



1 **ARTICLE INFORMATION**

2 **Article title**

3 The ‘SmartNIALMeter’ Electrical Appliance Disaggregation Dataset

4 **Authors**

5 Manuel Vogel<sup>1</sup>, Martin Friedli<sup>1</sup>, Martin Camenzind<sup>1</sup>, Guido Kniesel<sup>1</sup>, Christoph Klemenjak<sup>1</sup>, Gianni  
6 Gugolz<sup>1</sup>, Patrick Huber<sup>1</sup>, Alberto Calatroni<sup>1</sup>, Lukas Kaufmann<sup>1</sup>, Andreas Rumsch<sup>1</sup>, Andrew Paice<sup>1</sup>

7 **Affiliations**

8 <sup>1</sup>Lucerne University of Applied Sciences and Arts, Engineering and Architecture, iHomeLab, Horw,  
9 6048, Switzerland

10 **Corresponding author’s email address and Twitter handle**

11 \*Corresponding author: manuel.vogel@hslu.ch

12 **Keywords**

13 non-intrusive load monitoring (NILM); load disaggregation; NILM datasets; power signature; electric  
14 load

15 **Abstract**

16 Electrical disaggregation, also known as non-intrusive load monitoring (NILM) or non-intrusive  
17 appliance load monitoring (NIALM), attempts to recognize the energy consumption of single  
18 electrical appliances from the aggregated signal. This capability unlocks several applications, such as  
19 giving feedback to users regarding their energy consumption patterns or helping distribution system  
20 operators (DSOs) to recognize loads which could be shifted to stabilize the electrical grid. The project  
21 “SmartNIALMeter” brought together universities, companies and DSOs and involved the collection  
22 of a large data corpus comprising 20 buildings with a total of 100 electrical appliances for a period of  
23 up to two years at a sampling interval of five seconds. The variability of the loads, including heat  
24 pumps and a charging station for electric vehicles, and the presence of single-phase and three-phase  
25 devices make this dataset suitable for several investigations. The total consumption was collected  
26 through smart meters and each appliance’s consumption was measured with a dedicated sensor,  
27 providing sub-metering for all loads. The dataset can be used to tackle several open research  
28 questions, for example to investigate new NILM algorithms able to learn with a limited amount of  
29 sub-metered data.

30

31 **SPECIFICATIONS TABLE**

<b>Subject</b>	Energy Engineering and Power Technology
<b>Specific subject area</b>	Measurements of total electrical power consumption from 20 buildings with smart meters and 100 appliances with submeters or on-/off-sensors using a sampling interval of 5 s.
<b>Type of data</b>	Raw data Processed data Filtered HDF5 [1]
<b>Data collection</b>	<p>The aggregated power consumption of each building was measured with the Landis+Gyr (L+G) model E450 3-phase smart meter installed by the utility. Measurements were relayed through the L+G meter Consumer Information Interface (CII) to the Smart-me ‘L+G Module’, which uploaded the data to the Smart-me cloud every 5 s.</p> <p>Single appliance consumption measurements were done with 3-phase meters, single-phase meters or socket-plugs from the vendor Smart-me or with an on/off detection sensor developed by the iHomeLab (iHL). The Smart-me meters copied their measurements every 1 s to the Smart-me cloud and the iHL on/off sensor copied measurements of a 1 s interval to a locally installed Raspberry Pi computer.</p>
<b>Data source location</b>	20 Swiss residential homes in the areas of Zürich, Bern, and Lucerne.
<b>Data accessibility</b>	<p>Repository name: Zenodo</p> <p>Data identification number: <a href="https://zenodo.org/record/10875987">10.5281/zenodo.10875987</a></p> <p>Direct URL to data: <a href="https://zenodo.org/records/10875988">https://zenodo.org/records/10875988</a></p> <p>Instructions for accessing these data: Select link URL to data and download the compressed files “raw.7z” or “preprocessed.7z”. Further instructions can be found on <a href="https://github.com/ihomelab/snm-dataset">https://github.com/ihomelab/snm-dataset</a></p>

32

## 33 VALUE OF THE DATA

- 34 • The "SmartNIALMeter" (SNM) dataset has significant value to the scientific community  
35 because of the well-conceived measurement concept. Particularly in the case of load  
36 disaggregation, the sampling interval of five seconds and the availability of data of up to two  
37 years per household contribute towards training robust NILM models with desirable  
38 generalization abilities.
- 39 • Although SNM is not the first residential NILM dataset from Switzerland, the data provided  
40 by 20 Swiss households allows deep insights into regional characteristics, usage patterns and  
41 habits of residents.
- 42 • SNM can be used to analyse usage patterns of appliances and trends in electricity  
43 consumption, from which conclusions on conserving energy and increasing energy efficiency  
44 could be drawn.
- 45 • Besides wide-spread appliance types such as fridges and dishwashers, SNM provides energy  
46 consumption data of a few EV chargers and heat pumps, which is rarely found in related  
47 work.

