Methods for comparing digital applications in buildings and districts

Felix Rehmann, Falk Cudok and Rita Streblow
Technical University of Berlin, Germany

ABSTRACT
The so-called Energiewende is a complex task with a variety of stakeholders, regulations, technical infrastructure, and proposed solutions. Buildings are an important sector for increasing resource and energy efficiency, as in Germany around 35% of end energy usage can be attributed to them. Digital applications can help reduce these emissions through more efficient planning, operating, renovation, or demolition. Depending on the task and the parties involved, the complexity of descriptions and data models can vary from simple energy efficiency labels used in the labeling process of buildings to complex individual data models used for simulation. However, these specific information systems are often non-transferable, are hard to compare in their restrictions and requirements, and hence increase the overall costs of digital solutions. A good understanding is needed to preserve economic and ecological benefits while maintaining privacy and security aspects. To compare digital applications, a variety of solutions are identified. This paper first provides an overview of the identified solutions, namely frameworks, taxonomies, and ontologies. While the identified frameworks focus more on technological aspects and are complex to use, they provide an in-depth understanding. Taxonomies can be used for a scientific (e.g., classification of methods) comparison and provide simple relationships. Ontologies provide relationships and definitions while being task-dependent. We conclude that comparing the underlying data models of digital applications is a complex task and dependent on the application and its infrastructure. However, a variety of tasks refers to the same tools and data. After discussing these approaches, we then give an overview of digital applications developed by German researchers. Last, we give an insight on how to combine these aspects in our ongoing research. To summarize, in this paper we give an overview of the complexity involved in transferring digital solutions in the building sector, provide a method used for comparing applications and describe a solution to compare infrastructure and digital tools built for it.

1. Introduction

1.1. Introduction and motivation
Digital applications for districts and buildings provide a variety of functionalities and use-cases, from automating single aspects of modern living and working, to complex interactions with the energy system. Buildings play a crucial role in the transformation towards a decarbonized future, as they are responsible for 35% of end energy usage [1] and people tend to spend 86.9% of their life in them [2]. Digital applications can both lower emissions and increase user comfort in buildings throughout their life cycle [3]. According to the European Commission Information and Communication Technologies (ICT) as enablers of energy efficiency need to be explored and exploited [4]. However, the necessary ICT needs energy for operations and resources for production and triggers additional effects like economic growth, increasing the overall energy demand [5]. Digital applications for transforming the energy system are also part of a socio-technical system, where technical, economic, and societal aspects are affected [6]. In Germany, there are also a variety of buildings, technical infrastructure, and their use. The variety of infrastructure requires integral methods and solutions to manage...
structural complexity [7]. A good understanding of ICT, available tools, needed data, affected stakeholders, and possible benefits and risks is needed to evaluate digital applications. Applying energy-related applications to buildings is a very context-sensitive task [8]. To enable interoperability between digital tools and buildings, districts, and heating networks, methods are needed to compare digital solutions and make them understandable to all people affected. However, methods categorizing digital applications are lacking so far [6]. Such methods are being developed by our research.

The methods and findings proposed in this paper are based on an evaluation research project of the 7th Energy Research Program of the Federal Government of Germany related to energy and resource usage in buildings, districts, and heating networks\(^1\) [10]. The focus of the projects under investigation is on research topics in the building and district sector from Technology Readiness Level (TRL) 3 [1]. A TRL of 3 represents an experimental proof of concept is available. In the development of technologies and concepts, the focus is on a holistic view (technological aspects, economic aspects, and ecological aspects) of buildings and districts. While the focus in the building sector is on renovation, modernization, and further development of building materials, the focus in the district sector is on systematic interaction and thus on enabling the energy-efficient decentralized provision of heating, cooling, and electricity [9]. Digital applications developed in this area range a variety of approaches and tasks from the development of new simulation methods for the research of new filter materials in the context of room conditioning to reducing the energy demand of office buildings with gamification approaches and offering alternatives for renovation. An overview of short summaries, involved parties, funding, and time frame can be found using EnArgus [11]. A lot of projects must deal with the same challenges regarding data availability, interfaces, or reusability of results, methods, and data. The developed and described methods help to solve these problems, by identifying similarities in digital applications. Additionally, the methods help transfer applications from research to business solutions by providing a good understanding of the advantages and disadvantages. The insights in ongoing research can help other researchers, companies, and policymakers. However, different abstraction levels from energy flows for engineering perspective towards a more global and aggregated perspective for policymakers are needed [12]. Companies and policymakers might profit by understanding what benefits new technologies bring and how these technologies are related. To achieve this, we identify:

- Factors influencing digital applications in buildings
- State-of-the-art technology in research
- Challenges in ongoing research related to digital

1.2. Applications in Germany

- Future research directions

And based on these findings, we propose methods to:

- Identify and describe socio-technological impact factors on digital applications
- Compare and transfer digital applications
- A concept to reduce complexity in applying digital applications and increasing scalability and reusability

As a first step, we are focusing on buildings and districts, following the structure of the accompanying research. For further information, please visit [10]. The paper has the following structure: in the literature review, we define what we consider a digital application to be and what the relevant aspects are for our research. We then give examples of how these aspects are relevant in digital applications for districts and buildings. Additionally, we give an overview of methods of comparing and sharing digital applications. Further on, we describe our methodology and the dataset created by using a questionnaire. Then we discuss the results and explain what conclusions we draw from them. Based on this, we propose our framework for comparison of digital applications. In the end, a discussion and outlook follow.

1.3. Related work

This subsection summarizes related research. Due to the scope of our research, we will focus on projects related to digital applications in buildings and districts. The project National Research Data Infrastructure (NFDI) aims to develop and secure access to decentralized data of publicly funded research [13]. For this, standards in data management must be developed and established. Currently, a variety of disciplines are funded or signed letters of intent. The NFDI4Ing is the subchapter for engineering-related topics [14]. In a bottom-up approach, they developed seven archetypes of researchers, as an example of specific engineering-related tasks, where real objectives typically consist of two or three of them. A generalization of data management methods is planned

\(^1\) German Federal Ministry for Economic Affairs and Climate Action and the Federal Ministry of Education and Research funded 948 ongoing projects with around 103 million euros of cash outflow in 2020. In 2020, the ministries also approved 220 new projects with total funding of 127 million euros [9].
Figure 1. Illustration of the presented approach. In step 1 relevant aspects of digital applications.

