

A REVIEW ON COUNTRY SPECIFIC DATA AVAILABILITY AND ACQUISITION TECHNIQUES FOR CITY QUARTER INFORMATION MODELLING FOR BUILDING ENERGY ANALYSIS

A. Malhotra¹, J. Bischof², J. Allan³, J. O' Donnell⁴, T. Schwengler⁵, J. Benner⁶ and Gerald Schweiger⁵

¹RWTH Aachen University, Germany, ²Institut Wohnen und Umwelt, Germany

³EMPA, Switzerland, ⁴University College Dublin, Irland

⁵TU Graz, Austria, ⁶Karlsruher Institut für Technologie, Germany

ABSTRACT

This paper addresses the increasing number of disparate data resources used for urban modelling. The objective of this work is to provide a standardized approach for processing these resources for urban energy modelling studies. This paper details the approach of a collaborative project to standardize categorization, acquisition and processing of diverse datasets for energy modelling and simulations are explained. Furthermore, based on the data categorization, this research provides an overview of the country-specific data availability and sources (for Austria, Germany and Switzerland) required for urban energy simulations. The result is a standardized structure for information exchange which is published in an extendable online template.

KURZFASSUNG

Diese Arbeit thematisiert die wachsende Anzahl unterschiedlicher Datenquellen, die für urbane Modellierungen zum Einsatz kommen. Es wird eine standardisierte Methode entwickelt, welche die Verarbeitung und Nutzung dieser Daten in der energetischen Stadtmodellierung unterstützt. Dafür werden Details eines kollaborativen Projektes erläutert, das sich mit der Standardisierung der Kategorisierung sowie der Erfassung und Verarbeitung verschiedener Datensätze für die Energiemodellierung und Simulationen befasst. Basierend auf der Datenkategorisierung gibt diese Arbeit einen Überblick über die länderspezifische Datenverfügbarkeit und die für urbane Energiesimulationen erforderlichen Quellen für Österreich, Deutschland und die Schweiz. Zudem wird eine erweiterbare Online-Vorlage für eine standardisierte Austauschstruktur für Informationen zur Datenverfügbarkeit veröffentlicht.

INTRODUCTION

The proportion of the global population living in cities is expected to increase from 55% in 2018 to 68% in 2050 (UN DESA, 2018); therefore, the development of urban planning tools is of growing importance. The necessity to make cities cleaner and more resource efficient requires an in-depth understanding of the impact of energy production, distribution and usage,

not just across different energy sectors but also at different scales in order to compare and assess different energy retrofitting options and novel energy system configurations.

In the field of building energy demand modelling, past studies and literature distinguish between “top-down” and “bottom-up” approaches (Reinhart & Davila 2015, Sola et al. 2018). Top-down approaches determine the energy demand in an aggregated form, for example, the aggregated demand of a region is distributed across the buildings it contains; however such approaches are unsuitable for detailed investigations of integrated energy supply and demand scenarios. On the other hand, bottom-up modelling approaches, discussed in Swan & Ugursal 2009, are classified into two categories: 1) engineering and 2) statistical. Engineering models estimate the energy demand of buildings based on physical models of energy and mass flows in and around buildings, and profiles of occupancy and equipment usage. Statistical models on the other hand leverage regression analyses in order to determine correlations between building energy consumption, defined energy end-use categories and climatic conditions (Reinhart & Davila 2015, Sola et al. 2018).

A significant part of both bottom up modelling approaches require energy modelling of individual buildings at a city-quarter or city level. Meaningful or sufficiently accurate modelling of the thermodynamic behavior of the target system potentially requires significant quantities of information for each building. A large number of simulation inputs need to be derived from data resources; therefore, it is critical that these resources be categorized according to the desired modelling objectives.

This paper aims to detail an approach that standardizes the collection and processing of the diverse datasets for energy performance. The work is part of a collaborative project, under the umbrella of the International Building Performance Simulation Association (IBPSA) (Wetter, et al., 2019). The study introduces several universal data categories to facilitate and harmonize data collection for energy performance simulation at an urban scale. This data

can then be used to populate or define data models such as City Geographical Modeling Language (CityGML) (Gröger et al. 2012). This paper introduces a concept of primary, secondary and tertiary data categories: where primary data categories comprise the information that can be directly used in a building simulation (e.g. geometric features), whereas, secondary data includes classifiers, such as building usage and tenure that in combination with statistical information (tertiary data) compliment the primary data or inputs. Secondary data in combination with tertiary data can enrich individual data records. One main purpose of secondary and tertiary data is the creation of archetypes that can infer missing input data.

