

Integrating Data and Service Lifecycle for Smart Service Systems Engineering: Compilation of a Lifecycle Model for the Data Ecosystem of Smart Living

Henrik Kortum
DFKI,
Osnabrück, Germany
henrik.kortum@dfki.de

Jonas Rebstadt
DFKI,
Osnabrück, Germany
jonas.rebstadt@dfki.de

Simon Hagen
DFKI,
Osnabrück, Germany
simon.hagen@dfki.de

Oliver Thomas
DFKI,
Osnabrück, Germany
oliver.thomas@dfki.de

Abstract

In smart service systems engineering, where actors rely on the mutual exchange of data to create complex and holistic solutions, integration is crucial. Nevertheless, the management of data as a driving resource still lacks organizational structure. There is no holistic lifecycle approach that integrates data and service lifecycle and adopts a cross-actor perspective. Especially in data ecosystems, where sovereign actors depend on the mutual exchange of data to create complex, but transparent service systems, an integration is of crucial importance. This particularly applies to the smart living domain, where different industries, products and services interact in a complex environment. In this paper we address this shortcoming by proposing an integrated model that covers the different relevant lifecycles based on a systematic literature review and supplement it by concrete domain requirements from the smart living ecosystem obtained through semi-structured expert interviews.

1. Introduction

With the rise of data as an essential economic resource [1], data ecosystems are gaining relevance [2]. They hold great economic potential, such as creating new data-driven business models and services [3]. One of the fastest growing data ecosystem is the "smart living mega ecosystem" [4] by extending the smart home concept beyond simple home automation. On a technological level, the cross-actor exchange of data and its utilization in AI technologies are of key importance [5]. Speech recognition, object identification or energy management are exemplary application fields of AI in the smart living domain. The smart living ecosystem includes related areas such as health and education [6, 7] or energy management and retail [7], in addition to traditional solutions targeting comfort. This demonstrates the dependency on complementary domains and technologies [8]. Consequently, this results in a high degree of complexity and a need for cooperation

between a variety of actors with diverse competencies [9]. From a service science perspective, this sounds familiar. Despite the focus on data ecosystems this is subject of currently discussed smart service systems [10]: Utilizing smart products [11] as boundary objects to facilitate value-in-use [12] within an overarching (eco-)system [13] incorporating stakeholders from various domains [14]. It seems reasonable to therefore build upon approaches from established smart service systems engineering (SSSE) that allow a systemic perspective and can be linked to the lifecycle perspective of data ecosystems. Especially since SSSE is based on concepts such as product-service systems (PSS; [15]) and is thus part of a comprehensive history of engineering approaches to the integration of tangible goods and intangible services [16]. The use of process models and lifecycles is common in AI systems [17] and has already been used in product and service development, too. For example, [18] examined this by proposing a joint product and service lifecycle. However, there is a shortcoming of scientific work that explicitly focuses on the orchestration of data and service offerings across different actors [14] within one value offering. This results in a need for academic research that focuses on the coordination of the various lifecycles to support the engineering of smart service systems involving actors from different industries. The aim of this paper is to further close this gap and to provide a more holistic view on lifecycle assessment in SSSE, especially aiming at the data and service perspective. Even though the development of physical products as boundary objects of data-driven service systems is a core element of various classical procedure models, we abstract the product development process to the level of resulting data streams, decoupling the service from possibly underlying and exchangeable, short-lived products typical in the area of smart living. Within this paper we examine the status quo in the consideration of cross-actor lifecycle approaches using examples of the domain smart living and to pro-

pose an integrated view on a lifecycle model that combines all relevant perspectives with a specific focus on smart service systems. This goal is addressed by the following research questions, splitting the main aspect into two sub-questions.

Main-RQ: How can an integrated data and service lifecycle approach support the engineering of smart service systems in the domain of smart living?

RQ1: How can data and service lifecycle considerations existing in scientific literature be integrated?

RQ2: What are the practical requirements, implications and challenges influencing the SSSE in the application domain of smart living?

The paper is structured in five remaining sections, starting with the theoretical background in section 2 and the multi-methodical research approach in section 3. In section 4, the results of the structured literature review and the expert interviews are presented, resulting in a holistic lifecycle including insights from the theoretic perspective as well as from the smart living domain to answer the named research questions. Finally, in section 5, we discuss the potentials and limitations of our work and outlining the possibility of further research in section 6.

2. Theoretical Background

SSSE extends the scope of service engineering [19] beyond methods for developing singular value propositions, towards the design of integrated products, services and information technology [20]. Even though the pool of methods for the development of classical product-service systems is extensive and includes a wide range of perspectives [21, 22, 23], there is a lack of approaches that focus on the orchestration of multiple components with AI capabilities across different providers and on the integration of process models from the areas of software development and data science [9]. This is particularly important within data ecosystems. Data ecosystems describe the interaction of different actors, such as organizations and individuals, exchanging data in networks to utilize them as resources [24]. Big Data and AI technologies enable the processing of huge, unstructured amounts of data in near real-time, thus providing the technological basis for the emergence of data ecosystems [25]. Data ecosystems hold great potential in terms of generating economic value [1, 26], such as creating new data-driven business models and services [3, 27]. The three essential roles in a data ecosystem are the *data consumer*, the *data producer* and the *intermediary* [2, 24]. Due to this cross-actor utilization of data, cross-actor data lifecycle considerations play a crucial role in data ecosystems [24]. In general, lifecycles are separating

the design of systems in different phases like design, evaluation and documentation, trying to depict all steps relevant for the development of software, information or physical systems [28]. Due to its interdisciplinary nature, the engineering of a smart service system must take into account the software and product development lifecycle as well as the data lifecycle. One of the first models which can be assigned to the description of lifecycles is the waterfall model [29]. While the waterfall model is focusing on the classical development of software, [30] presented a model called CRISP-DM offering a standard for data mining projects including relevant implications from machine learning, which might be a relevant component of data-driven services as well. However, both models summarize all activities after the initial development in a single phase named “maintenance” or “deployment”, which has drawbacks due to increasing number of tasks that is required after the initial deployment [31]. In addition to these software oriented lifecycles, [32] have been one of the first to present a data lifecycle focusing on the maintenance of high quality. Setting up on this, [33] described lifecycles as set of stages describing how data passes through a system from entry to exit. [17] extended this understanding by explicating steps for data acquisition, processing and preservation. For a more detailed view on existing data lifecycles, [33] compared several models and started a first alignment of the contained phases. Lifecycle considerations are particularly essential in environments characterized by complex systems, such as those found in the smart living data ecosystem [34].

