



ENERGY STAR Connected Thermostats

CT Metrics Stakeholder Meeting Slides

March 24, 2021



Attendees

Abigail Daken, EPA

Abhishek Jathar, ICF for EPA

Alan Meier, LBNL

Leo Rainer, LBNL

Eric Floehr, Intellovations

Craig Maloney, Intellovations

Michael Blasnik, Google/Nest

Kevin Trinh, Ecobee

Michael Sinclair, Ecobee

Brad Powell, Carrier

Jason Thomas, Carrier

Theresa Gillette, JCI

Rohit Udavant, JCI

Diane Jakobs, Rheem

Carson Burrus, Rheem

Chris Puranen, Rheem

Glen Okita, EcoFactor

John Sartain, Emerson

 EPA Ko, Emerson

Albert Chung, Emerson

James Jackson, Emerson

Daniel Stephan, Emerson

Mike Lubliner, Wash State U

Charles Kim, SCE

Michael Fournier, Hydro Quebec

Dan Fredman, VEIC

Robert Weber, BPA

Phillip Kelsven, BPA

Casey Klock, AprilAire

Wade Ferkey, AprilAire

Kristin Heinemeier, Frontier Energy

Ulysses Grundler, Trane

John Hughes, Trane

Mike Caneja, Bosch

Sarathy Palaykar, Bosch

Mike Clapper, UL

Alex Boesenber, NEMA

Ethan Goldman

Jon Koliner, Apex Analytics

Hassan Shaban, Apex Analytics

Michael Siemann, Resideo

Arnie Meyer, Resideo

Aniruddh Roy, Goodman/Daikin

Jia Tao, Daikin

Dan Baldewicz, Energy Solutions for CA IOUs

Cassidee Kido, Energy Solutions for CA IOUs

Dave Winningham, Lennox

Dan Poplawski, Braeburn

Natasha Reid, Mysa

Peter Gifford, Mysa

Vrushali Mendon, Resource Refocus



Agenda

- Software updates
- Adjusting our choice on the spectrum of vendor anonymity
- Variable capacity lab test for Communicating Controllers
- CT use case discussion
- Topics from the floor



Software Updates: V2.0

- Version Updates
 - Updated to Python 3.9, numpy 1.20 and Pandas 1.2.x
 - Issue was not with eeweather, but how we were testing zero days of data that was failing.
 - Cleaned-up code to avert division-by-zero warnings no days of data
 - Python 3.5 is deprecated, so we have removed support to maintain compatibility with later releases of Python, Pandas, and Numpy.
- Better Logging
 - Thermostats that are kicked out are logged in a separate file (default: thermostat_import_errors.csv) along with the reason
 - We've silenced more logging errors so it shouldn't be quite as noisy about things that you can't control (zero days, averaging on zero, etc.)
 - Added thermostat ID to warning messages to help with tracing down issues



Software Updates: V2.0

- Where to get it?
 - Version 2.0.0a3 is available in PyPI
 - `pip install thermostat==2.0.0a3`
 - Latest releases are available at the following address:
 - https://github.com/EPAENERGYSTAR/epathermostat/tree/feature/epathermostat_2.0