Within this paper, the definition of different data categories is based on a case study covering available data identified in Germany (D), Austria (A) and Switzerland (CH). Collectively these are labelled DACH. The categorizations are more refined than a similar approach taken in (Reinhart & Davila, 2015), where the creation of building energy models for districts is subdivided into three tasks: Data input, thermal modelling and validation. In many countries and cities, digital building stock models exist but typically lack consistency in terms of formats and quality of data. Different data are gathered in multiple ways in the 'Data Acquisition' phase. Since much of the data sources can be inaccessible to energy modelers due to access restrictions them, this paper broadly defines the most prominent data publishers for the DACH nations (to the best of each authors knowledge). The next stage involves structuring the data into standardized 'Data Models' and the paper provides an overview of some of the commonly used data models for urban energy simulation. The information about the resolution of the data models along with their availability and acquisition (based on the identification of the data sources for the DACH countries) is discussed in this paper. Furthermore, as a part of the next steps, the authors aim to identify major gaps and problems with the DACH specific data for usage in urban building energy simulation. As a result of the current study, a comprehensive universal data categorization structure will be published as a template to store, manage and exchange information for multiple countries. This template can be extended for multiple countries, data types and its sources.

This paper is structured as follows: The Data Categorization section highlights the identified categories and datatypes associated with city quarter modelling; the Data Models section describes the data formats used for energy modeling and simulations; the Availability and Acquisition of Datatypes section discusses the availability and resources of these datatypes. The ultimate contribution of this work, a Universal Data Categorization Template is then presented. This is followed by the Conclusions and Future Work section.

DATA CATEGORIZATION

Starting from building level to an urban scale simulation, the amount of required relevant data for energy analysis can differ with respect to the considered data model and/or simulation environment. Thus, the data used for such simulations largely vary in resolution, usage and management. In order to define the data sources, this paper identifies three major data types:

- Primary (P)
- Secondary (S)
- Tertiary (T)

Table 1 provides an overview of identified relevant data categories, their descriptions, use cases and types.

DATA MODELS

For energy related simulations on city-quarter or city level, a lot of information about each building is needed to replicate the thermodynamic behavior of the system with sufficient accuracy. Such models require position, orientation, geometrical shape, data describing the building's constructional properties and physical state, the structure of its exterior shell (such as size and position of windows), its usage, internal energy gains and HVAC systems. Various data models are available for representing building related information in a machine-readable format, partly or fully supporting the aspects mentioned above. Besides, open and (nationally or internationally) standardized models, some proprietary models also exist but are not regarded in this paper. Data models originating from the Building Information Modelling (BIM) domain, such as Industry Foundation Classes (IFC) (ISO 16739, 2015, Bazjanac & Crawley, 1997) or Green Building XML (gbXML) (Dong et al. 2007), are able to store properties related to the energy performance of a building. IFC is mainly used in the architecture and construction domain but for energy related application it is mostly used in context of academic projects. In contrast, gbXML has been specially developed for energy related applications, and has functional interfaces to simulation environments such as TRNSYS or EnergyPlus (Malhotra et al. 2019, Geiger et al. 2019).