3. Research approach

3.1. Multimethod research design

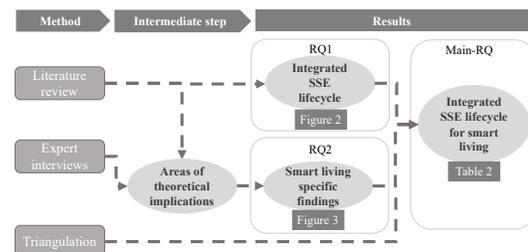


Figure 1. Research approach

To answer our main research question, we apply a multi-method approach [35], allowing us to consider the different perspectives of the question. Therefore, we divide our main RQ into the two sub-questions RQ1 and RQ2 and select appropriate methods to answer them. In order to capture the current state of research and derive a first integrated lifecycle model to

support SSSE, we conduct a structured literature review according to [36]. To ascertain concrete practical requirements, we use a combination of semi-structured expert interviews [37] and qualitative content analysis [38]. Subsequently, the two partial results are condensed into an overall result through triangulation [35]. Finally, we propose an integrated lifecycle model, complemented by specific considerations for the smart living ecosystem to answer our main RQ. Figure 1 visualizes our research approach.

3.2. Literature Review

In order to investigate how data and service lifecycle considerations can support the development of smart service systems, a literature search was conducted according to [39]. The aim of the review was to identify data and software related dimensions influencing the engineering of smart service systems. For this purpose, different search terms were evaluated, starting with the generalist search term (*“procedure model” OR “process model” AND data AND “Smart Living” OR “Smart Home”*). Due to the limited accuracy of the results, we differentiated between the various sub-areas of the data perspective and removed the smart living reference. These considerations and adjustments resulted in the new search term: (*“procedure model” OR “process model” OR lifecycle OR framework*) AND (*“data science” OR “data mining” OR “data engineering” OR “machine learning” OR “artificial intelligence” OR “knowledge discovery” OR “decision support systems”*). The term was applied to eight relevant databases in information systems research. The results were analyzed in several steps, based on the relevance of the publications in relation to RQ1 by first analyzing the title, then the abstract and finally the full text. The resulting 80 publications were analyzed and categorized based on [36], identifying nine main areas as presented in section 4.1.

Table 1. Overview: Literature search

| Database | Hits | on title | on abstract |
|--|---------|----------|-------------|
| IEEE | 239.760 | 118 | 42 |
| AISeI | 18.257 | 27 | 17 |
| SpringerLink | 176.995 | 82 | 29 |
| Science Direct | 133.311 | 46 | 21 |
| Wiley | 171.846 | 56 | 13 |
| Ebsco | 2 | 0 | 0 |
| Web of Science | 89.509 | 71 | 32 |
| JSTOR | 0 | 0 | 0 |
| Summed up | | | 154 |
| Duplicates, no access or out of scope | | | 74 |
| Total | | | 80 |

3.3. Expert interviews and qualitative content analysis

Due to the lacking application of these rather generic concepts, expert interviews have been conducted. As RQ2 aims to gain a deeper understanding of the requirements and challenges in SSSE in the smart living domain, we follow a qualitative empirical approach [40]. Representatives of the following company types were interviewed: Association of product manufacturers, IT service provider, product manufacturer, energy service provider, association of craftsmen, cloud service provider. All interviewed representatives are experts in the sense of [37] and hold a management position in their company. In order to obtain a broader overview, we also interviewed experts of industry associations who are able to speak for their sector on a representative basis. On average, an interview took 69 minutes. All questions were formulated in an open manner to motivate a broad, unbiased answer. The guide consisted of 23 questions in total and covered four topics, starting with demographic questions (1). The first major topic (2) consists of business-driven questions about the domain of smart living. Additional questions were aimed at obtaining assessments of the market's growth potential and perceived challenges. Another content block (3) focused on the SSSE within the data ecosystem and on the integration of AI into the companies' service offerings. Additionally, an overview of data sources currently and potentially used for this purpose should be revealed including potentials and risks resulting from it. The last block (4) consisted of questions about lifecycle analysis and process models in the context of SSSE to determine the relevance and acceptance of this methodologies for the respective company. The expert interviews were conducted personally, synchronously and in German, using the online conference tool Microsoft Teams. All interviews were recorded and transcribed using the tool MAXQDA. Care was taken to ensure a verbatim record, transcription and an unbiased translation without interpretation [41]. Following the approach of [38] for qualitative content analysis, a five-step evaluation concept was applied to the raw data in order to further transform the transcripts to answer the research questions and derive the relevant information to enhance our model. The evaluation concept consists of the following phases: (1) *Determination of research questions and analysis perspective*, (2) *Development of a category system*, (3) *Search texts for relevant information and assign them to corresponding categories*, (4) *Aggregation of information over different interviews*, (5) *Analyzing the results to answer RQ2*. Based on the concepts derived from the literature review, a category system was designed to encode the

transcribed interviews. The categories as well as the assignment of the core statements into these categories is shown in 4.2 (2,3,4). Finally, we answer RQ2 in section 4.2 (5).

