Plan Overview

A Data Management Plan created using DMPonline

Title: Twomes data collection Zwolle winter 2021-2022

Creator: Henri ter Hofte

Affiliation: Other

Template: DCC Template

Project abstract:

The purpose of the project Twomes is to find out which measurement data (with which precision and frequency) is needed to allow inverse grey-box modelling algorithms to learn crucial parameters of a specific home (a specific building, a specific heating installation and behaviour of a specific household) automatically, and accurately enough for more targeted advice for the heating transition in existing homes, an important part of the energy transition that is needed to avoid dangerous climate change.

ID: 65167

Start date: 01-02-2020

End date: 30-04-2022

Last modified: 24-06-2022

Grant number / URL: https://brains4buildings.org/about/

Copyright information:

Twomes data collection Zwolle winter 2021-2022

Data Collection

What data will you collect or create?

1. name + informed consent*
2. home address*; this is used
    1. as the destination address to deliver measurement devices
    2. as an index to retrieve public building data from online sources such as [PDOK](#), [3D BAG Viewer](#), [PICO (geodan.nl),](#) [EP-Online](#) and public weather data such as [KNMI Dataplatform](#)
3. properties of the home:
    1. internet connection*;
    2. wireless internet (Wi-Fi)*;
    3. whether the home is heated by gas-fired heating boiler* and/or other means;
4. make, model and photos of the following installation items:
    1. smart energy meter*;
    2. central heating boiler*;
    3. thermostat;
    4. ventilation system.
5. about the occupants:
    1. Does at least one occupant have a smartphone?*
    2. How many of the other occupants have a smartphone?
    3. Are all occupants prepared to leave on Bluetooth during the measurement period?

* the answers to these questions serve as inclusion/exclusion criteria

1. indoor climate (temperature, CO_2 concentration, relative humidity): several times per hour
2. smart energy meter readings (gas, electricity) via the P1-oport: a few times per hour
3. heating system data:
    1. supply and return temperature: several times per minute
    2. whether the boiler is firing and for what purpose (central heating / warm water): a few times per minute
    3. thermostat setpoints and programs (whenever they change);
4. ventilation data (a few times per hour)
5. presence data: whether smartphones respond and what the received signal strength indication is, in response to Bluetooth name requests to static Bluetooth addresses of the smartphones of the home occupants, from one or more measurement devices in the home (several times per hour

1. smart energy meter readings (gas, electricity) via the P3/P4 port and an EDSN-certified ODA:
    1. (5-minute) interval meter readings (up to 10 days in the past);
    2. daily meter readings (up to 40 days in the past);
    3. monthly meter readings (up to 13 months in the past)

The volume of data will be several GB for tens of homes measured for several months. We dimensioned the relational database at our Twomes backend server such that it can handle this with ease.

How will the data be collected or created?

We collect data via the following means:

- **data under 1-5 above:**
  will be collected via online mobile survey hosted on https://windesheim.eu.qualtrics.com/
- **data under I - V above:**
  will be measured via automated measurement devices in the home, enabled by the WarmteWachter app (published by Windesheim in the Google Play Store and Apple App store of Windesheim), uploaded via a secure connection over home Wi-Fi and internet to servers at energietransitiewindesheim.nl, hosted by Strato AG
- **data under A above:**
  will be measured via Enelogic.nl, an independent service provider ('ODA') certified by [Energie Data Services Nederland (EDSN)](#) that subjects authorize to access their smart meter measurements) to which respondents give us access.

Data quality is assured as follows; for:

- **data under 1-5 above:**
    - we use validated questions from the [WoON survey](#) wherever possible;
    - for brand and model of the installation devices, we allow users to select the most common brands and models;

moreover, we ask users to take pictures of the devices as an answer to the mobile survey.

