Information about Product Usage

A Method to Support Information Need Analysis in Product Development

Vom Fachbereich Produktionstechnik

der

UNIVERSITÄT BREMEN

zur Erlangung des Grades

Doktor der Ingenieurwissenschaften (Dr.-Ing.)

genehmigte

DISSERTATION

von

Dipl.-Wirtsch.-Ing. Stefan Wellsandt

Gutachter*innen: Prof. Dr.-Ing. habil. Klaus-Dieter Thoben

Prof. Dr.-Ing. habil. Iris Gräßler (Universität Paderborn)

Tag der mündlichen Prüfung: 22.08.2024

German Abstract

Mitarbeitende der Produktentwicklung stehen heute eine Vielzahl an Medien mit produktbezogenen Informationen zur Verfügung. Hierzu zählen insbesondere solche aus dem Web 2.0 (engl. „participative Internet“) und dem Internet der Dinge. Aus diesen Medien können sich Mitarbeitende Informationen über die Nutzung von Produkten beschaffen. In der englischen Fachliteratur sind diese als „product usage information“ PUI bekannt. Sie liefern Hinweise auf Produktverhalten, Nutzungsumgebung, Nutzerinteraktionen, Nutzerinteressen, und Kundenmeinungen. PUI sind nur dann hilfreich für Mitarbeitende, wenn sie eine für den jeweiligen Anwendungsfall angemessene Qualität haben. Aus diesem Grund müssen Unternehmen Prozesse, die Informationsqualität beeinflussen, verwalten. Sie benötigen dazu Methoden, die ihnen helfen den eigenen PUI Bedarf sowie die qualitätsbeeinflussenden Faktoren zu analysieren. Durch eine präzise Bedarfsanalyse kann ein entsprechendes Informationssystem für PUI beschafft werden. Diese Dissertation entwickelt eine Methode zur Bedarfsanalyse für PUI auf Basis einer qualitativen induktiven Inhaltsanalyse von 22 PUI Anwendungsfallbeschreibungen. Vier Ursache-Wirkungs-Diagramme, die durch gemeinsame Informationsqualitätsmerkmale und spezifische qualitätsbeeinflussende Faktoren detailliert sind, bilden den Kern der Methode. Jedes Diagramm bildet einen Teilaspekt der Qualität von PUI ab. Die entwickelte Methode wird in vier Schritten angewandt und dabei durch die eigens entwickelte webbasierte Anwendung „QualiExplore“ unterstützt. Zwei Anwendungsfälle mit Unternehmen demonstrieren und evaluieren die Ergebnisse dieser Dissertation im Rahmen von Workshops. Der erste Fall fokussiert die Erweiterung eines bestehenden Informationssystems für PUI und der zweite Fall die Entwicklung eines neuen Systems. Wesentliche Ergebnisse sind eine kritische Reflektion der Evaluation, die Erfüllung der gestellten Anforderungen an die Methode, und die Grenzen des Methodeneinsatzes und der Methodenentwicklung. Im Ausblick dieser Dissertation wird auf die Bedeutung einer besseren theoretischen Basis für die zukünftige PUI Forschung eingegangen. Weiterhin werden Verbesserungspotentiale für die Methode und QualiExplore erläutert. Der letzte Abschnitt stellt komplementäre Forschungsthemen vor. Hierzu zählen Wechselwirkungen mit dem Datenschutz und Dateneigentum, Auswirkungen von komplexer werdenden Datenanalysen, und den Risiken durch Desinformationen.

Abstract

Employees in product development have access to a wide range of media with product-related information. These include those from Web 2.0 (participative Internet) and the Internet of Things. From these media, employees can obtain information about the use of products. The literature refers to this information as “product usage information” (PUI). It provides information about product behavior, usage environment, user interactions, user interests, and customer opinions. PUI is only helpful to employees if it is appropriate for the particular use case. For this reason, organizations need to manage processes that affect information quality. They need methods that help them analyze their PUI needs and the factors influencing quality. A corresponding PUI information system can be provided through a precise needs analysis. This dissertation develops a method for PUI needs analysis based on a qualitative inductive content analysis of 22 PUI use case descriptions. Four cause-and-effect diagrams, detailed by common information quality characteristics and specific quality-related factors, form the method’s core. Each diagram depicts an aspect of PUI quality. The developed method is applied in four steps and supported by a new web-based application called “QualiExplore”. Two use cases with companies demonstrate and evaluate the results of this dissertation in workshops. The first case focuses on extending an existing information system for PUI and the second on developing a new system. The main results are the evaluation’s critical reflection, the fulfillment of the method’s requirements, and the limitations concerning the method’s scope and development process. The outlook of this dissertation discusses the importance of a better theoretical basis for future PUI research. Furthermore, it explains potential improvements for the method and QualiExplore. The last section presents complementary research topics. These include interactions with data privacy and ownership, implications of increasingly complex data analysis, and the risks of disinformation.

Table of Contents

Acknowledgment

This dissertation is my work, but it would not have been possible without the invaluable support of my advisors, colleagues, friends, and family.

First and foremost, I am profoundly grateful to Prof. Dr.-Ing. habil. Klaus-Dieter Thoben, my doctoral advisor. His enduring guidance over many years, especially in the early stages of my research, was instrumental in the development of both my topic and myself. His critical remarks and open-mindedness enabled me to see pathways forward where I initially saw none.

I also extend my thanks to Prof. Dr.-Ing. habil. Iris Gräsler, my second advisor, for her insightful remarks from a practitioner's perspective. Additionally, I am thankful to Prof. Dr.-Ing. Maren Petersen, Dr. Sylvie Gavirey, Marcel Wiechmann, and Lennart Lütjen for serving on my thesis committee.

My gratitude goes to my former colleague, Prof. Dr.-Ing. Thorsten Wuest, for the great experience of visiting his research lab at West Virginia University in 2017 and for our many deep and inspiring academic discussions.

My sincere thanks to all the researchers I have met at conferences over the years, who provided critical feedback and fresh perspectives on my topic. I am also grateful to my colleagues at BIBA for their personal and professional support, and to my students who contributed to my research through their study projects and Bachelor and Master theses.

I am especially thankful to my friends for the enjoyable moments that helped me recuperate after long work weeks.

My deepest gratitude goes to my wonderful wife, Mina, for her unwavering support, patience, love, and pride. Writing my thesis during the pandemic would have been infinitely harder without her—and our dog, Lola.

Finally, I am profoundly thankful to my incredible parents, Nanni and Kurt, for providing me with the opportunity to pursue a good education, and for their tireless support and unwavering trust in me. To my brothers, Tim and Tobias, thank you for being great siblings and for your genuine interest in my work.

1 Introduction

1.1 Motivation

Product development is essential for organizations because its employees determine every product's properties (Pahl et al., 2007, p. 6). Also, they significantly influence the production cost, the operation cost, and the product's quality. Their work is knowledge-intensive, and the employees must seek product-related information from other departments, such as production and customer care. They consult documents and their colleagues for this purpose. The majority of projects in the development apartment focus on product improvement. Employees identify weak spots using accessible information and address them by changing related product properties. Potential weaknesses can relate to, for instance, frequent product failures, missing functions, and over-engineering.

Advances in **information and communication technology** (ICT) led to the creation of new *communication channels* usable by employees working on development-related tasks. The participative Internet and the Internet of Things are two paradigms that influence these advances significantly.

The *participative Internet* empowers people to share their opinions, impressions, and beliefs about products. One example is the vast number of product reviews on e-commerce websites. Data from a study published in 2019 indicate, for instance, that Amazon hosted almost 21 million reviews for electronic products and 22 million for home and kitchen products (Ni et al., 2019). In addition, online discussion forums, such as Bosch's Bob Community (N.n., 2022b), maintain virtual environments where Internet users continuously create product-related information. Figure 1 outlines different types of Internet-based communication channels. It is not comprehensive but demonstrates channel diversity.

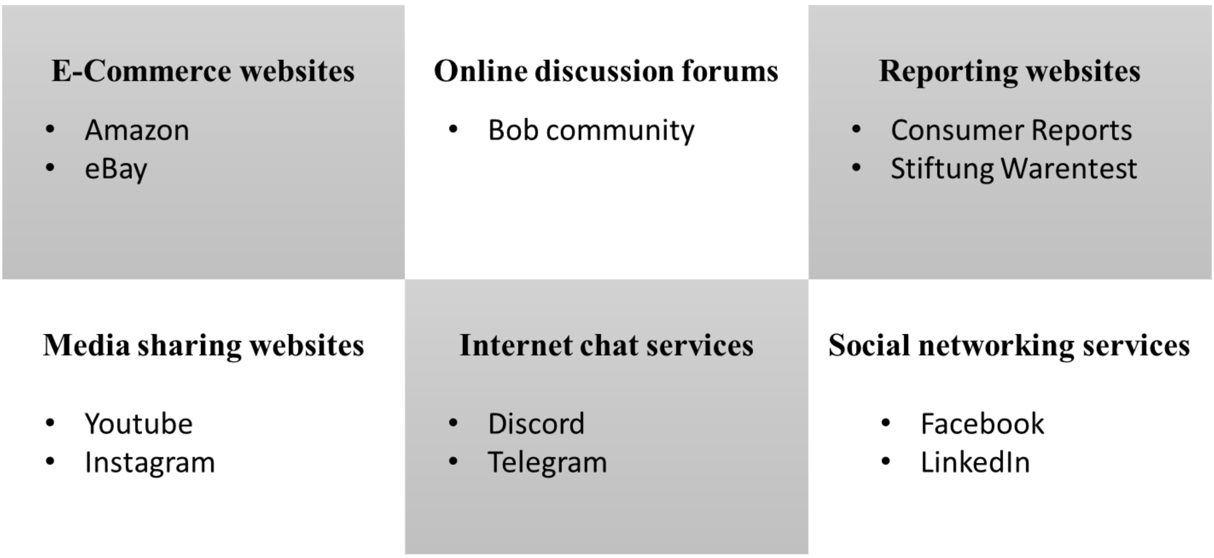

Figure 1: Internet-based communication channels with product-related information

The *Internet of Things* is a concept for connecting physical objects (Uckelmann et al., 2011, p. 8; Sánchez López et al., 2012). Product-embedded measurement systems are essential in this concept because they measure real-world phenomena in and around an object. Studies from 2016 and 2018 describe, for instance, the collection of product-related information in vehicles (N.n., 2016a; Plungis, 2018). Such information includes electric battery charging, kilometers driven on highway and urban roads, engine speed, and the use frequency of peripherals (N.n., 2016c, 2022d). Modern vehicles can send this information to the producer via mobile networks. Also, the Internet of Things enhances industrial goods with monitoring systems that enable concepts such as Prognostics and Health Management (Kim et al., 2017) and Computerized Maintenance Management (Cato and Mobley, 2002). Figure 2 illustrates the range of embedded sensors in vehicles.

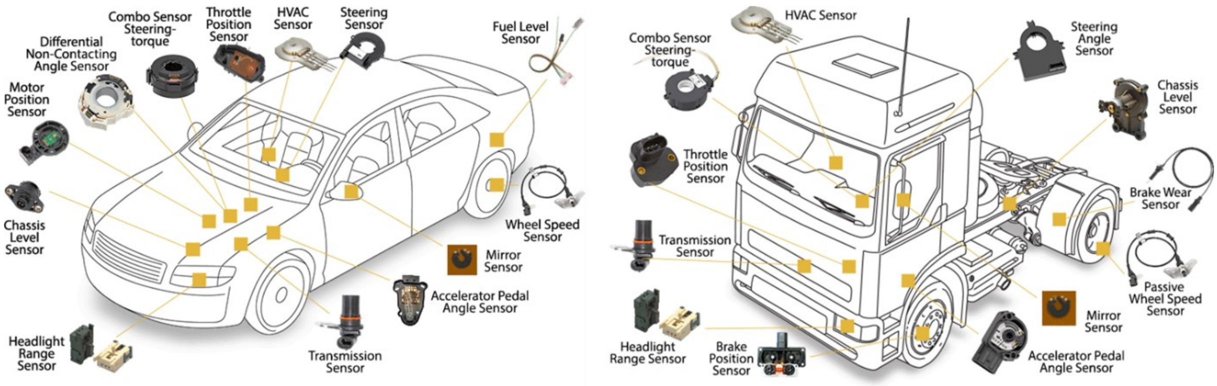

Figure 2: Vehicles with embedded sensors (N.n., 2022c, 2022a)

The communication channels' availability and their information about product usage have far-reaching consequences for the development department. This so-called **product usage information** (PUI) provides the employees an insight into, for instance, the product's behavior, user-product interactions, opinions about the product, the users' intentions in using it, and the product's environmental conditions (Deng et al., 2021). PUI can improve future product generations, strengthening the producer's competitive advantage (Abramovici et al., 2009). Several studies demonstrated the technical feasibility of this approach (Deng et al., 2019). Future challenges include non-technical problems that require interdisciplinary research between engineering and information management (Deng et al., 2021).

1.2 Problem statement

PUI is helpful in product development if it supports employees in performing their tasks. Preconditions for this are that this information is, for instance, free from error and accessible when needed – this means it has *high quality* or is "fit *for purpose*". Low-quality information can harm an organization and lead to additional costs. For this reason, organizations should influence information quality by controlling processes such as creating, storing, processing, and distributing information.

The starting point for systematically providing PUI is identifying employees' information needs. This process includes detailing what employees need in detail and what high-quality information means for them. A key challenge is that some needs are subject to compromises because the organization cannot control all processes affecting PUI quality, as the following situations indicate.

Customers write about products on websites, such as the Amazon marketplace. Producers cannot control how customers write these reviews because they do not design this communication channel. Furthermore, they cannot access all the needed information because the website provider does not share it. Potential reasons are missing incentives to share information and privacy restrictions.[1] Without access to all the needed information, producers can miss the context to interpret the PUI correctly.

The Internet of Things has other factors that influence PUI. For instance, organizations can purposefully design embedded measurement systems, allowing them to control the creation, storage, processing, and distribution processes. Once the physical product is with the customer, this control diminishes because the information flow typically stops (similarly Kiritsis, 2011). Factors such as sensor degradation, damage, and manipulation can go unnoticed. They can render the created information inaccurate and limit or delay an organization's access to it.

The academic product development literature hardly describes detailed information needs for PUI in product development. Besides, information quality is sometimes a concern, but articles do not propose comprehensive managed procedures. Organizations cannot fully benefit from PUI without having methods to detail PUI needs and a comprehensive understanding of factors that cause low-quality information.

1.3 Goal and scope

The main goal is to develop a **method** that supports organizations in **articulating precise PUI needs** for product improvement. This goal positions this dissertation at the intersection of information science and engineering. Target users are **information management** (IM) experts who design PUI systems.

The following sub-goals contribute to the main goal:

---

[1] The European General Data Protection Regulation (GDPR) defines requirements to sharing of personal data.

- The method development should *adopt acknowledged concepts from information science and engineering* because information need analysis is an IM task, while the method's application is in product development. Both domains have their vocabulary, definitions, and views, and some have different meanings, which could lead to misunderstandings.
- The method should *base on existing information need analysis practices*. IM literature describes practices that could significantly reduce the effort of method development. Basing the method on existing practices can increase its validity and user acceptance.
- The method should *formalize the information needs to support the deduction of testable requirements*. Information system design is cheaper if requirements are unambiguous and precise.

This manuscript revisits the goals above to evaluate the method. The next chapter introduces this dissertation's approach to reaching these goals.

1.4 Approach

The approach to achieve the goals above uses the **Design Science Research Methodology** (DSRM) proposed by Peffers et al. (2007). DSRM provides an acknowledged methodological basis for this dissertation because it focuses on information systems and the design of artifacts, such as constructs, models, or methods. Figure 3 illustrates the adapted DSRM used to structure this manuscript.

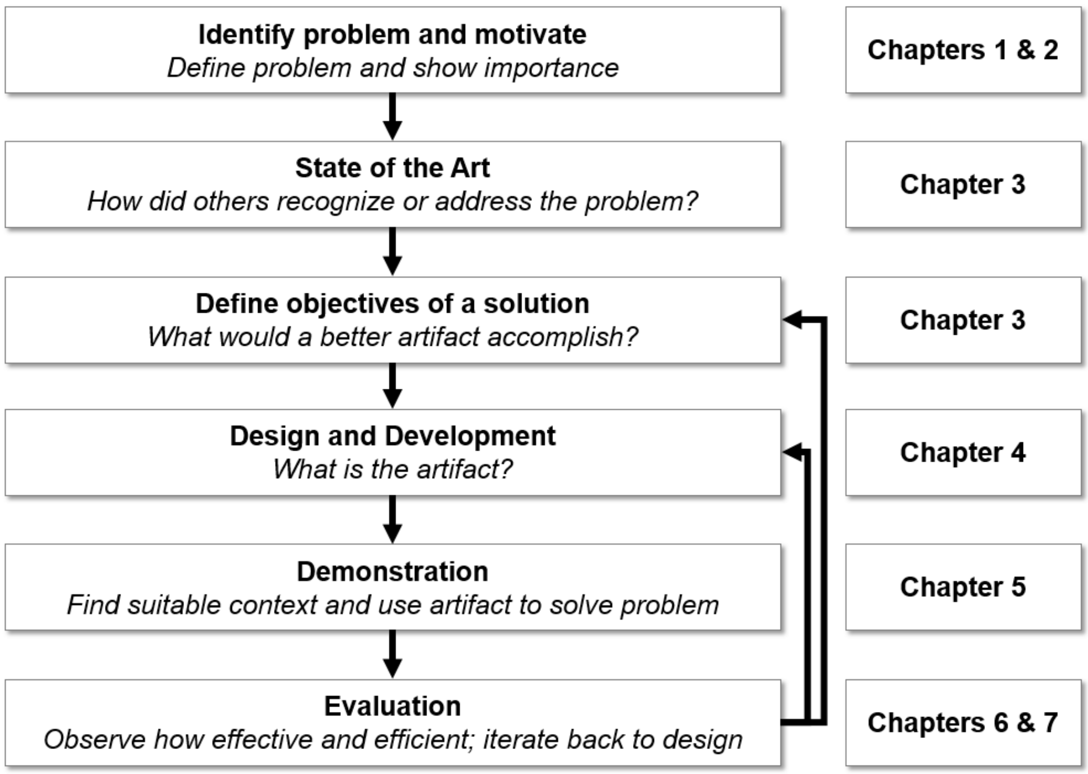

Figure 3: Research methodology and structure of this manuscript

Chapter 2 introduces the different core concepts related to this dissertation. *Chapter 2.1* begins with an outline of products in the engineering domain. The chapter introduces a product typology from literature and the concept of Product-Service Systems to specify this dissertation's scope. The following section uses Porter's value chain concept to explain how organizations create products. This explanation focuses on employees' work tasks and their relation to the information needed for such tasks. *Chapter 2.2* provides an overview of information flows in the product life cycle. It clarifies PUI and how it relates to product-related information. This sub-chapter closes with an information model specifying the relation between the information source, destination, and channel. *Chapter 2.3* introduces concepts of product-related IM. It uses Wollnik's layer model of IM to clarify the overall concept, provides definitions for the information system concept, and explains the meaning of ICT. Furthermore, it outlines relations between data, information, and knowledge management. The remaining part of this sub-chapter introduces branches of product-related IM and clarifies that this dissertation focuses on Product Life Cycle Management. *Chapter 2.4* highlights three essential challenges of product-related IM. It covers the creation of product-related information on the level of items, the combination of information from different sources, and the processing of product-related information into more helpful engineering information. *Chapter 2.5* concludes the state of research and indicates that a significant research gap is the precise understanding of PUI needs from the perspective of development-related work tasks.

Chapter 3 identifies and analyses existing literature about information need identification in work tasks. The chapter's broader scope is adequate because relevant literature directly related to PUI and product improvement is scarce. *Chapter 3.1* begins with clarifying critical concepts related to employees' information behavior. It introduces, amongst other concepts, an information need typology and sets this dissertation's focus on formalized needs. The remaining two key concepts are information seeking to satisfy information needs and using the information in a work task. *Chapter 3.2* details the use of product-related information in development-related work tasks. It covers general problem-solving, quality management, and reliability engineering. *Chapter 3.3* gives an overview of four methods used for the information need analysis of work tasks. *Chapter 3.4* concludes that an analysis method should identify information needs based on information characteristics. It should consider the specific work task and existing catalogs with potentially relevant characteristics. However, the authors of these catalogs did not develop them for PUI or apply them in a related product improvement activity.

Chapter 4 covers the method development process. *Chapter 4.1* identifies the method's requirements, while *Chapter 4.2* covers the literature analysis, which results in the method's theoretical foundation. They result in a catalog of relevant information characteristics, and factors influencing PUI quality. *Chapter 4.3* introduces cause-effect diagrams as an acknowledged method to visualize quality-related problems. Besides, it describes the developed method's procedure and syntax and integrates information characteristics. *Chapter 4.4* outlines the method's application assumptions, conditions, and procedure. *Chapter 4.5* presents the requirements and realization of a software tool supporting the method's application. Finally, *Chapter 4.6* concludes the method development.

Chapter 5 demonstrates the method's application in two use cases. The first case focuses on a software company offering a tool that collects and analyzes vehicle defect information. The second case covers a home appliances producer that wants to increase product quality with the information acquired from production machine usage. Bides, this chapter describes a workshop to apply the method.

Chapter 6 evaluates the demonstration's results qualitatively. It focuses on how users perceived the workshop and general observations. **Chapter 7** discusses the evaluation's results. In the beginning, *Chapter 7.1* critically reflects the workshop's evaluation results. *Chapter 7.2* revisits the requirements assessing how the method and tool fulfill them. Finally, *Chapter 7.3* focuses on the method's scope and development process limitations. The latter covers this dissertation's information quality model and the selection of the application cases.

Chapter 8 contains the summary and outlook of this dissertation. The latter identifies follow-up activities for academia and practice. These activities include the development of an acknowledged theoretical foundation, potential improvements for the method and tool, and various challenges in complementary research areas.

2 Concepts and current challenges

This chapter introduces the essential concepts of this dissertation and outlines the state of research related to information from a product's usage phase. *Chapter 2.1* clarifies the products this manuscript covers and associated concepts to scope this dissertation. Afterward, *Chapter 2.2* outlines the product life cycle – the superordinate concept of the usage phase – and product-related information. It identifies information types and relevant research domains that focus on the information from the usage phase. *Chapter 2.3* focuses on product-related IM from a product development perspective. It covers the information life cycle, management activities, and relevant information systems. Finally, *Chapters 2.4 and 2.5* concretize this dissertation's problem statement.

The following chapters adopt several concepts from international standards. They provide globally acknowledged frameworks providing a solid foundation for this dissertation. In addition, this manuscript refers to domain-specific references and the Oxford dictionary in the English language (Oxford University Press, 2017).

Furthermore, this dissertation sets several focuses. A *focus* means that this manuscript investigates a specific aspect of a topic in more detail. It does not mean that other aspects are less important nor that researchers should omit them.

2.1 Product and organizational context

This dissertation uses the term "product" frequently. Several research domains use this general term as well. Therefore, *chapter 2.1.1* clarifies what a product is in the context of this manuscript. *Chapter 2.1.2* introduces the role of information needs in the context of production.