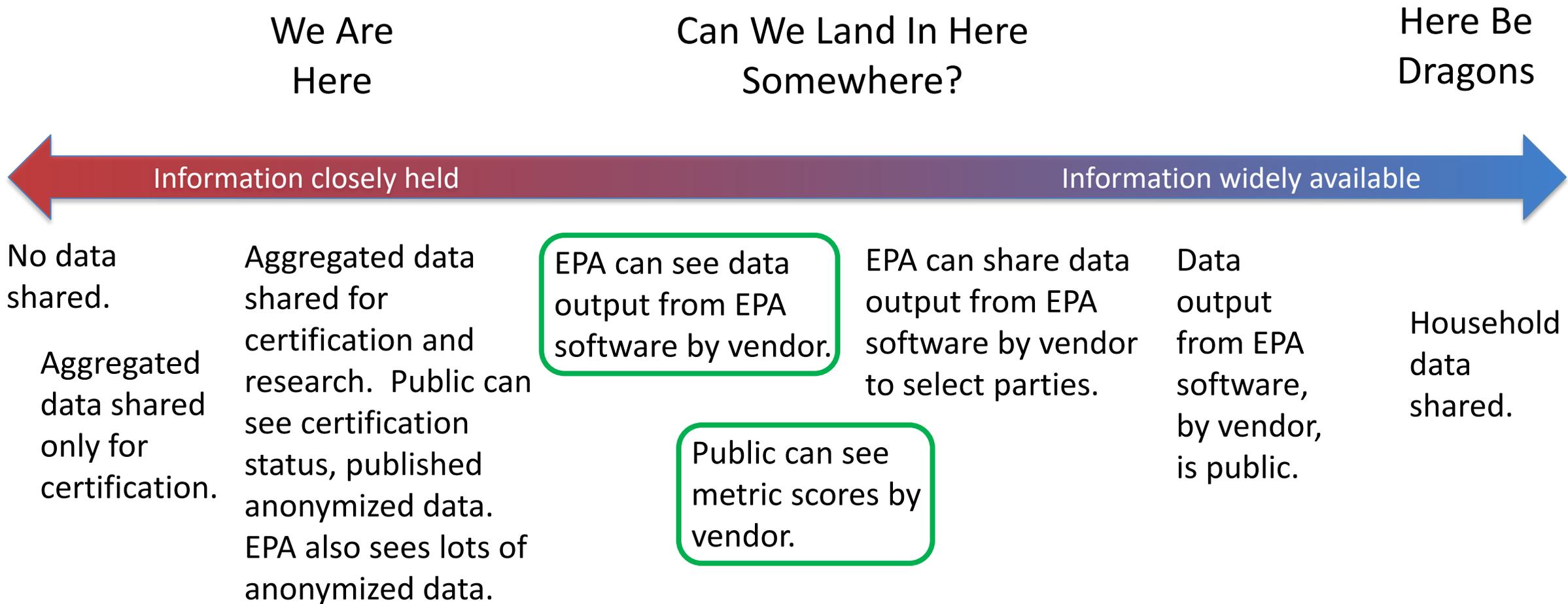


Discussion: Software Updates: V2.0

- All of these notes refer only to V2.0; V1.7.x are unchanged.



Vendor anonymity is a spectrum: WHAT can be seen by WHOM





Vendor Anonymity

- Right now, we are taking a very conservative stance on vendor anonymity
 - Only ICF folks (EPA contractors) discuss issues with data submission with the vendor; if EPA and LBL team participated, we would know who the vendor is
 - Recent real example: vendor has only 20% of sample of thermostats make it through to the statistics module – why?
 - Significantly slows metric progress
- We have also had questions from several vendors and other stakeholders about publishing metric scores
 - EPA has hesitated because we aren't confident metric scores accurately rank products for energy savings
 - However, publishing the scores might help us answer that question and improve the metric faster
- We would not recommend the extreme stance on the other end of the spectrum

Polls on Vendor Anonymity (the poll itself should be anonymous)

- Question 1 (for Vendors only, multiple select): What level of sharing is your organization comfortable with?
 - Please keep vendor anonymity where it is
 - Share metric scores publicly
 - EPA and team can see software output by vendor
 - EPA can share software output by vendor with select parties
 - Other
- Question 2: Are there any concerns you'd like to share anonymously that we should be considering for this decision? (Yes/No)
 - Please email to: connectedthermostats@energystar.gov