Category	Description	Use	Type
Building footprints	2D polygons of the building footprint	Geometric calculations and estimations involving individual and gross floor area	P
Building height	Height of the building. Often comprised of two measurements - eave height and ridge height	Used to construct 3D volumes and surfaces from 2D data	P
3D building model	3D volumetric model of the building. Varying levels of detail.	Provides geometric information about the outer shape of the building	P
Digital elevation model	Evenly spaced elevation values referenced to a projection or coordinate system	Determines the horizon line and location of shading objects for radiation calculations	P
Built form	Descriptive information about the relationship with other buildings	Determines the boundary conditions for simulation by identifying shared and exposed walls	P
Thermal envelope (wall, roof, floor, glazing types)	Detailed information on the construction and materials used in the elements of the building envelope	The thermos-physical properties of the building elements are used in the simulation to calculate heat flows	P
Building age	Age of the building or time since the last retrofit	Infer typical building materials based on regulations at the time of construction/refurbishment	S
Building use	Usage category of a building	Used to determine internal schedules, temperature and humidity set points required for simulation	S
Occupant Composition	Detailed information about the number, type, age-group of the inhabitants of the building	Used to determine load schedules specific to an occupant type	S
Tenure	Provides information on whether the house is owner occupied, rented or is a social housing	Used to infer information about construction, refurbishment status or scheduling	S
Building Energy System	Existence and parameters of energy systems (Renewables, HVAC, Lighting and DHW) in the building.	Used in calculations to determine the efficiency of the energy system and resulting demand profiles	P/S
Address Data	The street address of a building	Can be used for aggregation, clustering and linking. Address data could also be used to infer information about the age or use of a building.	S
Building Typology	Parameterized buildings models representing a specific category	Used to populate the parameters of missing or questionable data records	T
Stock Development	Rates for new building construction, building demolition and/or building refurbishment.	The stock development data acts as an independent parameter that 'drives' a possible stock development forecast.	T
Energy Generation and Distribution	The availability and description of district heating or external energy resources (e.g. CHP) as well as grid accessibility.	Used to evaluate the reliance and influence of the building on external energy systems	NA
Weather Data	Environmental variables experienced by the building over the simulated period	Directly used in the simulation to determine boundary conditions and heat flows	P

Table 1: Overview of identified relevant data categories, their descriptions, use cases and types for DACH countries.

Moreover, many different standardized building data formats originate from the Geographical Information System (GIS) domain. National cadastral data formats, such as the German Authoritative Cadastral Map Information System (ALKIS) (AdV 2012) or Norm based Exchange Platform (NAS) (AdV 2004), and the internationally used OpenStreetMap (OSM) format are frequently used but represent only a few building properties related to energy. The international GIS standard CityGML and the European standard INSPIRE Building are potential better suited in this regard. Besides the representation of geometric and non-geometric properties several energy related building properties like year of construction, building function, roof type or floor structure are also supported.

The CityGML standard is extensible and its functionalities can be extended through Application Domain Extensions (ADE) (Gröger et al. 2012, Biljecki et al. 2018). ADEs enable extensions of existing CityGML classes with pre-defined properties or definitions of completely new classes that relate to the basic standard. Currently, such an extension supporting all energy relevant aspects mentioned above is under development as the CityGML Energy ADE (Agugiaro, et al. 2018).

AVAILABILITY AND ACQUISITION OF DATATYPES

Despite having well defined data models, it is still a challenge to populate the attributes. Chen, et al. 2018 evaluated the availability of public datasets in San Francisco and found that there was a lack of common keys to enable mapping of data to CityGML. Rosser, et al. 2019 also details how datasets in the UK were mapped to the attributes of an EnergyADE file using two case studies. Nevertheless, the usability of CityGML is at times questionable. Braun, et al. 2018 provides an overview of some of the challenges, such as missing data, inaccuracies and geometric inconsistencies, faced when creating energy models from a CityGML model of Stuttgart, Germany.

The approach taken for data acquisition is highly diverse and depends on the kinds of data available in the country of focus. Although several data sets, with respect to their availability types, exist, some cannot be utilized for scientific usage as they cannot be accessed due to legal restrictions (usage agreements). Therefore, not only the data availability but also the legal data availability type restricts the availability to the user. Some of the most common availability types are summarized below.

Data availability types (DATs)

- Open - Dataset can be generally used free of charge for any purpose (Auer, et al., 2007)
- Public Sector Information (PCI) and Open Government Data - A charge may apply for certain usage e.g. commercial usage (Deutsche Post, 2019)
- Academic - Data can be used free of charge for research purposes (often the case for PCI data)
- Commercial - Data use is licensed for a fee
- Application restricted - Data can only be used with certain applications (e.g. Google Earth)
- Private - Data is not accessible to outsiders

Within the scope of this paper, data holders, for Germany, Austria and Switzerland, who are responsible for the geo-information data and its accessibility are discussed in the sections below. However, different countries may have different techniques and accessibility restrictions with respect to the data. It is therefore the responsibility of countries/users adopting this approach, to identify the accessibility restrictions that apply to them.