2.1.1 Products, services, and product-service systems

The terms "product" and "service" have multiple meanings in an organization's context. Organizations create products or services by performing processes (ISO, 2015a). A **product** is tangible, and a **service** is intangible. This dissertation focuses on products.

A product has distinguished features called **characteristics**. Characteristics can be inherent, assigned, quantified, or qualified (ISO, 2015a). Inherent means that a characteristic is permanent. Some domains group similar characteristics into classes, such as physical, sensory, behavioral, temporal, ergonomic, and functional. Table 1 provides examples of characteristics to clarify their nature.

Table 1: Examples of product characteristics

	Quantified	**Qualified**
Inherent	Length (100 mm)	Shape (square)
Assigned	Price (100 Euro)	Appearance (sportive)

A critical product characteristic in this dissertation is durability because it allows customers to use a product for many years. **Durability** is *"[...] the ability of a product to perform its required function over a lengthy period under normal conditions of use without excessive expenditure on maintenance or repair"* (Cooper, 1994). This characteristic means the product has a dedicated period where stakeholders can collect product-related information. A second key characteristic is the **product's use**. People can use products personally or for the production of other products.

Kotler and Keller (2012, pp. 325–328) identified several product classes considering the characteristics above. Their focus is on *marketing*, a management activity to meet customer needs profitably (Kotler and Keller, 2012, p. 5). Marketers define a product as anything that an organization can offer on the market to satisfy needs. Examples include physical goods, services, experiences, events, persons, places, properties, organizations, and information (Kotler and Keller, 2012, p. 18). Kotler and Keller's systematic is suitable for this dissertation because it is acknowledged and reflects the range of today's products, including those collecting information about product usage. Figure 4 summarizes Kotler and Keller's systematic for product classes.

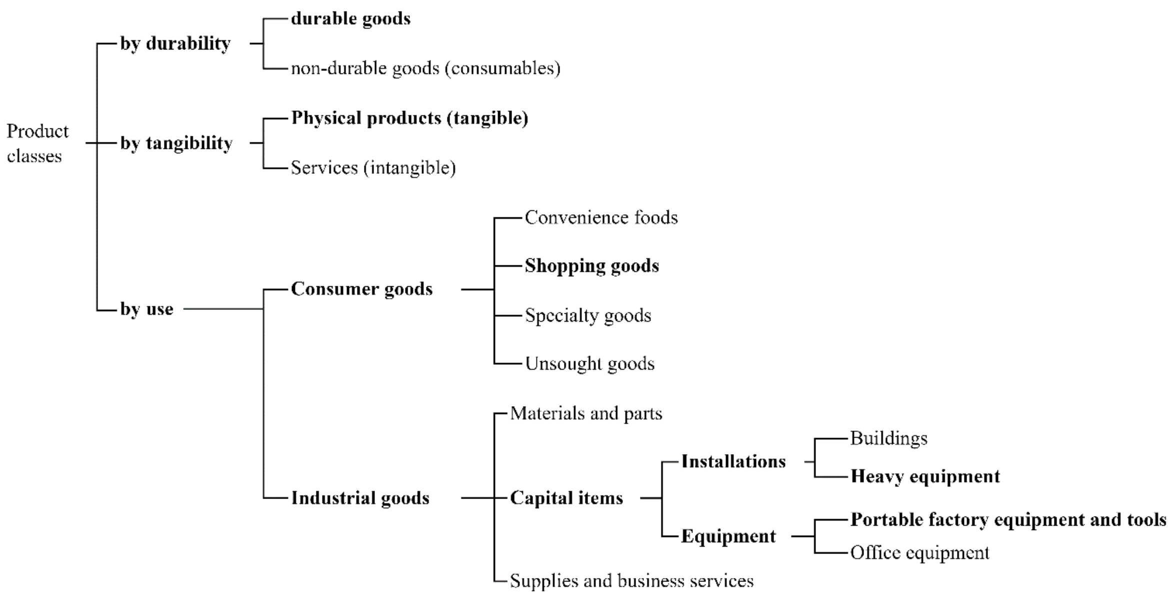

Figure 4: Product classification (based on Kotler and Keller, 2012, pp. 325–328)

This dissertation focuses on shopping goods and specific capital items. Shopping goods include, for instance, furniture and major appliances. Industrial equipment covers heavy equipment, such as generators, elevators, drill presses, and portable factory equipment and tools. These products are:

- durable goods with a use phase of several years,
- subject to maintenance services that create useful information, and
- they can contain functions that create product-related information.

The focused products can be part of a **Product-Service (PS)**, as illustrated in Figure 5. The PS concept is important for this dissertation because PS providers need to efficiently manage the usage phase to increase their PS's profitability. They use information about product usage for this purpose.

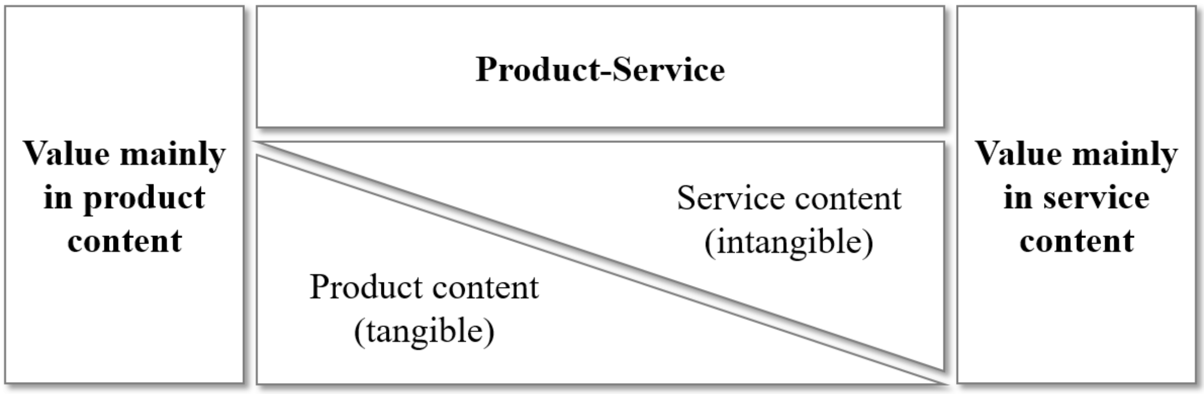

Figure 5: Continuum of product-service value propositions (adapted from Tukker, 2004)

The illustrated PS continuum covers value propositions, such as product sharing, maintenance, pay-per-service-unit, or the payment for a functional result. Providers should consider the following aspects to realize a PS:

- Value proposition (offer to the customer)
- Value network (partners needed to realize the value proposition)
- Revenue model (contractual and governance relations between partners)
- Technological infrastructure (hard- and software needed)

Tukker et al. (2006) define the entirety of these aspects as a **Product-Service System (PSS)**. This dissertation focuses on its *technological infrastructure*. It represents the hard- and software that enables and supports the provision of value for the customer. The technological infrastructure collects, stores, processes, and distributes product-related information. It can be the basis for services, such as product health monitoring and predictive maintenance (Mentzas et al., 2018).

2.1.2 Producers, work tasks, and information needs

A **producer** is an organization that creates value for the customer by combining purchased inputs and human resources to produce outputs (Porter, 1985, p. 61, 1998, 38–39). The smallest element in this process is a *value activity*, such as procurement, production, sales, marketing, product development, quality management, and service. If linked to other activities, they form the producer's *value chain*. Figure 6 illustrates how several value chains build a *value system*.

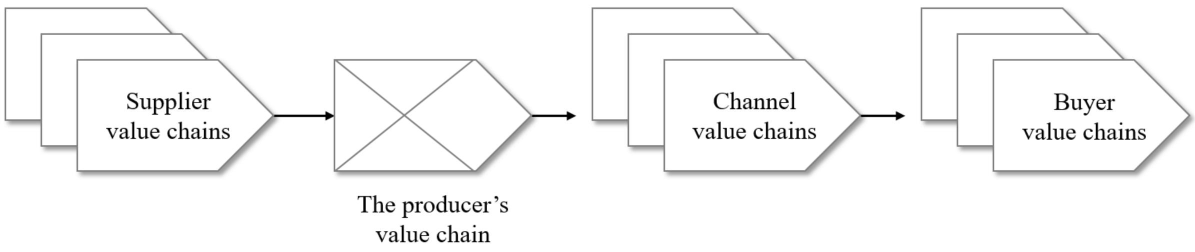

Figure 6: Value system (adapted from Porter, 1991)

A **work task** is an activity meant to accomplish a goal at work (adapted from Vakkari, 2003). It has a purpose, beginning, and end, and it results in something meaningful. Employees perform physical and cognitive actions during work tasks. For the latter, they *need information*, for instance, to understand problems or evaluate alternatives during decision-making.

The individual employee's **information need** can emerge from a personal feeling of uncertainty (Kuhlthau, 1993, p. 343; Case and Given, 2016, p. 371). It evolves from an unexpressed need to a formalized one (Taylor, 1968) and depends, for instance, on the individual's demographics, work roles, assigned tasks, and the environment (Byström and Järvelin, 1995; Leckie et al., 1996).

Identifying and addressing employees' information needs is challenging. It requires regular interaction with them, which is too costly to apply in organizations with hundreds of individuals (similar Voß and Gutenschwager, 2001, p. 136). Defining the needs of generalized *work roles* is one approach to addressing this issue.

This dissertation focuses on development-related work tasks and the information needs of related work roles. Table 2 provides examples to clarify these concepts and demonstrate their heterogeneity.

Table 2: Development-related work tasks, information needs, and work roles

Work tasks	Information needs	Work roles
Functional improvement	Product performance	Product developer
Root cause analysis	Product and user behavior	Quality engineer
Define market segment	Customer profiles	Marketing expert
Tailor a value proposition	Customer preferences	Sales expert

2.2 Product usage information

The producer's employees need information about products to perform development-related work tasks. Most of this information describes the product at a specific moment in time. *Chapter 2.2.1* introduces the Product Life Cycle, which describes the stages a product passes over time. *Chapter 2.2.2* outlines product-related information types and clarifies this dissertation's focus on them.

2.2.1 Product Life Cycle

Engineers use the life cycle concept to describe the product's evolution from conception to retirement (ISO, 2017). Both are examples of a broad range of *life cycle processes* (Wellsandt et al., 2016 for an overview). Kiritsis et al. (2003) proposed a model that aggregates life cycle processes into three phases. Figure 7 illustrates them, and their primary information flows.

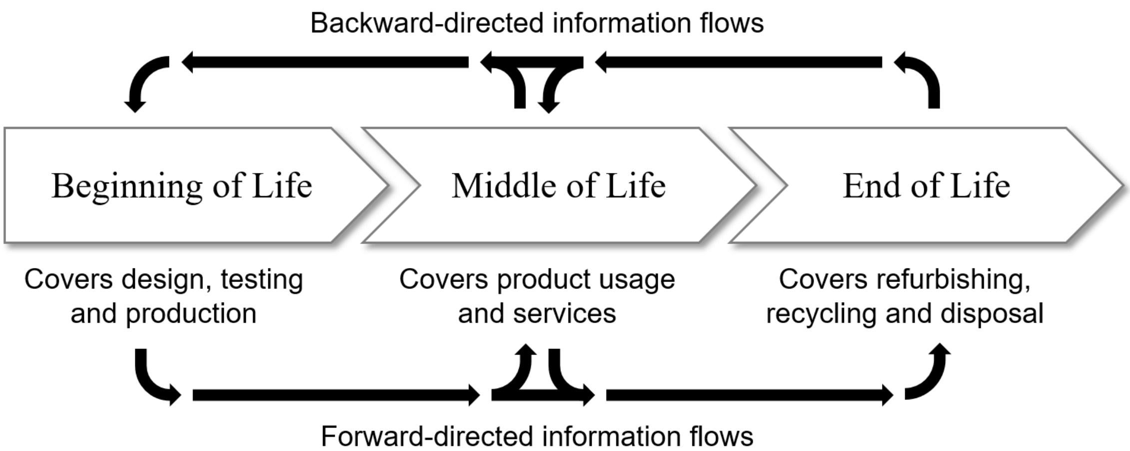

Figure 7: Product Life Cycle and information flows (Wellsandt et al., 2015b)

A phase's processes share information with processes in other phases. *Forward-directed* information flows indicate that a process provides its information to at least one subsequent process. Employees typically need this information to make progress in their tasks. *Backward-directed* information flows represent information feedback (Wellsandt et al., 2018). Employees can use it, for instance, to improve the reliability of future product generations.

Wellsandt et al. (2016) investigated 71 life cycle models and found that 86% of these models cover product usage – an indicator of this activity's relevance. The investigation also revealed usage aspects that future research should consider. *First*, a product's life cycle intersects with the life cycle of the machines that manufacture it. Operating (using) production machines affects the product's characteristics and, consequently, its performance during usage. Product improvement can benefit from an accurate understanding of manufacturing machine usage. *Second*, products are assemblies of components with individual life cycles. These life cycles coexist and affect each other; for instance, if a critical component breaks, the product may reach its end of life despite other components operating reliably. *Third*, organizations can remanufacture products that reach their end of life. A remanufactured product is usable again and has at least the performance specification of the original product (Ortegon et al., 2016). Figure 8 summarizes the aspects above in one illustration. The evaluation chapter uses it to position the use cases.

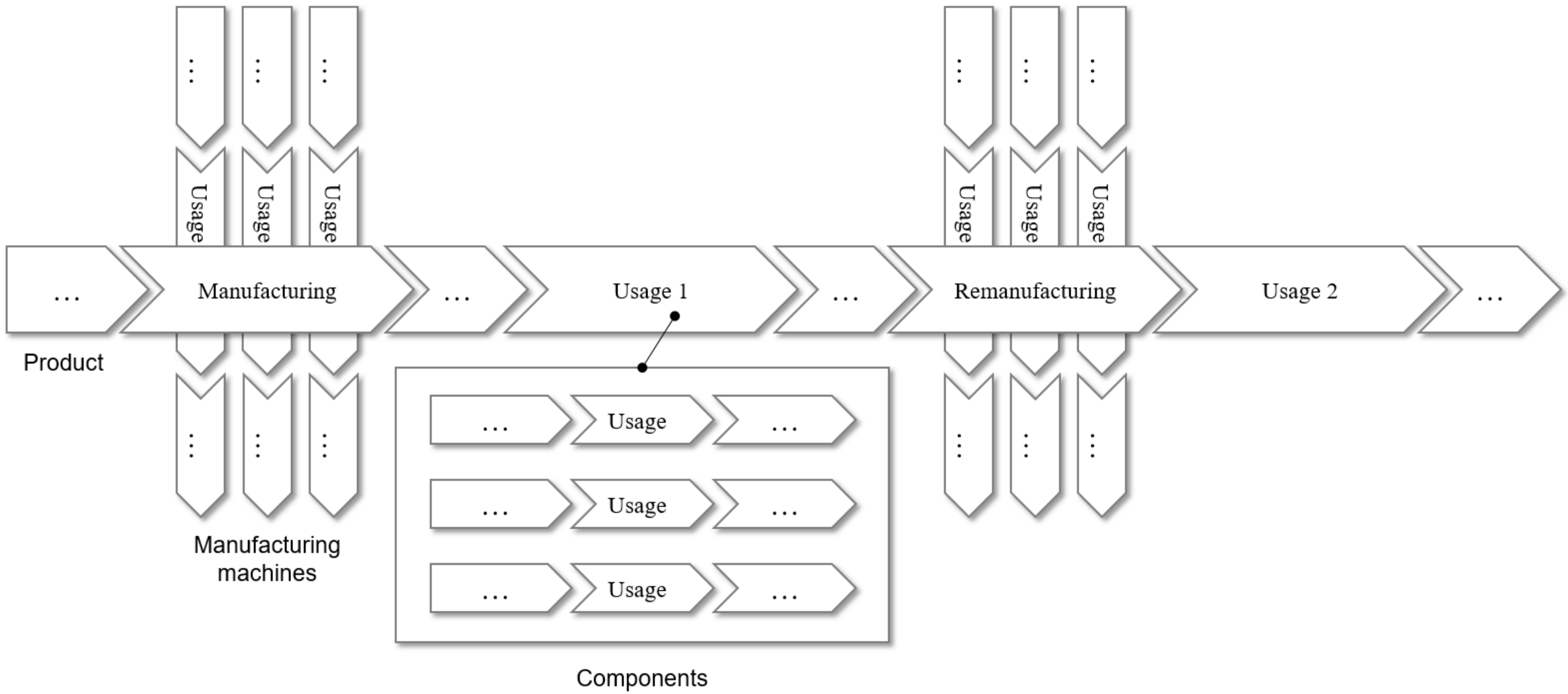

Figure 8: Aspects of usage in the product life cycle

The illustration above serves as this dissertation's conceptual model of the usage phase and emphasizes that one product's usage depends on and affects others.

2.2.2 Product-related information

Producers record various information about their value activities and related *entities*. Examples of the latter are products, employees, customers, suppliers, and facilities (ISO, 2012). **Product-related information** is meaningful data that describes a product. It resides in handwritten, printed, and electronic documents. Figure 9 illustrates a structure for product-related information that considers the Product Life Cycle.

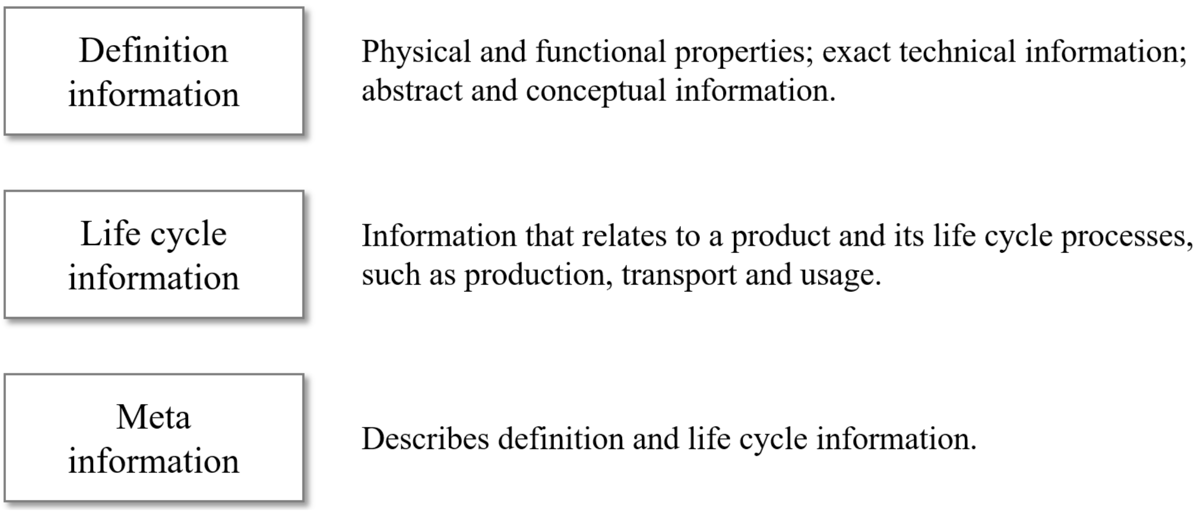

Figure 9: Types of product-related information (based on Saaksvuori and Immonen, 2008)

This dissertation focuses on life cycle information generated during the product's usage phase. It describes, for instance, how users interact with a product and how the product behaves during operations. This manuscript refers to this information as PUI.

Abramovici and his fellow researchers have a long track record of researching PUI in product development (Abramovici and Schulte, 2007; Abramovici et al., 2018). They defined it as measurable information accumulated during the product's usage and distinguished it from *subjective* information created by customers (Abramovici et al., 2009). Table 3 provides an overview of PUI synonyms in academic literature. Listing them is relevant because it provides an overview of related terms and the viewpoints of independent, international research groups.

Table 3: Synonyms for PUI in academic literature

Synonyms	Example references
In-service data	(Mey Goh and McMahon, 2009; Igba et al., 2015)
Field data	(Edler, 2001)
Installed-base information	(Rämänen et al., 2013)
Middle-of-life information	(Wellsandt et al., 2015b, 2015a)

Wellsandt et al. (2015b) suggested that middle-of-life information is product-related information created after the end customer buys the product and before it is no longer helpful for a user.[2] This dissertation defines PUI in this wider sense because measurement systems and humans, such as users and customers, can provide information about product usage.

Product-related information refers to specific items called **instances** or abstract concepts, such as product **classes and types**. In this context, a product instance requires a unique identifier, such as a serial number. If information refers to an item, literature refers to it as **item-level product information** (Hans et al., 2010). Table 4 clarifies these concepts for an imaginative shopping good.

Table 4: Product-related information types

Type	Product type "Top loader washing machine"	Product instance "1003001230"
Definition information	3D-model, Production instruction, Operations handbook	Installed spare parts, Warranty period
Life cycle information	Average energy consumption per day, Bathtub curve	Energy consumption per day, Laundry load, Reliability

Employees working on development-related tasks use PUI to satisfy some of their information needs. **Information channels** are critical in this process because they convey the needed PUI from the source to its destination. They include analog and digital channels. This dissertation focuses on digital channels because most product-related information exists on computers today.

[2] Customer and user can be different entities, especially in the case of industrial goods.

The **information's source** can be a human or a machine (adapted from Case and Given, 2016, p. 375). Humans typically author product information in development-related tasks and store it via spreadsheets, text files, images, videos, and databases. Machines create and store information automatically. An electronic control, for instance, can write information in log files in the case of product failure. The **information's destination** is a human or a machine. Figure 10 summarizes the concepts above in a communication model. This model serves as a reference in this dissertation.

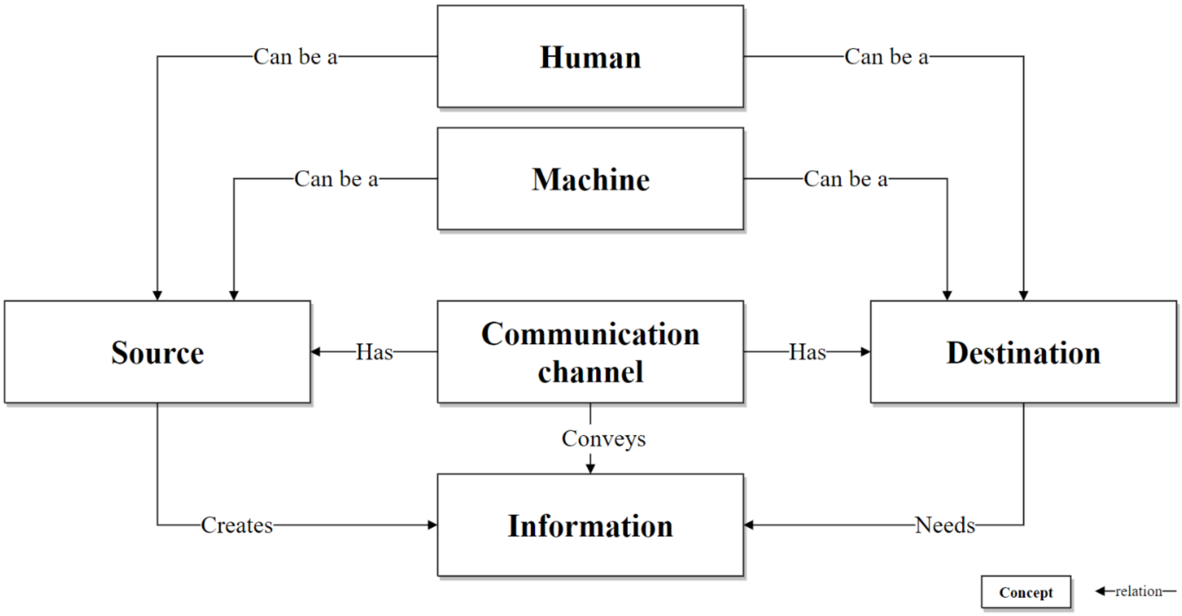

Figure 10: Communication model used in this dissertation

2.3 Product-related information management

This chapter introduces fundamental concepts related to the management of product-related information. *Chapter 2.3.2* gives an overview of IM. It is a discipline that, amongst others, designs the systems that provide information. *Chapter 2.3.3* defines the concept of product-related IM to clarify this manuscript's focus.