Identification of relevant aspects

Evaluation of literature and publicly available data to generate a structure and key topics, and to describe methods to typically describe digital applications in buildings and districts.

Survey of funded projects

Review of 134 projects and identification of the key aspects that are relevant to their developed digital application. Based on these findings we describe by the example of dashboards, what overlooked research questions are.

Conclusion

We conclude that heterogeneity and lack of information burden the scalability of digital applications and related research. Splitting digital applications into minor applications might help pace up the digitalization of energy-related tasks in buildings and districts.

by identifying use-cases for the archetypes. They further state, that a specificity of engineering-related research is the trade-off between open science and the economic interests of the involved parties, as most projects in this field consist of a variety of project partners investing together in technical assets.

To help researchers, authorities, and companies share their data, currently a few platforms are in development. The openMeter platform aims towards providing services for grid operators, researchers, municipalities, and service providers [15]. Currently, the energy consumption data of various objects and energy types can be uploaded, compared, and downloaded. The Open Energy Platform provides a more generalized service with the option to publish datasets and related factsheets to compare models and frameworks [16]. The focus of the platform is to increase transparency in energy system modeling. Additionally, an ontology is provided with definitions and tutorials explaining the concept of the platform.

The Open Research Knowledge Graph project has the goal of describing and comparing research papers in a structured manner [17]. This enables the comparison of papers in previously structured templates.

An analysis of digital applications in the previous energy research program is given in [18]. They provide a classification of software-based solutions, differentiating solvers, web applications, graphical user interfaces, (dynamic) simulation, accounting tools, geographic information systems, interfaces, and data schemas. According to the authors, systemic approaches for multiscale and multilevel simulation are still underrepresented.

While there is related work, so far there is no specific and up-to-date work on buildings and districts. The previously identified examples focus on providing data for a variety of energy sources or increasing the knowledge about energy research. So far, there are no projects in Germany published trying to identify relevant use-cases and provide data governance practices, baselines, and techno-economic and ecological overview in a systematically way.
2. Methodology

In this section, our methodology is described. We follow an approach in three steps for the basic identification of use-cases and relevant aspects of digitalization. In the first step, we review the literature and publicly available information about projects and methodologies related to energy in buildings and districts, and the methods which typically describe them. We first define digital applications, then review examples and methods for describing their relevant aspects. This enables us to develop a structure and a method for the ongoing research. In the following chapter, we review digital applications funded in the 7th Energy Research Program by using a survey. By using this method, we can give an overview of relevant aspects. In the section dealing with the questionnaire, we outline our questions and the answers that were given. We compare the written description with self-stated tags to describe the applications. We further elaborate our methodology by reviewing an example use-case. In the section results and discussion, we combine the review and evaluation and propose an identified framework for providing information to different stakeholders, which is then further described in the section proposal as the developed platform. This approach is summarized in figure 1.

3. Literature review

This section is structured in three subsections. In the first section, we state what we consider to be a digital application in buildings or districts. Then we give an overview of applications and current trends themselves. Lastly, we state methods to compare and describe given applications.

3.1. What is a digital application?

There are a variety of definitions of digitalization in usage [19]. To provide a clear definition, we define the aspects of digitalization considered as follows. Digital applications can be considered 'computer programs, procedures and possibly associated documentation and data about the operation of a computer system all or part of the programs, procedures, rules and associated documentation of an information processing system program or set of programs used to run a computer' [20]. Applied in buildings they can increase energy- and resource efficiency [3], but on the other hand need resources to produce the underlying ICT infrastructure and operate the applications [5]. Common goals include increasing (energy) efficiency, better policymaking through transparency, and enabling new business models [21]. For transferring technical solutions important aspects of them need to be made explicit to know their scope and applicability [22]. However, this aspect is often aggravated, as terms and definitions often are not agreed on. For example, smart, intelligent, and digital are simultaneously used [23]. Most digital technologies have interdependencies [24].

The kind of applications applied varies through the life cycle [25] and the scope, whether buildings, districts, or heating networks are affected [26]. The kind of necessary data aggregation to fulfill necessary tasks also varies from engineering to planning to entrepreneurial and political levels [12]. Due to this, we argue it is not suitable to provide a universally applicable definition of digital applications, rather we describe the aspects considered.

To us, digitalization is:
• Digitization of analog into machine-readable data,
• Utilization of machine-readable data,
• Processing of machine-readable data,
• Mapping of reality in virtual space,
• Linking physical, software, and human interfaces,
• Automation and documentation,
• Transformation of society.

Digitalization has the goal:
• Resource-efficient construction, operation, and redevelopment of buildings and districts,
• Enable communication between
  • Stakeholders,
  • Data,
  • Tools and methods,
• Traceability of
  o Processes
  o Decisions
Reliability and security of Information.
Aspects considered by us:
• Regulation
• Data governance
• Data sufficiency
• Stakeholders, especially users,
• Ecological effects
• Economic effects
• Cost
• Accessibility

Following these aspects, five main categories of relevance with interactions are identified by us, which need to be considered:
• Ecological aspects: ‘Does the digital application saves energy in total?’
• Economical aspects: ‘How much does it cost to develop and operate the digital application?’
• Regulatory aspects: ‘Which laws and governance criteria applying to the digital application?’
• Stakeholder-related aspects: ‘What information is needed about the digital application to understand it?’
• Technological aspects: ‘How does the digital application work?’

An example of a typical interconnection is for example the process of how to store sensor data related to the energy consumption of a flat. From an economic perspective, one can estimate what the installation, operation, and maintenance costs are and what related services can be developed based on this data. While the General Data Protection Regulation (GDPR) and hence the regulatory perspective is relevant for considering data processing and storage, which has direct impacts on the technical architecture and vice versa.