48

## 49 BACKGROUND

50 In the global push for decarbonization, Switzerland aims to enhance energy efficiency and transition  
51 to renewable energy sources (RES) by 2050, focusing on electrifying sectors traditionally dependent  
52 on fossil fuels such as transportation and heating. This shift introduces challenges due to large  
53 electrical loads and fluctuating RES production, complicating grid management. The SNM project  
54 addresses these issues by exploring NILM, a concept pioneered by Hart in 1985 [3], [4]. NILM  
55 techniques analyse aggregate electrical consumption from a central point, such as a smart meter, to  
56 identify individual appliance usage and energy consumption without requiring direct measurement.  
57 NILM unlocks two sets of applications: (a) Distribution system operators (DSOs) can infer the  
58 presence of specific appliances which might be suitable for peak shaving. For instance, activating a  
59 heat pump with a brief delay can go unnoticed, yet this timing adjustment can play a crucial role in  
60 lowering peak demand on the electrical grid. (b) Individuals, companies and building administrators  
61 can receive feedback about the energy consumption of certain appliances to help reduce the  
62 consumption or to diagnose faults. The SNM project contributed by creating a dataset reflecting Swiss  
63 energy peculiarities, facilitating the development and validation of NILM algorithms.

64

## 65 DATA DESCRIPTION

### 66 Data

67 For the residential data we chose the Hierarchical Data Format (HDF5), which has been developed for  
68 big datasets and fast access [1], [5]. We publish two versions of the SNM dataset - a raw version with  
69 minimal curation steps and a version with more extensive preprocessing applied. Both versions of the  
70 dataset are organized along the same structure: Appliances are saved individually as HDF5 and  
71 grouped according to the building they belong to. Measurements from individual phases are denoted  
72 by the ending L1, L2 or L3 in the file header (e.g. active power L1). This leads to the following file  
73 structure: `<type>/building_<x>/<appliance>.h5`, where:

- 74 • `<type>` denotes the type of the dataset, i.e. raw or preprocessed.



- 75 •  $\langle x \rangle$  is a unique integer assigned to the building.  
76 •  $\langle appliance \rangle$  is the name of the measured appliance. The naming follows the NILM metadata  
77 convention [6].

## 78 Metadata

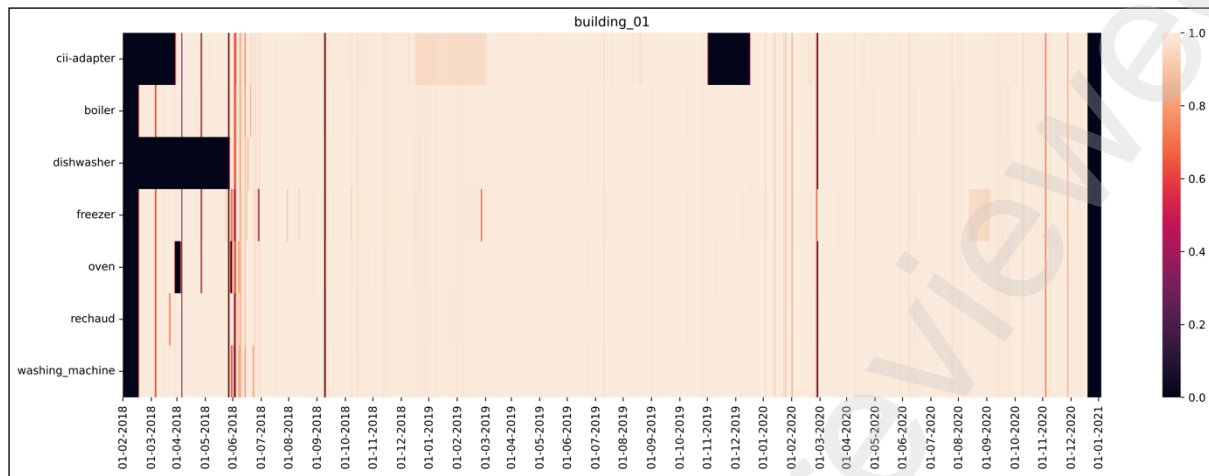
79 The provided metadata is identical for both datasets. It follows the NILM metadata convention [6],  
80 which is documented in [7]. The metadata is organized in multiple yaml files that can be found in the  
81 directory *metadata* in the repository *github.com/ihomelab/snm-dataset*. The metadata is distributed  
82 into three different file types with decreasing generality:

- 83 1. *dataset.yaml* contains metadata relevant to the whole dataset, e.g. the location, a short description,  
84 the publication date and the creators.
- 85 2. The file *meter\_devices.yaml* describes all sensors used in the measurement process. This includes  
86 information about the device model and the manufacturer, a short description, the sampling  
87 interval, the type and ratings of the measurements.
- 88 3. *building<x>.yaml* models the measurement setup of the respective building. Our measurement  
89 setup is described in the form of a tree, where at the root there is the main meter, and the leaves  
90 are the sub-meters. This file is organized as follows:
- 91 • *elec\_meters* contain the list of sensors installed in the building and how they are connected to  
92 the others. Every sensor has a number associated with it: The numbers one to three denote the  
93 three phases of the site meter, i.e. the smart meter measuring the building's aggregated  
94 consumption. These three metering points are marked by the variable *site\_meter = true*.  
95 Every other sensor is a sub-meter of one of these three sensors/phases. This is specified with  
96 the variable *submeter\_of*. The model used for metering and the path to the data file is also  
97 specified for every sensor. If the time where there is data present of the sub-meter deviates  
98 from the timeframe of the smart meter (see item timeframe below), this is also specified.
  - 99 • *appliances* contain the information about the sub-metered appliances. The NILM metadata  
100 convention issues a controlled vocabulary for the naming of the appliances which can be  
101 found on this GitHub page [8]. For every appliance the corresponding sensor is mentioned,  
102 together with the appliance type.
  - 103 • *timeframe* gives the period when data are available from the smart meter. The sub-meters  
104 inherit this timeframe, if not specified otherwise.

105 Please note that this setup builds on inheritance: For instance, values that are not provided in  
106 *building<x>.yaml* are inherited from *dataset.yaml*, i.e. the geolocation is the same as provided in the  
107 file *dataset.yaml* for all buildings.

## 108 Heatmaps

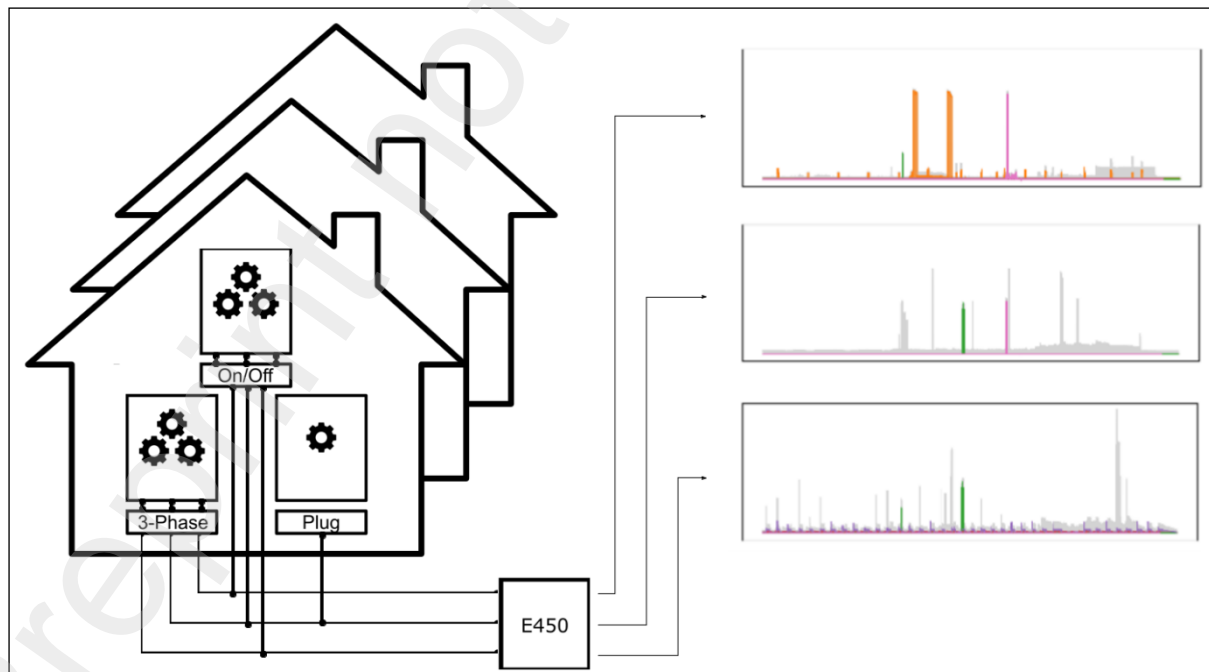
109 Heatmaps have been created for each building, showing the percentage availability of data for each  
110 device in the corresponding building. Figure 1 shows an example for building 01.