2.3.1 Information system

An **information system** (IS) contains computers and the associated organizational, technical, and financial resources to operate it (ISO, 2017). Employees use an IS to, for instance, store, search, and access information. These activities make such systems essential for satisfying information needs.

An IS uses information technology and communication technology. *Information technology* means resources that acquire, process, store, and disseminate information, while *communication technology* concerns resources that transfer information between stakeholders. (ISO, 2017)

2.3.2 Information management

IM manages an organization's information-related processes, resources, technologies, and policies (Choo, 2002, xiv; Krcmar, 2015, p. 109). It has three activity layers, as illustrated in Figure 11.

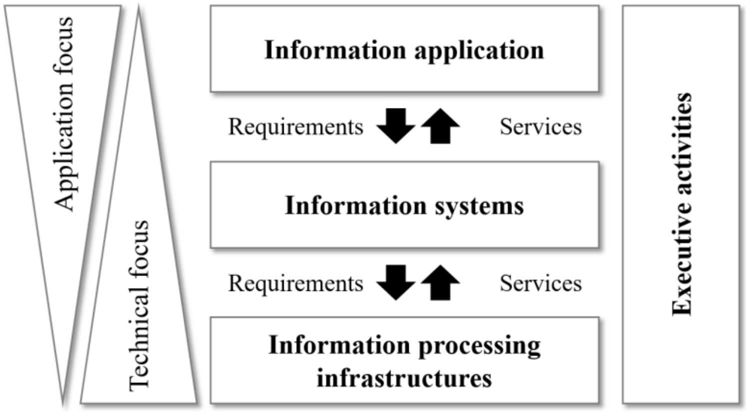

Figure 11: Model of information management (Wollnik, 1988, p. 38; Krcmar, 2015, 4.2.4)

The *topmost layer* focuses on the application cases that can benefit from the information. It includes activities such as understanding the information needs, available information sources, and channels.[3] The *middle layer* concerns the activities related to the IS and focuses on data, data-related processes, and software applications. The *bottom layer* is the most technical one. It focuses on the development and maintenance of information processing infrastructures. Each layer defines the requirements for the next lower layer, while the lower layers provide services to the upper ones. *Executive activities* cover decisions about strategy, governance, processes, personnel, controlling, and security in the context of information. They cover all three layers.

Data management and knowledge management are value activities that coexist with IM. There is no generally agreed definition for these activities, and their scopes overlap with IM in some areas. It is important to outline them in this chapter because IM should coexist with them. *Data management* focuses on technical aspects and the infrastructure's efficient use, while *knowledge management* mainly covers related information and what the organization's stakeholders know.

[3] Krcmar 2015 refers to information economics with information demand and supply.

2.3.3 **Product-related information management**

IM is not limited to a specific information type, but the literature mentions several activities similar to IM with a clear focus on products. The following list summarizes the most relevant ones for this dissertation to minimize ambiguity (similar Sendler, 2009, 3.4, 4.2; N.n., 2022e).

- *Product Data Management* (PDM) manages digital documents containing, for instance, technical drawings and 3D models. Its focus is on the documents and their management with software.
- *Product Information Management* (PIM) manages product-related information and its distribution via communication channels. Its focus is on the communication of definition information to stakeholders.
- *Product Life Cycle Management* (PLM) is a concept for managing the processes related to product development. It grounds on PDM and integrates PUI simulation and knowledge extraction services, among other services.

The concepts above indicate that the management of product-related information is a complex field. It covers different value activities, and software companies provide solutions for the related employees. This manuscript considers **product-related information management** (PRIM) as an IM branch and an essential aspect of PLM (similar Saaksvuori and Immonen, 2008).

2.4 PRIM challenges

Continuous developments in ICT and its progressing application in engineering tasks challenge the management of product-related information. New or improved technologies affect, for instance, how employees create, seek, and share information. Furthermore, they influence how humans and technical systems create and work with information. This chapter summarizes PRIM's critical challenges when it serves PUI for product improvement.

Chapter 2.4.1 provides an overview of the challenges identified in the literature. The following three chapters describe the different challenges. *Chapter 2.4.2* covers creating product-related information, with a focus on the developments regarding embedded systems and Social Media. *Chapter 2.4.3* continues with an overview of combining information from different information channels. Its primary assumption is that there is no generally agreed standard for storing PUI, which is a technical challenge in IS design. *Chapter 2.4.4* summarizes the challenge of processing PUI.

2.4.1 **Overview**

Wilberg (2019) identified *technical* and *organizational* challenges when exploiting product-embedded measurements. Technical challenges occur during data creation, transmission, storage, and analysis, while organizational ones concern the application areas and the integration of insights into decision-making. Specific challenges are creating high-quality data, data storage and processing, interdisciplinary collaboration among employees, responsibilities, and use case identification and selection. Addressing organizational challenges is critical for exploitation because producers can address them through their actions (King, 2014; Wilberg, 2019).

2.4.2 **Challenge 1: "creating item-level information"**

Creating product-related information for individual items can be expensive and time-consuming. Employees typically do it for prototypes and products used in product tests. This approach seeks to balance the costs of creating information for all product items and the benefits gained from less accurate, generalized information. The challenge is to <u>create accurate information for items with low effort</u>.

Technological developments in the last decade made hardware and software cheaper and easier to use (Porter and Heppelmann, 2014, p. 6). Today, there are two approaches to acquiring item-level information. *First*, producers can design products with **embedded measurement systems** or retrofit individual products after production. These systems measure the user's behavior, the product's behavior, and the surrounding environment. The resulting information typically contains a unique identifier which makes it item-specific. Table 5 contains products from three price ranges that create PUI through embedded systems. It indicates that embedded systems can create item-level information.

Table 5: Products creating PUI through embedded measurement systems

Products	**Prices**	**PUI examples**	**References**
Car, Boat, Machinery	>10k €	Engine temperature, geo-position, door status	(Wellsandt et al., 2015c; N.n., 2016b)
TV, washing machine	~1 k€	Number of revelations, the weight of cloths	(Lützenberger et al., 2016)
Power tools	<1 k€	Tool settings	(N.n., 2019)

Second, producers can utilize a variety of Internet-based communication channels. **Social Media** (SM) is the umbrella term for channels where Internet users can create and share information. Producers maintain these channels to intensify customer communication and acquire product-related information. The producers manage, for instance, online discussion forums, weblogs, and social networking service profiles (Wellsandt et al., 2015a). SM information must contain a serial number or similar identifier to reference an item, but this information is often unavailable.

2.4.3 Challenge 2: "combining information from different channels"

PUI typically resides in different databases curated by departments, such as the maintenance and repair department or customer service. These databases store structured and unstructured data. The former has a defined data model that indicates what the data means, while the latter lacks clear meaning. Examples of unstructured data are text, images, and videos. A significant problem for an IS is to combine the information from these sources[4] and provide the information users with a unified view on it (Lenzerini, 2002; Tao et al., 2019).

Information integration is the task that addresses this problem (Bernstein and Haas, 2008). It can apply, for instance, data warehouse loading, virtual data integration, message mapping, and object-to-relational mapper tools. Bernstein and Haas (2008) identified Extensible Markup Language (XML), data schema standards, data cleansing, schema mapping, schema matching, keyword search, information extraction, and dynamic web technologies as core technologies for information integration. Publications indicate that data integration solutions are sufficient to realize PUI applications (Deng et al., 2018). Besides, information integration is not exclusively technical and must address decisions regarding data storage and relevant views on data.

2.4.4 Challenge 3: "information processing"

Information processing is performing operations on information systematically (ISO, 2017). The application of PUI in work tasks grounds on computational models that incorporate logic and arithmetic rules. Model development and validation require knowledge about the focused work task, the available information, and the available processing methods (e.g., refer to Abramovici et al., 2009). A fundamental problem for the IM expert is formalizing engineering knowledge in a computational model.

Deng et al. (2021) outlined PUI applications in product development. They investigated the processed information and the application domain, indicating that information processing is an active research field with various processing approaches and methods. The investigated references have not yet suggested a methodology to approach the formalization of engineering knowledge – it remains a case-specific task.

[4] Sources in this specific context are the databases that contain the data to be integrated.

2.5 Conclusion

Embedded measurement systems and SM create product-related information. Information integration tools make this information accessible to software that processes it further. The identified achievements regarding the three challenges above indicate that PUI's application in product development is technically feasible for various tasks. However, case descriptions typically focus on the information systems and infrastructure layers. If they cover the planning and justification of the application, the procedure is brief and information needs hardly covered. This observation indicates a **research gap in the literature** because the precise understanding and description of the information needs is a prerequisite for further IM activities (Krcmar, 2015, p. 120). Besides, it is a critical task for an information manager (Detlor, 2010, p. 104). This criticality results from the potential follow-up cost caused by erroneous or inappropriate information. For the reasons above, this dissertation focuses on understanding and describing PUI needs to close the gap in the PUI literature.

The authors of PUI application case descriptions may have different reasons why their articles do not cover the application layer in detail. Relevant factors are:

- The study's focus was on the solution description.
- There was not enough writing time or space.
- The authors considered the provided details sufficient.
- The authors held back information to protect their results.
- The authors were unable to provide more detailed information.

Besides, I observed a lack of detailed information needs in two research projects focused on developing ISs that provide PUI to improve products. The projects Falcon[5] and Manutelligence[6] faced challenges during the design and development phase, where project partners analyzed relevant work tasks and derived requirements for the envisaged IS. Detailing the needs was a problem in these projects, and a systematic approach could have resulted in more accurate and precise IS requirements.

To my knowledge, there is no method to support the information need analysis for PUI in development-related tasks. The process models proposed by Wilberg (2019) and Meyer et al. (2022) are the most related work. The former has a step that focuses on data needs and the consolidation of use cases, including a template to assess the quality of needed and available usage information. However, the step's description is short and, therefore, hard to re-apply in other cases. The latter process model has one step to define data needs but its description is short and the

[5] EU-funded project (H2020, 2015-2017, Grant Agreement no. 636868).
[6] EU-funded project (H2020, 2016-2018, Grant Agreement no. 636951).

authors conclude that someone has to develop methodological approaches for this and the other steps. Both process models focus on measurements from embedded sensors and leave out a substantial part of PUI channels related to the participative Internet.

A related domain that can provide more detailed insight into existing methods is the “information need analysis” for work tasks.

3 Information need analysis for work tasks

Human action and its relation to information is a subject of information science. Researchers investigate intentional and planned actions in this context, as well as unintentional, serendipitous ones (Case and Given, 2016, p. 370). Important actions include information seeking, rejecting, and avoiding. Information science investigations further concern human thoughts and emotions related to these actions, and contextual elements, such as situation, time, affect, culture, and geography. *Information behavior* is their overarching topic.

Information needs are an intensely discussed concept within the information behavior domain, as Case and Givens (2016, 5) outlined. The recent literature in this domain is typically *person-centric*, which means that the research focuses on the persons that create, seek, interpret, and use information (Case and Given, 2016, p. 8). This dissertation adopts the person-centric view and structures this chapter accordingly.

Chapter 3.1 begins with human information behavior in a work task context. Its focus on work tasks emphasizes purposive, goal-oriented actions involving information and matching development-related task conditions. *Chapter 3.2* outlines what information use means in engineering. It outlines tasks where specific information needs are readily available in the literature. *Chapter 3.3* describes IM methodologies that cover information needs systematically. It identifies relevant activities and essential aspects to consider and, therefore, serves as the most critical part of this state-of-the-art. The conclusion in *Chapter 3.4* pinpoints this dissertation's research problem.

3.1 Information behavior

Behavior assumes that someone acts in response to a particular situation (Oxford University Press, 2017). In this dissertation, the situation is a *work task*, and the person is an employee working on this task. Employees typically collaborate to perform complex work tasks, such as the ones related to product development. Group-related information behavior can be significantly more complex than an individual's behavior. Therefore, this dissertation focuses on the individual's behavior to build a solid ground for further investigation.

Chapter 3.1.1 outlines the fundamental concept of "information and associated concepts, such as data and different types of knowledge. *Chapter 3.1.2* provides an overview of the information need concept, a key aspect of this dissertation. The following *Chapters, 3.1.3* and *3.1.4*, describe seeking suitable information to satisfy an information need and using this information. *Chapter 3.1.5* summarizes the findings.

3.1.1 **Information**

This dissertation's main object of research is information. The dictionary defines **information** as "*what is conveyed or represented by a particular arrangement or sequence of things*" or "*facts provided or learned about something or someone*" (Oxford University Press, 2017). Case and Given (2016, p. 58) summarize its meaning as twofold: information indicates either an item (message) or a process (informing). Both meanings are relevant to this dissertation. The following paragraph focuses on the information item and its related concepts. Later chapters will cover information use, which is closely related to informing.

An **information item** is a "*separately identifiable body of information that is produced, stored, and delivered for human use*" (ISO, 2017). Its meaning depends on the context in which a person uses information (Eppler, 2006, p. 106). An item may be, for instance, a sentence, paragraph, article, book, website, video, or report. Information relates to **data and knowledge**, but there is no generally agreed definition of these concepts and their relations (Zins, 2007). Furthermore, the subconcepts of explicit, implicit, and tacit knowledge are confusing (Wilson, 2002).

Table 6 summarizes this dissertation's interpretations of the concepts above. This summary is simple because this manuscript focuses on applying information rather than concisely discussing the concepts. References ground each interpretation with acknowledged literature.

Table 6: Data, information, and knowledge

Concepts	**Interpretations**	**References**
Data	Outside of the human mind; signs with no or little meaning; simple facts	(Wilson, 2002; Rowley, 2007)
Information	Outside of the human mind; organized data; data embedded in the context of relevance; meaningful	(Wilson, 2002; Rowley, 2007)
Explicit knowledge	"What I can convey about what I know"; synonym for "information"	(Wilson, 2002; Zins, 2007, p. 487)
Implicit knowledge	"What I know"; inside the human mind; what a person takes for granted and thus does not express	(Wilson, 2002)
Tacit knowledge	"We know more than we can say"; inside the human mind; not conscious; <u>cannot</u> be expressed[7]	(Wilson, 2002)

These concepts and their interpretations provide the ground for the following chapters. The next chapter outlines what an "information need" is.

[7] This constraint assumes verbal expressions as suggested by Polanyi (1958, p. 99).

3.1.2 **Information need**

Information science does not generally agree on what an information need is or how it relates to work tasks. Case and Given (2016) discussed several definitions and propose that an **information need** is "*a hypothesized state brought about when an individual realizes that they are not comfortable with their current state of knowledge*" (Case and Given, 2016, p. 371).

An information need resides within a person's head (Belkin and Vickery, 1985; Case and Given, 2016, p. 89). It is unsteady and temporary, making its description virtually the same as describing the person's current psychological state (Harter, 1992; Case and Given, 2016, p. 91). Taylor (1968, pp. 8–9) developed a model that describes how an information need evolves from a feeling to a question. He used the model to describe how and why people come to ask questions at library reference desks – which can be part of an IS. Figure 12 illustrates Taylor's model as a **continuum of formalization** (based on Case and Given, 2016, p. 84).

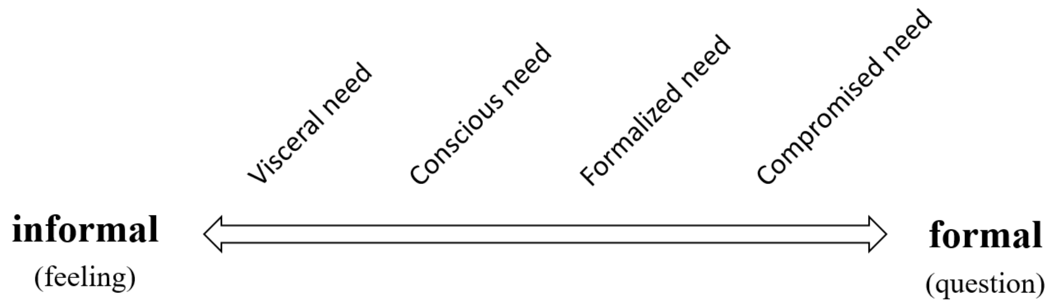

Figure 12: Taylor's information need typology in a work task context

The *visceral* information need is an unexpressed feeling, such as a vague dissatisfaction with the current state of knowledge. It evolves into a *conscious* mental description of the needed information. This description might require refinement to reduce ambiguity, for instance, through consultation with a colleague. The person might *formalize* the need and express it as a question. Finally, the user must recast the formalized need in the IS context (e.g., library reference desk). This recasting means that the need becomes a *compromise* between the user's actual need and the capabilities of the IS. The compromised need provides a link to the information retrieval domain. This domain investigates solutions humans use to search for information in documents and databases (Case and Given, 2016, p. 372).

Taylor's model is essential for information science because it implies "grades" of information needs. <u>This manuscript focuses on the formalized need</u> because an IM expert can use these unconstrained needs to develop requirements for an IS. Chapter 3.3 addresses this topic in more detail.

3.1.3 Information seeking

Employees satisfy their information needs through browsing, serendipity, and information seeking. *Browsing* is the information behavior of scanning the environment without a clearly defined goal, and *serendipity* is the action of encountering relevant information by accident. Both approaches are less structured and intentional (Case and Given, 2016), so this manuscript does not cover them further. This focus does not imply that browsing and serendipity cannot positively impact work tasks.

Information seeking is the human behavior that occurs when an individual identifies an information gap limiting the individual in satisfying a goal (Wilson, 2000; similar Case and Given, 2016, p. 372). This gap can occur when critical information is unavailable to the employee to proceed with a work task. Information seeking can occur multiple times during a work task, and each process may result in one or more information items.

Researchers use *models* to describe information seeking in a simplified way. Case and Given (2016, p. 144) identified several models and pointed out that they differ in their assumptions, structure, purposes, scope, and intended uses. They use flowcharts and depict factors that influence the information behavior. Case and Given's overview contains two models that address a work task context. This dissertation introduces these models in the following paragraphs for the following reasons:

- They detail what an information need is.
- They indicate how information needs relate to work tasks.
- They outline factors that influence information needs.
- They describe the role of information needs in information seeking.

Byström and Järvelin (1995) published the *first information seeking model* covered in this dissertation. Their focus is on work task complexity and its influence on information needs and information seeking. A task is complex if the human cannot determine its inputs, process, and outputs in advance. Figure 13 illustrates this information seeking model.

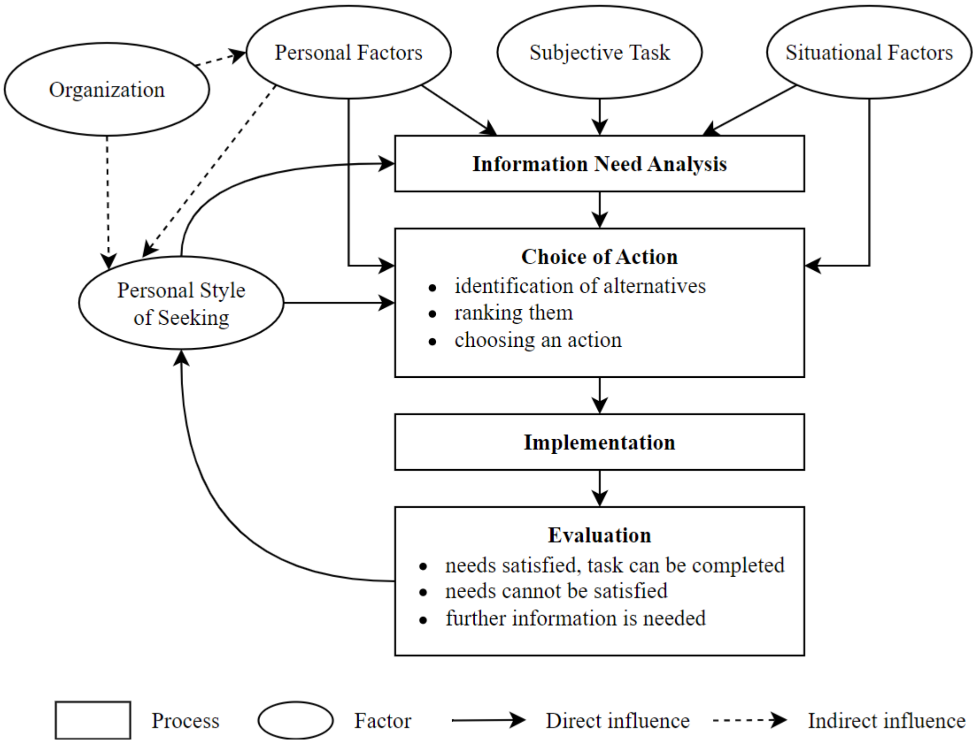

Figure 13: Information seeking model (based on Byström and Järvelin, 1995)

Case and Given (2016) summarize the model's structure as a feedback loop. It begins with the employee's understanding of the work task, personal factors (e.g., education, experience, and emotions), and situational factors (e.g., time available to perform the task). They determine what information the employee needs to perform the task. After analyzing the needs, the employee identifies possible actions and implements them to satisfy his/her needs. The employee evaluates the results of these actions and draws one of three conclusions:

- the needs are satisfied so he/she can complete the task
- the needs cannot be satisfied
- further information is needed

The evaluation result feeds back into determining the information needs mediated by the employee's seeking style. The organization affects personal factors and the personal seeking style.

Freund (2015) published the *second information seeking model* covered in this dissertation. She investigated how contextual factors influence the information seeking behavior of software engineers. Freund used a focus group and interviews to develop the model and noted that it is not comprehensive. Figure 14 depicts the model.