4. Results

4.1. Literature Review: Identification and integration of existing lifecycle approaches

As introduced earlier, data are an important resource in smart service systems. However, the implications arising from its (cross-actor) use are not yet adequately integrated into the engineering process. Our literature review identified 80 relevant publications that provide insights into 46 data- or software-related concepts in the context of SSSE. These concepts have been clustered bottom up resulting in eight meta-concepts. The first two meta-concepts are quite similar and are covering the processes around the life of considered systems, called *Process Model* and *Lifecycle*, extending the solely development process. In contrast to these, the other concepts do not describe concrete processes, but do influence the design and implementation of them. Three of these concepts are technically oriented, including procedures and frameworks for *Knowledge Representation*, several topics regarding specific *ML Methods*, as well as all aspects required for creation and integration of procedures and algorithms in a *ML Platform*. In addition to these technical aspects, “soft” factors are found to be of relevance. The first of these meta concepts, *Data Responsibility*, is summing up legal, ethical and organizational aspects. The second one, *User Integration*, includes all topics regarding the interaction between and the integration of humans in the considered systems. This section examines the first two concepts in more detail,

while the other five concepts will be considered in more detail in section 4.3. Several publications are characterizing different lifecycles or process models, focusing either on the development of software-systems or on acquisition and processing of data. In order to integrate the identified aspects and phases, data and software lifecycles were joined together, with a special focus on AI-based systems, as they are particularly dependent on data. On the side of software and services focusing on AI, CRISP-DM was identified as the model most frequently cited in the literature research. Due to this, CRISP-DM, formalized for example by [30, 42] and adopted by many others like [43], was used as a starting point for our integration. On the data perspective we were not able to identify one specifically dominant model. Due to this, all data lifecycle models identified in the literature review have been collected. This resulted after a first removal of duplication in [44, 45, 46, 47, 48] and [33] itself joining phases from eleven different data lifecycles. Setting on [33], the models have been deconstructed in the mentioned phases and compared based on their descriptions. Despite differentiating names, large overlaps have been found, which can be summed up in 16 distinguishable phases as one can see in Figure 2 (grey, yellow and red phases). Due to its intuitive structuring in three main phases “data acquisition”, “data processing” and “data preservation” and its high level of detail, the naming of most phases was oriented to [17] (grey and yellow in Figure 2). Additional phases from other models have been colored in red. These 16 sub-steps have been aligned with the phases mentioned by CRISP-DM using the extracted descriptions. The process starts with a problem specific business understanding [30] and an initial data collection from various data sources [17]. These data sources need to be integrated [33]. Before filtering, a quality check and

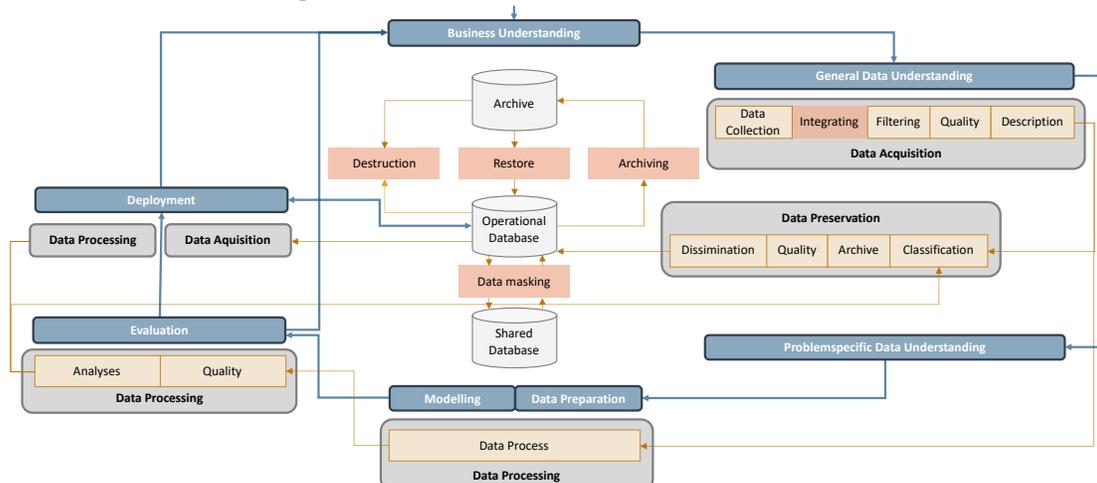


Figure 2. Integrated lifecycle model

first descriptions can be initialized [17] and aligned with a general data understanding [30]. After this sequential part, the process is split into two streams. On the one hand, the data preservation phase based on [17] and enriched by a data masking phase [33] for confidential data sources, is initialized. The data will be saved in some type of operational data preservation system, from which currently not used data will be archived in another system to be possibly restored afterwards [33]. During the lifetime of data, this can happen several times until the data is finally no longer required by any of the system actors. At this point the data will be destructed, as described by [33]. On the other hand, the problem specific data understanding has been exposed from CRISP-DMs data understanding, focusing on problem specific analysis and the specification of objectives and hypotheses as stated out e.g. by [46]. Using this definitions, the actual data processing [17] can be carried out, further differentiated by CRISP-DM in data preparation and modeling [30]. Following this, the evaluation phase [30] takes place, which is further specified by [17] in data quality and analysis. If the evaluation was completed with a positive result, the model can be deployed [30]. Even though in an automated manner, in the deployment for each prediction most of the data acquisition and processing steps [17] need to be applied. In addition to phase-related concepts, many general concepts were identified that focus on complete lifecycles or are applicable to all phases. Examples of this are the collection of process model and lifecycle and their differentiation in several characteristics as well as concepts applicable to specific phases. In section 4.3 the identified concepts will be joined with findings from the interviews and presented in detail.

4.2. Expert Interviews: Elevation of smart living specific requirements

In the following, the results of the expert interviews and the qualitative content analysis are presented to answer RQ2. Selected core statements (see Figure 3) are assigned to the appropriate phases of our integrated lifecycle model, where they serve as smart living specific requirements and extensions. The assignment is done via indexing, *LC#1* refers e.g., to the 1st entry of the lifecycle concept. **Business Understanding:** All participants agree that the customer and his needs play a crucial role over the entire lifecycle of smart living services. The goal of any smart living offer should therefore be the creation of real added value for the customer (*BI#6*). A frequently mentioned step in this process is the understanding of customers and their requirements (*LC#1*). In this context, platform