Data holders and accessibility

Several institutions in Germany, Austria and Switzerland hold data relevant to building performance simulation. These institutions can be divided into two groups; Governmental and Private Organizations.

Governmental Organizations

The following subsections describe the governmental data holders and initiatives in Germany, Austria and Switzerland.

Germany

In Germany three types of governmental data holders exist. These are the Federal Republic, the 16 Federal States and the more than 10,000 Municipalities (DESTATIS, 2020). For example, in the case of building polygons, the municipality data is aggregated to federal state level (AdV, 2020).

Federal Republic of Germany: Geo-datasets of the federal government are partly open, especially datasets with poorer resolution such as elevation models with a grid size of over 200m, or are commercially available (BKG, 2020). Two data portals offer access to geo-data on the level of the federal republic and the web access is available by the 'Bundesamt für Kartographie und Geodäsie' (BKG, 2020) and GovData (Open Government Data, 2016).

At the Federal State level, building-geo-datasets of Hamburg, North-Rhine-Westphalia, Thuringia, Berlin and Brandenburg are open but the datasets of the other states are commercial. These datasets can be accessed via each federal state's geo-data web portals. While some municipalities offer online access to their geo-data most do not. Data might be provided on request but each request is case specific.

Austria

Data holders are state, federal states, cities and municipalities. The state offices and their downstream agencies are responsible for national data sets, such as the digital cadastral map. Due to laws in the respective federal states (e.g. in the field of nature conservation) the federal states are responsible for a large amount of geodata. The initiative "geoland.at" started more than 15 years ago and is a geodata network across all federal states (Geoland, 2020). In 2011, the Federal Chancellery and the City of Vienna established Open Government Data in Austria. The federal states, some federal ministries and cities are the driving force behind this initiative. The basemap.at project initiated by the federal states is a prominent example of functioning Open Data in Austria (Basemap, 2014). In addition to the Open Government Data Portal of Austria (data.gv.at), the INSPIRE portal (inspire.gv.at) provides another access point to (partly but not completely open) geodata of Austrian administrations. Both portals are hosted by the European Open Data Portal opendataportal.eu (European Data Portal, 2019). Nevertheless, there is still a need for action in Austria, such as the provision of a nationally harmonized building model (LOD1,2,3), which is especially essential for energy-relevant analyses.

Switzerland

The largest Swiss datasets on buildings are the Federal Register of Buildings and Apartments (GWR) and the Federal Data of the Statistics of Buildings (GWS). In GWR, each individual building is geometrically represented by its centroid and associated with information for instance on usage, number of floors or heating systems. GWS summarizes building characteristics and information from the population and household statistics, such as occupation and age groups for hectare sized grid cells. Both datasets are compiled and controlled by the Swiss Federal Office of Statistics. Access to these datasets is provided only by request and can only be used for statistical work, research, planning and legal obligations (FSO, 2020). Detailed building footprints are part of the official cadasters, which are compiled and made available by cantons or municipalities. The data is often openly available from local sources. Less detailed building footprints are also openly available from user-generated contents such as OpenStreetMap. Most geographical datasets are controlled by the Swiss

Federal Office of Topography, there are multiple openly free datasets available but others, such as 3D building geometries are commercial and must be ordered (Bundesamt für Landestopographie, 2020). In 1979 the energy departments in each of the Swiss cantons came together and formed the *Engerie direktoren konferenz (EnDK)*. One of the main instruments of the EnDK is "Der Gebäudeenergieausweis" (GEAK), which is an energy performance assessment and certification tool for buildings (EnDK, 2020). The raw data used in the assessment and issue of GEAK certificates is not openly available and needs to be requested.