- **data under I - V above:**
  - clocks on measurement devices are synchronized several times per day over the internet via NTP;
  - we always make a distinction between the value 0 (zero) and no data;
  - measurements are timestamped using the device clock;
  - just before measured data is uploaded, the upload is timestamped using the device clock and just after arrival at the server, the upload is timestamped using the server clock, making timestamps during analysis more robust against device clock inaccuracies (like time skew);
  - data is buffered in persistent memory of the measurement device whenever data fails to upload, and will be uploaded from persistent memory once connectivity is restored, thus making the data collection more robust to connectivity failures between measurement device and server;
  - each device records 'heartbeat' measurements (every 10 minutes) and uploads them every hour; heartbeats of measurement devices are monitored by back-end researchers and action is taken via the front-end helpdesk to signal respondents, when no heartbeats are received for days, thus making the data collection more robust to device failures and long term connectivity issues;
  - we currently do not plan to introduce validity checks before uploading measurement data, other than standard security measures against buffer overflow attacks;
- **data under A above:**
  - except for the 13 month monthly interval back-service, these measurements are redundant compared to data item II and can serve as repeated sample.

Documentation and Metadata

What documentation and metadata will accompany the data?

All measurements consist of a device_id, property_id, timestamp, and value. Timestamps of measurements will always be recorded in Unix time. A device_id is always linked to a building_id and for each building_id we store a timezone name, from the list of tz database time zones. This allows us to establish both global (UTC) time for each measurement as well as the local clock time at which the measurement was established.

The properties and units for each measurement device type are encoded in the database and are published in the GitHub repository file twomes-backoffice-api/sensors.csv at master · energietransitie/twomes-backoffice-api (github.com)

We do not yet use an energy measurement ontology for the property.name and property.unit values, but are considering to use an ontology like Project Haystack (project-haystack.org) or Brick (brickschema.org), at the latest at the moment we publish measurement data online as open data (after proper anonimisation)

Below, please find the current definitions:

DeviceType.name,DeviceType.DisplayName,Property.name,Property.unit DSMR-P1-gateway-TinTsTr,Slimme meter monitor met temperatuursensoren,heartbeat, DSMR-P1-gateway-TinTsTr,Slimme meter monitor met temperatuursensoren,eMeterReadingSupplyLow,kWh DSMR-P1-gateway-TinTsTr,Slimme meter monitor met temperatuursensoren,eMeterReadingSupplyHigh,kWh DSMR-P1-gateway-TinTsTr,Slimme meter monitor met temperatuursensoren,eMeterReadingReturnLow,kWh DSMR-P1-gateway-TinTsTr,Slimme meter monitor met temperatuursensoren,eMeterReadingReturnHigh,kWh DSMR-P1-gateway-TinTsTr,Slimme meter monitor met temperatuursensoren,gMeterReadingSupply,m^3 DSMR-P1-gateway-TinTsTr,Slimme meter monitor met temperatuursensoren,eMeterReadingTimestamp, DSMR-P1-gateway-TinTsTr,Slimme meter monitor met temperatuursensoren,roomTemp,°C DSMR-P1-gateway-TinTsTr,Slimme meter monitor met temperatuursensoren,boilerTemp1,°C DSMR-P1-gateway-TinTsTr,Slimme meter monitor met temperatuursensoren,boilerTemp2,°C DSMR-P1-gateway-TinTsTr,Slimme meter monitor met temperatuursensoren,gMeterReadingTimestamp, DSMR-P1-gateway,Slimme meter monitor,heartbeat, DSMR-P1-gateway,Slimme meter monitor,eMeterReadingSupplyLow,kWh DSMR-P1-gateway,Slimme meter monitor,eMeterReadingSupplyHigh,kWh DSMR-P1-gateway,Slimme meter monitor,eMeterReadingReturnLow,kWh DSMR-P1-gateway,Slimme meter monitor,eMeterReadingReturnHigh,kWh DSMR-P1-gateway,Slimme meter monitor,gMeterReadingSupply,m^3 DSMR-P1-gateway,Slimme meter monitor,eMeterReadingTimestamp, DSMR-P1-gateway,Slimme meter monitor,gMeterReadingTimestamp, OpenTherm-Monitor,OpenTherm monitor,heartbeat, OpenTherm-Monitor,OpenTherm monitor,roomTemp,°C OpenTherm-Monitor,OpenTherm monitor,boilerSupplyTemp,°C OpenTherm-Monitor,OpenTherm monitor,isBoilerFlameOn, OpenTherm-Monitor,OpenTherm monitor,maxBoilerCap,kW OpenTherm-Monitor,OpenTherm monitor,maxModulationLevel,% OpenTherm-Monitor,OpenTherm monitor,minModulationLevel,% OpenTherm-Monitor,OpenTherm monitor,boilerReturnTemp,°C OpenTherm-Monitor,OpenTherm monitor,relativeModulationLevel,% OpenTherm-Monitor,OpenTherm monitor,isCentralHeatingModeOn, OpenTherm-Monitor,OpenTherm monitor,isDomesticHotWaterModeOn, OpenTherm-Monitor,OpenTherm monitor,roomSetpointTemp,°C OpenTherm-Monitor,OpenTherm monitor,listRSSI,[dBm] OpenTherm-Monitor,OpenTherm monitor,listHomePresence,[]

