Plan Overview

A Data Management Plan created using DMPonline

Title: AI4CarbonFreeHeating

Creator: Johanna Ganglbauer

Principal Investigator: Isabella Kolb, Kristina Dabrock, Konrad Andre, Stefano Coss

Data Manager: Stefan Janisch

Project Administrator: Johanna Ganglbauer

Affiliation: Other

Template: DCC Template

Project abstract:

The decarbonization of the Austrian building sector is essential for achieving climate goals. The Renewable Heat Legislation, which envisions the replacement of all fossil heating systems by 2035 (or 2040 in the case of non-renewable gas), is currently under discussion. This step presents financial, logistical, and societal challenges to the Austrian population, which must be addressed under significant time pressure.
Simultaneously, rising temperatures due to the relentless climate change are shifting the requirements for heating, ventilation, and air conditioning (HVAC) systems. The heating demand is trending lower, while the cooling demand in the summer is increasing. This has implications for the choice and sizing of heating systems and the need for air conditioning.
To facilitate the actual implementation of renovations and heating system replacements and to prevent hasty decisions despite the tight timeline, the support of decision-makers by experts and user-friendly calculation tools is highly advisable.
For this purpose, above all, a solution that enables a straightforward evaluation of the present condition of building exteriors across a vast number of buildings with minimal effort, is needed. Actual data is often not available or must be tediously compiled from various sources. In the AI4CarbonFreeHeating project, open-source methods for large-scale machine-based capture of building data will be developed for Austria. Using regions with good availability of building data, relationships between simple parameters, such as building footprints, which are widely available in Austria, and poorly recorded building characteristics, such as building heights or construction periods, are learned using machine learning methods. The application of these trained AI methods will subsequently contribute to the provision of a complete building dataset for Austria.
In addition, there is a need to conduct detailed assessments of climate change impacts or renovation measures in individual buildings to perform bottom-up scenario analyses for communities, cities, regions, or even all of Austria. In the AI4CarbonFreeHeating project, digital building models are being developed for this purpose, which can determine the effects of climate change and the effects of various energy efficiency measures on household energy demand profiles. Climate change scenarios and renovation scenarios are then developed and evaluated through large-scale simulations. The results are graphically analyzed and serve as guidance for the development of decarbonization strategies for communities, cities, energy planners, or infrastructure experts.

ID: 132104

Start date: 02-09-2024

End date: 31-08-2027

Last modified: 01-10-2023

Grant number / URL: https://www.ffg.at/ausschreibung/ai-for-green-ausschreibung-2023

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customise it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

AI4CarbonFreeHeating

Data Collection

What data will you collect or create?

Data set	Description	Coverage	Format
Eubucco	Building height (74 %), building type (45 %) construction year period (24 %)	200 million buildings across Europe	gpkg, csv
Open street map	Building footprints and occasionally building heights	Europe - complete	shp
Microsoft building footprints	Building footprints	World – Austria is complete	geojson
ZEUS	Energy certificates	~ 200 000 across Austria (not included: Vienna, upper Austria, Vorarlberg)	
LoD2	3D city model of vienna	Vienna	dxf, shp
STATatlas	Number of buildings per construction year period	raster of 250 m x 250 m	
STATatlas	Number of children per family, number of persons per family, etc.	on municipality level	
smart meter data	directly from users	timeseries data at specific points	csv
district heating demand profiles	directly from users or from heat grid operators	timeseries data at specific points	csv
energy bills	directly from users or from engineering offices	point data	jpg, pdf, csv
energy certificates	directly from users or from engineering offices	point data	jpg, pdf, csv
climate change scenario data	regional climate models	raster of 12.5 x 12.5 km	netCDF
training data set (building-AI)	building layout, building height, building type, construction year, household composition,...	subset of building stock in Austria	sql
input data set (building-AI)	building layout,...	building stock of Austria	
training data set (document-AI)	energy certificates and energy bills		jpg, pdf
input data set (document-AI)	energy certificates and energy bills		jpg, pdf
building data base	building layout, building height, building type, construction year, household composition,...	2e6 entries for 2e6 buildings in Austria	sql
test reference years (TRY)	outside temperature, global radiation, ...	rmunicipality level	netCDF
scenario analyses	annual energy demand, etc.	point data	sql

The building data base for Austria is of long-term value and is planned to be shared via Zenodo under the licence CC-BY 4.0 license, if this possible due to restrictions of input and training data of the applied AI.

The test reference years can be used project internally, but will belong to GeoSphere Austria and will not be accessible for free.

How will the data be collected or created?

The structure will be the following:

- building data:
  - gridded
  - municipality level
  - points
  - time series
- building-AI:
  - training data

    - input data
- document-AI:
    - training data
    - input data
- building data base
- test reference years
- scenario analysis

Each folder will contain a detailed description of the data and a logbook, keeping track of data versioning. The building data base and the training data sets will be incorporated in relational data bases like MariaDB (based on SQL).

Documentation and Metadata

What documentation and metadata will accompany the data?