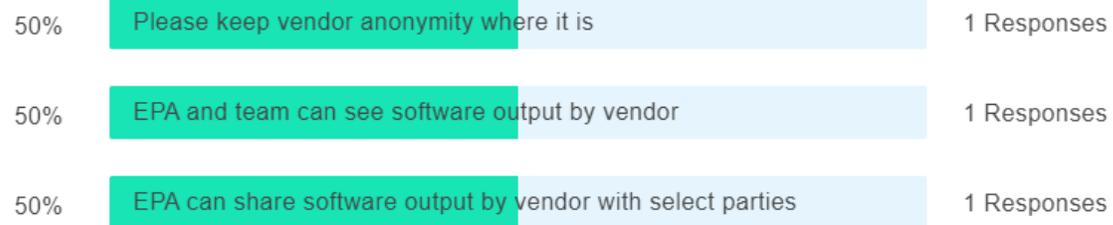


Discussion: Vendor Anonymity

- Poll results

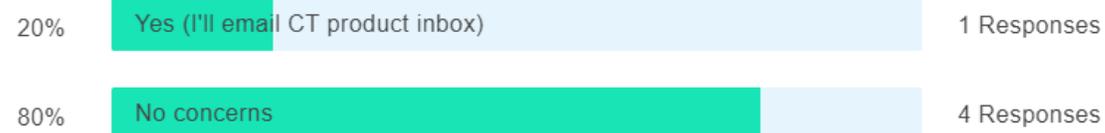
1 of 2. What level of sharing is your organization comfortable with?

Multiple choice with multiple answers



2 of 2. Are there any concerns you'd like to share anonymously that we should be considering for this decision?

Multiple choice with single answer





Discussion: Vendor Anonymity

- Poll results
- Would anyone like to argue for additional advantages of sharing more data?
 - As someone trying to make sense of submissions, being able to talk to vendors would make an extraordinary difference, and would have simplified the use case discussion significantly.
 - Would be useful to tie thermostat results to existing M&V reports
- Anything vendors or others are comfortable saying about this in this forum?
 - Struck by the trouble we've had extending the metric to cover a wider variety of savings measures, and this is likely to only increase given the way the heating market is going. Better able to develop better metrics with more data openness.
 - E.g. control scenario where you could run the fan to harvest additional heat – easier to correlate what's physically happening with thermostats in homes to metric scores in conversations with vendors.
 - Allows tying feature sets to savings, such as remote occupancy/temperature



Discussion: Vendor Anonymity

- Might be more comfortable sharing metric scores with public if there were more than one metric, e.g., comfort temperatures based against some fixed arbitrary comfort temperature
- Note that we're talking about sharing the "summary statistics file" (advanced_stats.csv), which has the detailed aggregated statistical information, but not information per thermostat
- A less thorough sharing would be just those summary statistics relevant to certification, which are in the certification.csv
- If properly anonymized (e.g., not including zip codes) might be able to share the per-thermostat data file as well.
- Will not be deciding until we've talked to each of our partners individually.

Discussion: Vendor Anonymity (Output File description)

- thermostat_example_stats.csv
 - Contains summary statistics for the thermostats. Submitted by vendors
- thermostat_example_stats_advanced.csv
 - Contains statistics with advanced filtering
- thermostat_example_certification.csv
 - Contains metric scores for filtered data of national weighted mean (all region)
 - https://epathermostat.readthedocs.io/en/feature-epathermostat_2.0/tutorial.html#certification-file
- thermostat_example_output.csv
 - Contains metrics that are passed onto the stats module. Not submitted by vendors
 - https://epathermostat.readthedocs.io/en/feature-epathermostat_2.0/data_files.html#thermostat-output
- thermostat_import_errors.csv
 - Contains Errors related to the inability to create thermostats
 - https://epathermostat.readthedocs.io/en/feature-epathermostat_2.0/data_files.html#thermostat-import-errors



Lab test for controllers for variable capacity systems

Context: Where we've been and how we're getting to this subject (Skipped due to absence of relevant vendors)

- We've been pursuing a method to evaluate the savings of communicating controls based on field data
- Stakeholders said "good" control of variable capacity systems means
 - Avoid cycling at higher capacity at lower loads
 - Intelligent setback and recovery to avoid use of high-capacity states and (for heat pumps) resistance heat
- Have gotten as far as we can on a field data method w/o more data
- Based on stakeholder interest in December, discussed the possibility of using a lab test instead in our February meeting
- Still not sure we'll use, but let's continue the discussion



Summary of February discussion: What factors do we need to consider?