Private Organizations

Two kinds of private organizations that hold building simulation relevant geo-datasets were identified in all DACH countries. The first are the 'Map Providers' and are for example OpenStreetMap (OSM) and Google. While OSM data is open, Google's data is application restricted. The second type of private organization that hold building simulation relevant geo-data-sets is the 'Building Service Providers'. Collectively these companies provide a number of different building related services. Examples are Insurances, Housing Companies, Utility companies or the postal service. These data sets are predominantly private but sometimes commercially available (Deutsche Post, 2019).

UNIVERSAL DATA CATEGORIZATION TEMPLATE

An online template presenting data categorization (as discussed in section data categorization and in Table 1) and country specific data availability (currently for Austria, Germany and Switzerland as discussed in the last section) can be found under: <https://ibpsa.github.io/project1-wp-2-1-cim-gis/>. One of the main aims of this template is to assist the processing country-specific data for both researchers and practitioners. The online template will include the information about the categories and sources of the data that can be used for energy simulations. Furthermore, the authors would like to encourage people to regularly update the database for different countries. A detailed description of how to contribute can be found on the website above.

CONCLUSIONS AND FUTURE WORK

Globally, the state of being available of the open and country-wide building polygon data is an important requirement for city modelling but is seldom available. Such building geo data should at least be available free of charge for research and scientific purposes. The quality and reliability of the building geo data provides further cause for concern. Another big problem faced by the community regards the non-

homogenous nature of the data attributes provided by different federal states. Currently, every federal state has its own standard for the available information. A factor that can result any number of ad-hoc properties or concept definitions such as building usage categories which in many cases do not provide a clear classification of the buildings under consideration. Nationwide and EU-wide standards should be created in order to create more reusable geographical building information attributes. The formulation of the Universal Data Categorization Template as presented in this paper provides an important step forward on the way to standardization. This template includes different data categories, datatypes and data sources for urban scale information modeling and simulations. However, further work foreseen by the authors is required both on national and international level.

ACKNOWLEDGEMENT

This work emerged from the IBPSA Project 1, an international project conducted under the umbrella of the International Building Performance Simulation Association (IBPSA). Project 1 will develop and demonstrate a BIM/GIS and Modelica Framework for building and community energy system design and operation.

REFERENCES

- Agugiario, G., Benner, J., Cipriano, P., & Nouvel, R. (5. March 2018). The Energy Application Domain Extension for CityGML: enhancing interoperability for urban energy simulations. *Open Geospatial Data, Software and Standards*.
- Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland (AdV). (2020). Die amtlichen Hausumringe Deutschland (HU-DE). Abgerufen am 14.02.2020 von <http://www.adv-online.de/AdV-Produkte/Liegenschaftskataster/Amtliche-Hausumringe/>
- Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland (AdV). (2012). *ALKIS® Grunddatenbestand und länderspezifische Inhalte*. AdV.
- Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland (AdV). (2004). *Erarbeitung eines Qualitätssicherungssystems für die Geodaten des amtlichen Vermessungswesens*. Arbeitskreis Informations- und Kommunikationstechnik.
- Auer, S. R., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., & Ives, Z. (2007). DBpedia: A nucleus for a web of open data. *Lecture Notes in Computer Science, vol 4825* (S. 722). Heidelberg: Springer, Berlin, Heidelberg.
- Basemap. (2014). *basemap.at*. Von https://www.basemap.at/index_en.html abgerufen
- Bazjanac, V., & Crawley, D. B. (1997). The implementation of industry foundation classes in simulation tools for the building energy. *Building Simulation* (S. 203-210). Prague, Czech Republic: IBPSA.
- Biljecki, F., Kumar, K., & Nagel, C. (27. August 2018). CityGML Application Domain Extension (ADE): overview of developments. *Open Geospatial Data, Software and Standards volume 3*.
- Braun, R., Weiler, V., Zirak, M., Dobisch, L., Coors, V., & Eicker, U. (2018). Using 3D CityGML Models for Building Simulation Applications at District Level. *International Conference on Engineering, Technology and Innovation (ICE/ITMC)*. Stuttgart, Germany: IEEE.
- Bundesamt für Kartographie und Geodäsie (BKG). 2020. *Produkte und Services*. Abgerufen am 14.02.2020 von <https://gdz.bkg.bund.de/>
- Bundesministerium für Digitalisierung. (kein Datum). *Open Data Österreich*. Abgerufen am 11. 02 2020 von <https://www.data.gv.at>
- Bundesamt für Landestopografie (2020). swisstopo online shop. Available at: <https://shop.swisstopo.admin.ch/de/> / (Accessed: 20th February 2020)
- Chen, Y., Hong, T., Luo, X., & Hooper, B. (November 2018). Development of city buildings dataset for urban building energy modeling. *Energy and Buildings 183*, S. 252-265.
- DESTATIS. (2020). Gemeindeverzeichnis-Informationssystem GV-ISys. Abgerufen am 14.02.2020 von <https://www.destatis.de/DE/Themen/Laender-Regionen/Regionales/Gemeindeverzeichnis/Administrativ/Archiv/GV100ADQ/GV100AD2902.html>
- Deutsche Post. (2019). Data Factory. Deutschland. Abgerufen am 06. February 2020 von <https://www.deutschepost.de/de/d/deutsche-post-direkt/datafactory.html#>
- Dong, B., Lam, K., Huang, Y., & Dobbs, G. (2007). A comparative study of the IFC and gbXML informational infrastructures for data exchange in computational design support environments. *IBPSA- International Building Performance Simulation Association*, S. 1530-1537.
- EnDK. (2020). Der GEAK. Available at: <https://www.geak.ch/de/der-geak/>. (Accessed: 20th February 2020)
- European Data Portal. (2019). *European Data Portal*. Abgerufen am 06. February 2020 von <https://www.europeandataportal.eu/en/providing-data/goldbook/open-data-nutshell>
- FSO. (2020). Federal Register of Buildings and Dwellings. Available at: <https://www.bfs.admin.ch/bfs/en/home/register/federal-register-buildings-dwellings.html>