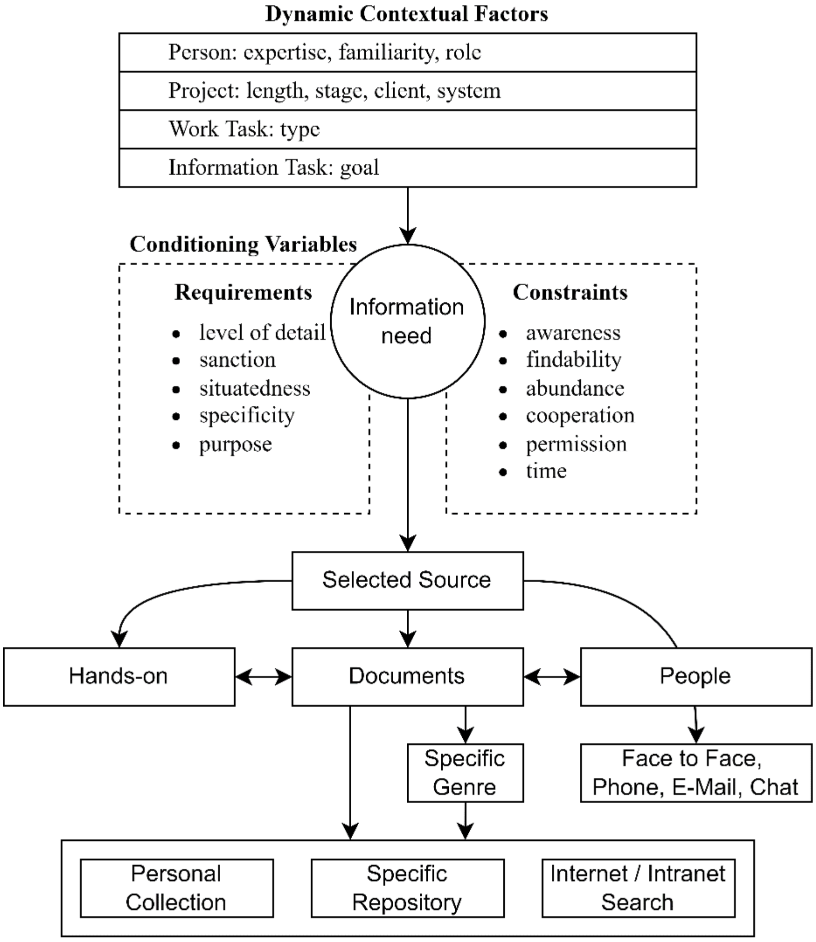

Figure 14: Information seeking model (based on Freund, 2015)

Dynamic and interrelated contextual factors motivate the information need. These factors relate to the person, project, work task, and information task. The latter includes learning, deciding, doing, fact-finding, and problem-solving. Conditioning variables emerge from contextual factors. They concretize the information need and make it more likely to satisfy it. The conditioning variables consist of "*[...] requirements on the nature of the information sought [...]*" and constraints on the seeker and the information sources (Freund, 2015, p. 1601). Appendix A explains the conditioning variables. These variables cover aspects related to information quality. The following action is the selection of an information source[8]. Freund classified the sources as follows:

- **Hands-on learning** is one approach to satisfy an information need via trial and error. This behavior means that the seeker learns something by, for instance, building a functional prototype of a product to perform experiments.

[8] Here, information source refers to actual sources and communication channels.

- **Documents** reside in, for instance, personal collections and document repositories in an intranet or the Internet. The seeker can access them via computer-based search functions.
- **People** provide information via a communication channel (e.g., face-to-face, telephone, and Email). Using these sources depends on personal relationships and the awareness of other peoples' expertise.

In conclusion, the "information need" concepts above represent compromises between the requirements and the constraints that result from personal and non-personal factors.[9] *Personal factors* include, for instance, education, experience, work role, awareness of information sources, and the seeker's emotions. *Non-personal factors* include the available resources to perform a work task (e.g., available time and access to sources) and the organization's work culture. Chapter 3.1.4 continues with the information behavior after the employee finds an information item.

3.1.4 **Information use**

Case and Given (2016, p. 93) point out that "using information" is an ambiguous behavior. Kari (2007) identified two meanings of information use: (1) *what a person does with information* and (2) the *effects of information on a person*. The first meaning represents an instrumental view of information, where the information user applies information to some task. The second one represents a human-centered view concerned with, for instance, sense-making and how information changes the feeling of uncertainty (Kuhlthau, 1993). This dissertation aims to support applied research. Therefore, it focuses on the instrumental view of information use. It uses the term "**information user**" as a synonym for an employee because they use information in work tasks.

3.1.5 **Summary**

Chapter 3.1 provided an overview of essential definitions, concepts, and theories about information behavior in work tasks. It is not comprehensive, but it ensures that this dissertation's interdisciplinary topic has a sound and solid basis from the information science perspective. Chapters 3.1.1 and 3.1.2 introduced pragmatic definitions for data, information, and knowledge, and they explained that information needs evolve from informal feelings to formal questions. Chapter 3.1.3 focused on how employees search for information. Besides, it summarized factors influencing the information need and the seeking process. Chapter 3.1.4 covered the meanings of "information use" that Chapter 3.2 will detail for engineering.

[9] This extends Taylor's definition of compromised information needs.

3.2 Use of product-related information in work tasks

Experience with products is instrumental for *learning and improvement* in an organization (similar Choo, 2002, pp. 12–14). Both activities closely relate to information feedback and work task iteration, as Wellsandt et al. (2018) discussed. Feedback and iteration are ambiguous concepts without generally agreed definitions. They occur naturally during engineering work and concern processes, such as correcting errors and coordinating work tasks (Wynn and Eckert, 2017).

Chapter 3.2.1 outlines general problem-solving, which is the foundation of most engineering work tasks[10]. It introduces the basic concepts that affect the task and explains why engineers seek information. Afterward, *Chapters 3.2.2* and *3.2.3* provide an overview of the specific information needs related to product development. They focus on quality management and reliability engineering because textbooks and international standards cover their information needs explicitly. For this reason, they are good examples to outline the state of knowledge about information needs related to engineering work.

3.2.1 General problem-solving

Employees seek and use information during **problem-solving**. The latter is the subject of a research domain rooted in psychology (Dörner, 1974; Duncker, 1974; Funke, 2003).[11] Researchers such as Duncker, Dörner, and Funke, focused on constituents of complex problems and the activities to solve them.[12] Engineering textbooks and industry guidelines used their findings to create frameworks for systematic engineering (Pahl et al., 2007; Ehrlenspiel and Meerkamm, 2017; Verein Deutscher Ingenieure, 2019). This dissertation uses general problem-solving to clarify work tasks in the engineering domain.

Figure 15 illustrates Frensch and Funke's (1995b) description of problem-solving. The authors introduced it as a "European perspective" to indicate that European problem-solving research differed from similar efforts in North America.

[10] Engineering tasks are a subset of development-related work tasks.
[11] Duncker's first edition of the cited book was published 1935.
[12] Few of their articles are in English.

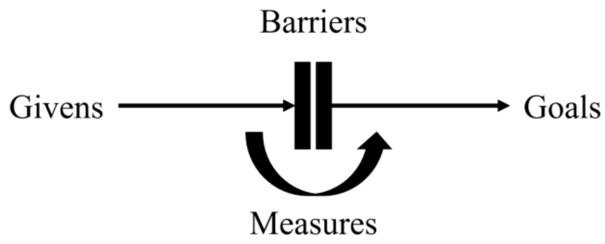

Figure 15: Elements of problem-solving (based on Frensch and Funke, 1995a)

A problem-solving situation consists of givens, goals, barriers, and measures. The *givens* describe the problem solver's current situation, and the *goals* define the desired situation. *Barriers* inhibit that the problem solver reaches the goals without further thinking. There are interpolation, synthesis, and dialectic barriers (Dörner, 1979; Betsch et al., 2011). An *interpolation* barrier exists when the solver knows the goals and the measures well but needs to select suitable measures to solve the problem. A *synthesis* barrier exists when the solver knows the goals well but does not know the measures. Overcoming this barrier requires the identification or development of a suitable measure. A *dialectic* barrier exists when the solver knows the measures but lacks a clear understanding of the goals. A problem solver in a work task context could clarify the goals through a dialogue with stakeholders. Complex problems typically have multiple barriers simultaneously.

Givens, goals, and barriers are complex, change dynamically, and are *intransparent*. The latter means that the solver does not know their exact properties in advance (Byström and Järvelin, 1995; Funke, 2003). The intransparency of a problem-solving situation provides links to information needs and information use. Byström and Järvelin (1995) identified:

- needs related to problem formulation and
- needs related to problem-solving.

The first need type concerns information supporting the employee to make the givens, goals, barriers, and the available measures transparent. It includes, for instance, information about methods in the work domain or the resources available to solve a specific problem. The second need type covers information that supports selecting suitable measures and their application to overcome the barriers.

The previous outline summarizes that using information in a work task context means (1) to reduce the intransparency of the givens, goals, barriers, and measures of the problem situation at hand, and (2) to support the selection and application of suitable measures to overcome barriers.

3.2.2 Quality management tasks

This sub-chapter grounds on a section in (Wellsandt et al., 2018).

Quality is the *"degree to which a set of inherent characteristics of an object fulfils [sic] requirements"* (ISO, 2015a). A *requirement* is a statement that describes a stakeholder's needs and expectations. The customer is a critical stakeholder influencing a product's requirements. An increase in quality improves, amongst others, customer satisfaction. In this context, *customer satisfaction* means the gap between the customer's expectations and the perception of the product (ISO, 2015a). Satisfied customers are loyal to the producer and contribute to higher profitability (Anderson et al., 1994). Employees need information about customer satisfaction to manage product quality.

A *quality management system* (QMS) is a tool to improve product quality continuously. It affects the organization's norms and encourages information feedback in product development. This dissertation covers QMS for the latter reason.

The ISO 9001:2015 standard suggests a structure for a QMS (ISO, 2015b). It is internationally acknowledged and, therefore, focused on in this dissertation. The QMS has five elements: planning, operation and support, performance evaluation, improvement, and leadership.

- **Planning** requires, amongst others, the definition of the objectives for quality management. It further concerns preparing actions to address risks and opportunities for the organization.
- **Operation and support** manage the quality-related aspects of the development process and customer communication. The *development process* should involve customers and users, and its input information should be adequate for development purposes, complete, and unambiguous. Inputs include the functional and performance requirements, information from previous development activities, and potential consequences of product failures. The *customer communication process* should obtain customer feedback. This information includes *"opinions, comments, and expressions of interest in a product [...]"* (ISO, 2015a), and the customer service typically acquires it.
- **Performance evaluation** concerns the monitoring, measurement, analysis, and evaluation of the QMS. Also, the organization shall monitor the customers' satisfaction. It should determine the methods for obtaining, monitoring, and reviewing customer-related information.
- **Improvement** refers to the actions that capacitate products to meet requirements and address future needs and expectations. Also, it shall correct, prevent, and reduce undesired effects.

- **Leadership** promotes, for instance, improvement and ensures that the QMS remains effective.

ISO 9000 has two complementary standards covering complaint handling, and monitoring and measuring customer satisfaction. Both tasks share product-related information with engineering work tasks to resolve complaints and improve customer satisfaction. The following paragraphs cover these standards because they provide details about the information needs of development-related tasks.

ISO 10002 focuses on complaints, i.e., expressions of dissatisfaction provided to the organization. They relate to the product and require a response or resolution. A complaint contains information about, for instance, the product and the encountered problem. The analysis of complaints aims to identify systematic, recurring, and single-incident problems and trends. It seeks to help eliminate the root cause of the complaints also. The more serious, frequent, and severe the complaint is, the more effort the organization should spend on its investigation.

ISO 10004 proposes a model for the monitoring and measurement of customer satisfaction. Four processes build its basis: identify customer expectations, gather customer satisfaction data, analyze customer satisfaction data, and provide feedback for improvement. The organization that wants to comply with this standard should consider the investigated sample size and manage errors, completeness, and accuracy of the gathered data. Also, the organization should examine the data sources[13] for validity and relevance. The organization should use the acquired information to improve its products, processes, or strategies. Possible contents of the examined customer satisfaction data are, for instance, information about trends, comparative or competitor information, weaknesses and strengths of the organization's product, challenges, and potential opportunities.

3.2.3 Reliability engineering

Reliability is "*[…] the ability of a product or system to perform as intended (i.e., without failure and within specified performance limits) for a specified time, in its life cycle conditions*" (Kapur and Pecht, 2014). A more straightforward notion defines it as the number of failures over a period (O'Connor and Kleyner, 2012, p. 1). It is typically impossible to make an exact forecast when a product fails. Employees use probabilities to describe how likely failures occur for this reason.

Reliability engineering is an interdisciplinary activity involving employees from, for instance, product development, manufacturing, testing, transport, and service departments. Its basis is the identification of a failure's root causes (Kapur and

[13] Here, data source refers to actual sources and communication channels.

Pecht, 2014, p. 339). A **root cause** is a *"source of a defect such that if it is removed, the defect is decreased or removed"* (ISO, 2017). An employee creates failure-related information during the root cause analysis. This information serves as input for subsequent tasks, such as measuring, predicting, accommodating, and preventing failures (Condra, 2001, pp. 13–15). The following paragraphs briefly describe these tasks and their related information needs and use.

- The **measurement of failures** bases on statistics applied to historical information about a product population and its failures (Kapur and Pecht, 2014, p. 19). Failure information describes, for example, when a product failed and under which load conditions it happened. *Loads* are essential because their nature, magnitude, and duration can cause stress in the product. The latter can result in failures which reduce the product's reliability. Employees in reliability engineering apply various mathematical functions and visualization methods to create and interpret failure information. An example is a histogram that visualizes the amount of failed products with the operation time until the failure occurred. Quantified reliability-related information supports the employee in identifying areas that need measures to accommodate or prevent failures.
- **Prediction of failures** uses mathematical functions to describe the probability that a system will work as intended during its lifetime (Kapur and Pecht, 2014, pp. 82–83). It uses probability models (e.g., normal and exponential distributions) and makes assumptions about the product. The prediction's result depends on their quality and, for instance, how the employee defines a failure. Besides, predictions need to consider non-statistical factors, such as design changes and unforeseeable production or service problems (O'Connor and Kleyner, 2012, p. 7).
- **Accommodation of failures** assumes that failures will occur during the product's lifetime, and employees should take measures to minimize their effects (Condra, 2001, p. 14). Effective measures include redundancy of system components, warranties, and maintenance. A more complex measure is, for instance, using a Functional Product – a type of PSS – that transfers the responsibility of failure accommodation from the customer to the service provider (Markeset and Kumar, 2005). Failure accommodation benefits from a comprehensive understanding of the product and its use. The latter includes information about load conditions.
- **Prevention of failures** seeks to design and produce products that are unlikely to fail. It requires that employees thoroughly understand the product, its behavior during use, failure mechanisms, and the capability to control the product, its operation, and the operating environment (Condra, 2001, p. 15).

Prediction, accommodation, and prevention tasks benefit from information about load conditions. A *life cycle profile* describes the load conditions and their context (Kapur and Pecht, 2014, p. 150). Employees can use three steps to develop one:

- Describe the expected *events*[14] from manufacturing to end of life, and identify the product requirements.
- Identify natural and induced *environmental conditions,* their combinations, and related load conditions for each expected event.
- Detail and quantify the expected *load conditions*.

Employees can acquire the information for a life cycle profile through market surveys and standards, analysis of similar products, field trials, service records, and product performance monitoring (Kapur and Pecht, 2014, p. 165).

3.2.4 Summary

Chapter 3.2 outlined why and how employees use product-related information in engineering work tasks. It is not a comprehensive overview of engineering tasks, but it covers typical tasks that could benefit from PUI. This chapter introduced a framework for general problem-solving, quality management, and reliability engineering to clarify different work tasks and their information needs. Most generally, problem-solving needs information to raise the transparency of the givens, goals, barriers, and measures of a problem situation. Quality management needs information about the fulfillment of functional and non-functional customer requirements. Reliability engineering needs, for instance, information about load conditions to measure, predict, accommodate, and prevent failures through the product life cycle.

3.3 Methodologies and methods for information need analysis

As outlined above, an information need is dynamic. It begins as a vague feeling and evolves into a specific understanding of the information items needed to perform a work task. When needs are formal, the employee can communicate them to an IM expert. The expert analyzes these needs to develop requirements for an IS capable of satisfying the stated information needs.

Literature proposed methodologies to structure the analysis of information needs. *Chapter 3.3.1* outlines a general methodology for need assessment, which provides basic concepts for the analysis (similar to Dorner et al., 2015, p. 46). It covers crucial assumptions, process steps, and examples of methods to collect infor-

[14] Examples are testing, qualification, storage, transportation, operation and service.

mation about needs. *Chapter 3.3.2* describes more specific methodologies to perform an information need analysis. These methodologies differ in their approach, structure, and scope, as summarized in Table 7.

Table 7: Overview of information need analysis methodologies

Authors	**Approach**	**Structure**	**Scope**
Dorner et al.	Prepare, gather, analyze, report	Four steps; iterative	Information organizations
Voß and Gutenschwager	Scoping, as-is analysis, to-be analysis, act	Four phases and seven steps; phase three is iterative	Economic perspective on information needs
Lundqvist et al.	Scoping, participative modeling, analysis, evaluation, documentation	Four steps; information demand pattern	Not specific (explicitly holistic)
Weber et al.	Identify interaction, analyze work task, identify needs, clarify characteristics	Four steps	Product-related information & engineering work tasks

3.3.1 A general methodology of need assessment

Watkins et al. (2012, p. 20) define a *need* as the gap between the current and desired results in a work task. They consider need assessment a performance improvement method focusing on individual, group, organizational, or societal needs. This dissertation uses the term *assessor* to refer to the person that conducts a need assessment. Watkins et al. (2012, pp. 46–50) identified three steps to assess needs:

- **Identify needs**. The assessor identifies the partaking stakeholders, needed information, potential information sources, and arrangements to collect or create the information. This step focuses on the results rather than the activities and resources to achieve them.
- **Analyze needs**. The assessor prioritizes the needs, analyses the givens and the barriers[15], identifies the relationships between the needs, and collects information about the barriers' root causes.

[15] Watkins et al. vaguely refer to "[…] what is working [and] what is not working, […]".

- **Decide what to do next**. The assessor uses the information from the previous steps to identify actions that may contribute to the desired result. Partaking stakeholders and the assessor evaluate how each option contributes to the intended improvement. A subsequent activity prioritizes the needs according to the cost of meeting them and not doing so.

The assessor can use many methods to acquire information about needs, such as document review, data review, guided expert review, focus group, interviews, and task analysis. Other methods support analysis and prioritization tasks, such as multi-criteria analysis, cause-effect diagrams, scenarios, root cause analysis, and fault tree analysis.

3.3.2 **Methodologies for information need analysis**

Information need analysis is an essential task in IM (Voß and Gutenschwager, 2001, p. 74; Choo, 2002, p. 24; Krcmar, 2015, p. 121). It shares ideas and methods with need assessment though it is not explicitly related. This dissertation identified four methodologies to perform an information need analysis in a work task context. It further provides an overview of related supportive methods. The methodologies and supportive methods specify: (1) the *activities* of IM experts during information need analysis and (2) what *aspects* they should consider. The first methodology is the most general, and the last focuses on product-related information and engineering work tasks.

3.3.2.1 Dorner et al.'s methodology

Dorner et al. (2015, p. 67) proposed an analysis based on the context of general need assessment. Their methodology contributes to three goals: (1) determine what needs exist, (2) identify how to meet the needs, and (3) provide benchmark criteria to determine how well the investigated entities meet the needs.[16] The procedure has four activities that occur recursively and concurrently. The following list briefly summarizes them:

- **Preparation**. This activity defines the purpose of the analysis, identifies questions, persons to ask, supportive methods, available resources, and a schedule of the entire procedure.
- **Information gathering**. The IM expert plans how to gather information about information needs. It includes defining research questions and identifying the methods to acquire information about the needs. The IM expert must consider the analysis' validity, reliability, and ethical questions.

[16] Dorner et al. merged the activities one and two. I separated them for easier understanding.

- **Information analysis**. The analyst tries to answer the research questions by analyzing qualitative and quantitative data using, for instance, qualitative content analysis and statistics. This activity is iterative and may require the collection of additional information.
- **Reporting the results**. This activity creates one or more reports about the information need analysis. It provides a permanent record of the previously described activities and communicates the findings to the target audience.

Iterations between the activities above are common, and the activities influence each other significantly.

3.3.2.2 Voß and Gutenschwager's methodology

Voß and Gutenschwager (2001) proposed a methodology grounded on an economic perspective on information needs.[17] The German economist Arnold Picot stated that an information need *"[...] is defined according to the type, quantity and quality of the information that persons require in order to complete tasks within a certain timeframe"* (Picot et al., 2001, p. 81, 2008, p. 68).

The analysis of information needs provides the IM expert with an accurate understanding of the information that an IS needs to provide (similar Krcmar, 2015, p. 121). Figure 16 illustrates Voß and Gutenschwager's methodology.

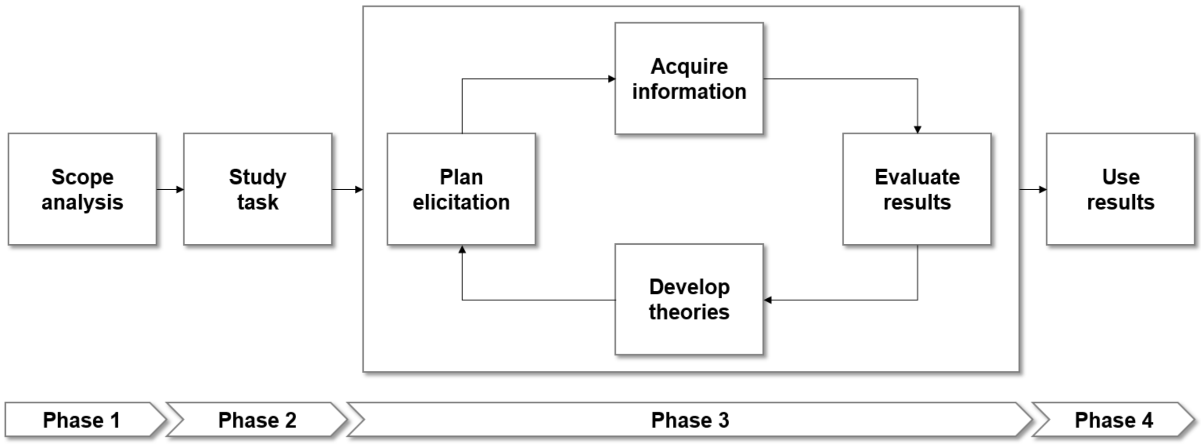

Figure 16: Model of information need analysis (based on Voß and Gutenschwager, 2001)

The methodology above has four subsequent phases. The **first phase** is an initial assessment of the investigated work task, and its primary purpose is to scope the analysis. This activity is necessary because the IM expert typically cannot focus on all employees' information needs at once (Voß and Gutenschwager, 2001, p. 136). During the **second phase**, the analyst captures, describes, analyses, and

[17] Krcmar et al. (2015) reference this methodology in their recent text book.

evaluates the current information the focused employees use. The analysis covers the purpose of the information's use, its impact on the work task, and the influence of existing IS. Process models of the task support the IM expert during this phase. The **third phase** includes a methodology to elicit new information needs. The following list describes its steps briefly:

- **Plan elicitation**: Identify methods to elicit information needs and plan the elicitation, e.g., schedule appointments.
- **Acquire information**. Use the identified methods to collect information from documents and during employee appointments.
- **Evaluate results**. Identify the needs through, for instance, statistical methods. Discuss the methods' application and the acquired information.
- **Develop theories**. Structure the information needs and their relations.