111  
112 **Figure 1:** Overview of the data available within the recordings of building\_01. The colour coding expresses the percentage  
113 of data which are present for a certain time slot. The brighter a region the higher the percentage of the available data. All 20  
114 heatmaps are available in the repository [github.com/ihomelab/snm-dataset](https://github.com/ihomelab/snm-dataset) in the *heatmaps* directory.  
115

## 116 EXPERIMENTAL DESIGN, MATERIALS AND METHODS

117 The data collection has been carried out in two phases. We first deployed the measurement setup to  
118 seven buildings (numbers 1 to 7). During this phase we improved the measurement and data  
119 acquisition chain. More installations were subsequently rolled out in the pilot phase (buildings 8 to  
120 20). The following sections give a detailed description of the measurement equipment, the data  
121 collection processes and the data preprocessing steps. Figure 2 shows a schematic overview of the  
122 study design.



123  
124 **Figure 2:** Schematic overview of the study design. The power consumption of various households in Switzerland has been  
125 metered with a L+G E450 smart meter. This meter records the power consumption on all three phases individually. Various  
126 appliances have been sub-metered to record either their respective power consumption on a single or all three phases or their  
127 on/off state.

128 **Measurement Equipment**

129 In each building, we measured the overall electrical power consumption and that of specific  
130 appliances. The list of pieces of equipment used to collect the dataset, along with detailed  
131 specifications and images, is presented in Table 1 and in Figure 3.

Device name	Details	Measures	Rate
L+G E450 + Smart-me L+G Module	The L+G E450 is the smart meter used to measure the aggregated power, whereas the Smart-me L+G module acts as interface to integrate these readings into the Smart-me cloud. It uploads the root mean square (rms) of the current (Irms), voltage (Vrms) and power factor ( $\cos \phi$ ) measurement for each phase every five seconds to the Smart-me cloud.	Vrms, Irms, $\cos \phi$	0.2 Hz
Smart-me plug	A single-phase meter from Smart-me. It is inserted between the application and the socket and measures active power (P) and power factor ( $\cos \phi$ ) every second and uploads the data to the Smart-me cloud.	P, $\cos \phi$	1 Hz
Smart-me single-phase	A single-phase meter from Smart-me that is mounted on a DIN rail. It measures active power (P) and power factor ( $\cos \phi$ ) every second and uploads the data to the Smart-me cloud.	P, $\cos \phi$	1 Hz
Smart-me 3-phase	A three-phase meter from Smart-me that is mounted on a DIN rail. It measures active power (P) and reactive power (Q) every second and uploads the data to the Smart-me cloud.	P, Q	1 Hz
iHL on/off sensor	A low-cost magnetic field sensor developed at the iHomeLab for this project. It measures the magnetic field around the power cable every second and sends it to a Raspberry Pi. This data is used to determine if the application is on or off.	1/0	1 Hz

132 Table 1: Overview of the deployed sensors used for data sampling. All Smart-me devices upload data directly to the Smart-me cloud. The on/off sensor relays the data to a Raspberry Pi that uploads it to a Google cloud. Additional information on  
133 Smart-me devices and the L+G smart meter can be found at [9], [10]  
134  
135



136  
137 Figure 3: Employed sensors: a) left Smart-me L+G module, right L+G E450 smart meter. b) The two single-phase sensors  
138 (left Smart-me single phase, right Smart-me plug) c) Smart-me three phase sensor. d) the developed iHL on/off sensor.  
139

## 140 **Aggregate Consumption**

141 The aggregate power consumption of each building was measured with the smart meter installed by  
142 the utility company. All smart meters were Landis+Gyr model E450. This type of meter provides rms  
143 voltage and current values and the phase angle for all three phases every five seconds. Measurements  
144 are relayed through the L+G Consumer Interface (CII) to the Smart-me 'L+G Module', that feeds the  
145 data to the Smart-me cloud. The module is connected to the internet through a Wi-Fi link.

## 146 **Appliance Consumption**

147 Measurements of single appliance consumption (also called sub-metering in the NILM jargon) were  
148 performed with different sensors. For single phase appliances, we deployed three types of devices:

- 149 • Smart-me single-phase meter. This is a commercially available device which is mounted in the  
150 electrical panel, attached to the DIN rail. This device is placed in series with the current flow and  
151 measures the active power and phase angle.
- 152 • Smart-me plug. This device is very similar to the previous one but is mounted in series to the  
153 power cord of the appliance between the plug and the wall socket. It also measures the active  
154 power and phase angle.
- 155 • iHL on/off sensor. We developed this cheap sensor system to detect whether an appliance is in the  
156 ON or OFF state. Generally, due to the proximity of Phase and Neutral in an AC power cord, the  
157 magnetic field cancels itself in the far field. However, close to the power cord there exists a  
158 residual magnetic field due to the asymmetry of the cable assembly. Our sensor is equipped with  
159 three small inductors which measure the residual magnetic field at three different locations close  
160 to the surface of the power cord. This magnetic field is proportional to the current flowing in the  
161 cable and therefore – assuming a fixed grid voltage – proportional to the apparent power of the  
162 appliance. By setting a manually configured threshold we were able to estimate the ON and OFF  
163 state of an appliance. The sensor connects via Bluetooth Low Energy (LE) to a nearby Raspberry  
164 Pi which takes care of further processing.