3.2. Digital applications in the context of buildings and districts

As previously described, projects that receive funding in the research program and are reviewed by us are mainly focused on buildings and districts. A small number of projects have the focus on monitoring or digital fundamentals like cloud development. The following paragraphs aim towards giving a brief introduction to digital applications in buildings and districts and their respective relevant aspects in general.

3.2.1. Digital applications in buildings

In the German building sector districts and heating, networks are affected by diverse characteristics of buildings properties and the technology used for heating and cooling [27, 28]. In total, there are around 19.05 million residential buildings, of which around 3.22 million are multi-family homes [27]. Additionally, there is an estimation of 21.12 million nonresidential buildings, from which around 1.98 million are heated or cooled [29]. Typical digital applications focusing on buildings include increasing energy and resource efficiency and automating building diagnostics and maintenance [30]. Conservative estimations of cost-saving potentials for building automation are 1%–1.5% of total building cost and save around 10% of costs for energy usage [31]. To achieve this, most digital applications that deal with energy usage in buildings typically need data about the climate zone, the building envelope, its energy systems, and information about user behavior [32–34]. Specific data types can be collected by different methods [35].

A prominent example of a digital application in buildings is the Building Energy Management System (BEMS). It can be used to monitor and control the energy consumption of buildings [36]. [37] compares BEMS to Home Energy Management systems (HEMS) and concludes that both are applied to reduce electricity costs and maintain the efficiency of appliances. A HEMS however, focuses on the use of energy within a household, while BEMS are applied for both residential and commercial buildings to them [37]. A review of HEMS is given in [38] with the conclusion that due to increased accessibility and affordability, the popularity of HEMS has increased. [39] provides a review of strategies for BEMS.

In recent years, the focus of research on digital applications in the building sector has been mostly on Building Information Modeling (BIM), digital twins, and IoT [24]. BIM is a common method applied in the building context. Currently, there are no common definitions of what kind of information the model should provide, it rather depends on the context the method is applied [40]. A common open-source standard used for BIM is IFC, which is used to represent architectural and constructional data [41]. Currently, IFC is still lacking the dimensions needed for building automation [42]. The usage of BIM in existing buildings is still limited in contrast to newly planned and built buildings [43]. While building data could or even should be enriched with the BIM method during the life cycle, it is often not the case, due to various reasons like stakeholders missing knowledge, unclear process definitions and standards, or non-existing collaboration [44].
Outside of the building life cycle, information about the operation of buildings is typically stored in automation systems or monitoring solutions, for example in the naming schemes or metadata of Building Automation Systems [45, 46]. Currently, these systems are highly diversified and can lead to vendor lock-in. The interoperability of IoT devices could be achieved by a common communication standard, which can be achieved by using the same protocol or using a gateway that can translate the protocol. Additionally, semantic interoperability is needed. Especially in older buildings, the renovation and rewiring of older systems is cost intensive. Internet-of-Things (IoT) based architectures tend to lower costs.

The digital twin (DT) concept after Michael Grieves consists of three parts [47]:

- A physical representation of the product in reality,
- A virtual representation,
- The connection of data and information which ties them together.

Outside of the building life cycle, information about the operation of buildings is typically stored in automation systems or monitoring solutions, for example in the naming schemes or metadata of Building Automation Systems [45, 46]. Currently, these systems are highly diversified and can lead to vendor lock-in. The interoperability of IoT devices could be achieved by a common communication standard, which can be achieved by using the same protocol or using a gateway that can translate the protocol. Additionally, semantic interoperability is needed. Especially in older buildings, the renovation and rewiring of older systems is cost intensive. Internet-of-Things (IoT) based architectures tend to lower costs.

In the literature, DT is linked to a variety of abilities from prediction, simulation, and optimization to IoT and BIM [48, 49]. And while BIM is often applied during the planning and construction of a building, it is a valuable source of data for the construction of the DT itself, where IoT can tie virtuality and reality [48]. Well-documented digital resources can so increase efficiency and enable further positive effects.

3.2.2. Digital applications in districts

Districts consist of more than one building, residential, non-residential, or a mixture of both. The term is often synonymously used with neighborhood or block [50]. In Germany, there is so far no official or regulatory definition of a district [51]. As districts are consisting of several buildings, the individual problems or approaches of buildings can often be transferred to districts. A variety of data points are both needed for buildings and district-related digital applications [52]. However, approaches and technologies for single buildings are in the case of districts interrelated and influence each other and affect each other [53]. In districts, two effects additionally come into place. First, districts often have a variety of different stakeholders, which further complicates agreement on definitions and contracts [37]. Second, combining several energy sources increases the amount of needed data [54]. In practical application, this is often solved by combining several data sources [52]. For example, load profiles can be approximated by methods described in research or standards [55, 56]. In total, simulating the energy demand at an urban level is more complex than at a building level, due to more diversity in usage, the huge amount of information needed, and the effects of the urban environment [57].

Accordingly, districts also have the option of energy management systems that provide options to integrate a variety of energy sources and aim to decrease the total amount of emissions [58]. To coordinate the buildings in a district a robust coordination mechanism has to be in place [59] provides an overview of approaches and methods with the conclusion that reports show that the electricity bill can be reduced by 5% to 30% by coordinating HEMS.

Typical digital applications include modeling the energy use in buildings and enabling analyzing retrofit potentials or gathering infrastructure data [30]. For these digital applications often, GIS-based models are used. One common model is CityGML [60]. So-called application domain extensions (ADE) can enrich the data models, where the energy ADE provides the extension for energy-related applications [61]. A recent more web-developed-based alternative is CityJSON [62], a JSON-based encoding. While these models provide a variety of options for clustering information, there is a lack of multi-scale and multi-level approaches to evaluate data [18]. The increased complexity of districts and their potential for increasing energy efficiency make such tools almost essential.