The IM expert can repeat any of the steps above. These recursions can be necessary if, for instance, the results are inconsistent or turn out to be incorrect. According to the analysis results, the **fourth phase** typically means changing an existing IS architecture.

3.3.2.3 Lundqvist et al. methodology

Lundqvist et al. (2011) define an information demand as the "*[...] constantly changing need for relevant, current, accurate, reliable, and integrated information to support (business) activities, when ever [sic] and where ever [sic] it is needed.*". Their notion of an information demand is so closely related to an information need that this dissertation does not differentiate the terms.[18] Figure 17 illustrates a simplified version of Lundqvist et al.'s methodology.

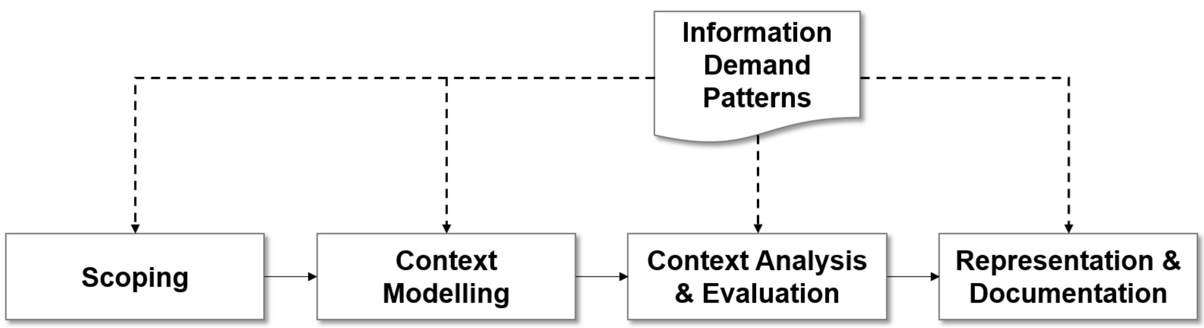

Figure 17: Information demand analysis (adapted from Sandkuhl and Seigerroth, 2018)[19]

The analysis has four process steps. Each one produces a description of an information need's context. The methodology relies on participative modeling with

[18] Case and Given (2016) discuss differences between need and demand in detail.
[19] The primary source was not accessible.

stakeholders (Lundqvist et al., 2011, p. 17). These stakeholders determine **information demand patterns** consisting of the following elements (Sandkuhl, 2011, pp. 3–4; Sandkuhl and Seigerroth, 2018, p. 9):

- A statement about the *organizational context* where the pattern is useful
- The *problem description* of the role that the pattern addresses
- The *solution description* that resolves the problem covers:
    - Specific parts of the role's information demand
    - Quality characteristics, such as accuracy, completeness & timeliness
    - Timeline when the information parts should be available at the latest

Each pattern includes an impact assessment covering, for instance, potential economic consequences, efficiency gains or losses in performing the task, and the influence on the quality of the work result.

3.3.2.4 Weber et al. methodology

Oehlmann et al. (1997) proposed a methodology to analyze product-related information exchange in engineering work tasks. The authors named it the Formal Interaction Analysis (FIA). An **interaction** represents the situation when an employee uses a communication channel to satisfy an information need. The FIA helps information managers to create awareness about challenges and potentials concerning the availability of information. Weber (2005) improved the FIA methodology and evaluated its applicability. The following list outlines the steps of his version of this methodology that cover information needs.

1. The first step **identifies the problematic interaction**. Key indicators for problems during an interaction are unsatisfying communication times and inappropriate information characteristics.
2. The second step **characterizes the work tasks**. It creates a description of the relations between each task and the involved employee profiles.
3. The third step **identifies the information needs** in terms of the needed contents to perform the respective task. Weber points out that this step focuses on the <u>personal</u> information need rather than a hypothesized objective need (Weber, 2005, p. 116). The IM expert typically clarifies the information need through keywords or short statements describing specific information items contained in a recognizable information container (e.g., a bill of materials or datasheet).
4. The fourth step **clarifies the subjectively needed characteristics** of the previously identified information items. The FIA user should focus on the relevant information characteristics. Weber provides a guideline for this purpose that covers, for instance, readability, relevance, novelty, reusability, accessibility, and scope of content (Weber, 2005, pp. 146–147).

Weber applied the methodology in several engineering companies. He concluded that the evaluation's participants understood the concept of using information characteristics well. They were able to identify relevant characteristics and define new ones where necessary.

3.3.2.5 Supportive methods

The IM expert can use several methods to support an information need analysis (Voß and Gutenschwager, 2001; Vuori, 2006; Krcmar, 2015). Krcmar differentiates subjective, objective, and mixed methods.

- **Subjective methods**, such as questionnaires and interviews, focus on employees and their personal information needs. The IM expert can generalize individual needs and define the *information user's roles* for the investigated work task. This approach is helpful because the IM expert may not have enough resources to apply these methods to all employees.
- **Objective methods** derive information needs analytically. They include the analysis of strategic goals, task analysis, and document analysis.
- **Mixed methods** complement analytical methods with subjective ones. Examples are the Critical Success Factors method (Rockart, 1979) and the Key Intelligence Topics method (Herring, 1999). Both focus on strategic management and require a skilled researcher familiar with the method (Vuori, 2006).

Information need analysis typically includes several methods (Watkins et al., 2012, 3). This approach is helpful because employees can have subconscious needs (Vuori, 2006). The IM expert cannot elicit them using a questionnaire but may consult work-related documents to identify cues for these needs. The expert could use these cues in an interview with an employee to validate the findings from the documents.

3.3.3 Comparison

The methodologies above outline how IM experts can analyze information needs. Chapter 3.3.1 describes the steps of a need assessment. It outlined the assessor's methods to acquire and analyze need-related information. Chapter 3.3.2 introduced methodologies that include information need analysis. Table 8 summarizes the activities and concrete aspects to consider for each methodology. The aspects provide insight into how IM experts could concretize information needs.

Table 8: Activities and aspects of information need analysis

Methodologies	Activities	Aspects
Dorner et al.	Preparation, information gathering, information analysis, reporting the results	Validity, reliability, and ethical aspects of the analysis
Voß and Gutenschwager	Scope analysis, study task, plan elicitation, acquire information, evaluate results, develop theories, use results	Type, quantity and quality of information
Lundquist et al.	Scoping, context modeling, context analysis & evaluation, representation & documentation	Relevance, currentness, accuracy, reliability, integration of information; impact assessment
Oehlmann et al. / Weber	Identify problematic interaction, characterize work task, identify information needs, clarify subjectively needed characteristics	Readability, relevance, novelty, reusability, accessibility, and scope of content

Dorner et al.'s methodology grounds on the general *need assessment* approach. It suggests, amongst others, that IM experts should consider the analysis' validity, reliability, and ethical aspects. Voß and Gutenschwager base their methodology on the *economic perspective* that assumes an information demand and an information supply in organizations. Their methodology includes similar activities compared to Dorner et al.'s methodology. A difference is that Voß and Gutenschwager clearly distinguish between currently used information (second phase) and new information needs (third phase). Lundqvist and Sandkuhl ground their methodology on the economic perspective as well. Their solution uses *participative modeling* to identify *role-centric* information demand patterns. The patterns include an *impact assessment* that clarifies how meeting the demand affects the work task. Weber's improved FIA methodology focuses on the design of ICT solutions for engineering-related interactions. It includes four steps that cover the analysis of the needed information. These steps are largely similar to the steps of the other methodologies. A difference is that the FIA user specifies the needed information on the *level of information characteristics*. Weber evaluated this approach and found evidence that the FIA users understood the approach and could identify relevant and new information characteristics.

3.4 Conclusion

The chapters above do not explicitly cover PUI. Therefore, the following paragraphs transfer the findings above to this specific context. Each paragraph has a caption to summarize the main conclusions and key assumptions for the following method development.

PUI seeking occurs to improve task results.

The main driver of information seeking during work tasks is the information gap that employees experience. For development-related tasks and PUI, this gap does not necessarily mean an employee can perform a task only if PUI is available. Instead, the employee seeks PUI to produce results of higher quality. Examples are making product requirements and models more accurate and complete (Oriol et al., 2018; Tao et al., 2019).

Assist in formalizing information needs using information characteristics.

The introduced concepts above outline what an information need is and how it relates to work tasks and information behaviors. Taylor's information need typology provides the critical assumption that an employee can express an information need by describing it formally. Formalizing information needs is complex, and the employee typically cannot do this alone. The methodologies presented in Chapter 3.3.2 provide the IM expert with systematic activities to formalize information needs. Subjective, objective, and mixed methods support the expert in this process. Furthermore, Weber's findings indicate that using information characteristics is helpful in concretizing which information items the employees need.

Derive information characteristics from the specific work task.

Literature related to product development describes that employees need product-related information to identify or better understand load cases, failure reports, or customer requirements. The investigated literature in Chapter 3.2 seldomly mentions specific information characteristics, such as validity and accuracy (e.g., Kapur and Pecht, 2014, p. 82). One explanation is that some characteristics depend on the context of the individual work task (Strong et al., 1997). It is, for instance, impossible to determine the currentness of information precisely without considering the specific work task. For instance, failure information that is several years old could be valid if an employee needs to understand a failure that occurs only once per year. In this case, the employee balances the limited number of failure events and the information's "timeliness" requirement. This behavior is in line with the compromise between the sought information's requirements and constraints emerging from personal and non-personal factors introduced in Chapter 3.1.3.

The latter two statements are the main assumptions to scope the following chapters. Information need analysis should describe the needs precisely on the level of information characteristics and within the context of a specific work task. The methodologies and methods above provide a general toolbox for this task but little support for identifying relevant information characteristics – and less for PUI. In addition, technological progress challenges the application of PUI in product development. Employees can have little knowledge about the application potential and the constraints. Therefore, they may be less capable of describing realistic information needs.

Both gaps increase the number of iterations of need analysis until the employee can precisely describe the needed information. This is a problem because it takes longer or is more costly to design the IS.

Chapter 4 addresses the problems above with a need analysis method for development-related work tasks and PUI.

4 Method development

The previous chapter outlined important concepts and terms about the emergence of information needs and their analysis in a work task context. It concluded that the IM expert should specify the needs on the level of information characteristics and for specific work tasks. This chapter describes a method for PUI need analysis applicable to development-related work tasks.

According to Mayer et al. (1995, p. xi), a **method** has a definition, a discipline, and many uses. The method's definition contains its concepts, motivations, and theoretical foundations. Its discipline component covers the procedure and syntax. These elements represent the method's user interface, whereas the former describes what to do and the latter how to communicate it. The syntax can be computer-interpretable or graphical to communicate the method to humans. The use of a method means its application in specific situations. **Method engineering** is the discipline of designing, constructing, and adapting methods for IS development (Brinkkemper, 1996). Its central proposition is to adopt an existing method whenever possible. If this is not suitable, it suggests adapting an existing one or developing a new method if it is not feasible (Mayer et al., 1995). This dissertation applies the method engineering approach above to organize this chapter.

Chapter 4.1 covers the method's concepts and motivations. It begins with revising some concepts and their organization in a framework. This framework is the conceptual basis for the method engineering process. Requirements specify the method's motivation in the second part of this sub-chapter. *Chapter 4.2* provides the method's theoretical foundation and focuses on application cases that used PUI in development-related work tasks. The related literature mentions or outlines relevant information characteristics and the phenomena that affect them. *Chapter 4.3* identifies the application procedure and derives the method's syntax from the theoretical foundation. *Chapter 4.4* introduces the assumptions, pre-conditions, and activities to apply the method. *Chapter 4.5* introduces a tool that supports the method application, and *Chapter 4.6* concludes this method development process.

4.1 Concepts and motivations

4.1.1 Framework

Chapter 3 introduced various concepts relevant to this dissertation. The academic discourse about them is complex and could make this manuscript incomprehensible. Figure 18 provides an overview of its most relevant information and communication concepts. It adapts Weber's meta-model of interaction because his model is similar in structure and content (Weber, 2005).

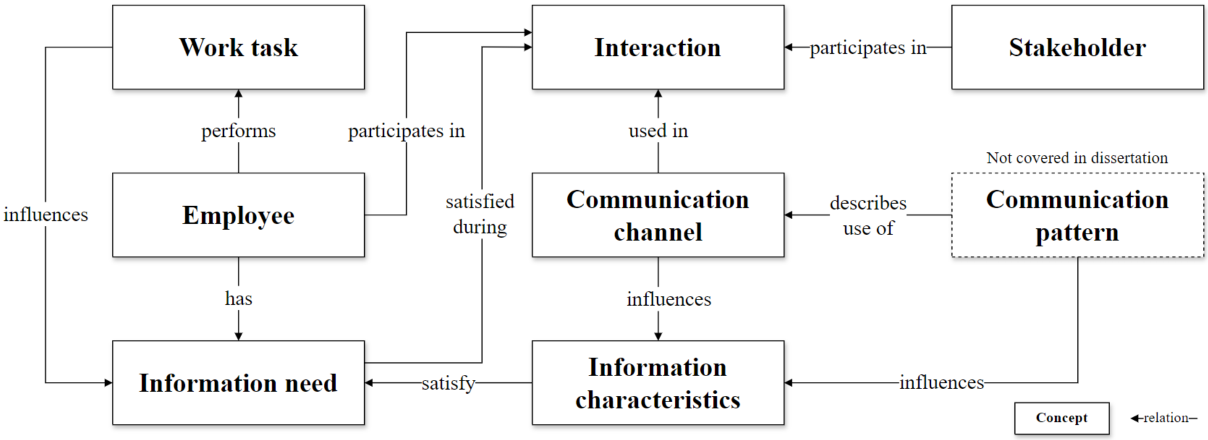

Figure 18: Framework of concepts in development-related work tasks

The following list describes the adaptions made to Weber's meta-model.

- Weber's model indicates that work tasks <u>have</u> information needs. According to the identified research about information needs, tasks <u>influence</u> information needs (e.g., Leckie et al., 1996). The need itself belongs to the employee that performs a work task.
- Weber's model differentiates information characteristics and information content. The latter relates to the information's relevance. Wang and Strong (1996) argue that relevance is a characteristic; therefore, this framework collapses Weber's two concepts into "information characteristics".
- Weber's model defines that a work position participates in an interaction. PUI-related interactions can involve other parties, for instance, product users or service technicians. Therefore, this framework includes stakeholders as participants in interactions to account for this broader scope.
- Communication patterns are not considered further in this dissertation because they would likely require a deeper investigation of social sciences.

Chapter 4.1.2 summarizes the requirements of the method as identified in this dissertation. They provide the basis for the method's development and serve as evaluation criteria later.

4.1.2 Requirements

A **requirement** is a "*need or expectation that is stated, generally implied or obligatory*" (ISO, 2015a). This dissertation derives the method's requirements from *literature and experience* applying PUI in development-related work tasks. Besides, the research projects Manutelligence (N.n., 2018) and Falcon (N.n., 2017a) provided insight into related IS design tasks. All requirement descriptions in this manuscript follow an acknowledged phrasing schema to improve their consistency (Hull et al., 2011, 4.8).

The following paragraphs use the term "**user**" to refer to the person applying the method. Typical users are persons with IM knowledge responsible for designing a new PUI system or improving an existing one. Besides, users include stakeholders and employees working on development-related tasks – they partake in the information need analysis.

Requirement 1: The user should be able to identify relevant information characteristics systematically.

Applying all available PUI in development-related work tasks is still a comparably new field of research. Chapter 3 pointed out that knowledge about information characteristics is important in IS design. The research projects above, for instance, identified the needed characteristics without using a systematic process.[20] This shortcoming led to an incomplete view of the employees' information needs, resulting in potentially avoidable IS changes. For this reason, the method should allow the user to identify relevant quality characteristics according to a fixed plan or system, i.e., systematically. The systematic approach could allow the user to formalize information needs faster and lower the chance of missing essential characteristics. A collection of potentially relevant characteristics would be helpful in this context.

Requirement 2: The user should be able to spot constraints for information needs.

Some phenomena influence information quality and make it necessary that the employee accepts, for instance, a *tradeoff* between two needed characteristics or that the IS cannot meet some characteristics in an entirely satisfactory way.[21] An example of a tradeoff is the relation between accuracy and timeliness – reaching high accuracy can require time-consuming data processing, e.g., managing outliers and data points with empty values. The user should be able to identify tradeoffs to formulate more real information needs. For this, they need to understand phenomena influencing information quality.

Requirement 3: The user should be able to apply the method in situations with multiple PUI types.

PUI can be machine-generated or human-authored. The literature typically focuses on one of these types, but employees often use a combination because their characteristics are complementary. For instance, measurements are typically more

[20] Falcon, for instance, elicited the overall requirements systematically but did not specifically investigate information characteristics. The project's main Ontology does does not contain information quality (N.n., 2017b).

[21] A constrained information need is not the same as Taylor's compromised information need. The former is a more realistic formalized need, while the latter refers to how a person ultimately presents their need to the IS.

precise than a product user's report, while a report can incorporate critical context for abrupt changes during measurement. The user should be able to apply the method in situations where machine-generated and human-authored PUI types are relevant. This approach ensures support for a broader range of application cases.

Requirement 4: The user should require minimum effort to use the method.

A common goal in work tasks is that the applied methods have low costs and high benefits. Costly efforts to apply the method concern, for instance:

- **Preparation**. User training or the preparation of supporting materials, such as forms, descriptions, illustrations, and presentations.
- **Application**. Time users must spend during the method application and the time for the documentation of contributions.
- **Post-processing**. Analysis of the supporting materials and the documentation to derive results and conclusions for further action.

A measure to avoid costs is to design the method to be self-explanatory, providing all information to quickly prepare, apply, and post-process the method. Preparation and post-processing steps should be avoided.

4.2 Theoretical foundation

This chapter uses literature to identify and systematize PUI characteristics and their influencing factors.

4.2.1 Identification of quality characteristics

The fastest approach to identifying quality characteristics is to use an existing collection. Several researchers described potentially relevant characteristics as indicated by the following paragraph – ordered by publication time.

Augustin (1990) outlined the information need in terms of the right information, at the right time, in the right amount, at the right place, and with the required quality. These characteristics are comparably unspecific and thus difficult to adopt. He distinguishes quality from other characteristics, such as content (right information), timeliness, quantity, and location. This manuscript does not adopt Augustin's collection because it lacks empirical grounding. Wang (1996) specified empirically which data characteristics the data users need.[22] He identified 15 data characteristics as most relevant. Wang's research coined the concept of data quality and often is the foundation of more recent collections of characteristics.

[22] Wang refers to data but the characteristics consider information as well.

Weber (2005) developed a software tool with a catalog of information characteristics to support the application of the FIA methodology in product development. He mentions that this catalog includes approximately 130 characteristics, but this dissertation cannot consider it further since the software was not published. Eppler (2006) proposed 12 information characteristics based on literature analysis. He refined previous work in the light of information rather than data.

None of the researchers above and no other researcher focused on product usage and applying PUI in development-related work tasks. Therefore, this dissertation cannot reuse an existing collection of information characteristics and will instead focus on creating a custom collection.

Liu and Chi (2002) identified three approaches to determine the characteristics of information. The *intuitive* approach grounds on an expert's experience and intuitive understanding of relevant information characteristics. A focus on data users and their information needs is a characteristic of the *empirical* approach. The *theoretical* approach derives characteristics from a theory. Liu and Chi identified the mathematical theory of communication, information economics, operations research, and ontological mapping as theories. The selection of the right approach depends on the available resources, e.g., time, access to experts, and the availability of suitable theories.

This dissertation uses document analysis to identify relevant information characteristics. The documents originate from academic literature and contain **application case descriptions**. This approach is a mix of the intuitive approach and the empirical approach. It grounds on the researchers' experiences and opinions (intuitive) that applied PUI in development-related work tasks (empirical). Application case descriptions are readily available in the literature, while experts or users that could identify information characteristics are not. A downside of this document analysis is that many case descriptions do not focus on information quality. This lack of scope means that the identification partially relies on the interpretation of text passages, which may introduce interpretation errors. This dissertation addresses this downside as follows:

- The document analysis covers multiple application case descriptions. This approach avoids that a potential misinterpretation of one reference's text passage has too much impact on this dissertation's conclusions.
- The identification process utilizes an acknowledged information quality model. It acts as a reference point to classify text passages and provides an initial structure this dissertation can adapt.

Chapter 4.2.1.1 introduces information quality models that are potentially relevant to support the identification of information characteristics.

4.2.1.1 Information quality models

Eppler (2006, p. 47) identified various models describing the information characteristics an organization needs to consider. His study lists 70 potential characteristics. Technical specifications and international standards, such as the ISO 8000 series (master data quality) and ISO 25012 (software product quality), propose general quality characteristics. Appendix B compares the information characteristics of three acknowledged information quality models to substantiate the decision for one of the models in this thesis. This comparison indicates that the three models are similar.

One controversial aspect of the models concerns the characteristic that describes the fit between the information's content and its purpose. Information and library sciences use the concepts of aboutness and relevance to clarify this fit (Saracevic, 2016). *Aboutness* relates to information organization through classification codes, subject headings, and index terms. *Relevance* means the match between a query and the retrieved information items. It is a central information characteristic in computer-supported search. This dissertation does not focus on aboutness and relevance or the complex "search and information retrieval" domain. Furthermore, this dissertation adopts the 15 characteristics proposed by the ISO 25012:2008 standard. It is the most suitable model for this manuscript because:

- the standard's scope includes persisted data of all data types,
- the ISO is an internationally reputed organization,
- the ISO reviews its standards regularly, and
- alternative models share many of its characteristics (refer to Appendix B).

Chapter 4.2.1.2 describes this dissertation's methodology to identify information quality characteristics within PUI application case descriptions.

4.2.1.2 Methodology

Academic literature about the integration of PUI in development-related work tasks has different scopes (Hou and Jiao, 2020; Deng et al., 2021; Meyer et al., 2021). This dissertation focuses on application case descriptions and text passages that mention needed information's characteristics. Content analysis is an acknowledged research method to analyze texts systematically. It can be quantitative or qualitative. The quantitative analysis builds metrics to make sense of a text by counting, for instance, specific words. It is less appropriate when the research question is explorative (Mayring, 2000), as in this dissertation. A **qualitative content analysis** systematizes text into *content categories* (Mayring, 2000). Elo and Kyngäs (2008) identified two approaches to performing a qualitative content analysis:

- The *inductive* approach focuses on specific observations and develops general conclusions. It is useful when the knowledge about a phenomenon is fragmented. An open categorization process is a feature of this approach.
- The *deductive* approach transfers an existing theory or model to a specific case to test hypotheses. It is useful when previous knowledge about the content is available. A strict categorization procedure with coding instructions is a feature of this approach.

This dissertation investigates application case descriptions, which are specific observations of the integration of PUI in development-related work tasks. The identified descriptions are mainly[23] independent of each other and follow no generally agreed-on terminology. This knowledge fragmentation makes an inductive approach most suitable for this dissertation.