and service providers in particular criticized the overall orientation of many smart living solutions, which are often gadget-like and product-centric rather than really contributing to the satisfaction of customer needs (*IS#1, IS#2, IS#6*). Therefore, currently available smart living services tend to offer only partial solutions and could not be linked to more sophisticated systems (*IS#1, IS#3, CH#3*). However, the customer expects solutions to be thought of holistically and to integrate various IOT devices and services (*IS#7*). Several participants indicated that the development, operation, and maintenance of such ecosystem-wide smart service systems require the establishment of data value chains that integrate all potential stakeholders and take into account their economic interests (*LC#4, BI#1*). The interviewees generally agree that a combination of different core competencies and actors is necessary to create those holistic solutions (*IS#4*). First, companies are needed which have a direct touchpoint to the customer, use it to generate data and provide it to the ecosystem as a resource (*CO#2*). In the next step, a central intermediary is needed to bundle this data and make it available to others (*CO#2*). One of the tasks of this intermediary should be to create transparency about the ecosystem's data sources (*DR#7*). In addition, value creation within the ecosystem is dependent on data scientists and service engineers who can develop intelligent services based on the data. These services could be end-user services as well as middleware services, which are provided to the ecosystem to build on. While there is basic agreement about the tasks within a data ecosystem, there is no agreement about what the framework of such an ecosystem should look like (*CO#1*). The platform providers surveyed generally advocate more open solutions, while product manufacturers prefer cooperation within established, trustful consortia (*DR#4, CO#5*). However, potential disadvantages in the form of dependencies were also pointed out. Some stakeholders fear a loss of their autonomy (*CH#11*) and the risk of having no influence on the availability and quality of the data provided by the partners (*CH#2, DS#2, DS#5*). Many interviewees also predicted the emergence of completely new business models (*BI#1*). These could, for example, consist of the disposal or transfer of data to third companies (*BI#3*), which would trigger a synergy driven value creation process (*BI#2*). **Data Acquisition & General Data Understanding:** All interviewees agree on the heterogeneous nature of the smart living data ecosystem. It is characterized by many different actors, a lack of standards (*CH#1, CH#15*), monolithic systems (*DS#6*), and short lifecycles of IOT devices (*LC#6*). For smart service systems to offer holistic value, a cross-actor integration of various data sources, products and technical services is necessary

(IS#7). To integrate products and data extending the original scope of system providers, an opening of the ecosystem is indispensable (DS#5, IS#8). In particular, the respondents demanded transparency about the data generated in the ecosystem (DR#7) and a clarification of data ownership (DR#1). Furthermore, equal access should be possible for all actors authorized by the data owner (DR#8, DR#2). The basic data availability was assessed quite unequally by the participants. A representative of a platform provider stated that many companies do not generate the data necessary to operate their service themselves (DS#1) and are consequently dependent on data producers (CH#2). This opinion overlaps with the statement of a cloud provider who argued that the design of a large number of the products currently available in the smart living market does not allow sufficient data generation (DS#1). While these statements indicate a general lack of data (CH#6), other experts are of the opinion that the alleged lack is used as an excuse. The problem is not caused by a too small population of data, but by its fragmentation into isolated solutions (DS#6). It was found that the problem of this isolation exists on two levels (CH#12). On the one hand data silos are created between different companies (cross-actor). On the other hand, silos also exist within a company, where they arise between departments and organizational units. Even within a single company, data flows are not continuous and there is a lack of transparency and access. The experts frequently cited widespread silo thinking and budget conflicts as the reason for this. (BI#7). **Data Preparation/ Processing, Modeling & Evaluation:** According to the general assessment of the participants, currently only a minority of the smart living solutions uses real AI components (LC#7). At the same time, the consensus opinion prevails that this will and must change massively in the near future, also in order to meet customer requirements (LC#7). Machine learning methods must be increasingly applied so that smart living solutions can make ever more complex decisions autonomously. However, all interviewees also emphasized a responsible approach to the use of AI (DR#9), especially when it comes to linking different data sources to gain knowledge about the customer (DR#3). Consequently, the phase of data preparation/processing, modeling & evaluation will become increasingly important. All surveyed companies that currently apply AI use data scientific process models, even if there is no uniform standard (LC#12). Companies which have not integrated explicit AI functionalities into their solutions have not mentioned any concrete process models (LC#12). The importance of transparency in the use of data was emphasized by all participants. It is the foundation for establishing a

trusting relationship, both between individual companies as well as between customers and companies (CO#5). Consequently, companies should be transparent about which data sources they combine for which purposes and the models they use (DR#9, DR#6). **Data Preservation:** A prerequisite for an increased use of AI methods is the availability of historical data to initiate the necessary training of such models. For this purpose, data should be persisted throughout the entire value chain and annotated with appropriate labels. The understanding of the necessity to build up such data histories is consistently present in the interviewed companies (DS#8), even if the creation of such annotated data sets has rarely been implemented operationally. In order to adequately address these challenges, on one hand, conditions must be created to ensure that the data and systems of the individual actors in the ecosystem become exchangeable and interoperable. Currently, the linking of data represents a great challenge (DS#3). To cope with rapid change, data and interface management as well as data processing must be continuously developed and subjected to lifecycle management (LC#3). On the other hand, solutions transparency (DR#6, DR#7), data quality (DR#5) and the guarantee of availability (CH#4) are critical factors for the willingness to share data in an ecosystem and to integrate third party data into own processes. A representative of a platform provider stated that some of his customers were skeptical about integrating external data into their products to avoid a loss of control (DS#2). He suggested the establishment of a quality label for data and its availability to provide potential users with more security. **Deployment:** In addition to the initial collection of customer requirements, a strong integration of the customer is also of crucial importance after the launch of a solution (LC#2). In this context, representatives of the electrical trade see an expansion of their tasks, which in future no longer end with the initial installation of the systems but will increasingly include accompanying and supporting the customer during runtime. From the perspective of product manufacturers and the electrical engineering industry, the adaptability of smart living solutions will lead to a further shortening of product life cycles in the ecosystem (LC#6). Smart living services should be able to recognize and react to changes in user behavior and customer preferences. To enable such monitoring, solutions in the data ecosystem must possess an appropriately designed data management and cover a holistic data lifecycle. There is a need for continuous adaptation and updating, especially regarding machine learning models. In the context of the business understanding phase, changing customer behavior and needs have already been addressed. Several respondents emphasized the impact on the