Data versioning and the use of available software for data versioning will be discussed at the project Kick-Off and a common approach will be chosen. Regular data check-ups (every two months) will be made by either the data manger Stefan Janisch or the coordinator Johanna Ganglbauer, to check if descriptions and versions are up to date and if needed the process is residesigned in such a way that data quality remains high while keeping the effort manageable.

Ethics and Legal Compliance

How will you manage any ethical issues?

Within the project energy bills, energy certificates, and energy demand profiles from participants will be collected and stored. As the data needs to be connected to certain buildings (address) to be useful, anonymization is impossible in that case. But the sensitivity of data can be reduced by extracting only certain features. E. g. the annual sums of energy demand profiles could be computed and incorporated in data bases instead of the original data, which contains much more personal information.
Generally, such sensitive data will only be used within the project with the allowence of the participants.

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

The test reference years will be owned by GeoSphere Austria. The building data base will be made available for everyone. The most critical point is the integration of user data in the training of AI - it is planned to publish only the outputs of AI-processing and not the training data sets.

Storage and Backup

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

The data will be centrally stored in Graz at the office of 4wardEnergyResearch GmbH, and regularly back-uped by company-internal standards.

How will you manage access and security?

The highest risk, is that protected user data could be seen by third parties not having the allowence. To minimize risks the original data will not be shared at all. Processed user data (e. g. the annual energy consumption, which can be extracted from energy demand profiles) will be shared only for users with the right authentification. For such restrictions most probably MariaDB authorization and permission process for SQL Server Users will be implemented.

Selection and Preservation

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

The following data sets are of long-term value:

- training data set of building-AI and document-AI (restricted)
- complete building data base (open)
- results of scenario-analyses (open)
- test reference years (restricted)

The latest versions of the following data sets will be stored locally at the hard disc of 4wardEnergy Research, shared via Zenodo, or stored locally at the hard disc or server of GeoSphere Austria. Especially the open data sets are forseen to be reused in bottom-up scenario analyses of municipalities, or further R&D projects, therefor the data will be preserved for at least five years after project finalization.

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

The long term preservation plan for the open data sets is to maintain it on Zenodo, which is free. The long term plan for the preservation of the restricted data sets is to keep them securely stored in company-internal infrastructure.

Data Sharing

How will you share the data?

Potential users find out about our data via (i) scientific publications and conferences, (ii) through oral dissemination (especially among relevant experts in Austria), (iii) keyword search in Zenodo.

We will make every effort to overcome barriers in order to ensure universal access to the building database for Austria. We will publish the dataset via Zenodo. The data set is planned to be published on 31.05.2027. A DOI will be used as persistent identifier and relevent keywords suitable for the sector will be assigned to the data set to guarantee findability.

Are any restrictions on data sharing required?

Within the AI4CarbonFreeHeating project three types of restrictions need to be considered:

(i) resharing protected data of participants (energy certificates, energy bills, measured heat load or smart meter profiles) will be restricted and we are committed to use the data and not to share it.

(ii) rules of open data sets contributing to the buliding data base (STATatlas from statistic austria, open street map, etc.) may impose certain conditions for the reuse and publication of the data. Special care will be taken to ensure legal reuse.

(iii) the intellectual property of the test reference years (TRYs) belongs to project partner GeoSphere, which will be respected by the consortium of AI4CarbonFreeHeating.

Responsibilities and Resources

Who will be responsible for data management?

Johanna Ganglbauer and Stefan Janisch are responsible for the DMP and ensure that the data descriptions and versioning is up to date every two months throughout the project.

Each partner will be responsible for the subset of data he or she is working with and must follow the common rules in terms of versioning and documentation.

Certain data sets, like e. g. energy demand profiles, energy certificates or energy bills need protection and will be made available only to those who have the rights for it.

The test reference years will be owned by GeoSphere Austria and only shared with those who have the explicit permission. The need for contractual agreements will be discussed at the start of the project

What resources will you require to deliver your plan?

Software for data versioning needs to be investigated, tested and applied within the project. Common standards and automated data quality tests will be applied for data to be incorporated in the relevant data bases. For that purpose best practises will be researched and discussed with experts.

For the most relevant data bases, most probably MariaDB will be used, which is free software.

Maybe additional hard disks will be necesary, but it won't be a big financial effort.

Planned Research Outputs

Event - "building-AI"

AI algorithm developed to complete the building data base for Austria

Collection - "test reference years"

Test reference years for Austria on municipality level considering climate change impacts.

Collection - "Scenario analyses"

Results of the scenario analyses

Collection - "building data set"

This data base will contain basic building information like building layout, building height, building type, building age and residential household composition

---

Planned research output details

Title	DOI	Type	Release date	Access level	Repository(ies)	File size	License	Metadata standard(s)	May contain sensitive data?	May contain PII?
building-AI		Event	2027-01-31	Open	None specified		MIT License	None specified	No	No
test reference years		Collection	2026-09-30	Restricted	None specified		None specified	None specified	Yes	No
Scenario analyses		Collection	2027-06-30	Open	None specified		Creative Commons Attribution No Derivatives 4.0 International	None specified	Yes	No
building data set		Collection	2027-03-31	Open	None specified	1 TB	Creative Commons Attribution No Derivatives 4.0 International	None specified	No	No