The **inductive approach** has preparation, organization, and reporting phases (Elo and Kyngäs, 2008). It begins with the *preparation phase,* where the analyst selects the unit of analysis and makes sense of the literature as a whole. The unit of analysis defines which items the analyst focuses on. Example items are words, sentences, and paragraphs. The *organization phase* incorporates coding, grouping, categorization, and abstraction. Coding is an activity where the analyst makes notes and writes down headings/categories for the unit of analysis. For the inductive approach, coding is open because there are no pre-defined headings/categories and no rules on assigning them. Grouping places the notes under higher-order headings – this creates categories. The categorization step compares the items with the created categories, while abstraction generalizes the categories into, for instance, models and concepts describing the investigated phenomenon. The *reporting phase* summarizes the procedure and the findings of the content analysis.

This dissertation bases its analysis on the three phases above. It focuses on the sentences (unit of analysis) in application case descriptions to identify information characteristics. Paragraph analysis provides insight when authors reference characteristics indirectly, i.e., when they do not name the characteristic. I identified relevant passages within the case descriptions and coded them. ISO 25012:2008 and its 15 information characteristics provide the starting point for this process. The ISO model grounds the coding process on current knowledge from information quality research.[24] Using the ISO model introduces a deductive component because the model is an existing theory. Figure 19 summarizes the analysis methodology above.

[23] Sometimes authors refer to each other or they described several, similar application cases.
[24] ISO 25012:2008 was last reviewed in 2019 and remains current.

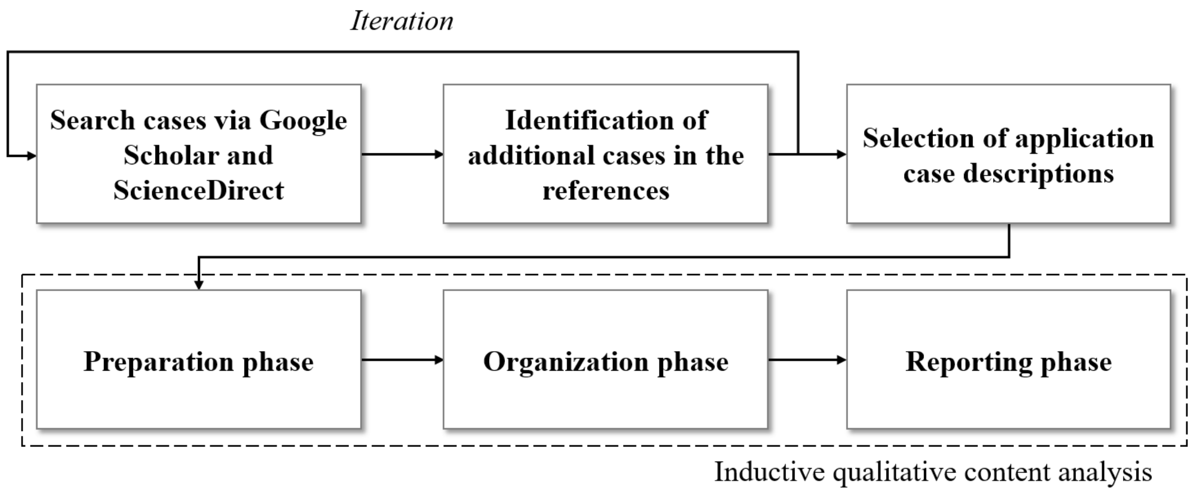

Figure 19: Dissertation's methodology to identify characteristics of needed information

The application case identification took two years (2016-2017), and after several iterations and continuous extensions, the collection contained 22 case descriptions. The earliest is from 1997, and the latest is from 2016. Appendix C summarizes each case's main objectives, products, and communication channels. The main challenge in collecting the case descriptions was that their authors did not use shared vocabulary to describe their work. For this reason, a literature search with a fixed set of keywords was not possible.

Chapter 4.2.1.3 presents the results of the reporting phase of this content analysis.

4.2.1.3 Findings

The content analysis identified 55 text passages related to information characteristics mentioned in the ISO 25012:2008 standard. They concern completeness, credibility, currentness, accessibility, confidentiality, efficiency, precision/accuracy, understandability, and availability of PUI. I merged accuracy and precision because both concepts were difficult to distinguish in the texts, and the application cases' authors did not differentiate clearly between them. The analysis did not identify consistency, compliance, traceability, portability, and recoverability in the case descriptions. Furthermore, the analysis identified 13 text passages that indicate characteristics not included in the ISO standard. Table 9 summarizes the findings above. Appendix D and Appendix E provide the entire categorization of the relevant text passages.

Table 9: Summary of the identified information characteristics in PUI application cases

Mentioned in	Information characteristics
Standard & case	Accessibility, accuracy, availability, completeness, confidentiality, credibility, currentness, efficiency, precision, understandability
Standard only	Compliance, consistency, portability, recoverability, traceability
Case only	Value, relevance, richness

The list of identified characteristics is not comprehensive. However, it provides a first systematic overview of the characteristics of PUI needed for development-related work tasks.

4.2.1.4 Conclusion

The content analysis of the application case descriptions identified 13 characteristics that indicate PUI quality.[25] Value, relevance, and richness are complex and ambiguous characteristics investigated in literature: Moody and Walsh (1999) discuss the value of information, Saracevic (2016) summarizes research about relevance in the library and information sciences, and Daft and Lengel (1984) developed the notion of information richness. This dissertation focuses on the characteristics mentioned in the application case descriptions <u>and</u> the ISO 25012:2008 standard. This scope is helpful because the standard provides pragmatic definitions adequate for applied engineering science. Table 10 concludes the definitions of information quality characteristics used in this dissertation.

Table 10: Concluded meanings of product-related information characteristics

No.	Characteristics	Information <…> (the meaning of characteristic)
1	Accessibility	provides an opportunity to be used when needed
2	Accuracy	is without syntactic and semantic problems; describes the true value of something
3	Availability	exists and the employee knows about it
4	Completeness	describes a population of events, products, or persons
5	Confidentiality	can be accessed by authorized data users only
6	Credibility	is assessed as true or believable
7	Currentness	has the right age for a specific context of use
8	Efficiency	does not contain meaningless data
9	Precision	contains details about a specific context
10	Understandability	can be interpreted by data users

These ten elements specify how, i.e., in which concrete aspects, the needed and provided information should match. Chapter 4.2.2 identifies phenomena that influence PUI quality in development-related work tasks.

[25] Ten from standard & case plus three mentioned in the cases only.

4.2.2 Systematization of causes influencing characteristics

Quality management uses root cause analysis to identify the main causes of quality deviations. Liu and Chi (2002) investigated this topic from the information quality perspective. They base their work on life cycle thinking, ensuring that the analysis scope covers multiple related activities from data creation to data usage. This dissertation adapts their findings and transfers them to PUI.

Chapter 4.2.2.1 introduces Liu and Chi's work. It outlines the idea of using cause-effect diagrams for information quality problems and provides the basis for the subsequent literature analysis.

4.2.2.1 Liu and Chi's evolutional data quality concept

Liu and Chi (2002) argue that data has meaning if theories exist that put it in a context. They defined a theory as a "*[...] general designation for any technique, method, approach, or model [...]*" that someone employed during the data[26] life cycle. The latter describes activities that focus on data. The following paragraphs introduce Liu and Chi's data life cycle and their *evolutional data quality* concept.

Figure 20 illustrates Liu and Chi's data evolution life cycle using the name "**data life cycle model**" because the original figure does not explicitly illustrate the "evolution" feature. The following paragraphs describe this feature separately.

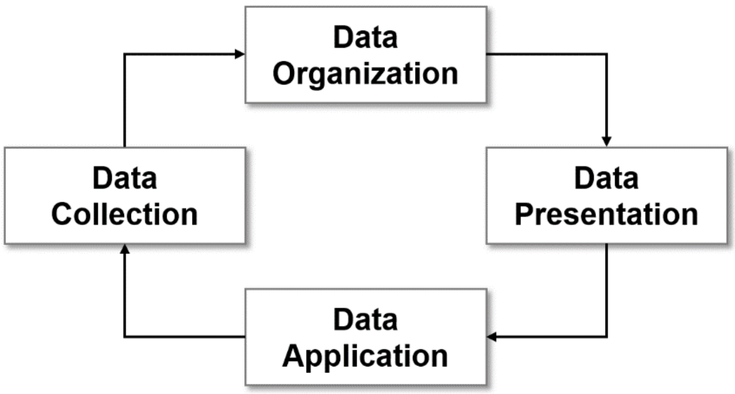

Figure 20: Data life cycle model (Liu and Chi, 2002)

The data life cycle model has four steps, which this dissertation interprets from an applied engineering science viewpoint. Its definitions can be ambiguous, but they are sufficient to describe PUI's application in development-related tasks.

[26] I use data and information interchangeably in this part to align with Liu and Chi's notions.

Data collection incorporates the creation of data. It means that someone or something puts data into existence. Manual text writing and automatic measurement are two methods used in this activity. *Data organization* aims to ease data search, access, processing, updating, and destruction. It structures data according to theories, such as relational data models, Ontology models, and document or file formats. *Data presentation* makes the organized data comprehensive and specifies, for instance, the layout, style, and language of information items. *Data application* means using data for a purpose. This activity covers, for instance, data analysis and the management of security, privacy, and confidentiality problems. An application of data typically results in further collection, organization, and presentation activities – thus, the notion of a cycle.

The **evolutional data quality** concept assumes that theories define data's needed characteristics. For instance, the theories used during data collection result in different needs than those used during data organization. Liu and Chi's second assumption is that data quality at earlier life cycle steps contributes to the data quality of the subsequent ones. Figure 21 illustrates it as a pyramid.

Figure 21: Evolutional data quality concept (Liu and Chi, 2002)

Each level represents the meaning of data quality based on the theories of a data-focused activity. The *vertical dimension* of the pyramid (height) illustrates that some quality aspects are based on others. Collection quality is the basement. It considers the theories used during the data collection activity. Organization quality considers the ones used during data organization, and also it has to respect the theories of data collection. This pattern repeats from bottom to top. Application quality is at the top of the pyramid and must consider all theories of previous activities. Liu and Chi argue that the *horizontal dimension* of the pyramid (width) represents the specificity of the data quality meaning. Lower levels are broader in scope and less restrictive.

Liu and Chi suggest using a diagram to visualize the causes and effects of data quality problems. Their proposal contains the four data quality concepts above as effects with one or more causes. In this dissertation, a **cause** is any phenomenon that can contribute to a data quality change. The relevance of specific causes depends on the application case.

Analysts use cause-effect diagrams to understand quality problems better. The diagrams help them to identify root causes systematically. ISO 8000-61:2016 defines root cause identification as a task of data quality improvement (ISO, 2016). Figure 22 illustrates Liu and Chi's cause-effect diagram example.

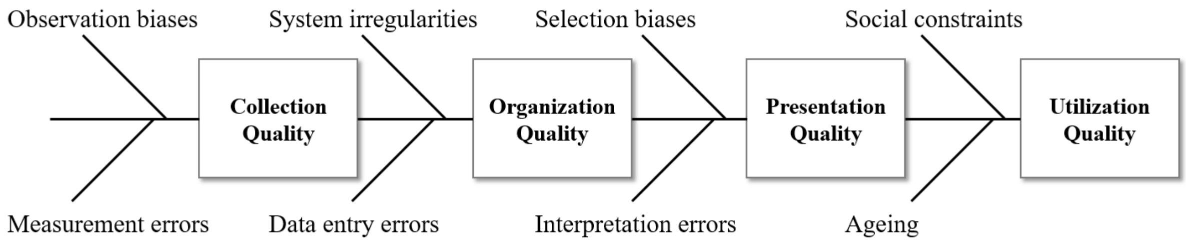

Figure 22: Cause-effect diagram for evolutional data quality (based on Liu and Chi, 2002)

This dissertation uses Liu and Chi's findings to systematically identify and visualize causes influencing PUI quality (requirement 2). It adopts Liu and Chi's data life cycle model but omits the term "cause" and uses "**factor**" instead. This change is because in Liu and Chi's diagram, the quality concepts are no effects in the narrow sense. They represent neutral concepts, and "factor" accounts for this because they can positively or negatively affect quality.

4.2.2.2 Methodology

The identification of the factors grounds on the methodology introduced in Chapter 4.2.1.2. The findings begin with a list of biases that occur in several data life cycle steps, followed by findings related to the four data life cycle steps above.

4.2.2.3 Findings

This content analysis identified more than 60 text passages describing data quality-related factors. The grouping activity identified biases and other phenomena during data creation, organization, presentation, and application. Appendices A to K list the identified text passages and the derived factors, while Figure 23 provides an overview of their structure.

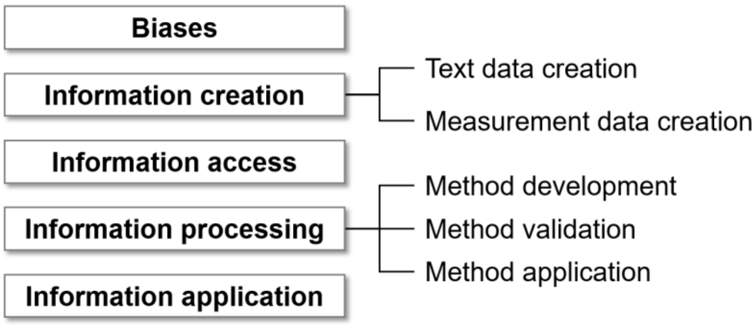

Figure 23: Overview of the structure of the identified quality-related factors

Chapter 4.2.2.3.1 begins with the identified factors that relate to biases. They are an essential factor type because they influence the accuracy of data analysis but can be difficult to identify and avoid.

4.2.2.3.1 Factors related to biases

According to the Oxford dictionary, a **bias** is a concentration on one particular area or subject. Social sciences and medical studies cover biases extensively (Higgins and Altman; Podsakoff et al., 2012). One reason is that biases in the related study design could lead to conclusions that can harm social groups or patients. This dissertation focuses on the biases mentioned by the authors of the application case descriptions. Some list potential biases, while others describe specific ones in detail. The following paragraph outlines the latter to explain how the authors encountered them.

Abrahams et al. (2015) describe several biases in the context of Social Media information and the identification of product defects. In a similar context, Aldaz et al. (2013) mention that *recall bias* affects the quality of self-reports provided by product users. The authors further mention a self-selection bias. Finch (1999) states that *non-response bias* will always be present in Internet conversations. He argues that this is not a significant problem for producers because they aim to collect as much feedback as possible rather than information about all product users. Furthermore, Finch and Luebbe (1997) mention a *sampling bias*. An example is a customer feedback sample with a biased age distribution, such as an underrepresentation of the elderly. Hajian et al. (2016) pointed out that algorithms used to analyze data can be biased. None of the investigated papers explicitly mentions this factor, but there are cues that it is relevant for PUI. For instance, Raghupathi et al. (2015) selected the validators of their proposed analysis method to contain people of different ages, genders, and business areas.

Table 11 summarizes the biases identified in the application case descriptions, while Appendix F provides an overview of the corresponding references. This dissertation concluded the biases' meanings from the context of the referenced articles.

Table 11: Identified biases related to information about product usage

Biases	Meaning in a PUI context
Self-selection	Internet users decide about their contribution to a conversation.
Non-response	Internet users that provide PUI substantially differ from users that do not provide them.
Negativity	It is more likely that product users share negative experiences.
Single source	Selection of the investigated communication channels is biased.
Common method	The measurement method's characteristics affect the findings.
Selection/exclusion	Selections regarding the sample are biased.
Product popularity	Internet users mention popular products more often.
Recall error	Individual reconstruction of episodic memories affects the accuracy of reports of experiences.
Sampling	The sampling process produced a dataset that misrepresents relevant product user groups.
Response	Very satisfied or very dissatisfied product users respond more often.
Bradley effect	In an opinion survey, "*[...] responders are unwilling to provide accurate answers, when they feel such answers may reflect unpopular attitudes or opinions [...]*" (Raghupathi et al., 2015, p. 203)
Algorithmic	An algorithm used to process PUI produces biased information.

The next chapter summarizes the factors that relate to information creation.

4.2.2.3.2 Factors related to information creation

Application case descriptions mention several factors that concern creating information about product usage. This content analysis identified factors related to human-generated (authored) text and machine-generated measurements.

Human-generated text. Aldaz et al. (2013) mention that people reporting about products via the Internet may be more *sophisticated* and product-aware than the average product user. This statement indicates that the *author's background* is a relevant factor. Similarly, Finch (1999) mentions that a product user's *motivation* may determine the value of information. He argues that very interested or motivated users might provide more valuable information than the average user. Lim and Tucker (2016) mention different *writing formats* and *constraints of communication channels*. An example of the latter is Twitter, which limits text length to 140 characters. Raghupathi et al. (2015) state that Twitter's constrained content length limits the authors to give binary answers about product satisfaction. Aldaz et al. (2013) indicated that users might *provide false information* when reporting product usage. One reason for this behavior is social desirability (Podsakoff and Organ, 1986). It means that a respondent creates a report that presents them in a good light. Finch and Luebbe (1997) described a similar factor. They argue that false information could be willfully "planted" into an Internet conversation. Also, Internet users could be motivated to provide inaccurate information in response to privacy issues. Finch and Luebbe mentioned that users could distribute false information about companies they do not like. The existence of computer programs that automatically contribute to Internet conversations indicates that some of Finch and Luebbe's thoughts became a reality (Messias et al., 2013; Ferrara et al., 2016). Raghupathi et al. (2015) mentioned that *incentives, pressure, fear, and intimidation* affect information creation. The authors do not provide details about these factors, but they claim that Social Media encourage people to express their opinions without the influence of these factors. Abrahams et al. (2015) identified that *lags* limit product defect discovery from PUI. One lag occurs between the purchase of the product and the user's observation of a defect. A second lag exists between the observation and the communication of a defect to the producer. Magniez et al. (2009) identified *reporting errors* in service reports as a factor to consider. Errors include, for instance, wrong part numbers and fault codes. Producers can outsource services to an external provider. Thiruvenkadam et al. (2008) argued that outsourcing helpdesks leads to unstructured information. They mentioned that their focus is to settle customer claims satisfactorily rather than create information that meets the needs of employees working on development-related tasks. Finch (1999) reported a technical issue with data creation. He mentions that a server change caused eleven days of *missing data* from Internet conversations.

Machine-generated measurements. Vichare et al. (2007) described a method that measures thermal loads on electronic products with sensors. The employee identifies the load parameters and sensors with the needed *accuracy*. Vichare et al.'s method includes a monitoring plan that defines the measurement frequency and the setting of thresholds. The *frequency* determines how often the measurement process stores load data. A higher frequency than needed can result in irrelevant data points. *Thresholds* can increase the efficiency of data creation because the measurement only stores data that pass the threshold. Bleda et al. (2012) state that producers should *integrate sensors ubiquitously* to provide realistic PUI. Ubiquitous means that the product's user does not recognize the sensor (similar Weiser, 1999).

Magniez et al. (2009) identified a factor that does not belong to the factor classes above. They mentioned that *oversimplification* and a lack of accuracy in customer complaints could render PUI useless for a root cause analysis. They refer to fault codes that represent symptoms rather than causes.

Table 12: Identified factors related to information creation

Factors	Meaning in a PUI context
Author's background	An author's skills and experience
Author's motivation	An author's reasons to create information
Writing format	The structure that determines the content of text information, e.g., comment, answer, or report
Constraints of the communication channel	Characteristics that determine how a person can create and share information via a channel
Provide false information	An activity whose purpose is to create inaccurate information
Incentives	A thing that motivates or encourages someone to create, update, or delete information (based on Oxford University Press, 2017)
Pressure	The use of persuasion or intimidation to make someone create, update, or delete information (based on Oxford University Press, 2017)
Fear	An unpleasant emotion caused by the threat of danger, pain, or harm (Oxford University Press, 2017)
Intimidation	Action to frighten or overawe (someone), especially in order to make them do what one wants (Oxford University Press, 2017)
Lags	The time between an event and the creation of related information The time between creating information and communicating it
Reporting errors	Inaccurate information created during reporting
Outsourcing	Involving a third party to create and share information
Missing data	Information not created as planned or expected
Accuracy of load parameters	The measure of the match between identified load parameters and the actual load situation
Accuracy of sensors	A quantitative measure of the magnitude of error (ISO, 2017)
Measurement frequency	A measurement characteristic that influences how often the measurement system creates information
Measurement thresholds	A measurement characteristic that influences which information the measurement system creates
Integrate sensors ubiquitously	A measurement characteristic that influences the product user's awareness of a measurement system
Oversimplification	Simplification of something to such an extent that it gives a distorted impression (Oxford University Press, 2017)

Chapter 4.2.2.3.3 summarizes the identified factors that relate to information access.

4.2.2.3.3 Factors related to information access

Finch and Luebbe (1997) mentioned that *control over an information channel* is essential. In an uncontrolled channel, i.e., one not maintained by the producer, the information user may not know an information item's author. In controlled channels, the producer may have metadata about the authors. Magniez et al. (2009) found that development-related employees are *unable to verify the accuracy* of information because other departments or organizations typically provide it. Shin et al. (2015) mentioned that for some products, access to information is only possible if the owner or user requests a maintenance service. One reason for this limitation is that a product has no *functionality to send data remotely*. The in-situ retrieval of data is necessary, for instance, during maintenance. Abrahams et al. (2013) pointed out that they asked for permission from Internet forum owners to evaluate their PUI analysis approach. Getting *permission to collect raw data* is particularly relevant for Internet channels and personal data. The latter is highly critical because of the General Data Protection Regulation (GDPR) of the European Union (European Commission, 2018). The GDPR creates several obligations for producers, including providing transparent information on why it wants to create and process personal data. Finally, Igba et al. (2013) pointed out that service providers must be *willing to share PUI* with the product's producer.

Table 13: Identified factors related to information access

Factors	**Descriptions**
Information channel identification	Identify the most suitable information channels for the focused application case
Information channel control	Employee's possibilities to access information of a channel
Verifiability of information accuracy	Employee's capability to assess the information's accuracy
Remote data access	Capability to read the data from a distance
Permission to collect raw data	Laws may require the data owner's permission before someone or something collects them
Stakeholder's willingness to share PUI	Stakeholders must see value in sharing information they own

Chapter 4.2.2.3.4 presents the findings related to information processing. The latter transforms input information into even more helpful information.

4.2.2.3.4 Factors related to information processing

This content analysis identified factors related to developing, validating, and applying information processing methods.

<u>Development</u>. The investigated articles typically describe custom-made methods for information processing as a support for one or more development-related activities. This dissertation identified two method categories. First are methods that

use *machine learning* (ML), which means that a computer program can learn to perform a task with increasing accuracy. ML bases on techniques that often apply statistics (refer to Kubat, 2015). One ML approach uses manually selected information to train the computer program. Second, methods that use *hard-coded explicit knowledge* about product-related phenomena. It consists of mathematical formulas written into the software code of the analysis.

