UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
WASHINGTON, D.C. 20460

OFFICE OF
AIR AND RADIATION

December 10, 2015

Dear Connected Thermostat Provider or Other Interested Party:

The U.S. Environmental Protection Agency (EPA) requests your assistance in assembling data about connected thermostat (CT) performance as measured by the draft metric. A draft of the data elements was released and discussed with stakeholders in early December, informing this final version. The objectives of this data assembly are to inform:

- Selection and improvement of a single metric to be used in the ENERGY STAR CT specification
- Specific requirements for field savings in Draft 3 of the CT specification

This effort will help ensure the ENERGY STAR mark helps consumers find connected thermostat products that deliver cost-effective energy savings and emissions reductions. **Data submitted to Doug Frazee at ICF international by January 8, 2016** will be considered in choosing statistics and setting proposed levels for Draft 3 of the specification. The data requested is described below. Stakeholders may contact EPA directly to discuss questions and concerns.

Data Confidentiality

EPA recognizes there may be concerns with respect to data confidentiality. Data submitted will be shared with EPA and the public in the form of a masked data set, with the submitter's identity removed and contingent upon the receipt of data on at least (3) CT Products. If service providers have additional concerns about data confidentiality, please contact EPA to discuss.

Requested Data Submission

EPA encourages all interested CT manufacturers and service providers to participate by submitting an edited output file from the ENERGY STAR CT Field Savings software for each applicable CT product. The open-source software tool is available for download from < https://github.com/impactlab/thermostat/ >. Software documentation with instructions for installation and use is available at < http://thermostat.readthedocs.org/en/latest/ >. For input data please use a statistically significant data set consisting of CT data collected from January 1 through December 31, 2014 for cooling and between July 1, 2014 and June 30 2015 for heating. Please include at least 250 thermostats randomly selected from customers in each Building America climate zone for each season if possible. Aside from the regional selection, follow the instruction in the Draft 1 Method for Demonstrating Connected Thermostat Field Savings for building a data set. Run the software tool using the Building America Climate Zone/Zipcode Database, supplied by EPA, such that the output file includes data broken out by Building America climate zones. As detailed below, the output file may be edited to include only averages by Building America climate zone.

The format of output data file from statistics module is a tabular CSV file with the following format (there is in fact only one header row):

	Number of day with both heating and cooling					*Number of days with insufficient data*		
	Mean	*Standard error*	*10Th percentile of*	*20th percentile of*	*etc*	*Mean*	*Standard error*	*etc*
National - heating	#	#	#	. . .		#	#	. . .
National - cooling	#	#				#	#	
Region 1 - heating	.					.		
Region 1 - cooling	.					.		
Region 2 – heating	.					.		
Region 2 - cooling								
etc								

Columns:

The data elements (e.g. number of days in a season) are in columns, and most data elements will have 11 columns: one each for mean, standard error and 9 decile bins. The result is a data file with hundreds of columns. We only request the mean and standard error for most data elements, and we do not need every data element. We request decile bins only for the percent seasonal run time savings for each method. There should be a total of 106 columns in the file responding to this request, as detailed in the table below.

data element				*columns*
# thermostat seasons				1
# days in season				2
# days between first and last days of the season				2
# days with both heating and cooling				2
For each modeling method:			*columns per method*	
	Parameters of the fit as applicable (mean and std err for 2 parameters)		4	
		Method 1 (ΔT linear fit): slope, intercept		
		Method 2 (daily average HDD/CDD): Alpha, Delta T base		
		Method 3 (hourly average HDD/CDD): Alpha, Delta T base		
	Mean squared error		2	
	Predicted baseline daily run time		2	
	Predicted baseline seasonal run time		2	
	Seasonal avoided run time		2	
	Seasonal savings (include decile bins for these three data elements only)		13	
		subtotal	23	69
Baseline comfort temperature				2
Actual daily run time				2
Actual seasonal run time				2
All 12 RHU bins				24
			total	**106**

Rows:
There are two rows for each region: one row for heating, and one row for cooling. We are interested in the rows for national data, and the rows for the eight BA climate zones only, which (depending on how much data you have in each zone) could be up to 16 rows. Your raw output data file is likely to contain many more rows than that, but files sent in response to this request need only have up to these 18 rows.

Thank you for gathering data and **submitting it to Doug Frazee at ICF International by January 8, 2016**. If you have any questions regarding it, please feel free to contact me at daken.abigail@epa.gov and 202-343-9375, or Doug Frazee, ICF International, at dfrazee@icfi.com and 443-333-9267.

Sincerely,

Abigail Daken, Product Manager
ENERGY STAR for HVAC