- (Accessed: 20th February 2020)
- Geiger, A., Benner, J., Häfele, K., & Hagenmeyer, V. (2019). Building Energy Simulations at Urban Scale Based on Standardized Data Models Using a Transparent Enrichment Process. *Building Simulations - International Conference and Exhibition*. Rome: IBPSA.
- Geoland. (2020). *geoland.at*. Accessed on 09. 02 2020 from <https://www.geoland.at>
- Gröger, G., Kolbe, T. H., Nagel, C., & Häfele, K. H. (2012). *OGC City Geography Markup Language (CityGML) En-coding Standard*. Open Geospatial Consortium.
- International Organization for Standardization (2013), ISO/PAS 16739:2005 Industry Foundation Classes, Release 2x, Platform Specification (IFC2x Platform). Central Secretariat, Geneva, Switzerland.
- Malhotra, A., Frisch, J., & van Treeck, C. (2019). Literature review concerning IFC, gbXML and CityGML data models for energy performance simulations. Aachen: RWTH Aachen University.
- Open Government Data (2016), Available at: <https://www.govdata.de/> (Accessed: 25th February 2020)
- Population Division of the UN Department of Economic and Social Affairs (UN DESA) in 2018 Revision of World Urbanization Prospects (2018)
- Reinhart, C. F., & Davila, C. C. (2015, December). Urban Building Energy Modeling – A Review of a Nascent Field. *Building and Environment* 97, pp. 196-202.
- Rosser, J. F., Long, G., Zakhary, S., Boyd, D. S., Mao, Y., & Robinson, D. (29. March 2019). Modelling Urban Housing Stocks for Building Energy Simulation using CityGML EnergyADE. *International Journal of Geo-Information* 2019, 8(4), S. 163.
- Sola, A., Corchero, C., Salom, J., & Sanmarti, M. (23. November 2018). Simulation Tools to Build Urban-Scale Energy Models: A Review. *Energies*.
- Swan, L. G., & Ugursal, V. I. (2009). Modeling of end-use energy consumption in the residential sector: A review of modeling techniques. *Renewable Sustainable Energy*, S. 1819-1835.
- Wetter, M., van Treeck, C., Helsen, L., Maccarini, A., Saelens, D., Robinson, D., & Schweiger, G. (September 2019). IBPSA Project 1: BIM/GIS and Modelica framework for building and community energy system design and operation - ongoing developments, lessons learned and challenges. *Earth and Environmental Science*, Vol. 323, S. 012114.