3.3. Approaches to comparing digital applications

There are a variety of approaches to describe and compare the heterogeneity of digital applications, their requirements for both users and infrastructure, and applicable regulations. The following sections give a brief overview and relevant examples.

3.3.1. Frameworks/models

A framework is a concept to model aspects of a specific domain, by representing the domain in abstract classes and their allowed collaboration [63]. While the term is mostly used in software engineering, the concept can also be applied to structure a system and enable a systematic view. There are currently only a few frameworks that can be used for comparing digital applications in the building sector. Additionally, frameworks are focusing on smart grids. The Smart Grid Architecture Model (SGAM) is a framework to enable the Europa standardization of smart grids [64]. It provides several layers (Components, Communication, Information, Function, and Business), Domains, and Zones to study the system and understand interoperability between systems. The SGAM helps users in tasks such as visualizing, analyzing, planning, and modeling the architecture. An overview of related methods and their theoretical fundamentals can be found in [22].
The GridWise Architecture Council defines another framework applicable in the context of smart grids [65]. They define three contextual layers, including organizational, informational, and technical.

Another approach based on enterprise architecture is presented by [66]. The focus of this project is on processing heterogeneous data sources in residential buildings and apartments. The architecture framework considers Data Sources, Technologies, Data Space, Applications and Data Processing, Business (Virtual Enterprises), Services, and Context. [67] provides a more general framework for IoT in smart homes. They propose the layers of smart objects, hubs, cloud, third-party, and an overall security layer.

3.3.2. Taxonomies
A taxonomy can be described as a classification mechanism [68]. Taxonomies provide the description and relationships between hierarchical levels, which eases knowledge sharing, better understanding between relationships, identifying knowledge gaps in a field, and supports the decision-making processes [68]. Related to digital applications in buildings and districts there is a variety of different taxonomies.

A common aspect to describe digital applications that focus on energy usage prediction is the aspect of classifying the modeling aspect. Digital applications focusing on energy prediction are often structured in physical, hybrid/gray, statistical, and artificial intelligence approaches [69]. These can then be further classified, for example, by [70] providing a machine learning taxonomy that distinguishes occupant or energy/device-centric approaches.

[52] for example, provides a taxonomy for classifying papers on urban building energy, based on input data, simulation tools and simulation results, validation, and verification. They argue that around 95% of studies cannot be reproduced due to missing data or workflows.

[35] provides a taxonomy for building energy use data based on static and dynamic parameters. Based on these findings, they discuss data collection methods and their advantages and disadvantages.

[71] categorizes energy system models by comparing the applied methodology, assessment criteria, analytical and mathematical approach, reusability, and challenges to categorize multi-energy system models in mixed-used districts. They argue that inconsistent use of terminology in approaches and literature leads to a problem in the topic of modeling and optimization of energy systems.

[72] provides a summary of 24 data tools used in the building life cycle. They use six distinct categories to describe and summarize the tools. They differentiate between tool categories, data coverage, data relations, flexibility, extensibility, the stage of application, and adoption.

Other approaches to finding clusters on digital applications are for example, combining the building and urban scale with the previously discussed modeling aspect [26]. Other approaches cluster digital applications by the use [6] or the building life-cycle phase [25].

3.3.3. Ontologies
Ontologies are semantic models which describe concepts and their relationships by establishing a domain’s concept [8]. They can differ from domain-independent, more general approaches to domain-specific focused approaches [73]. Diverse backgrounds might lead to different assumptions about the usefulness of categorizations. About buildings, there are several ontologies, an extensive review can be found in [8, 72]. In contrast to taxonomies, ontologies enrich the relations from taxonomies by providing more detailed relationships and definitions [74]. Applying this definition, taxonomies are simpler ontologies [74, 75].

The Smart Applications Reference Ontology and its extensions define core concepts in the ‘smart applications domains’ based on the fundamental principles of reuse and alignment, modularity, extensibility, and maintainability [76]. SAREF4BLDG extends the ontology by parts of the previously discussed IFC standard [77].

[78] describes the Smart Energy Domain Ontology (SARGON) which extends the SAREF ontology with smart grid and automation use-cases. It focuses on cross-sectoral information for IoT devices in 5G research projects related to energy in Germany.

4. Survey of funded digital applications
To evaluate the findings from the literature and compare them to the German research landscape, a survey was developed and sent to the funded projects. The questionnaire contained questions, focused on monitoring, technical readiness level, technologies used in districts, buildings, heating networks, user participation, and digital applications. Questions were a mixture of dichotomous questions, multiple-choice, rank order, and open text. It was sent out to 271 project leaders, and the project was ongoing in August 2021. The survey was previously tested with project leaders and research associates. If a project consists of a few subprojects, each project leader represented a joint project. In total, 179 individual joint projects answered the survey. Not every project had to answer every question, as the survey was individualized through categorization questions. For
example, in the subsection digitalization, the entry question was: ‘Do you use or develop digital applications in your project?’ If the editor picked neither, this part of the survey was skipped. Of 179 projects asked, 114 answered they were developing digital applications and 116 answered they were using digital applications. In the following, we focus on the aspects described in the given answers. The main goal was to understand what support researchers need in the context of digitalization. By this, we apply qualitative content analysis to gain further insights and textual representation of digital applications [79].

In this section, we first describe the process of reviewing a written description of the developed applications. These results are called categories. Second, we provide an overview of the chosen selection of provided descriptions. These criteria are called tags in the following, as the interviewees were asked to select fitting labels. Third, we compare the self-selected tags and given categories based on the descriptions. Last, we discuss these findings.

4.1. Overview of manually selected categories identified by using descriptions of digital applications
The survey aimed to get an overview of digital applications in use or developed. However, as previously described there is a lack of commonly used or accepted terms. To get a better understanding of the terms in use, we asked an open question: ‘What kind of digital tools are you currently using or developing?’ 131 projects stated an answer to this question, of which 128 were selected as valid answers. Nonvalid ones included for example ‘—’ or ‘Project is currently still in the concept phase’. Thereafter, the given answers were split into sub-answers. For this, separators such as ‘,’ or an indication such as a new line or German grammar were used. In total, 339 sub-answers were identified. As previously discussed, we argue digital applications are rather complex and should be split into sub-applications to provide a better understanding. For a visual example, refer to figure 2.