CO2-meter-SCD4x,CO_2-meter,heartbeat, CO2-meter-SCD4x,CO_2-meter,CO2concentration,ppm CO2-meter-SCD4x,CO_2-meter,roomTemp,°C CO2-meter-SCD4x,CO_2-meter,relativeHumidity,%RH CO2-meter-SCD4x,CO_2-meter,listRSSI,[dBm] CO2-meter-SCD4x,CO_2-meter,listHomePresence,[]

Presence-Detector,Aanwezigheidsdetector,heartbeat, Presence-Detector,Aanwezigheidsdetector,listRSSI,[dBm] Presence-Detector,Aanwezigheidsdetector,listHomePresence,[]

Ethics and Legal Compliance

How will you manage any ethical issues?

1. Data collection plans are reviewed by the Windesheim ethics board
2. We inform subjects about the privacy policy repeatedly, which is continuously available:
   [Privacyverklaring WarmteWachter (energietransitiewindesheim.nl)](energietransitiewindesheim.nl)
   This has a short summary as well as more complete coverage and we give information how to contact us (the front-end helpdesk, principal researcher or privacy officer of Windesheheim) if they have questions
3. Informed consent will be asked from subjects:
   - by informing subjets about the privacy policy before they click on the online survey link;
   - by informing subjects again about the privacy policy at the beginning of the survey and ask explicit consent for data collection phase 1 (online survey), which is recorded as one of the first answers to the survey, together with the name;
   - by informing them of the privacy policy again when they are invited to take part in phase 2 (automated data collection);
   - for each of the measurement devices that subjects install with help of the WarmteWachter app, by shortly summarizing the data this particular measurement device collects, on the screen of the WarmteWachter app, after they scanned the QR-code of the measurement device with the WarmteWachter app.

How will you manage copyright and Intellectual Property Rights (IPR) issues?

We apply the following open licenses to our work:

- **CERN-OHL-P v2** open hardware license to all hardware (designs) for measurement devices;
- **Apache License 2.0** open source license to all:
  - firmware for Twomes measurement devices;
  - source code for the Twomes WarmteWachter app source code, including IONIC5/Capacitor plugin for Espressif Unified Provisioning;
  - software for the Twomes API and backend
- **CC BY 4.0** copyright license to any open data and report that we publish.

... unless we cannot avoid re-using available components, source code and/or libraries that have an incompatible license, such as the GNU GPLv3 source code license. In that case, we seek to publish our work under the least restrictive license possible.

All our open source software, open hardware designs and open data will be published on GitHub under [Research group Energy Transition at Windesheim (https://github.com/energietransitie)](https://github.com/energietransitie)

Storage and Backup

How will the data be stored and backed up during the research?

We keep the following data separate, linked via a pseudonym (a unique, random number per subject)

1. directly identifiable data (such as name, street address, e-mail address)
2. indirectly identifiable data (all other data, which may include a pseudonym)

The online survey responses contain a combination of both directly and indirectly identifiable data (hosted by Qualtrics LLC, located in Provo, UT, USA).

The online survey responses will be exported and directly identifiable data will be replaced by a pseudonym. This pseudonymized data may also be stored in data stores provided by Office 365 services and OneDrive for Business services under the terms agreed by Windesheim with Microsoft, headquartered in Redmond, WA, USA.