- **Machine learning**. Some machine learning techniques rely on a training dataset that determines how the computer program will process other data. Consequently, this training dataset's characteristics affect the information characteristics after the processing. Abrahams et al. (2013) described a methodology to identify vehicle defects in online discussion forum posts. Paid domain experts evaluated posts manually to identify text passages related to vehicle problems – these served as training data. The *evaluators' selection influences this dataset's characteristics* and the analysis result. One reason is that the evaluation grounds on the expert's interpretation of the content. A specified *content evaluation procedure* can mitigate this effect. For this purpose, Abrahams et al. (2015) used a classification schema for safety issues in the automotive industry. Another factor related to machine learning is the *quantity of training data*. Abramovici et al. (2011) identified this factor during the training of a Bayesian Network for product fault analysis. Abrahams et al. (2013) estimated the size of the training dataset by defining a margin-of-error and a confidence level.
- **Hard-coded explicit knowledge**. Information processing can rely on one or more measures that describe a product behavior or a product-related phenomenon (Deng et al., 2021). For instance, Shin et al. (2015) focused on degradation, a common phenomenon for durable products. They used the remaining lifetime of the product to represent degradation. A mathematical model calculated this indicator through parameters, such as the product's usage time. The *selection of the indicator* and the *mathematical function* determine the processing and characteristics of the resulting information.

<u>Validation</u>. The validation of processing methods is important because it ensures that the method provides accurate results. ISO 9000:2015 defines **validation** as the *"confirmation through the provision of objective evidence [...], that the requirements [...] for specific intended use or application have been fulfilled"*. In the scientific domain, the meaning of validation is more sophisticated. Mitchell and Jolley (2013) differentiate, for instance, internal, construct, and external validity. This dissertation does not focus on these sub-concepts because the case descriptions cover validity on a more general level. Table 14 summarizes application cases that mention method validation.

Table 14: Application cases describing method validation related to information processing

No.	References	Proposed methods[27]
1	(Raghupathi et al., 2015)	Automatic analysis of product reviews with sentiment rating
2	(Tucker and Kim, 2011)	Forecasting product feature trends through a predictive model
3	(Abrahams et al., 2015)	Multivariate explanatory analysis relating defects to quantitative measures derived from text
4	(Ghosh et al., 2016)	Joint analysis of product user survey and embedded sensor data
5	(Stone and Choi, 2013)	Automated analysis of product reviews with sentiment classification
6	(Lim and Tucker, 2016)	Product feature extraction from microblog content

This dissertation identified the following validation approaches within the application case descriptions:

- **Estimated data vs. other data**. Stone and Choi (2013) compared their estimates for Smartphone users' sentiment with Amazon's product rank. The authors selected the product rank because they considered it an approximation of sales data. They had no access to actual sales data.
- **Predicted data vs. observed data**. Tucker and Kim (2011) used a mathematical formula to validate the accuracy of their proposed forecasting method. They quantified the difference between the predicted and the observed data.
- **Automated method vs. manual method**. Abrahams et al. (2015) compared the proposed automated application of their method with a manual application. Its objective is to confirm that automation is more accurate and faster in identifying relevant content than manual analysis by experts.
- **Method vs. method**. Ghosh et al. (2016) validated the proposed method through a Confirmatory Factor Analysis. They used eight statistical tests, such as the Chi-squared test, to compare their sensor-enhanced method's performance with a method that entirely relies on user perceptions. Abrahams et al. (2015) applied Multivariate Logistic Regression to compare three method variants.

The approaches above validate specific aspects of processing methods. First, they ensure that the method provides the information the analyst needs. This goal is similar to the meaning of accuracy used in ISO 25012:2008. The remainder of this dissertation will refer to it as *accuracy validation*. Second, the approaches ensure

[27] Not all references suggest a name for their developed method.

that the method is the most suitable from a selection of similar methods. This manuscript refers to this aspect as *best-fit validation*.

The *accuracy validation procedure* affects the characteristics of the resulting information. Raghupathi et al. (2015) pointed out that they selected their validation participants to reflect different gender, ages, and business areas. This decision indicates that the *selection of the validators* affects the validation result.

Application. The following list describes the factors related to applying an information processing method.

- Raghupathi et al. (2015) mentioned the *robustness of the analysis method* as a factor. Robustness means the degree to which the method can function if it receives invalid input data (similar Institute of Electrical and Electronics Engineers, 2008). Examples are datasets with missing data and outliers.
- Raghupathi et al. (2015) and Ghosh et al. (2016) stated that data pre-processing removes the *noise* within raw data. Lim and Tucker (2016) argued that different writing formats or constraints of online media platforms create noise. It can cause false-positive and false-negative errors.
- Several authors mentioned that text characteristics affect data processing methods. Amongst them are the *language* (Finch, 1999), *language structure* (Lim and Tucker, 2016), *acronyms* (Raghupathi et al., 2015), *semantic defects* and *syntactic defects* (Abramovici and Lindner, 2011), *term dependence*, *term disambiguation,* and *term distribution* (Lim and Tucker, 2016), *sarcasm and ironic expressions* (Raghupathi et al., 2015), and the *content length* (Lim and Tucker, 2016). Appendix G explains these factors briefly.
- Finch (1999) raised an argument against revealing information sources. He mentions that a revealed source, e.g., a user, may stop providing information. Finch wrote this statement in a section named *Confidentiality*. This dissertation adopts the same name for this factor.
- Aldaz et al. (2013) mentioned that a *software bug* affected their PUI analysis. It prevented their method from recognizing a specific product usage scenario.

Chapter 4.2.2.3.5 describes the identified factors that emerge when the employee applies the needed information to solve a problem.

4.2.2.3.5 Factors related to information application

Igba et al. (2013) and Kuestner and Wartzack (2015) mentioned that developers might not be *aware of information* about previous products. This dissertation assumes it is valid for information about product usage as well. The authors above further pointed out that the *availability of context information* is important for the information user because the context affects the meaning of information. Igba et al. (2013) stated that this is especially relevant for long-lived products where employees may use service information after decades. Magniez et al. (2009) indicated that information about a product's environment and user profiles might be relevant to verifying field failure information. The *amount of data* affects the statistical validity of the conclusions drawn from information about product usage. Thiruvenkadam et al. (2008) reported this in the context of customer feedback.

4.2.2.4 Conclusion

The content analysis of the application case descriptions identified factors in five areas. They include a variety of phenomena associated with potential root causes of information quality problems in the context of PUI. Application cases describe the factors sometimes extensively, and sometimes, they list them without further details. The latter case makes the interpretation of factors complex and can result in inaccurate or superficial definitions. This dissertation uses the findings of this second content analysis to develop an initial list of factors. Despite its preliminary character, it could support the IM expert in identifying where a compromise between the needed information and the potentially provided one is inevitable. For instance, PUI extracted from online discussion forums could be available to employees but with a warning that it likely contains inaccurate information.

4.3 Procedure and Syntax

This chapter uses the findings of Chapter 4.2 to develop the method's procedure and syntax. The procedure's basic structure grounds on this **assumption**:

> *The IM experts design a useful IS by (1) defining the characteristics of the needed information and by (2) considering the factors that influence them.*

The IM methodologies outlined and concluded in Chapter 3 back the first part of this assumption (Voß and Gutenschwager, 2001; Krcmar, 2015). ISO's data quality management framework and its embedded continuous improvement process substantiate the second part (ISO, 2008, 2016). This chapter transfers the main assumption above to PUI.

The basis for the procedure and syntax is Liu and Chi's cause-effect diagram variant for information quality. *Chapter 4.3.1* describes it along with the adaptations made in this dissertation. *Chapters 4.3.2* and *4.3.3* integrate the theoretical foundation above into the cause-effect diagram.

4.3.1 Diagram Structure

A cause-effect diagram is a typical quality management method to analyze potential causes of quality-related problems. It is an acknowledged method in practice, allowing the analyst to adapt it to the situation at hand. Liu and Chi (2002) substantiated its relevance in information quality management. This dissertation focuses on the cause-effect diagram for these reasons.

The diagram structure has the shape of a fishbone with a central bone representing the investigated quality-related effect. Diagonal bones illustrate causes that contribute to it. This structure repeats for sub-causes until the analyst decides that the level of detail is sufficient for the analysis task. Figure 24 summarizes the **basic structure** of a cause-effect diagram.

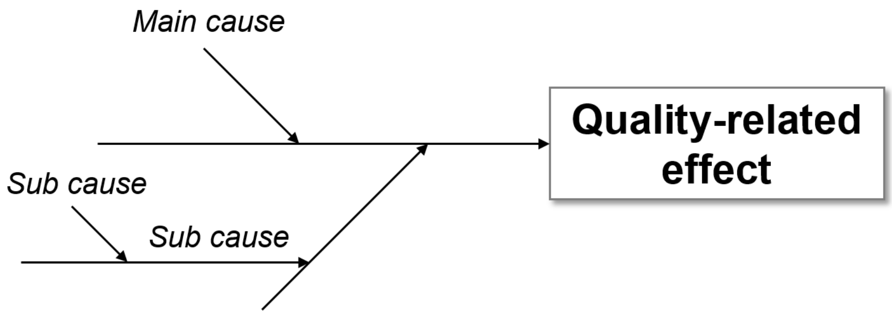

Figure 24: Cause-effect diagram structure as used in quality management

The method uses a neutral version of this cause-effect diagram that focuses on cause-effect relations but interprets them from the perspective of a quality concept rather than a quality problem.

Quality management literature suggests **categories** to structure the causes. One acknowledged category system is the 6M structure used by the Six Sigma quality management approach. It proposes Man (people), Machine, Material, Method, Measurement, and Mother Nature (environment) as categories (Graves, 2013). The IM expert can use them as a starting point for the analysis or change them as needed (similar to Liu and Chi, 2002).

This dissertation identified two purposes for using categories in a cause-effect diagram. First, they support the analyst in *comprehending complex diagrams*. Complexity means that the diagram contains many causes and several layers for sub-causes. Categories structure the causes, such as the 6Ms proposed by Six Sigma. Second, they *guide the analyst* during a root-cause analysis. The analyst can systematically check each category for causes and ensure that the root-cause analysis misses no important domain. This manuscript focuses on categories as a guide for IM experts.

Some category systems, such as the 6Ms, could be a basis for an adapted cause-effect diagram. Transferring one of these category systems to PUI quality bears the risk that the categories need extensive changes (e.g., machine and material) and may become less intelligible. For this reason, this manuscript directly develops categories from the application case descriptions.

Liu and Chi (2002) suggest chaining cause-effect diagrams to account for the data life cycle. As described in Chapter 4.2.2.1, this approach integrates a life cycle model into the cause-effect analysis. Figure 25 illustrates this dissertation's cause-effect diagram structure for information quality.

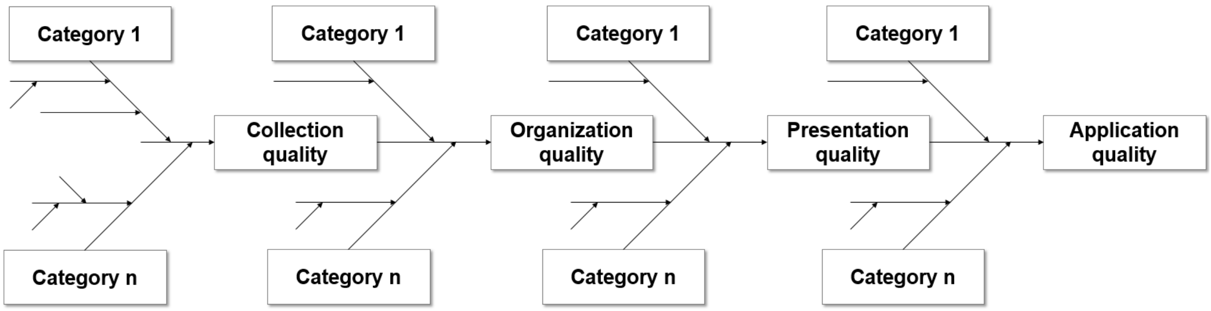

Figure 25: Structure of an adapted cause-effect diagram for information quality

Chapter 4.3.2 describes integrating the identified PUI characteristics into the cause-effect diagram. It develops categories from the findings of Chapter 4.2.

4.3.2 Integration of the identified characteristics

The starting point of this chapter is a statement in the ISO 25012 standard. It suggests that data quality characteristics can categorize data quality requirements and data quality measures (ISO, 2008, p. 1). This dissertation adopts the ISO approach. It develops a categorization system for factors based on the identified PUI characteristics. The arguments for this approach are these:

- The characteristics divide information quality into elements. This approach ensures that diagram users do not omit potentially relevant aspects.
- It is not clear which factors create PUI quality problems frequently. Defining key factors, such as human, measurement, or method, may introduce a bias to the diagram that over- or under-values these factors' actual importance.
- The diagram user can track factors for specific characteristics across the data life cycle. This feature could provide insight into where factors might neutralize or amplify each other.

The following paragraphs describe the design decisions for the adapted cause-effect diagrams.

The relevance of the identified information characteristics

A first design decision is which of the identified information characteristics the cause-effect diagram should include. The findings indicate that application case descriptions mention ten information characteristics described in the ISO 25012 standard. They are *accessibility, accuracy, availability, completeness, confidentiality, credibility, currentness, efficiency, precision, and understandability*. This dissertation focuses on these because they are less ambiguous and complex than relevance, richness, and value.

Identified characteristics' relevance concerning the four quality concepts

A second decision is how the relevant characteristics relate to the four data quality concepts. One approach considers them potentially relevant for all concepts; the other is to assess their relevance per data quality concept. The first approach assumes that the IM expert decides upon their relevance while developing a cause-effect diagram for a specific PUI application case. The second approach means that this dissertation must assess and decide upon the relations.

This dissertation applies the first approach because the IM expert will likely know the envisaged IS design task and the PUI application case. A pre-selection in this manuscript could result in unnecessary constraints and limitations of the method's applicability.

Extension of the diagram structure with the categories

Figure 26 illustrates the integration of information characteristics as categories. It focuses on collection quality to demonstrate this extension. The categories' distribution is alphabetical. Their positioning above or below the central bone means each category can contribute to collection quality. It does not imply a systematic (Chapter 4.3.3 revisits this aspect).

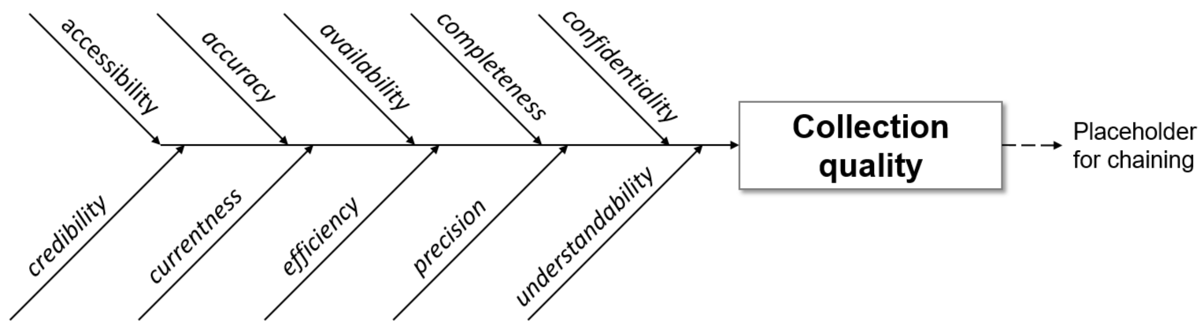

Figure 26: Cause-effect diagram structure with information characteristics as categories

The IM expert could use this cause-effect diagram to systematically identify factors related to information characteristics. Its scope is broad, and the IM expert could use it for IS design in many domains.

Chapter 4.3.3 extends the diagram structure further by integrating the identified factors of PUI quality. This step scopes this method to designing IS providing PUI for development-related work tasks.

4.3.3 Integration of the identified factors

This chapter begins with a revision of the categories for the cause-effect diagram. It argues how the identified categories and the identified factors relate.

Extension of the diagram to align it with ISO 9000's quality definition

ISO 9000 defines quality as the "*degree to which a set of inherent characteristics of an object fulfills requirements*" (ISO, 2015a). An object can be an information item, which makes this definition applicable to PUI. The ISO 9000 quality definition implies two roles for information characteristics:

- Characteristics specify what an information item (e.g., comment) is like.
- Requirements use characteristics to specify what stakeholders need.

This dissertation derives two categories of quality-related factors from them:

- Factors changing the information characteristics of information items
- Factors changing the requirements of stakeholders regarding these items

Both factor categories result in information quality changes. An example of the first category is a user's ambiguous information about product usage published during an Internet conversation. The factor is user behavior that affects the interpretability of the information item. An example of the second category is an employee who needs more precise information than the available one. In this case, the characteristics of the existing information items remain the same.

The cause-effect diagram for PUI quality should consider both factor categories because the IM expert has to address them differently in an IS design. An IS could, for instance, mitigate factors that change the characteristics of an information item through suitable measurement practices, information processing, or presentation techniques. The IM expert could mitigate requirements' changes by spending sufficient time with the stakeholders during the requirements elicitation phase of an IS design. Figure 27 illustrates the cause-effect diagram structure for collection quality with the factor categories derived from ISO 9000.

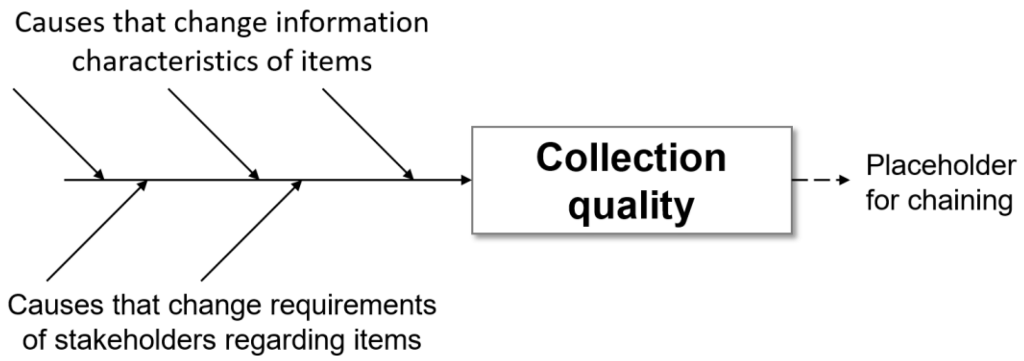

Figure 27: Cause-effect diagram structure with factor categories derived from ISO 9000

The diagram visualizes factors above the central bone if they change the characteristics of needed information. Factors that relate to requirement changes are below it. These categories are complementary to the ones used in Chapter 4.3.2.

This dissertation focuses on the factors that change information characteristics. The content analysis results did not allow further differentiation without introducing additional subjectivity.

Identified factors' relevance concerning the four quality concepts

One design decision is how relevant the identified factors are for Liu and Chi's four quality concepts. The findings from Chapter 4.2.2.3 indicate that quality-related factors belong to one or more activities in the data life cycle. This aspect means that the cause-effect diagram should assess a factor's relevance per quality concept.

Chapter 4.2.2.3 identified text passages that provide cues about activities related to factors. Table 15 summarizes these findings concerning the four quality concepts. Appendices A to K contain the respective text phrases from the application case descriptions.

Table 15: Summary of the identified factors concerning the four quality concepts

Quality concepts	**Information quality-related factors**
Collection	*Biases*: self-selection, non-response, negativity, single source, common method, selection/exclusion, product popularity, recall error, response bias, sampling, Bradley effect *Channel*: allowed message length, information channel control, directedness, free text, self-report design *Content*: depth of discussion, diversity of discussion, language, mentioning of a specific product, oversimplification, presence of acronyms, reporting errors, semantic defects, syntactic defects, typographical errors

	Information creation context: accuracy of sensors, placement of sensors, unobtrusive monitoring *Measurement frequency*: measurement frequency, defect detection lags, reporting lag, speed of data collection, the time frame of data *Missing data*: hardware problem, software bug *No sub-category*: validity of information, customer sophistication, availability of context information, providing disinformation, duplicates
Organization	*No sub-category*: Data format, data reduction
Presentation	*No sub-category*: Condensation of information, presentation format, prioritization
Application	*Content*: customer context information, discussion length, information with privacy-related constraints *Information access context*: access permission, commercial factors, community membership, cultural factors, political factors, willingness to share data *Machine learning*: domain expert, number of classes, tagging rules *Method*: manual classification by internet users, method sensitivity, robustness of algorithm, sample size, speed of algorithm, unit of analysis, validation, language structure, false negative error, false positive error *No sub-category*: access during a service event, availability of context information, noise, use of secondary sources, source/author exposure, employees' awareness of data existence, right format for use

Most of the identified factors relate to collection quality and application quality. I defined sub-categories to group similar factors. The sub-categories provide a systematic visualization of the cause-effect diagrams and make the diagrams more comprehensible.

Identified factors' relevance concerning the information characteristics

The application case descriptions sometimes indicate which quality characteristic a factor may change. These text passages are important for this dissertation because they provide cues for a factor's relevance concerning the mentioned characteristics. In other cases, this relation is not apparent. I concluded it from the respective text passage, the application case, and the quality characteristics' definitions provided in Chapter 4.2.1.3. Appendix D provides details about these text passages and the assigned characteristics.

Figure 28 illustrates the resulting cause-effect diagram for collection quality. Its main categories are the information quality characteristics, and Table 15 provides sub-categories to structure it further. If a diagram's sub-category contains only one element (e.g., information channel control), the diagram omits this category in favor of a reduced diagram complexity.

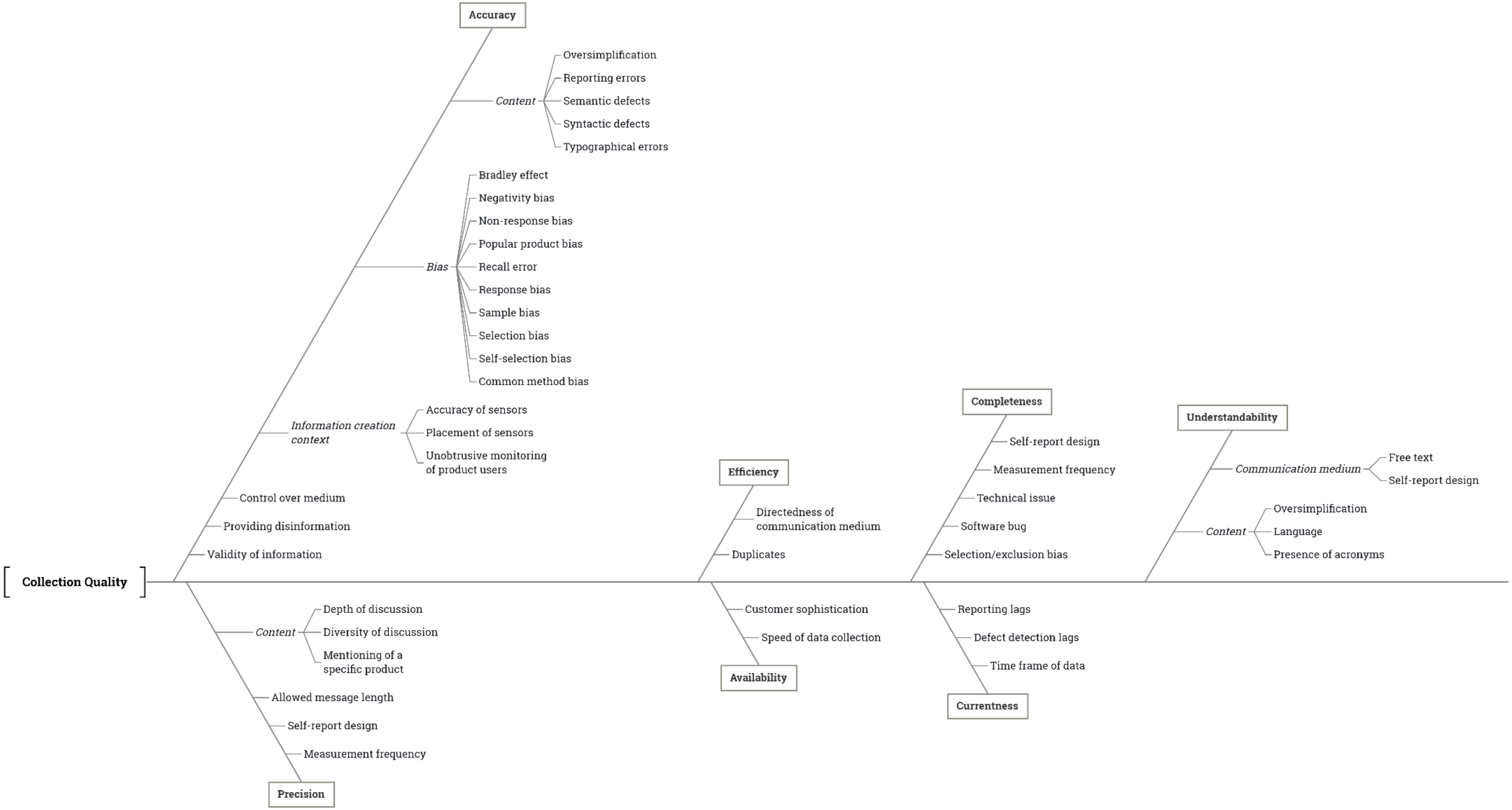

Figure 28: Cause-effect diagram for the collection quality of PUI

4.4 Application

The previously developed diagrams aim to support IM experts during IS design. They contain potential factors of information quality changes with a focus on PUI. Klesse et al. (2004) apply cause-effect diagrams with a similar purpose but a different application domain. They designed an IS that supports a bank's customer investigation process. The authors describe no general procedure to apply the cause-effect diagrams. In addition, this dissertation's literature survey did not identify references that explain how an IS design process could integrate cause-effect diagrams. For this reason, this chapter outlines assumptions and conditions for using the diagrams above and a procedure to apply them in IS design.