| | | | | | |
|---------------------------------|---|--|--|---|--|
| Lifecycle (LC) | <ul style="list-style-type: none"> (1) Explicit consideration of customer requirements (2) Support of the customer during use (3) Data management, connection and processing must be continuously refined. (4) Development and maintenance of a cross-actor data value chain | <ul style="list-style-type: none"> (5) AI models must be continuously retrained in order to react to a changing environment (6) Short-lived and gadget character of many IOT products (7) AI not widespread but gaining in importance to enable required functionality | <ul style="list-style-type: none"> (8) Currently many Smart Living products are replaced rather than serviced (9) Optimization during runtime hardly exists until now (10) Only isolated consideration of individual phases and steps of various process models | <ul style="list-style-type: none"> (11) No uniform lifecycle approach (12) No uniform use of process models (13) There is currently no lifecycle management for buildings | |
| Data responsibility (DR) | <ul style="list-style-type: none"> (1) Clarification of data ownership (2) The data owner should at all times have control over who he shares it with (3) Caution when merging and aggregating data (4) Establish a relationship of trust within the ecosystem (5) Quality of data shared with the ecosystem must be ensured (6) Transparency about the use of the data in all phases (7) Transparency of all data available in the ecosystem (8) Equal access to data sources for all actors in the ecosystem (9) Responsible use of AI | Data sources (DS) | integrated solutions (IS) | Cooperation (CO) | Business Integration (BI) |
| Challenges (CH) | <ul style="list-style-type: none"> (1) Heterogeneous, non-interoperable system landscape (2) Dependencies on other ecosystem actors in preceding instances (in value chains) (3) Dominated by proprietary solutions (4) Maximum requirements for data availability (service level for data) | <ul style="list-style-type: none"> (1) Many companies in the ecosystem do not generate data themselves (2) Cautious use of external data, due to dependency fears (3) Linking the different data sources is a challenge (4) Many products from the Smart Living environment do not yet generate data (5) The idea of integrating external data is widely spread among companies (6) Data silos and monolithic solutions hinder the development of smart living solutions (7) General willingness to pay for external data (8) Creation of data histories as a basis for AI integration must be driven (9) Use of external data sources varies | <ul style="list-style-type: none"> (1) Ecosystem only provides partial solutions today (2) Need for customer-oriented thinking in the development of holistic solutions (3) Lack of compatibility of proprietary solutions (4) Development of holistic solutions requires the competencies of different actors (5) Current degree of integration is limited to the most necessary (6) Product-centered view inhibits the development of holistic solutions (7) Holistic solutions require stronger integration of external systems and data (8) Holistic solutions require collaboration across industries and domains | <ul style="list-style-type: none"> (1) Various types (Cooperation, Platforms, Data cooperatives, ...) (2) All roles of a classic data ecosystem represented (Consumer, Producer, Intermediare) (3) No distinct role assignment (4) Cultural fit is of great importance (5) Cooperative data exchange requires a basis of trust | <ul style="list-style-type: none"> (1) High potential for cross-actor business models (2) Cooperation in the ecosystem holds the potential for huge synergy effects (3) Monetisation of data as a new business model (4) Networking with related domains such as energy or mobility to generate cross-domain business models (5) Demand for integrated, holistic solutions (6) Business models must reward the data owner (7) Internal company incentive systems must change in order to promote cooperation and data-driven thinking |
| | <ul style="list-style-type: none"> (5) A large number of very diverse actors (6) General lack of data (7) Need for initial investment in shared infrastructure (8) Need for guaranteeing secure data exchange | <ul style="list-style-type: none"> (9) Lack of data-oriented and cross-product thinking (10) Only sporadic understanding of digital business models (11) Fear of losing autonomy (12) Internal and external data silos | <ul style="list-style-type: none"> (13) Unwillingness to share data (14) Unsuitable mechanisms systems in companies (15) Lack of standards (16) Requires the combination of different core competences | | |

Figure 3. Core statements derived from the expert interviews

maintenance of AI models, as they need to be continuously trained with new data to reflect the current environment. (LC#5). A representative of a platform provider demanded that an update policy must be established for AI components, which brings the machine learning models used up to date in the same way as software updates. Describing the requirements and implications mentioned by practitioners from the smart living ecosystem, RQ2 has been answered partially.

4.3. Triangulation: Adjusting the lifecycle for the smart living domain

In the following, the results of the literature review and the expert interview are aggregated using the method of triangulation [35]. The lifecycle model derived from the literature is supplemented by practice-oriented requirements and specifications from the smart living domain. In both sub-steps, many concepts were identified that focus on complete lifecycles or are applicable to all phases. Some publications specify metrics and ideas driving the systems. Exemplary for this are [49], [50] or [51] introducing return-on-investment-, risk- or goal-driven development, as well as [52] and [53] presenting human- and business-centered development. Also, some of the authors extend models and lifecycles with explicit role specifications such as [54, 55, 56]. The expert interviews also show a quite differentiated picture of the process and lifecycle models used and phases covered in the development of smart service systems. We found that product

manufacturers rely on traditional, sequential processes, while service and platform providers are increasingly applying agile, iterative, and more holistic development methods (LC#12). Product manufacturers often focus on their own devices when developing service offerings (IS#6), in contrast to service and platform providers whose goal is to create universally compatible solutions. Nevertheless, product manufacturers are also recognizing that, in order to develop holistic, customer-oriented systems, it will become necessary to integrate into the ecosystem and collaborate with other actors (IS#4, IS#7, IS#8). In the context of this co-creation, a representative of an energy service provider emphasized that a first step in establishing and implementing shared process models is to build a collaborative business understanding that give equal consideration to the economic interests of all actors involved (BI#1). Through our work we were able to determine that the consideration of different lifecycles plays a crucial role in the smart living data ecosystem. On the one hand, this differentiation includes various types of lifecycles with regard to their scope of observation: like data, service or software lifecycles. On the other hand, lifecycles exist in every company and, possibly, multiple times, which further increases the complexity of their integration. The different lifecycles cannot be considered independently of each other but are interrelated and mutually dependent. The interviewed companies were well aware of this complex interaction, even if the explicit consideration of resulting implications in the engineering processes is not yet

far advanced. This resulted in a very inconsistent picture of the lifecycle models used (*LC#11*). The revealed shortcoming implies a practical need for an integrated lifecycle model to support ecosystem businesses in developing cross-actor, data-driven service systems. In contrast to these more general aspects, most concepts from literature and expert interviews can be assigned to one or more specific phases. These are presented in Table 3, assigned to the best matching phases. Concepts identified in the literature search are assigned to “general concepts” and aspects collected from experts in the smart living domain are mentioned under “Smart living specifications”.