165 We rolled out as many commercial devices as we had budget for. During the pilot phase, some of the  
166 single-phase measurements were conducted with the iHL on/off sensor.

167 For three-phase appliances, such as stoves, ovens, boilers and some washing machines, we used the  
168 'Smart-me 3-phase meter', which is mounted in the electrical panel exactly like its single-phase  
169 counterpart.

170 Table 2 shows an overview of the devices and meters used in each building for different appliances.

171

172

buildings	total consumption	washing machine	fridge	freezer	tumble dryer	boiler	stove	dish washer	heat pump	oven	EV charging station	comfort ventilation	coffee machine	dehumidifier	total
1	L+G	3		P		3	3	3		3					6
2	L+G	P	P			3	3	P						P	6
3	L+G	3	P		3	3	3	3		3					7
4	L+G	P	P		P		3*	P		*	3, 3				7
5	L+G	P	P		P		3	P				1	P		7
6	L+G	3	P		P	3	3	P				P			7
7	L+G	3*	P, P	P	*	3	3	3							7
8	L+G	P	P	P	P										4
9	L+G		P	P			O								3
10	L+G	P	P	P	P		O								5
11	L+G	P	P	P	P	O									5
12	L+G	P		P			O	P	O						5
13	L+G		P	P							P				3
14	L+G	P	P	P	P										4
15	L+G	P		P	P			P							4
16	L+G	P	P	P	P				O						5
17	L+G			P	P										2
18	L+G	P	P	P	P				O						5
19	L+G	P	P	P	P			P	L+G						6
20	L+G			P	P										2
total		16	16	15	14	6	10	10	4	2	3	2	1	1	100

**Table 2:** Overview on the appliances measured in each building and on the employed meter. Each row corresponds to a building, columns list the available appliances. The naming of the applications follows the NILM metadata convention [6]. Letters in the table have the following meaning: L+G → Landis+Gyr E450 smart meter combined with the Smart-me L+G module; P → Smart-me plug; 1 → Smart-me single phase meter; 3 → Smart-me three phase meter; O → iHL on/off sensor. Buildings 4 and 7 have two appliances of the same type indicated by comma separated letters. The last column and row give the total numbers of the sub-metered appliances. For devices marked with \*, the same submeter was used for two appliances in the building, whereby only the first occurrence is counted in the table.

### 181 Technical Validation

182 The measurement instruments provided by L+G and Smart-me fulfill the requirements enforced by  
 183 the Swiss regulations on the measurement of electrical energy and power, which define three  
 184 precision classes ‘A’, ‘B’ and ‘C’ [11]. The reliability of our measurements can be categorized as  
 185 follows:

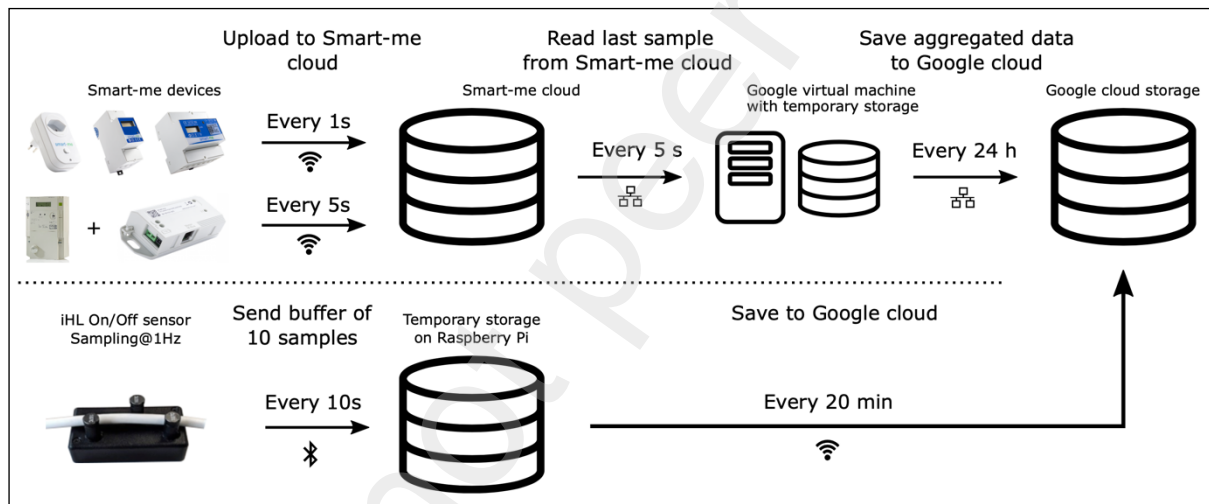
- 186 • The L+G E450 smart meters fulfill class ‘B’ precision and have therefore a maximum relative  
 187 error of 1 %.
- 188 • All Smart-me meters fulfill class ‘B’ requirements and have a maximum relative measurement  
 189 error of 1 %.