After this, the stated answers were ‘categorized’. For example, the answer powerful time-series databases [Original Quote: ‘Leistungsstarke Zeitreihen Datenbanken’] was categorized as database [original category: ‘Datenbank’]. If a sub-answer was inconclusive, both categories were added. Then the categories were reviewed and consolidated where possible (e.g., web-based visualization [Original Quote: ‘Web-basierte Visualisierung’] and modular web application [Original quote: ‘modular aufgebaute web application’] were consolidated as web application]. The categories were reviewed and counted by frequency. If a category was only counted once, it was added to the category other. In total, 33 categories were identified, by which we now provide an overview. A full discussion of all categories is currently still under development. We provide an overview of the applied methodology in section five. The identified categories, their frequency, and a related example answer are stated in table 1.
Table 1. Manually by description identified categories for digital applications and their frequency and a translated example answer.

| Category                        | Frequency | Example                                                                 |
|---------------------------------|-----------|-------------------------------------------------------------------------|
| Simulation                      | 57        | Simulation for thermal-hydraulic-mechanical coupled processes            |
| Monitoring                      | 30        | Monitoring data for the further development of indirect measurement methods |
| Planning                        | 27        | Tool for optimizing the planning of long-distance heating networks      |
| Operation optimization          | 22        | Smart operation optimization method for heat supply system based on machine learning |
| Energy management               | 18        | Energy management system for connected and optimized operation          |
| Control methods                 | 17        | Model predictive controller (MPC) for energy management system          |
| Framework                       | 15        | A software framework that can handle the highly fragmented database    |
| Web application                 | 13        | Modular web application as a counseling tool for municipalities         |
| Other                           | 12        | Blockchain                                                              |
| Data management                 | 10        | Data collection and evaluation                                          |
| BIM                             | 10        | BIM models for PV planning                                              |
| Artificial intelligence/Machine learning | 9          | Automated data evaluation (machine learning)                           |
| Evaluation tools                | 8         | Tool for hourly evaluation of CO₂ emissions and primary energy consumption of districts |
| Database                        | 8         | Database on the non-residential building stock                          |
| Digital twin                    | 8         | A DT is being developed                                                 |
| Fault or pattern recognition    | 7         | Development of algorithms for the detection of faulty operation conditions |
| Cloud                           | 7         | Cloud application to visualize and analyze heating circulators in the field |
| Platform                        | 7         | District data exchange platform to support the exchange of data between different planners within the neighborhood |
| Dashboard                       | 7         | Energy monitoring-dashboard for user information                       |
| Accounting                      | 5         | Account of indoor environmental quality                                 |
| Visualization                   | 4         | Cloud application to visualize and analyze heating circulators in the field |
| Prognosis                       | 4         | Development of load forecasts using machine learning methods            |
| GIS                             | 4         | Partially automated scripts for processing geodata in a database system are |
| Mixed reality                   | 4         | Virtual and augmented reality being developed                           |
| Assistance- or expertsystem     | 3         | Expert system for generation of function lists                         |
| Experimentation                 | 3         | Validation of models through laboratory and field testing              |
| Internet of things              | 3         | Evaluation of the performance of IoT-specific radio technologies       |
| Data scheme                     | 3         | Common and updatable data basis                                        |
| Predictive maintenance          | 3         | Planning of maintenance measures through simulation                    |
| Digitalization                  | 2         | Digital conversion                                                    |
| Prosumer                        | 2         | Models for the bidirectional sale of heat                              |
| Grid friendly operation         | 2         | Control of heat sinks (buildings) for grid-friendly operation           |
| Measurement systems             | 2         | Smart meter                                                            |
| District model                  | 2         | 3D district model                                                      |

Table 2. Overview of the self-stated tags for the projects and their frequency.

| Tag                               | Frequency |
|-----------------------------------|-----------|
| Simulation                        | 67        |
| Operation optimization            | 65        |
| Monitoring                        | 63        |
| Energy management                 | 49        |
| Planning tools                    | 47        |
| Digital twin                      | 32        |
| Sector coupling                   | 24        |
| Platform                          | 21        |
| Data models                       | 19        |
| User behavior                     | 17        |
| Model predictive control          | 17        |
| BIM                               | 14        |
| Fault detection and diagnostics   | 14        |
| Interconnectivity                 | 13        |
| Other                             | 13        |
| User E                            | 12        |
| Generic data                      | 10        |
| Social aspects                    | 1         |

4.2. Overview of self-stated description by using tags

The following question referred to the self-assessment of the projects. To not confuse the self-stated assessment and our categories identified from descriptions we name these self-stated aspects tags. The projects were asked
Table 3. Comparing the frequency of self-stated tags and manually identified categories.