Table 2. Supplementation of the lifecycle model

| Phases | General concepts | Smart living specifications |
|--|---|---|
| Business Understanding | Requirements Engineering (Altarturi et al. 2017) | <i>BI#1, BI#2, BI#3, BI#6, CH#2, CH#3, CH#11, CO#1, CO#2, CO#5,</i> |
| | Integrating Domain specific Implications (Cao 2008) | |
| | Process Mining (Van Der Aalst and Damiani 2015) | |
| | Monetisation of Data (Martinez-Plumed et al. 2019) | <i>DS#2, DS#5, DR#4, DR#7, LC#1, LC#4, IS#1, IS#2, IS#3, IS#4, IS#6, IS#7</i> |
| Data Acquisition & General Data understanding | Standardisation of interfaces and data (Salman et al. 2019) | <i>BI#7, CH#1, CH#2, CH#6, CH#12, CH#15, DR#2, DR#7, DR#8,</i> |
| | Data Quality (Buchheit et al. 2000) | <i>DS#1, DS#5, DS#6, LC#6, IS#7, IS#8</i> |
| | Linking Data Sources (Pandey and Shukla 2018) | |
| | Epistemology (Rennolls 2005) | |
| | Data validation (Li et al. 2016) | |
| | Time validity of data (Wang et al. 2016) | |
| Problem specific Data Understanding | Objectives & Hypotheses derivation (Rotondo and Quilligan 2020) | see general data understanding phase |
| | | |
| Data Preparation, Processing, Modeling and Evaluation | Machine Learning based (Martinez 2014) | <i>DR#5, DS#3, DS#2, DS#8, LC#3</i> |
| | Model Selection (Rotondo and Quilligan 2020) | |
| | Compute-time optimization (Fernandez et al. 2014) | |
| | Model validation (Zhang et al. 2020) | |
| | Business-Metrics-based Evaluation (Sharma and Osei-Bryson 2010) | |
| | Expert-Integrated development (Buchheit et al. 2000) | |
| Data Preservation | Non-Expert-Integrated development (Tuovinen 2016) | |
| | Data Management (Martinez 2014) | <i>CO#5, DR#3, DR#6, DR#9, LC#7, LC#12</i> |
| | Data preservation technologies (Zhao et al. 2018) | |
| | ML-Platform specifications (Guido et al. 2017) | |
| | Knowledge-classification (Schuh et al. 2019) | |
| | Knowledge Fusion (Wang et al. 2016) | |
| Deployment, Data Acquisition & Data Processing | Human Computer Interaction (Bernabè-Moreno and Wildberger 2019) | <i>LC#2, LC#5, LC#6</i> |
| | Architecture of Systems and Platforms (Zhao et al. 2018) | |
| | Decision-Support-oriented Systems (Osei-Bryson 2012) | |
| | Usability/ Acceptability (Elam and Henderson 1983) | |
| | User-accepted reaction-time (Wang et al. 2016) | |

5. Discussion

Even though the integration of lifecycles is already reasonable inside a single company, only the consideration of intercompany relations allows the potentials to be fully exploited. The integration of smart living specific requirements into the developed lifecycle offers a good starting point to further elaborate this interaction. An integrated lifecycle model must therefore promote synergies and value creation between actors, but also to guarantee actors a maximum of security and autonomy so that they are willing to participate in the system. The heterogeneity of smart living offers the opportunity for cooperation between small and medium-sized companies, as a counterbalance to industry giants like Amazon or Google and promote

the development of independent, transparent smart service systems. A cross-company, holistic view of lifecycles can support SSSE by making complex interrelationships transparent and reducing the hurdles for cooperation projects. In terms of general data availability, it was found that the design of many products currently offered on the smart living market do not allow sufficient generation of data. An integrated lifecycle optimizes the access to data in the ecosystem and enables the companies to take collaborative data-based aspects into account already during raw data generation. Furthermore, in many companies the problem is not caused by a too small population of data, but by its fragmentation into isolated solutions. This isolation exists both on an intra-company and on a cross-company level. The consideration of an integrated lifecycle model for the engineering of smart service systems can help ecosystem actors to make these data silos transparent and support their resolution. Statements from companies that are also active in other domains indicate that this is not an exclusive problem of smart living, which indicates a use of our findings for other ecosystems as well. Especially based on the statements regarding data responsibility we propose a creation of concrete mechanisms to promote transparency and dissolve data silos. Such mechanisms should make transparent what data is available in the system and transparently document its use across the individual value creation stages within the lifecycle [57]. This provides data producers and end users with greater explainability and assurance of what is happening with their data. Promoting data transparency in this way can help to address concerns about the crucial factor of data quality identified in the interviews and discussed in the literature. In addition, creating systems to increase transparency around the use of data can be the basis for more equitable monetization of data, as raised in the interviews with the need to reward data owners.

6. Conclusion and Outlook

This paper presents an integration approach of data and service lifecycle to support SSSE in the smart living data ecosystem. Through a structured literature review, we provided an overview of the current state of research in the field of lifecycles. Based on these theoretical findings, we derived and proposed an integrated lifecycle model to answer RQ1. Then, through expert interviews and qualitative content analysis, we identified practical requirements, implications and challenges that influence SSSE in the data ecosystem of smart living, thus answering RQ2. Finally, we extended our initial model with the findings from RQ2 to answer our main RQ. Our work contributes to the research areas of data ecosystems, SSSE and lifecycle

research. More explicitly it focuses on the coordination of the various development processes and lifecycles within an ecosystem. Using this, the development of coordinated systems involving actors from different industries can be supported, resulting in a more holistic view on lifecycle assessment in SSSE. Nevertheless, our work is not free of limitations. Although care was taken to include all ecosystem actor types in the selection of interviewees, it cannot be guaranteed that all domain-specific requirements were identified. However, this work and the lifecycle model proposed in it provide a starting point for further evaluations. We encourage others to transfer our model to and evaluate it in other domains. In addition, further research should address how mechanisms to promote data sharing in smart service systems can be implemented in a transparent and secure manner.