- 190 • The iHL on/off sensor was thoroughly tested and the threshold on the analog-to-digital converter  
191 (ADC) values recorded from the magnetic field measurements was carefully adjusted until the  
192 data did not show any mistakes in the on/off signal.
- 193 • The time synchronization of the devices attaching timestamps to the samples was obtained by  
194 synchronization with network time protocol (NTP) servers. Typical variations in the  
195 synchronization are in the order of a few milliseconds and rarely reach a few dozens of  
196 milliseconds.
- 197 • Despite our best efforts to avoid it, several data are missing. The fraction of available data is  
198 depicted in heatmaps for each building (see example for building\_01 in *Figure 1*).

## 199 Data Collection

200 In this section we describe the hardware part of the measurement infrastructure. We now illustrate  
201 how the data were gathered. The initial part of the data path is specific to the sensor type, while  
202 afterwards all data get centralized to the Google cloud storage. A schematic overview is provided in  
203 Figure 4.



204  
205 **Figure 4:** Schematic overview of the data collection process. The upper half depicts the data flow for the Smart-me devices  
206 and the lower half for the iHL on/off sensor.  
207

## 208 Smart-me Devices

209 The Smart-me L+G module uploads one measurement point every five seconds to the Smart-me  
210 cloud. The sensors for sub-metering at the appliance level upload one measurement point per second.  
211 In the Smart-me cloud, the data points are stored together with the UTC timestamp measured upon  
212 reception of each sample. The Smart-me cloud offers a live view of these data through a web  
213 interface. However, historical data are stored only with 15 minutes temporal resolution. To store the  
214 higher-rate data, we set up a virtual machine on the Google cloud that requests continuously the  
215 newest samples from the Smart-me cloud using their API. Samples from all Smart-me devices are  
216 pulled every five seconds irrespective of the original sampling frequencies. On the virtual machine,  
217 samples are appended to comma-separated value (CSV) data files. Once per day, the CSV files are  
218 compressed and stored in a Google bucket. The file names contain the measurement date.

219 As it appears clear from the setup, samples obtained from individual sensors are not synchronized:  
220 The virtual machine always requests the newest measurement from the Smart-me cloud, which can be

221 up to five seconds in the past in the case of the L+G module and one second in the case of the Smart-  
222 me sub-meters.

223 The described data collection process exhibited occasional connection issues: First, if a sensor loses  
224 its wireless connection, no new sample is uploaded to the Smart-me cloud and the newest value on the  
225 cloud remains unchanged. Therefore, the virtual machine repeatedly reads the same value(s), resulting  
226 in duplicated samples in the CSV file. These values can be easily filtered out since the timestamp  
227 itself also remains unchanged. In the raw dataset, these connection issues show up as periods with  
228 missing data. Second, whenever the virtual machine loses the connection to the Smart-me cloud, it  
229 cannot read any new value, resulting in missing values in the CSV file. To detect these problems, we  
230 set up a script on the virtual machine that checks for duplicate and missing values and sends a daily  
231 report. From the reports received, we were able to see if there was a need to reboot any of the wireless  
232 routers or to reestablish the connection to the Smart-me cloud. Figure 1 illustrates the data availability  
233 using building 01 as an example.

234 The virtual machine reboots every morning at 02:15 local time (CET). This ensures that all the tasks  
235 that read the data from the Smart-me cloud are rebooted and are running properly for the next day.  
236 However, since no new data can be read from the Smart-me cloud while rebooting, this leads to a data  
237 loss of about 30 seconds. As time stamps of the released dataset are given in UTC time, this reboot  
238 happens at 00:15 UTC during the Central European Summer Time (CEST) between the last Sunday of  
239 March and the last Sunday of October and at 01:15 UTC during Central European Time (CET).