| Category/tag                                      | Manually identified category | Self-stated tag |
|---------------------------------------------------|------------------------------|-----------------|
| Simulation                                        | 57                           | 67              |
| Monitoring                                        | 30                           | 63              |
| Planning                                          | 27                           | 47              |
| Operation optimization                            | 22                           | 65              |
| Energy management                                 | 18                           | 49              |
| Control methods                                   | 17                           | —               |
| Framework                                         | 15                           | —               |
| Web application                                   | 13                           | —               |
| Other                                             | 12                           | 13              |
| Data management                                   | 10                           | —               |
| Database                                          | 9                            | —               |
| Artificial intelligence/machine learning          | 9                            | —               |
| Evaluation tools                                  | 8                            | —               |
| Digital twin                                      | 8                            | 32              |
| Fault or pattern recognition                      | 7                            | —               |
| Platform                                          | 7                            | 21              |
| Dashboard                                         | 7                            | —               |
| Cloud                                             | 7                            | —               |
| Accounting                                        | 5                            | —               |
| Visualization                                     | 4                            | —               |
| Prognosis                                         | 4                            | —               |
| GIS                                               | 4                            | —               |
| Mixed reality                                     | 4                            | —               |
| BIM                                               | 3                            | 14              |
| Assistance- or expert system                      | 3                            | —               |
| Experimentation                                   | 3                            | —               |
| Internet of things                                | 3                            | —               |
| Digitalization                                    | 2                            | —               |
| Prosumer                                          | 2                            | —               |
| Grid friendly operation                           | 2                            | —               |
| Predictive maintenance                            | 3                            | —               |
| Measurement systems                               | 2                            | —               |
| District model                                     | 2                            | —               |
| Sector coupling                                   | —                            | 24              |
| Data models                                       | —                            | 19              |
| User behavior                                     | —                            | 17              |
| Model predictive control                          | —                            | 17              |
| Fault detection and diagnostics                   | —                            | 14              |
| Interconnectivity                                 | —                            | 13              |
| User engagement                                   | —                            | 12              |
| Generic data                                      | —                            | 10              |
| Social aspects                                    | —                            | 1               |

to select tags that fitted their project. ‘Which tags fit the most to the digital application in your project?’. They were provided a preselection that was considered relevant, either by the candidates pre-testing the questionnaire or by the project team for identifying relevant aspects. In total, 18 terms were considered relevant. The projects could pick up to five relevant tags. 498 tags were picked by the 128 projects. Each project could use up to five different tags. The results are summarized in table 2.

4.3. Comparing categories identified from the description and self-stated tags
In the next step, the self-stated tags and analyzed categories can be compared. The comparison can be only done partially, as the terms for the self-stated tags were selected previously together with test projects. In contrast, the categories were identified after reviewing the descriptions. So, table 3 only displays the comparison, where it is directly possible. Terms that are open to interpretations, e.g., how fault detection and diagnostics might or might not be related to predictive maintenance, are in this first step not considered. Table 3 lists the frequency of each of the identified categories and the tags.

4.4. Discussion of the survey
While the survey is an important tool to get an overview of the landscape in digital applications either developed or used in the research area, it has some problems. The first is related to knowledge and language. See
the example that occurred between testing and the final survey. Two people with different backgrounds are referring to the same digital application (‘What kind of digital application are you using or developing?’):

(a) A tool for optimizing the technology and energy carrier mix in a heating network is developed and applied. For example, a CO2 limit value can be specified that must not be exceeded. In this case, a cost-optimal generator park is modeled. The tool uses load profiles of the concrete neighborhood. There is thus an interface between a GIS-based building tool and the simulation tool.

(b) A method for determining an economically optimized heat supply approach. In the process, the operation of different supply technologies is already optimized.

We see this as an example that the background and knowledge of a person influence language and terms used to describe digital applications developed by them. Hence, two people referring to the same applications might use completely different terms, which complicates looking for similarities. Different terms referring to the same object or vice versa lead to confusion or not finding the knowledge a person is looking for. Together with the sheer amount of data and research produced, one often cannot see the forest due to too many trees. For example, weather data is publicly available provided by federal initiatives [80]. However, different formats, interfaces, or necessary accounts pose challenges that make it difficult to make them usable. Additionally, it is often not suitable in our proposition to only look at the bigger ‘applications’ as ‘simulation’ or ‘monitoring’. For example, the term user behavior was counted 17 times when analyzing the self-stated description. Manually reviewing the given answer on ‘What kind of digital tools are you currently using or developing?’ the term ‘user’ only turned up seven times. However, these answers rather described user behavior as an influencing factor or even in the case of user-management of the project database. This leads to the conclusion that often terms are used differently. We tried analyzing the tools before the survey and developed the categories together with test projects, however, the bottom-up processes yielded new perspectives and insights on how terms are used.

While the team is not prone to misinterpreting descriptions themselves, we try to validate our assumptions through further interviews and workshops. This is a necessary step to validate preliminary results and gain finer granulated results.

5. Example description of a cluster

To provide insight on how the process of gaining and creating more information about these previously described clusters looks, this section describes it with the example of dashboards. Seven projects have been manually identified as developing or using a digital application that can be called a dashboard. We manually reviewed publicly available information about the projects and interviewed one project to further understand the possible challenges and connections to other digital applications.

Dashboards can be characterized in a variety of ways, from characteristics like purpose, audience, or visual features to data semantics [81]. They are used to visualize data and support users in fulfilling their goals [82]. An example project by [83] visualizes seven templates with the goals of helping the user retrieve information about environment perception, the global energy demand in real-time, and one’s energy consumption. The user can then define rules for automatic actions. They conclude that the visualization and analytics of data are important for the automation process.

An overview of the dashboards developed related to energy in buildings or districts is given in table 4.

An example cluster description can now be derived from these research projects by reviewing published documents, interviews, questionnaires, or workshops. We conducted an example interview with project number 1 to further refine the developed approach. While the main goals of the project are providing monitoring of the previous refurbishment of land marked buildings, implementing energy management, and providing a dashboard for the users are also part of the project. The interview partner stated that: ‘the dashboard aims to
make data available to the user [the inhabitant] and to derive suitable recommendations for action'. Their main interest in Dashboards developed by other research projects were the frameworks applied by them. We argue that's the first indication of the success of the here proposed approach in splitting digital applications into subtasks.

6. Results and discussion

The number of approaches and methods used to develop digital applications in buildings and districts is immense and so are the terms, stakeholders, and infrastructure. As shown in the literature review, there is a variety of digital applications and research. Additionally, approaches are being developed aiming to increase the synergies between digital applications and related research. While digitalization promises to manage complexity and increase transparency, it is often lacking these factors themselves, through unclear definitions. For example, the concepts of taxonomies and ontologies are also used similarly [75]. Ontologies in buildings are often highly specialized approaches, no longer being updated, with minimal guidance and minimal alignment between each other [8]. However, there is a need for flexibility in the scope of data management and a lack of options to integrate heterogeneous data during the life cycle and different applications we argue to understand the underlying problems and develop approaches, one needs to first understand the different applications, terms, and data categories. Describing them helps develop solutions for scalability. If the correct subtask is known, there are a variety of approaches to further gather the information needed. Hence, we argue, it is sufficient to split digital applications into subtasks and applications and provide clear definitions for these.