7. Acknowledgement

This research was conducted in the scope of the research project ForeSight, which is funded by the German Federal Ministry for Economic Affairs and Energy (BMWi). We would like to thank them for their support.

8. References

- [1] H. Chen, R.H.L. Chiang, and V.C. Storey, "Business intelligence and analytics: From big data to big impact", *MIS Quarterly: Management Information Systems* 36(4), 2012, pp. 1165–1188.
- [2] C. Cappiello, A. Gal, M. Jarke, and J. Rehof, "Data ecosystems--sovereign data exchange among organizations", *Dagstuhl Rep* 9(9), 2019, pp. 66–134.
- [3] F. Kammler, P.C. Gembarski, and H. Kortum, "Leveraging the Value of Data in the Continuum of Products and Services: Business Types in the Function-Oriented Offerings Model", *Proceedings of the MCPC 2021 [ACCEPTED]*.
- [4] Meyer, S., *Analyse der Mieteranforderungen und Akzeptanz vernetzter Systeme in vermieteten SmartHome-Wohnungen. Sozial und Technikforschung*, 2019.
- [5] J. Bauer, H. Hoffmann, T. Feld, et al., "ForeSight - Platform Approach for Enabling AI-based Services for Smart Living", *How AI Impacts Urban Living and Public Health - 17th International Conference, ICOST*, Springer (2019), pp. 204–211.
- [6] S. Solaimani, W. Keijzer-Broers, and H. Bouwman, "What we do – and don't – know about the Smart Home: An analysis of the Smart Home literature", *Indoor and Built Environment* 24(3), 2015, pp. 370–383.
- [7] R. Yasirandi, A. Lander, H.R. Sakinah, and I.M. Insan, "IoT Products Adoption for Smart Living in Indonesia: Technology Challenges and Prospects", *8th International Conference on Information and Communication Technology*, 2020.
- [8] J. Bauer, A. Ketschau, M. Michl, and J. Franke, "Die intelligente Wohnung als Baustein im Internet der Dinge", *Erste transdisziplinäre Konferenz zum Thema „Technische Unterstützungssysteme, die die Menschen wirklich wollen"*, (2014).
- [9] H. Kortum, L.S. Gravemeier, N. Zarvic, T. Feld, and O. Thomas, "Engineering of Data-Driven Service Systems for Smart Living: Application and Challenges", *IFIP Advances in Information and Communication Technology*, Springer (2020), pp. 291–298.
- [10] D. Beverungen, C.F. Breidbach, J. Poepelbuss, and V.K. Tuunainen, "Smart service systems: An interdisciplinary perspective.", *Inf. Syst. J.* 29(6), 2019, pp. 1201–1206.
- [11] M.E. Porter, and J.E. Heppelmann, "How smart, connected products are transforming competition", *Harvard Business Review* 92, 2014, pp. 64–88.
- [12] S.L. Vargo, and R.F. Lusch, "Service-dominant logic: continuing the evolution", *Journal of the Academy of marketing Science* 36(1), 2008, pp. 1–10.
- [13] P.P. Maglio, S.L. Vargo, N. Caswell, and J. Spohrer, "The service system is the basic abstraction of service science", *Information Systems and e-Business Management* 7(4 SPEC. ISS.), 2009, pp. 395–406.
- [14] R.C. Larson, "Smart service systems: Bridging the silos", *Service Science* 8, 2016, 359–367.
- [15] S. Hagen, and O. Thomas, "Expectations vs. reality - benefits of smart services in the field of tension between industry and science", *Wirtschaftsinformatik Proceedings*, 2019.
- [16] M. Boehm, and O. Thomas, "Looking beyond the rim of one's teacup: a multidisciplinary literature review of Product-Service Systems in Information Systems, Business Management, and Engineering & Design", *Journal of Cleaner Production* 51, 2013, pp. 245–260.
- [17] A. Sinaeepourfard, J. Garcia, X. Masip-Bruin, and E. Marin-Torder, "Towards a comprehensive data LifeCycle model for big data environments", *Proceedings - 3rd IEEE/ACM International Conference on Big Data Computing, Applications and Technologies*, 2016
- [18] S. Wiesner, M. Freitag, I. Westphal, and K.D. Thoben, "Interactions between service and product lifecycle management", *Procedia CIRP* 30, 2015, pp. 36–41.
- [19] Leimeister, J.M., *Dienstleistungsengineering und -management*, Springer Berlin Heidelberg, 2020.
- [20] D. Beverungen, H. Lüttenberg, and V. Wolf, "Recombinant Service Systems Engineering", *Business and Information Systems Engineering* 60, 2018, pp. 377–391.
- [21] D. Metzger, C. Niemöller, and O. Thomas, "Design and demonstration of an engineering method for service support systems", *Information Systems and e-Business Management* 15(4), 2017, pp. 789–823.
- [22] L. Trevisan, and D. Brissaud, "Engineering models to support product-service system integrated design", *CIRP Journal of Manufacturing Science and Technology* 15, 2016, pp. 3–18.
- [23] S. Hagen, F. Kammler, and O. Thomas, "Adapting product-service system methods for the digital era: requirements for smart PSS engineering", *In Customization 4.0*. Springer, 2018, pp. 87–99.