#### 240 ***iHL on/off Sensors***

241 The iHL on/off sensors use a different data logging process. Each sensor records the magnetic field  
242 strength around the cable every second. Every ten seconds, the buffer containing ten samples  $x_0 \dots$   
243  $x_9$  is sent over a Bluetooth LE connection to a nearby Raspberry Pi. On the Raspberry Pi, the  
244 corresponding timestamps  $t_0 \dots t_9$  are set so that the newest timestamp  $t_9$  is the time measured upon  
245 reception, and all previous timestamps are decreased, i.e.,  $t_{i-1} = t_i - 1 \forall i \in [1, 9]$ . To reduce  
246 synchronization issues, the Raspberry Pi is synchronized with a NTP server. The time/value pairs are  
247 then appended to a local CSV file. A periodic process (cronjob) on the Raspberry Pi synchronizes the  
248 locally stored files every 20 minutes with the Google cloud through the rsync Linux command. Data  
249 from the iHL on/off sensors are not compressed. To minimize the risk of data loss, a copy of the data  
250 is also saved locally on the SD card of the Raspberry Pi.

#### 251 ***Privacy Preservation***

252 To ensure privacy, all sensors are identified with a random id on the Google cloud. This id can only  
253 be mapped to the according metadata such as sensor type and building with the help of a local  
254 configuration file. The data storage is organized in a hierarchical manner: Data from every building is  
255 collected in a distinct directory, wherein every sensor has its own subdirectory, containing all the  
256 daily CSV files.

#### 257 **Processing of Measured Data**

258 We published two versions of the dataset: The raw version was subjected to basic data curation steps,  
259 while the preprocessed version underwent outlier removal and some further feature extraction.  
260 Because of the large amount of data, we use dask [12] to efficiently parallelize the data processing.  
261 The code used to generate the preprocessed version is available on [github.com/ihomelab/snm-dataset](https://github.com/ihomelab/snm-dataset)  
262 in the *src* folder.

263 **Raw Dataset**

264 We performed the following steps to obtain the published raw dataset:

- 265 • In some buildings, the measurement infrastructure was changed during the data collection. In such  
266 cases, the sensor id changed and the data before and after the change were saved in two different  
267 folders, even if belonging to the same appliance. These cases were dealt with by manually  
268 copying the data to the appropriate directory.
- 269 • All duplicated, incomplete and not-a-number (NaN) values are deleted from the data records.
- 270 • The raw measurements of the iHL on/off sensor, which are acquired through an ADC from the  
271 magnetic field, are turned into a binary on/off signal and the original ADC values are discarded.  
272 This is done by applying a fixed threshold to the ADC data. The threshold was tuned to be  
273 slightly above the noise floor and proved to be effective in the development of the sensor.
- 274 • The data are saved to multiple HDF5 files as discussed in the section Data Description.

275 **Preprocessed Data**

276 The preprocessed dataset is obtained from the raw data with the following procedure:

- 277 • The different Smart-me sensors provide different outputs (see Table 1). We used the following  
278 equations to calculate the active power P, reactive power Q and apparent power S for all sensors:

279 
$$P = S \cdot \cos \phi \qquad Q = \sqrt{S^2 - P^2} \qquad S = P / \cos \phi = V_{\text{rms}} \cdot I_{\text{rms}}$$

280 where  $V_{\text{rms}}$ ,  $I_{\text{rms}}$  and  $\cos \phi$  are rms voltage, rms current and power factor respectively.

- 281 • We perform outlier detection and remove the corresponding data points. We flag as outliers all  
282 measurement values exceeding the sensor rating and negative active power values.
- 283 • The data are then saved to HDF5 files, see the section Data Description.

284 **LIMITATIONS**

285 Limitations exist in the sense of data outages due to the temporary failure of individual routers, meters  
286 or Smart-me devices during recording. Therefore, heatmaps have been created for each building,  
287 showing the percentage availability of data for each device in the corresponding building. All 20  
288 heatmaps are available in the repository [github.com/ihomelab/snm-dataset](https://github.com/ihomelab/snm-dataset) in the heatmaps directory.

289 **ETHICS STATEMENT**

290 The authors confirm that they have read and follow the ethical requirements for publication in Data in  
291 Brief. They also confirm that the current work does not involve human subjects, animal experiments,  
292 or any data collected from social media platforms.

293

294 **CRedit AUTHOR STATEMENT**

295 **Manuel Vogel:** Data curation, Formal analysis, Software, Validation, Visualization. **Martin Friedli:**  
296 Data curation, Validation, Software, Writing – Original Draft, Writing – review & editing. **Martin**  
297 **Camenzind:** Investigation, Software, Writing – review & editing. **Guido Kniesel:** Project  
298 administration, Writing – Original Draft, Writing – review & editing. **Christoph Klemenjak:** Writing  
299 – review & editing. **Gianni Gugolz:** Data curation, Formal analysis, Software, Visualization, Writing  
300 – Original Draft, **Patrick Huber:** Writing – Original Draft, Writing – review & editing. **Alberto**  
301 **Calatroni:** Writing – review & editing. **Lukas Kaufmann:** Investigation, Project administration.