Based on the street addresses, specific building data will be downloaded from sources such as PDOK, 3D BAG Viewer, [PICO (geodan.nl),](geodan.nl) and EP-Online and stored as indirectly identifiable information along with the pseudonym.

During phase 2, only indirectly identifiable measurement data will be uploaded to energietransitiewindesheim.nl (hosted by Strato AG, located in Berlin, Germany). During phase 2, we will perform regular backups of measured data (several times per week) on SURFdrive (hosted by SURF B.V., located in Utrecht, the Netherlands).

Windesheim has data processing agreements with Qualtrics LLC, Strato AG, SURF B.V., and Microsoft.

How will you manage access and security?

Directly identifiable information will be stored in a separate 'key file', which links each pseudonym with directly identifiable information such as name, street address and e-mail address.
Access to directly identifiable information is severely restricted, i.e. to:

- front-end helpdesk arranged by Stichting 50 Tinten Groen Assendorp (who are bound to confidentiality based on the contract Windesheim has with Stichting 50 Tinten Groen Assendorp).
- selected key research staff at Windesheim University of Applied Sciences (who are bound to confidentiality based on their employment contract). At Windesheim, this file will be encrypted and stored in a secure location. The encryption key is only avalaible to the principal researcher, one researcher responsible for data management and one person from support staff of the research group Energy transition.

Even though Windesheim is responsible for data processing, we strive to minimize directly identifiable available to Windesheim personnel as much as possible:

- the Front-end helpdesk at Stichting 50 Tinten Groen handles all direct communication with respondents via e-mail addresses, street addresses and translates requirements and questions for Windesheim researchers in terms of pseudonyms.
- the backend device preparation and data analysis are done at Windesheim. The required sets of measurement devices are defined per pseudonym, labelled with a pseudonym and handed off to the front-end helpdesk, who make sure they arrive at the proper street address.

Selection and Preservation

Which data are of long-term value and should be retained, shared, and/or preserved?

An anonymized version of measured data will be published online as open data. This involves:

- making sure that ALL key files that link directly identifiable information with pseudonyms are deleted;
- assessment whether remaining data cannot reasonably be linked to natural persons or individal homes;
- applying additional anonymization techniques when deemed necessary;
- documenting metadata in a way similar to ictinnovaties-zorg/dataset-diabetes-adolescents-time-series-with-heart-rate (github.com).

The remaining anonymized data will be published as open data on https://github.com/energietransitie and/or DANS KNAW under the **CC BY 4.0** copyright license.

What is the long-term preservation plan for the dataset?

See the answer to the "Selection and Preservation" question.

Data Sharing

How will you share the data?

See the answer to the "Selection and Preservation" question.

Are any restrictions on data sharing required?

The default IPR clauses in TechForFuture project contract need to be honoured.
All firmware, software and hardware designs are made by Windesheim so far, giving Windesheim the liberty to choose the licences mentioned earlier:

- **CERN-OHL-P v2** for open hardware (designs);
- **Apache License 2.0** for open source software and firmware;
- **CC BY 4.0** for open data.

These licences all have a permissive nature (without strong or even weak reciprocal licensing obligations), which allows for re-use and adaptation in practice, even in a commercial context.

Even though Windesheim is at liberty to choose a license as we see fit, we will seek approval by the project steering board for this policy explicitly and more generically, to apply this licencing scheme to all results produced in the Twomes project.

Responsibilities and Resources

Who will be responsible for data management?

As principal researcher and project leader, Henri ter Hofte is responsible for implementing the DMP.

He will delegate various data management tasks to Derk-Jan Winkel.

Front-end helpdesk tasks and dealing with directly identifiable information will be delegated to Stichting 50 Tinten Groen, where Kirsten Notten will be the primary person responsible.

What resources will you require to deliver your plan?

The resources are planned as part of the Twomes project. To automate database backup from measurement databases at energietransitiewindesheim.nl (hosted at Strato AG) to SURFdrive (hosted at SURF B.V.), we will seek advice and assistance of the Windesheim IT department.