New and established methods, tools, and standards, e.g., IFC, CityGML, or ontologies, help to structure data and standardize underlying processes. However, they are rarely fully implemented, dependencies are not fully understood, or scalability is not possible due to different information provided. Additionally, digitalization also enables portability of minor tasks, like the implementation of a dashboard. We are arguing that researchers can profit from this, as they can split the whole task into minor tasks. They then find substitutes that might integrate with existing structures and data.

As our research so far shows often lacking interconnections and differently used terms might be a reason why there is often overseen potential benefit. For example, almost any research project must deal with correct visualization strategies somehow, if they want the user to take correct actions. From a technical perspective, a variety of different formats, used programs, or different monitoring patterns complicate data sharing. Different usage of terms and abbreviations hinders research. From a social perspective, different forms of data aggregation are needed to supply researchers, politicians, or other stakeholders with the information needed.

We argue that a lot of these aspects need the same technological parameters, in different detail. For example, trustworthy weather data is the basis for energy usage prediction. In practical applications, the individual investigation of data, especially real building parameters, is cost and time intensive. Re-usage of already determined data can help to lower those costs and thus further spread the extent to which useful digital applications are used in buildings, districts, and heating networks. To do so, we argue it is necessary to build bridges between heterogeneous data, individualized tasks, or different stakeholders.

To achieve greater scale, we aim to identify and cataloging digital applications in the building sector and hence towards a greater understanding of the scale and benefits of digital applications. A variety of digital applications refer to the same data, which is either present in methods, raw data, or prepared form. Such data and methods should be shared, where no restriction (e.g., data protection or company secrets) applies. Context is often necessary and should be provided where needed (e.g., assumptions in modeling or possible measurement errors). However, for example, small steps on how to access a public API are often overlooked and still help researchers or third parties trying to understand results or developing similar applications [84].

Digitalization promises managing complexity and transparency but often is lacking these factors themselves through unclear definitions.

From our point of view, the digitalization of the building sector is a sea full of islands, and to reach its full potential, the islands should be connected. While this research is currently limited to Germany, we have no clear indication of why the results should be affected by geographical reasons. However, one possible limitation correlated with the region might be the funding structure and whether it also accounts for platforms connecting the research. For example, in the USA the IBPSA provides the BEST directory, which provides further information on software [85]. Functionality provided by BEST, such as a comparison of selected software and searching for suitable approaches by ‘Capabilities’ or ‘Building type’ leads to an easier comparison of approaches, tools, and related tasks. With EnergyPlus, there is also an accordingly in collaboration developed software [86], which might increase knowledge sharing.

To summarize, we first give an overview of digital applications in buildings and districts and what aspects can be relevant to them. We then continue by stating different methods for comparing digital applications.
Table 5. Overview of the information about to be implemented in the knowledge platform for the in section four identified categories and its possible sources.

| Source               | Examples                                                                 | Examples                                      |
|----------------------|--------------------------------------------------------------------------|-----------------------------------------------|
| Publicly available data | * Related literature                                                   | A curated list of relevant data (see for example [87]) |
|                      | * Git repositories Regulatory data, e.g., GDPR                         |                                               |
|                      | e.g., GDPR                                                               |                                               |
| Project-related data  | * Survey data                                                            | Description of use-cases, developed in workshops |
|                      | * Workshop                                                               |                                               |
|                      | * Interviews                                                             |                                               |
| Interconnections      | * Analysis                                                               | Visualization of use-cases                    |
|                      | * Own software                                                           |                                               |

However, these methods are not feasible so far, as they either focus on data or compare the methods themselves. We then gave an overview of digital applications and their content funded in German research by the related research program. From the comparison of self-stated tags and manually reviewed descriptions, we conclude that aspects are being overlooked. Based on this, we propose a system to connect the gaps between research and display similarities with the plan to accelerate and scale the development of digital applications.

7. Outlook—knowledge platform

This project plans to offer low-threshold, stakeholder-relevant information about technological, regulatory, economic, and ecological aspects of the digitalization of buildings and districts. In this paper, we first wanted to present the theses that, in addition to data availability and complex methods, other factors make digitization difficult. Another constraint is the unclear description of digital applications and the lack of overviews.

The necessary information (see chapter 4) must be published in a useful and visual way. This paragraph is about combining methods and data sources to achieve this goal. As our research suggests so far there is a variety of digital applications, with different necessities in data and applied models. We identified several categories as described in section 4.1. Through the identification of these categories, the following goals are to be implemented now by a detailed description:

- Structure for the description of the aspects (technical, regulatory, economic, and ecological) of the digitization of buildings, quarters, and heat networks
- Toolchains and breaking points
- Presentation of relevant examples
- Provision of relevant data and models
- Collection and presentation of existing knowledge in a new breadth

This section describes the aspects of the framework and the wiki. An overview and examples are stated in table 5. As described in the introduction describing socio-technological impact factors, enabling scalability, and reducing complexity are the main goals of the knowledge platform.

The previous sections show that there is a variety of concepts that lack clear definitions. With a definition of terms, data sets, and methods related to German research, we aim to identify knowledge gaps and getting researchers fast access to the methods already known. As a first step, we are currently collecting tools, public data, methods, and further information related to digital applications developed in the research area. Connecting these to the building life cycle and the data life cycle is the foundation for further work. To achieve this, we identified several data sources, listed in table 5. In general, there is already a lot of knowledge available, however, often it is distributed on different channels. A collection of relevant datasets, terminology, and methods developed by third parties is a steppingstone to a broader perspective, outside of the research in Germany. As those projects are the focus, the integration of surveys, workshop results, and interviews developed by projects is then used to describe the use-cases and their impact of them in Germany. Lastly, the interconnections can be drawn after investigating in use-cases. For example, all different infrastructure (e.g., residential vs non-residential) concepts and used methods (e.g., machine learning vs physical models) can then be described and their effects visualized.