- Coordination with activities for testing variable capacity products (at DOE, CSA, and elsewhere) will be important; may delay development.
- Several participants noted that one possible problem is that labs with psychometric chambers are essentially fully booked for the next 2-4 *years*; there would need to be a pretty big payoff to get lab time.
- Test points drive expense: More test points is more expensive but more representative. If we go this way we'd want something simple.
- ASHRAE 90.1 potentially interested in looking at control of mini-splits



Several proposals for what such a test might look like came up

- **Round robin with equipment:** Round robin with communicating control being swapped out for non-communicating control with the same type of equipment to test the communicating control (needs psychometric chamber).
- **Small chamber + simulation:** Small chamber with a building model temp profile and look at output of the signal it would be sending to equipment, rather than looking at the behavior of the equipment itself. Could simulate response of home by changing small chamber. (no psychometric chamber, but those signals typically proprietary)
- **Building simulation with thermostat emulator:** Examine whether the control works well with a variety of buildings and equipment combinations (Now we have way more variables, because of the complexity of the home as simulated. How would we handle that variety? Also would rely on simulation of equipment being accurate.)
- **Field Hybrid:** Define a test sequence that units could run through in field installations to evaluate behavior in specific circumstances, without affecting user experience



But I'd like to backup to look at what we're trying to do: what to test?

- What would we want to test in a lab? What can we test? Possibilities include:
 - Recover from setback (or decide not to set back) in a way that saves energy.
 - Validate set points for DOE testing as per the VRF CVP.
 - Avoid cycling at higher capacity to meet lower loads w/in turndown ratio
 - Others?
- What would the lab test leave un-tested?
 - Interaction between user behavior and control capabilities.
 - What else?



Discussion: What would we want to test? What would we be unable to test in a lab?

- notes



Discussion: Specific test ideas

- notes



Use Cases: Summary of Discussions

- Goal was to identify use-cases where current metric would give misleading results or unfairly penalize a vendor
- We identified >10 use-cases and considered:
 - Fraction of CTs in this category
 - Impact on metric, inside temperatures, runtimes
 - Problems in samples
 - Other factors

#	Use Case
1	SF detached home (1 thermostat)
2	Vacation home
3	SF home (>1 thermostat)
4	Multiple thermostats on a single account
5	SF home with multiple temperature sensors
6	Small commercial with own HVAC
7	Apartment with own HVAC
8	Duplex home, multiple thermostats, different accounts, same dwelling
9	Variable capacity heating or cooling
10	2-stage system
11	Dual fuel



High-Level Conclusions

- Limited data available on use-cases
 - Anecdotal or limited metadata
 - Vendors often used creative methods to infer presence of different use-cases
- Most non-standard use-cases:
 - 1) increased variance but not clear if they biased results
 - 2) were already filtered out in the software
 - 3) were a small fraction of the samples
- No significant use-case behaved so differently as to require revisiting metric. (No “show-stoppers”)
- More metadata is needed to reliably identify different use-cases



Next Steps

- EPA will not, at this time, make any adjustments to the metric in order to account for different use-cases
- EPA may increase sample size in order to take into account increased variability caused by different use-cases
- EPA will periodically re-visit use-cases to ensure the credibility of the metric
- EPA encourages further investigation of use-cases
 - Vendors are invited to report on internal studies
 - Targeted investigations by researchers, universities, utilities, and other groups



Discussion: CT Use cases

- Some use cases (e.g. vacation homes) did have significantly metric scores – if some vendors have much different proportions of these use cases, EPA would need to somehow take that into account.
- An A/B test with representative sample vs. specific use case and it showed a statistically different result
- Could also see whether the CVRMSE for installations with the specific use case were very high
- However, we'd then need to understand how CTs are classified into different use cases, which either means a standard for what you can pull out of the data or living with whatever vendors get their users to report
- Would it be relatively simple to ID the units that are being controlled by geofencing?
 - Know when geofencing is active isn't hard, but how much of that year counts as "using geofencing"?
 - How does fencing distance play into that?
 - Note that whether or not a home relies on geofencing may also correlate to other aspects that affect thermostat use, like whether a family has kids



