This data quality summary is from Rob Pratt at PNNL:

All data collected by the project is now included in the data set at the request of the sponsor. Some observations at some sites will not pass PNNL’s rigorous data quality sum-check, in which the total load is compared to the sum of each measured end use on an hourly basis. The instrumentation protocol for the project was to measure every circuit, either individually or in combination with others, so the sum of the end use loads are at all times equal to the measured total. It is left to the user to decide whether to include data from any time period for any specific site.

There are four basic reasons why the sum-check may be violated:

1. The circuit panel was modified. This could be due to the addition of a new (unmonitored) circuit. Or, during the course of routine maintenance a circuit became incorrectly monitored because a circuit wire was not replaced correctly through the current transformer. This could leave it out altogether, monitor it twice, etc. This type of error only manifests itself when the load on the incorrectly monitored circuit is large, i.e. a significant fraction of the total (more than a few 1/256 of the total circuit scale). All but the other end uses are valid loads; it can be very difficult to ascertain which one is incorrect, however.
2. The A-to-D converters were all 8-bit, i.e. 1 part in 256 resolution of full scale. Thus, each circuit in the data logger required a proper size current transformer and scaling resistor to be installed, based on the amperage rating of the circuit. As a result, over-current conditions resulting in a reading of more than 256 parts would overflow into adjacent registers of the data logger and “pollute” readings for other end-uses. This may only occur intermittently during high load conditions. Similarly, reversal of a current transformer would result in negative readings that would likewise overflow into adjacent registers. This will occur whenever there is any significant load on the circuit involved. Such problems were automatically detected and field crews dispatch to fix the problems. But, the data should be considered as potentially corrupt, albeit it intermittently, until such problems no longer manifest themselves.
3. Also because the A-to-D converters were 8-bit, an offset of about 3 counts was used to define a reading of zero. This, actual resolution of data was more like 1/253 of full scale. In particular, some offset drift from 3 counts was noted for some circuits at some sites over time. Thus, the user should take care in detecting and interpreting end uses that never go to zero, but should. An example is a water heater. Conversely, other circuits, such as a furnace, heat pump, refrigerator, or air conditioner, will have continual or occasional small loads associated with control transformers, anti-sweat or anti-frosting modes, or other parasitic power draws. User judgment is required, depending on the nature of the analysis being pursued. (The write-up in the attached report discusses the issue further.)
4. Again, because the A-to-D converters were all 8-bit, the criteria for passing a sum-check test can only be considered accurate to approximately 1/253 of full scale for the total end-use. Occasional, random errors can be even larger (1/253 is 1 count for each circuit, several of which may be summed to form end uses at any given site); anything large and systematic is symptomatic of a problem. Thus, when total loads are small, the error in percentage terms can be large. (Note that many circuits may never draw anything close to their rated amperage for an entire hour. The full-scale ratings for the total end use are, unfortunately, not part of the data set.) A scatter plot of the error in percent vs. the total load should produce a symmetric “horn” shaped plot with the wide mouth of the horn at the left, in a healthy data logger. The user is strongly urged to perform this test on each site, on a monthly basis, and use that as a basis for including data from that site for that month in any analysis. Such tests can be automated with some effort.

Despite these issues, and partly because the protocol allows them to be examined with some rigor, the ELCAP data set remains the premier metered end use data set for many purposes. All analyses published by PNNL only used data that passed automated versions of these tests.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

q:\cg\main\data\elcap\raw data from tape\elcap data quality summary.docx