Our goal is not to create new terms, but rather to provide access and interconnection to the terms already in use. Increasing efficiency in the built environment is a lot of research around digital applications about and for buildings and districts, which often is struggling with subtasks that are non-related to the main task. For example, is the task of data visualization in dashboards often seen as a subtask, where the accompanying
research can elevate the process of creation. We aim towards splitting complexity into several subtasks and investigating them in their technical, regulatory, economic, and ecological effects and interrelationships.

**Acknowledgments**

The authors would like to acknowledge the financial support of the German Federal Ministry of Economic Affairs and Climate Action within the funded project 03EB004A ‘BF2020 Begleitforschung Energiewende- bauen - Modul Digitalisierung; Teilvorhaben: TU Berlin: Wissensplattform.’. The authors would like to express their appreciation to Institute for Ecological Economy Research (IOW) and Berlin University of the Arts (UdK) for many useful inputs and valuable comments.

**Data availability statement**

The data generated and/or analysed during the current study are not publicly available for legal/ethical reasons but are available from the corresponding author on reasonable request.

**ORCID iDs**

Felix Rehmann [ORCID](https://orcid.org/0000-0002-6124-3300)

Falk Cudok [ORCID](https://orcid.org/0000-0002-8916-4209)

Rita Streblow [ORCID](https://orcid.org/0000-0001-7640-0930)

**References**

[1] Bundesministerium für Wirtschaft und Energie (BMWi) 2018 7. Energieforschungsprogramm der Bundesregierung (Bundesministerium für Wirtschaft und Energie (BMWi) https://erneuerbare-energien.de/EE/Redaktion/DE/Downloads Hintergrundinformationen/7-energieforschungsprogramm-der-bundesregierung.html (accessed 29 September 2021)

[2] Klepeis N E, Nelson W C, Ott W R, Robinson J P, Tsang A M, Switzer P, Behar J V, Hn S C and Engelmann W H 2001 The National Human Activity Pattern Survey (NHAPS): a resource for assessing exposure to environmental pollutants J. Expo. Sci. Environ. Epidemiol. 11 231–52

[3] IEA (International Energy Agency) 2017 Digitalisation and Energy [Online] available: https://iea.org/reports/digitalisation-and-energy

[4] European Commission 2009 ICT for a Low Carbon Economy Smart Buildings (European Commission) [Online] available: https://ec.europa.eu/information_society/activities/sustainable_growth/docs/sh_publications/smartbuildings-lb.pdf

[5] Lange S, Pohl J and Santarius T 2020 Digitalization and energy consumption. Does ICT reduce energy demand? *Ecol. Econ.* 176 106760

[6] Weigel P and Fischedick M 2019 Review and categorization of digital applications in the energy sector *Appl. Sci.* 9 5350

[7] Kohler N and Hassler U 2002 The building stock as a research object *Build. Res. Info.* 30 226–36

[8] Przomil M, Paine D, Pierro G, Mosimian C, Poplawski M, Saha A, Bender J and Granderson J 2021 Metadata schemas and ontologies for building energy applications: a critical review and use case analysis *Energies* 14 2024

[9] Bundesbericht Energieforschung für die Energiewende (Federal Ministry for Economic Affairs and Climate Action) März 2021 [Online] available: https://bmwi.de/Redaktion/DE/Publikationen/Energie/bundesbericht -energieforschung-2021.pdf?__blob=publicationFile & v=14

[10] Projektträger Jülich (PtJ) 2022 Aktuelles - energiewendegebäude.de https://energiewendegebäude.de/energiewendegebäude.de/ (accessed 07 November 2022)

[11] Oppermann L, Hessel S, Guldner A, Heiwolt K, Krassowski J, Schade U, Lange C and Prinz W 2021 Finding and analysing energy research funding data: the EnArgus system *Energy AI* 5 100070

[12] Krüger A and Kolbe T H 2012 Building analysis for urban energy planning using key indicators on virtual 3D city models—the energy atlas of BERLIN Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. XXXIX-B2 pp 145–50

[13] Deutsche Forschungs-Gemeinschaft (DFG) 2021 National Research Data Infrastructure https://www.dfg.de/foerderung/programme/nfdi/ (07 November 2022)

[14] Bronger T, Demandt E L, Heine I, Kraft A, Preuß N and Schwarz A 2021 *Die Nationale Forschungsdateninfrastruktur für die Ingenieurwissenschaften (NFDI4Ing)* (Bausteine Forschungsdatenmanagement 20212) 110–23

[15] Energieagentur Rheinland-Pfalz, Fraunhofer ITWM and Discovergy (openMeter) 2021 [Online] available: https://openenergy-platform.org (accessed 17 November 2022)

[16] Otto-von-Guericke-Universität - Institut für kooperierende Systeme, Reine Lemoine Institut, Europa-Universität Flensburg Flensburg et al (Open Energy Platform) https://openenergy-platform.org (accessed 17 November 2022)

[17] Technische Informationsbibliothek—German National Library of Science and Technology (TIB), Leibniz Universität Hannover and L3S Research Center https://orkg.org/orkg/ https://orkg.org/orkg/ (07 November 2022)

[18] Kirnats L, Joost J-N, Berg S, Frisch J and Van Trecck C 2018 Tiefenbohrung - Digitale Werkzeuge und softwarebasierte Lösungsansätze (Wissenschaftliche Begleitforschung Energiewendegebäude) [Online] available: https://forschungsnetzwerke-energie.de/lw_resource/datasetfiles/aggregationspublication95162103DAA51D19E0539A695E862DF6/live/document/Tiefenbohrung_Digitale_Werkzeuge_und_softwarebasierte_L%3C%3Esansamt%2C%3Driec.pdf