Discussion: CT Use cases

- What kind of metadata could be collected that might actually help? Some vendors knew some of it, others vendors knew nothing
 - Many vendors can identify when multiples of their thermostats are in a single home
 - Could also have yes/no/NA if some vendors don't know
 - Where is the occupancy detection (on-device/geofencing/remote-devices)
 - Presence of remote temperature sensors and whether the system is ever relying on them
 - Housing type (multifamily/single family/etc)
- For ecobee customers, they can decide how to use remote temperature sensors to be used for run time decisions – blended temperature is what ecobee submits for “indoor temp”. Likely adds more noise, but no systematic variance.
- Are we interested only in hard metadata (that can be derived from data) or also soft metadata (reported by the user)? Given that users are often just wrong, and vendors could be asking questions differently or categorizing things differently. Zip code is entered by users, but likely more accurate than fuel type, but a few folks clearly avoid giving the right one.
- All of these small cases become more important if EPA establishes a minimum sample size.



Discussion: CT Use cases

- Housing type is generally soft report, but at least one vendor gets self-report which they believe to be basically accurate, and others have a sense of what market they are mostly selling into
- Other implicit but not explicit: do users use a mobile app, do they interact with thermostat itself, do customers set a schedule – but does it give us any data that might be used for the metric?
- Going back a step, looking for metrics that might have a significant impact on metric scores, and what's being dropped or thrown out
- Do some of the data collected reflect on the usability of the thermostat? Do people not touch the tstat because they don't know how to use it? Not clear how you'd tease that out.
- We've also talked about holds in the past – users who like holds are easy to find. Not clear that holds always lead to more energy use.



Topics from the floor

- We'd also like to hear what you are interested in
 - What questions would you like us to explore in these meetings?
 - How do you wish the ENERGY STAR thermostat program was different?



Discussion: Topics from the floor

- Software version differences
 - In one vendor's submission, there were some discrepancy even though the samples were the same.
 - EPA currently has V2 samples from 2 vendors, haven't been able to share.
 - May be differences based on which thermostats get included.
- How to get credit for energy-saving features that don't get credit in the metric
 - Features that impact comfort temperatures: relatively easy to develop a metric, maybe harder to understand the impact
 - Fan over-run: running the fan a little longer after the compressor goes off
 - Britain has decided to use a metric reflecting an average indoor temperature, which will be better correlated to energy use rather than energy waste – Michael will connect us
 - How often do you also capture fan run time (not harvesting, set by customers); seconded. Collect fan-only run time at first. More and less efficient ways to use fan to circulate air or destratify. Can be a noticeable amount of energy.



Discussion: Topics from the floor

- Fan usage: is there a ratio of power consumption from the fan to the compressor?
 - Now use of ECM motors may increase fan power some (not enough to wash out their savings) because they push harder for higher ESP
 - For older less efficient PSC fans, ratio is 1 to 4 or 5 relative to the compressor
- A/B for a specific feature, or some other randomized control trial that might be added to the metric score
- Revisit the metadata for whether we can pull out use cases for monitoring
- Tracking sub-sets of product families, at least after the products have been on the market for a while.
- Did RHU change in 2020 due to the pandemic? Less setback, but also milder winter. EPA will look at data we have, which is limited because there's no heat pump oversample.
- What about standby power? Reconsidering? Yes, that's on the table for V2.0 – haven't talked about it here because it's not a metrics topic. Look for the V2.0 discussion guide in Q2.
- Reconsider climate zone weighting? Yes, in the discussion guide.



Discussion: Topics from the floor

- Some of these topics we might come to a different conclusion if we decide to make metric scores public, because then the metrics might be used differently. Publish metric scores by climate region? Middle ground: fallback for small sample sizes to combine with other climate zones? Or as in UK, calculate national metric from whatever regions have enough data.
- Is any thought being given to weather adjustment? Any such effect is currently washed out by statistical noise – perhaps if we go to larger samples, perhaps we'd see it more clearly and could consider addressing in version 3.
 - We probably have all the data we need to take this into account, if we want to
 - Interesting to look at average % saving score vs. the difference between comfort and average temperature – this will be affected by weather and by climate. (% savings per degree of setback.) Fairly well behaved. Does annual weather variation affect ΔT enough to matter?