary	BaselineMaterialCost	5	4	80%	
6	dbo_CustomBaselineSummary	BaselineOrExisting	5	4	80%	
7	dbo_CustomBaselineSummary	BaselineSystemDescription	5	4	80%	
9	dbo_CustomBaselineSummary	ProposedDescription	5	4	80%	
10	dbo_CustomBaselineSummary	ProposedLaborCost	5	4	80%	
11	dbo_CustomBaselineSummary	ProposedMaterialCost	5	4	80%	
12	dbo_CustomProjectDescription	ProjectDescription	9	8	89%	
13	dbo_Participants	AccountNo	310	308	99%	
14	dbo_Projects	AccountNo	377	376	99.7%	Are Account Numbers beginning in '99999' valid? One record with an account number of '9999999999'
16	dbo_Projects	ContactEmail	377	374	99%	
17	dbo_Projects	IncentiveAmount	377	322	85%	
18	dbo_Projects	kWHsave	377	216	57%	
19	dbo_Projects	kWsave	377	216	57%	
20	dbo_Projects	PayeeAddress	95	18	19%	
21	dbo_Projects	PayeeCity	95	18	19%	
22	dbo_Projects	PayeeEmail	95	18	19%	
23	dbo_Projects	PayeeFax	95	18	19%	
24	dbo_Projects	PayeeName	95	18	19%	

N	Table	Field Name	Pop N	# Populated	% Populated	Data quality issues
25	dbo_Projects	PayeePhone	95	18	19%	
26	dbo_Projects	PayeeState	95	18	19%	
27	dbo_Projects	PayeeTaxID	95	18	19%	
28	dbo_Projects	PayeeTaxStatus	95	18	19%	
29	dbo_Projects	PayeeZip	95	11	12%	
30	dbo_Projects	PrintedDate	95	11	12%	
31	dbo_Projects	PrintedName	95	18	19%	
32	dbo_StandardHVAC	TotalKWSavings	3	2	67%	
33	dbo_StandardHVAC	UnitKWSavings	3	2	67%	
34	dbo_StandardLighting	MeasureCode	120	119	99%	
35	dbo_StandardLighting	MeasureDescription	120	116	97%	
36	dbo_StandardLighting	TotalIncentive	120	112	93%	
37	dbo_StandardLighting	TotalKW	120	114	95%	
38	dbo_StandardLighting	UnitKW	120	114	95%	
39	dbo_StandardRefrigeration	Description	74	73	99%	
40	dbo_StandardRefrigeration	TotalIncremental	74	73	99%	
41	dbo_StandardRefrigeration	TotalKWHSavings	74	73	99%	
42	dbo_StandardRefrigeration	TotalKWSavings	74	73	99%	
43	dbo_StandardRefrigeration	UnitIncremental	74	73	99%	
44	dbo_StandardRefrigeration	UnitKWHSavings	74	73	99%	
45	dbo_StandardRefrigeration	UnitKWSavings	74	73	99%	

Table 6: Listing of Variables that are Not Fully Populated, but desired by Evaluation Team

N	Table	Field Name	Pop N	# Populated	% Populated	Data quality issues
46	dbo_Allies	Address1	241	235	98%	
47	dbo_Allies	City	241	236	98%	
48	dbo_Allies	ContactTitle	241	140	58%	
49	dbo_Allies	Email	241	175	73%	
50	dbo_Allies	Website	241	146	61%	
51	dbo_Allies	Zipcode	241	235	98%	
52	dbo_Participants	RateCode	310	0	0%	
53	dbo_Projects	ActualDateReceived	377	357	95%	
54	dbo_Projects	AllyID	377	240	64%	
55	dbo_Projects	BuilderAccount	?	141	?	'Description' field is recorded inconsistently so we can not currently identify new construction projects and therefore the population size, but we think this is probably a useful field.
56	dbo_Projects	ContactTitle	377	366	97%	
57	dbo_Projects	DateDenied	98	71	72%	
58	dbo_Projects	DSCode	377	179	47%	
59	dbo_Projects	FacilityAC	377	375	99%	
60	dbo_Projects	FacilitySQFeet	377	340	90%	
61	dbo_Projects	FacilityType	377	375	99%	
62	dbo_Projects	ProjIncremental	377	35	9%	
63	dbo_StandardRefrigeration	Facility	74	57	77%	