- [24] M.I.S. Oliveira, and B.F. Lóscio, “What is a data ecosystem?”, ACM International Conference Proceeding Series, Association for Computing Machinery (2018).
- [25] Y. Demchenko, C. De Laat, and P. Membrey, “Defining architecture components of the Big Data Ecosystem”, 2014 International Conference on Collaboration Technologies and Systems, pp. 104–112.
- [26] H., Kortum, J. Rebstadt, L.S. Gravemeier, and O. Thomas, “Data-based Customer-Retention-as-a-Service: Induktive Entwicklung eines datenbasierten Geschäftsmodells auf Basis einer Fallstudie der Automobilbranche”, HMD 58(3), 2021, pp. 537–551.
- [27] F. Kammler, S. Hagen, J. Brinker, and O. Thomas, “Leveraging the value of data-driven service systems in manufacturing: a graph-based approach”, ECIS 2019, 2019.
- [28] M. Kramer, “Lifecycle: an analyses based on the waterfall model”, Review of Business & Finance Studies 9(1), 2018, pp. 77–84.
- [29] W.W. Royce, “Managing the development of large software systems: concepts and techniques”, Proceedings of the 9th international conference on Software Engineering. IEEE Computer Society Press, 1987.
- [30] P. Chapman, J. Clinton, R. Kerber, et al., “CRISP-DM 1.0: Step-by-step data mining guide”, 2000, pp. 1–76.
- [31] Y. Bassil, “A Simulation Model for the Spiral Software Development Life Cycle”, International Journal of Innovative Research in Computer and Communication Engineering 03(05), 2015, pp. 3823–3830.
- [32] A. V. Levitin, and T.C. Redman, “A model of the data (life) cycles with application to quality”, Information and Software Technology 4, 1993, pp. 217–223.
- [33] Arass, M. El, I. Tikito, and N. Souissi, “Data lifecycles analysis: Towards intelligent cycle”, 2017 Intelligent Systems and Computer Vision, ISCV 2017, 2017.
- [34] A.L. Kor, M. Yanovsky, C. Pattinson, and V. Kharchenko, “SMART-ITEM: IoT-enabled smart living”, FTC 2016 - Proceedings of Future Technologies Conference, 2017, pp. 739–749.
- [35] J. Brewer, A. Hunter, et al., “Foundations of multi-method research: Synthesizing styles”, Sage, 2006.
- [36] J. Webster, and R.T. Watson, “Analyzing the Past to Prepare for the Future: Writing a Literature Review.”, MIS Quarterly 26(2), 2002, pp. xiii–xxiii.
- [37] A. Bogner, B. Littig, and W. Menz, “Introduction: Expert Interviews - An Introduction to a New Methodological Debate”, In Interviewing Experts. Palgrave Macmillan UK, 2009, pp. 1–13.
- [38] J. Gläser, and G. Laudel, „Experteninterviews und qualitative Inhaltsanalyse als Instrumente rekonstruierter Untersuchungen“, VS Verlag, 2009.
- [39] J. vom Brocke, A. Simons, K. Riemer, B. Niehaves, R. Plattfaut, and A. Cleven, “Standing on the shoulders of giants: Challenges and Recommendations of literature search in information systems research.”, Communications of the association for information systems 37(9), 2015, pp. 205–224.
- [40] M. Myers, “Qualitative research in business and management”, SAGE Publications Ltd., London, 2019.
- [41] E. McLellan, K.M. MacQueen, and J.L. Neidig, “Beyond the Qualitative Interview: Data Preparation and Transcription”, Field Methods 15(1), 2003, pp. 63–84.
- [42] R. Wirth, “CRISP-DM: Towards a Standard Process Model for Data Mining”, Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining, 2000, pp. 29–39.
- [43] Overgoor, G., M. Chica, W. Rand, and A. Weishampel, Letting the computers take over: Using Ai to solve marketing problems, 2019.
- [44] M. Zambetti, R. Pinto, and G. Pezzotta, “Data lifecycle and technology-based opportunities in new product service system offering towards a multidimensional framework”, Procedia CIRP 83, 2019, pp. 163–169.
- [45] Y. Zhang, S. Ren, Y. Liu, T. Sakao, and D. Huisingh, “A framework for Big Data driven product lifecycle management”, Journal of Cleaner Production 159, 2017, pp. 229–240.
- [46] A. Rotondo, and F. Quilligan, “Evolution Paths for Knowledge Discovery and Data Mining Process Models”, SN Computer Science 1(2), 2020, pp. 1–19.
- [47] Y. Li, M.A. Thomas, and K.M. Osei-Bryson, “A snail shell process model for knowledge discovery via data analytics”, Decision Supp. Systems 91, 2016, pp. 1–12.
- [48] F. Aijing, and F. Aiwan, “A supporting framework for real-time data mining”, Key Engineering Materials 439–440, 2010, pp. 1499–1504.
- [49] L. Cao, “Domain Driven Data Mining (D3 M)”, Proceedings - IEEE International Conference on Data Mining Workshops, ICDM Workshops 2008, pp. 74–76.
- [50] N.W. Grady, J.A. Payne, and H. Parker, “Agile big data analytics: AnalyticsOps for data science”, Proceedings - 2017 IEEE International Conference on Big Data, Big Data 2017 2018-Janua, 2017, pp. 2331–2339.
- [51] F. Martinez-Plumed, L. Contreras-Ochando, C. Ferri, et al., “CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories”, IEEE Transactions on Knowledge and Data Engineering 4347(c), 2019.
- [52] L. Cao, Y. Zhao, H. Zhang, D. Luo, C. Zhang, and E.K. Park, “Flexible frameworks for actionable knowledge discovery”, IEEE Transactions on Knowledge and Data Engineering 22(9), 2010, pp. 1299–1312.
- [53] K.M. Osei-Bryson, “A context-aware data mining process model based framework for supporting evaluation of data mining results”, Expert Systems with Applications 39(1), 2012, pp. 1156–1164.
- [54] M. Valadkhani, “Knowledge Discovery in Data (KDD) Process”, LinkedIn, 2016, pp. 1603–1608.
- [55] J.S. Saltz, and N.W. Grady, “The ambiguity of data science team roles and the need for a data science workforce framework”, Proceedings - 2017 IEEE International Conference on Big Data, pp. 2355–2361.
- [56] J. Saltz, and A. Suthrland, “SKI: An Agile Framework for Data Science”, Proceedings - 2019 IEEE International Conference on Big Data, pp. 3468–3476.
- [57] J. Rebstadt, H. Kortum, S. Hagen, and O. Thomas, “Towards a transparency-oriented and integrating Service Registry for the Smart Living Ecosystem”, INFORMATIK 2021, 2021.