302 **Andreas Rumsch:** Conceptualization, Funding acquisition, Project administration, Supervision,  
303 Writing – review & editing. **Andrew Paice:** Funding acquisition, Supervision, Writing – review &  
304 editing.

## 305 ACKNOWLEDGEMENTS

306 The project was supported by the pilot and demonstration program of the Swiss Federal Office of  
307 Energy (SFOE) [13].

308 The following additional stakeholders were part of the project:

- 309 • Bits to Energy Lab at ETH Zurich
- 310 • EKZ (One of Switzerland's largest DSOs operating in the canton of Zurich)
- 311 • Energie Thun (local DSO in and around the city of Thun)
- 312 • Landis+Gyr (provider of smart meters and complete energy solutions)
- 313 • Smart-me (provider of smart meters and cloud solutions for energy metering and billing)

314

## 315 DECLARATION OF COMPETING INTERESTS

316 The authors declare that they have no known competing financial interests or personal relationships  
317 that could have appeared to influence the work reported in this paper.

## 318 REFERENCES

- 319 [1] “The HDF5® Library & File Format,” The HDF Group. Accessed: Mar. 19, 2024. [Online].  
320 Available: <https://www.hdfgroup.org/solutions/hdf5/>
- 321 [2] “SmartNIALMeter – Smarte Lastaufschlüsselung mit Smart-Meter - Texts.” Accessed: Mar.  
322 27, 2024. [Online]. Available: [https://www.aramis.admin.ch/Texte/?ProjectID=40224&Sprache=en-](https://www.aramis.admin.ch/Texte/?ProjectID=40224&Sprache=en-US)  
323 [US](https://www.aramis.admin.ch/Texte/?ProjectID=40224&Sprache=en-US)
- 324 [3] “Hart1985.pdf.” Accessed: Mar. 19, 2024. [Online]. Available:  
325 <https://www.georgehart.com/research/Hart1985.pdf>
- 326 [4] G. W. Hart, “Nonintrusive appliance load monitoring,” *Proceedings of the IEEE*, vol. 80, no.  
327 12, pp. 1870–1891, Dec. 1992, doi: 10.1109/5.192069.
- 328 [5] “The HDF Group - ensuring long-term access and usability of HDF data and supporting users  
329 of HDF technologies,” The HDF Group. Accessed: Mar. 22, 2024. [Online]. Available:  
330 <https://www.hdfgroup.org/>
- 331 [6] J. Kelly and W. Knottenbelt, “Metadata for Energy Disaggregation,” in *2014 IEEE 38th*  
332 *International Computer Software and Applications Conference Workshops*, Jul. 2014, pp. 578–583.  
333 doi: 10.1109/COMPSACW.2014.97.
- 334 [7] “Welcome to NILM Metadata’s documentation! — NILM Metadata 0.2.0 documentation.”  
335 Accessed: Mar. 25, 2024. [Online]. Available: <https://nilm-metadata.readthedocs.io/en/latest/>
- 336 [8] “nilmtk/nilm\_metadata.” nilmtk, Jan. 12, 2024. Accessed: Mar. 25, 2024. [Online]. Available:  
337 [https://github.com/nilmtk/nilm\\_metadata](https://github.com/nilmtk/nilm_metadata)
- 338 [9] “Smart Energy Metering Technology - smart-me AG.” Accessed: Mar. 25, 2024. [Online].  
339 Available: <https://web.smart-me.com/en/technology/>
- 340 [10] “E450 Haushaltszähler mit integriertem PLC-Modem,” Landis+Gyr. Accessed: Mar. 25,  
341 2024. [Online]. Available: <https://www.landisgyr.ch/product/landisgyr-e450/>
- 342 [11] “SR 941.251 - Verordnung des EJPD vom 26. August 2015 über...,” Fedlex. Accessed: Mar.  
343 25, 2024. [Online]. Available: <https://www.fedlex.admin.ch/eli/cc/2015/578/de>
- 344 [12] “Dask | Scale the Python tools you love.” Accessed: Mar. 25, 2024. [Online]. Available:



345 <https://www.dask.org/>  
346 [13] S. F. O. of E. SFOE, "Homepage." Accessed: Mar. 27, 2024. [Online]. Available:  
347 <https://www.bfe.admin.ch/bfe/en/home.html>  
348

Preprint not peer reviewed