4.4.1 Assumptions and conditions

The method's application assumes that one or more IM experts received the task of designing an IS that provides PUI for employees working on development-related tasks. They *identified the application case* where PUI should support product improvement. Deng et al. (2019) identified several application cases in the literature that can inspire experts to identify a use case. Besides, Wilberg (2019) suggests a procedure to design a PUI strategy for connected products. It contains a step to identify application areas and use cases.

The experts should apply a *collaborative design* approach, where they frequently share information with the involved stakeholders. This approach is essential because the developed diagrams do not contain complete lists of all potential or relevant factors. If the IS's users are unavailable, the IM experts should involve domain experts who know user expectations and requirements.

All methodologies presented in Chapter 3.3.2 suggest involving multiple stakeholders. A *workshop* is appropriate for integrating users and domain experts in the IS design process. Other participants should be employees working with PUI, e.g., IT, sales, customer service, and after-sales service.

4.4.2 Application procedure

This dissertation adapts the need assessment procedure of Watkins et al. (2012) to ground the application procedure. The developed method adapts it to provide the structure for a workshop with IM experts, system users, domain experts, and others working with PUI. Table 16 summarizes the original and the adapted procedure.

Table 16: Application procedure

No.	Watkins et al.'s procedure	No.	Adapted procedure
1	Identify needs	1	Identify the needed information and its characteristics.
2	Analyze needs	2.1	Identify relevant factors amongst the ones mentioned in the diagrams.
		2.2	The users can define custom categories and factors.
3	Decide what to do next	3	Identify how the IS design should manage the relevant factors.

The adapted procedure has a smaller scope than the original to account for this dissertation's focus on product-related information and its quality.

4.5 Tool development

The developed cause-effect diagrams contain many findings of this dissertation, which can cause information overload for the persons using them. Eppler and Mengis (2004) identified over 50 measures to counter information overload. They structure them through the following categories: personal factors, information characteristics, task and process parameters, organizational design, and information technology application. The latter category fits best for this dissertation. It suggests offering information organization options and decision support systems to reduce a large set of information into a manageable size. This dissertation develops a tool to realize these measures. A **tool**, in this context, is a software system that supports the method application (Mayer et al., 1995, p. xi).

Chapter 4.5.1 identifies what the tool has to provide, and *Chapter 4.5.2* explains how the tool addresses these requirements.

4.5.1 Requirements

The following paragraphs describe the tool's requirements. They are "should" requirements, and each description contains at least one argument for justification.

R1: The user should be able to access the software easily

A workshop for information need analysis will most likely occur where the development department is. Therefore, the tool must work anywhere with Internet access on a business notebook.

R2: The user should be able to see the cause-effect relations

Visualization is an approach to avoiding information overload (Eppler and Mengis, 2004). The cause-effect diagrams are a visualization already, and the user should be able to see them in the tool.

R3: The user should be able to customize the cause-effect diagrams

This dissertation provides an initial list of cause-effect diagrams. The user should be able to add or remove items from them quickly and without specific expertise.

R4: The user should be able to receive relevant factors

This dissertation developed the cause-effect diagrams from multiple application case descriptions. They may contain information that is not relevant from the perspective of specific development-related work tasks. The users should receive relevant factors to avoid confusion and information overload during a workshop.

R5: The user should be able to discover relevant factors by accident

The software might not be able to identify all relevant factors because a filtering mechanism can be, for instance, biased, too narrow, or too broad. Therefore, the tool should allow users to discover relevant factors (serendipity).

R6: The user should be able to track the method's application progress

If the tool identifies many relevant factors, it can create a situation where the users lose track of the factors they covered. They should receive information on which factors they have covered already.

R7: The user should be able to receive a description of the factors

The users may not know the meaning of each factor, which limits the method's applicability. Therefore, the user should be able to receive information that explains a factor's meaning.

4.5.2 Realization

The tool's design addresses one or more of the previously identified requirements with a solution characteristic. A bracket [**R1**] indicates the relation to the specific requirement. The tool's name is *QualiExplore*.

QualiExplore is Internet-based and accessible as a website via the browser. This feature avoids installing software clients, which are more challenging to update and might not be allowed on the target computer [**R1**]. The tool visualizes the cause-effect diagrams through a tree structure [**R2**]. It stores the cause-effect relations in a text file in JSON format – a standard format for exchanging information via the Internet. The user can change the text file to edit the cause-effect information [**R3**]. I decided against using a database to simplify the software's realization. The text file contains the tree structure with one root and several branches. The four main branches represent the data quality concepts presented in Chapter 4.3, each ending in one or more quality-related factors as leaves. To control the visualization's complexity, the user can collapse or expand a branch to hide or reveal its leaves and sub-branches. Figure 29 illustrates the file structure and the corresponding visualization.

Figure 29: JSON file structure (left) and tree visualization (right)

A filter mechanism supports the user in identifying relevant factors for the targeted application case. The user has to adjust the filter and then proceed to the next step to visualize the matching factors. Figure 30 illustrates the filter function.

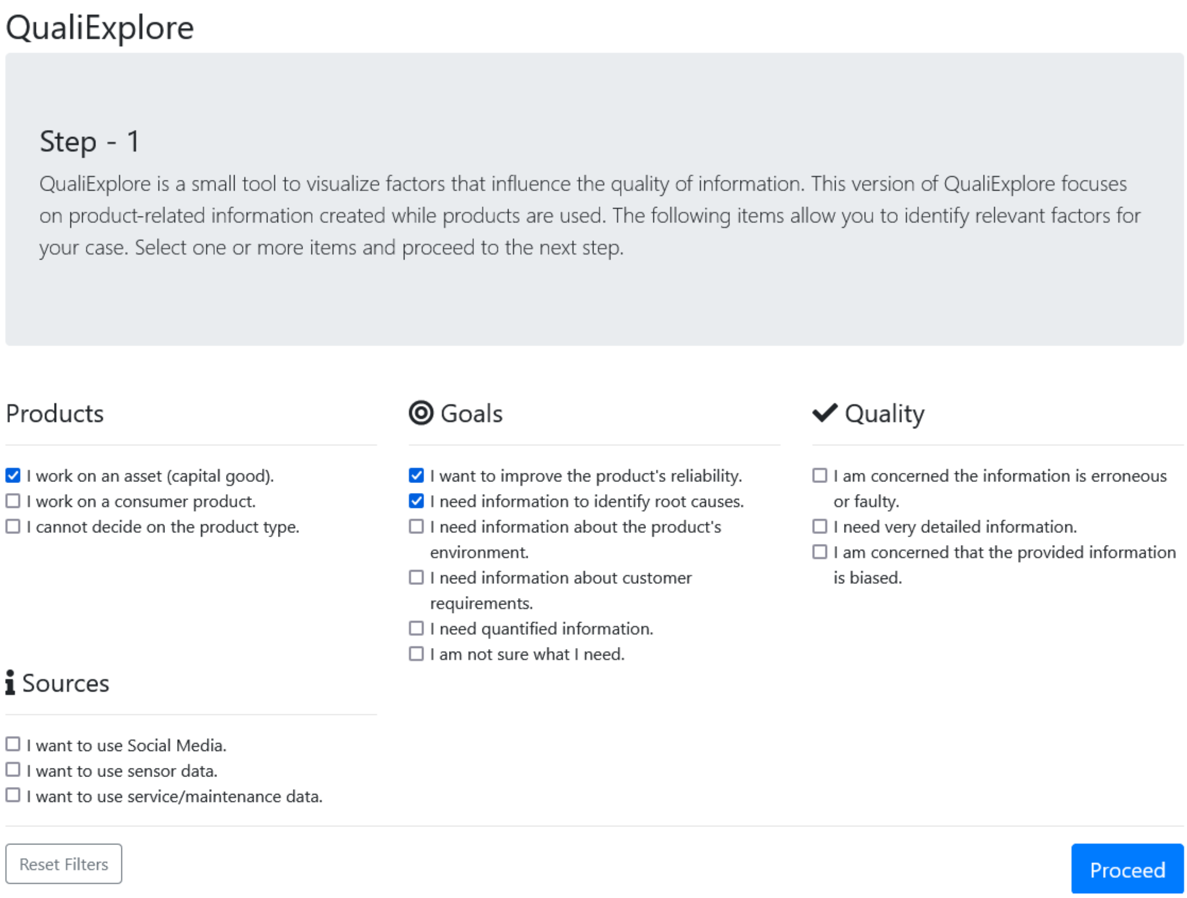

Figure 30: Filter function for cause-effect relations in QualiExplore

Short statements represent the application case characteristics and refer to the product, the PUI users' potential goals, information quality, and sources[28]. The user can tick the boxes of a statement to indicate that the statement is relevant to the targeted case. I assigned the identified quality-related factors to the statements in the JSON file above. A second JSON file contains the statements. In the final structure, each factor references one or more statements.

Once the user proceeds to the next step, the tool visualizes the factors that match the filter [**R4**]. It shows all available factors [**R5**] but highlights the relevant ones with a flag symbol. This decision is a compromise between the reduction of information overload and the beneficial effects of serendipity. Figure 31 illustrates the visualization of the quality-related factors.

[28] This term has to be understandable for tool users and therefore avoids using "channels".

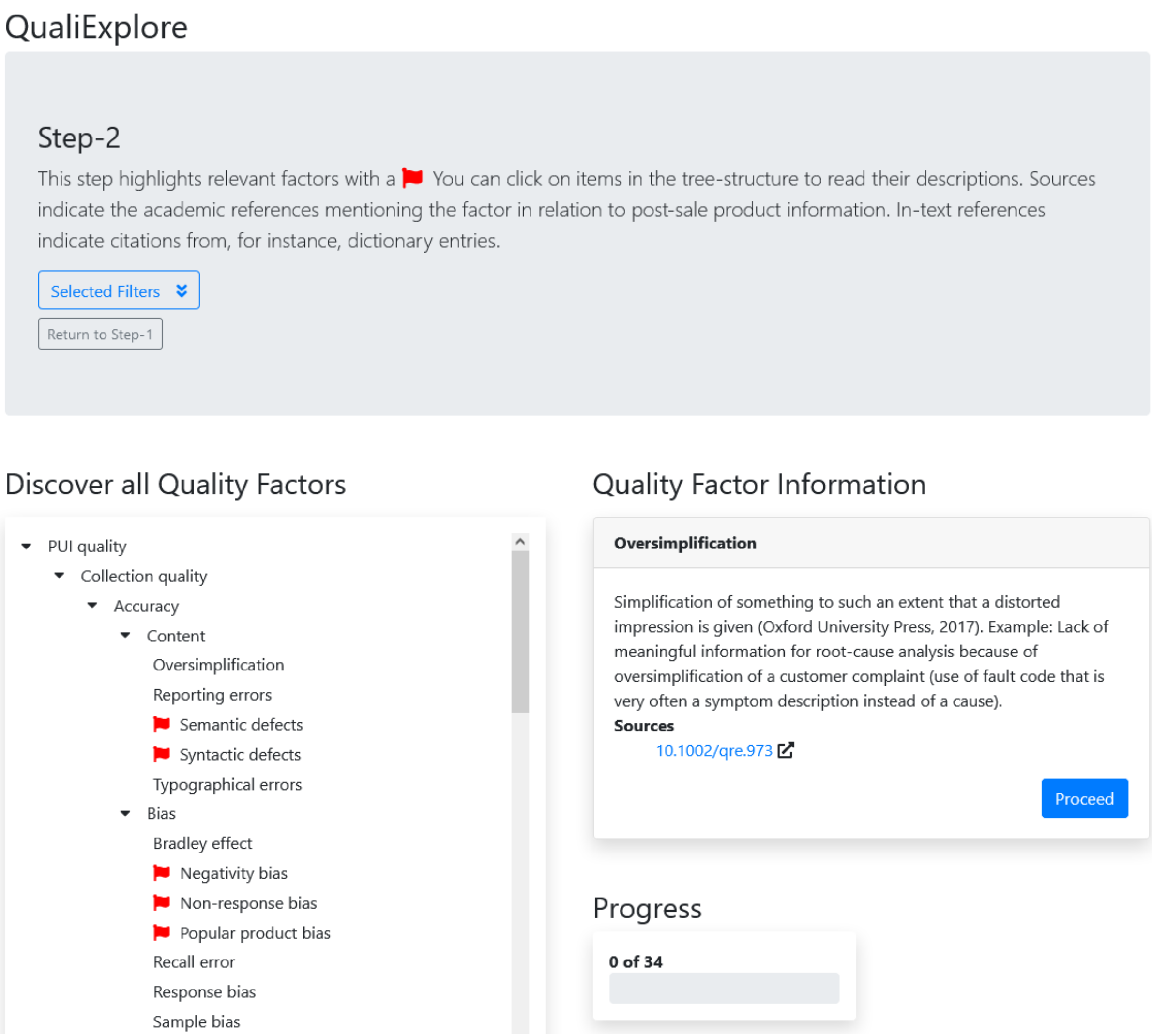

Figure 31: Visualization of the cause-effect relations in QualiExplore

The user can confirm through a button that he or she has read the factor description. Confirmation changes the flag's color from red to green, and a progress bar indicates how many factors the user confirmed – this represents the method's application progress [**R6**]. Users can access a factor's description by clicking on the factor in the tree structure [**R7**]. The software illustrates related information in a textbox containing a definition or a description of an example. If a definition was available in the application case descriptions, I preferred it over an example. A link to related references allows the user to check the source of a definition or to seek context information about the factor.

4.6 Conclusion

The cause-effect diagrams above base on application case descriptions. This approach grounds the diagrams on the practical experience of researchers from the product development domain. The ISO 25012:2008 standard and the application cases provide the basis for the identified categories. They structure potential quality-related factors and provide guidance in which areas users can search for additional ones. The identified quality-related factors provide a starting point for IM experts to design an IS that satisfies the employees' PUI needs. The persons who use the developed tool should extend the filters, categories, and factors over time. This content will concretize the hurdles and pitfalls of applying PUI in product development.

Chapter 5 demonstrates the developed method and the tool and serves as the basis for the evaluation and the following discussion. The latter will revisit the method and tool requirements and their fulfillment.

5 Method demonstration

The next step in this dissertation's approach is demonstrating how the developed method addresses the focused problem. This manuscript uses case studies for this purpose. **Case studies** are a qualitative research method focusing on examining individuals and groups facing specific challenges (Lazar et al., 2017, 7). One of their goals is to demonstrate the successful use of new software tools. In contrast to quantitative research methods, case studies focus on investigation depth rather than a representative sample. Case studies are the most suitable method for this dissertation because the number of accessible cases where organizations use or want to use PUI in product development is comparably low. *Chapter 5.1* introduces the use cases and the related companies, while *Chapter 5.2* covers the workshops to apply the developed method and the QualiExplore tool in these use cases.

5.1 Use cases

This chapter summarizes the use cases with an outline, a paragraph focusing on the potential benefits of the related usage information, the targeted user group, and a description of the needed content. Table 17 summarizes the cases' main characteristics.

Table 17: Summary of the use cases

Characteristics	**Use case 1**	**Use case 2**
Company role	Software service provider	Producer & information user
Product type	Automobile component	Home appliances
Information system exists	Yes	No
Information sources	Human and machine	Human and machine
Communication channels	Online discussion forums, weblogs, public administration databases, internal recall databases, warranty case databases	Statistical quality control spreadsheets, spreadsheets with identified product defects, product quality test reports, consumer service*
Main benefits of usage information	Fulfill legal requirements, early warning system, detect unjustified warranty claims, benchmarking	Increase production process quality, increase product quality

* Channel is available but has not yet been exploited. Its usage is considered a long-term goal.

Each of the following cases represents a specific archetypical user of the developed method. Case one focuses on users with an existing IS collecting, organizing, and presenting usage information to clients. Case two covers producers that want to have such an IS to improve their production process quality and, as a result, the product quality. Figure 32 illustrates the positioning of the above-use cases along the product life cycle presented in Chapter 2.2.1.

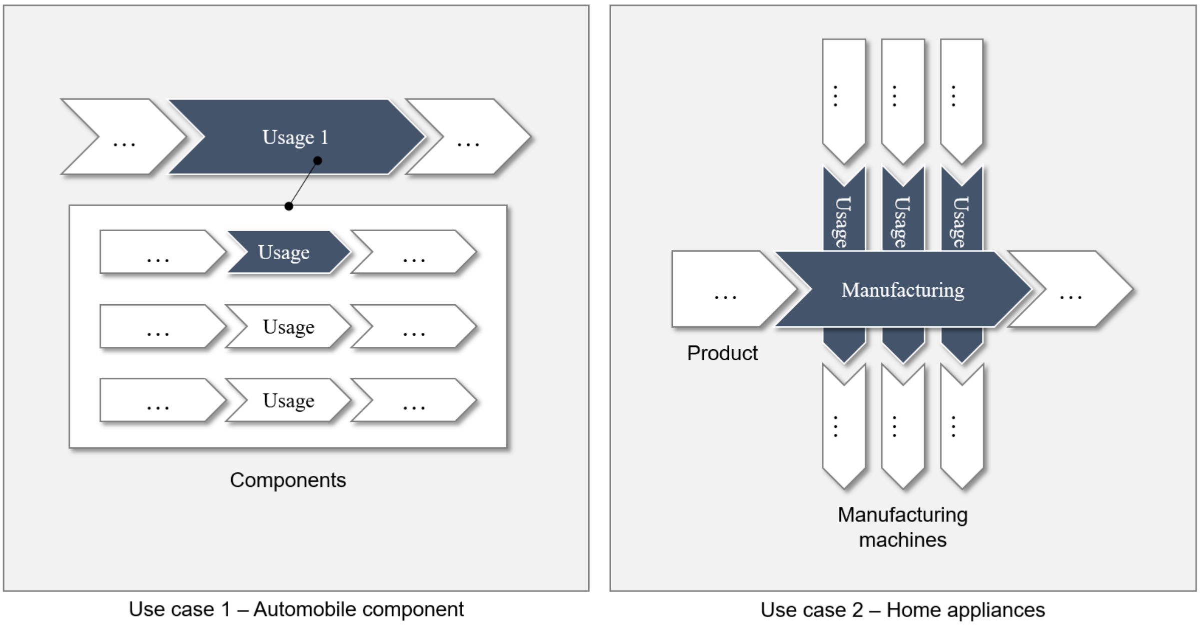

Figure 32: Positioning of the use cases in the product life cycle

Use case one covers the usage phase of a specific component in a product. In contrast, use case two focuses on a product's manufacturing phase and the usage of several production machines. This dissertation does not cover the usage of machines during the remanufacturing phase because no suitable use case was accessible.

5.1.1 Case 1 – Extending an existing information system

Company 1 is a software company specialized in data analytics and information management. It operates the Internet service "Carwatch.net" (Pumacy GmbH, 2021), which provides information about car recalls, investigations, and complaints. Carwatch extracts information periodically from different channels, organizes this data, and presents the results in different visual formats, such as tables and graphs. *Information channels* include the National Highway Traffic Safety Administration (NHTSA) database (U.S. Department of Transportation, 2017), the client's recall databases, the client's warranty datasets, and public online discussion forums. The main *benefits* of the collected information are as follows:

- **Fulfill legal requirements to monitor product defects**. For instance, in the United States, producers are legally obliged to report products that pose a substantial safety risk to the Consumer Product Safety Commission (U.S. Consumer Product Safety Commission, 2021).
- **Early warning system**. Defect information can indicate latent or emerging systematic product issues that require action to avoid escalation and a costly vehicle recall.
- **Detect unjustified warranty claims**. Resolving warranty claims can be costly, and producers try to identify unjustified claims to minimize costs.

Details about defects can help employees in the warranty department identify unjustified claims.
- **Benchmarking**. Producers can use defect information to compare their products with competitors' products and identify areas for improvement.

Company 1 identified one of its clients as a good example for evaluating the method and QualiExplore. This client is an *automotive supplier* developing and manufacturing a critical vehicle component. Car users can directly experience defects and other component behavior and may write about this on the Internet. The selected supplier is an example of an organization that could improve its product design through usage information.

The focused user profile is the *warranty manager* responsible for assessing warranty cases and deciding their resolution. These managers can share information with the product development department to contribute to product improvement. The focused warranty managers *need information* about the following aspects:

- Alerts about critical events
- All product-related problems
- Background to understand problems better
- Customer experience with the product
- Related or precedent (legal) cases

This case covers tasks similar to the Quality Management and Reliability Engineering tasks outlined in Chapter 3.2.

5.1.2 Case 2 – Designing a new information system

Company 2 produces home appliances in Europe. The company is interested in software that helps increase the quality of production processes and products.[29] It wants to automatically predict product quality during production. This solution would minimize costly manual quality tests at the end of a production line. In this case, humans and machines involved in the production process create PUI. Available *information channels* include spreadsheets with statistical quality control information and identified product defects. Besides, the test laboratory operators report quality test results. Consumer service data from the product's usage phase is available, but company 2 considers its integration in the prediction mechanism a long-term goal. The collected information's main *benefit* is **minimizing the costs of quality tests**. A virtual product quality testing process based on quality prediction could reduce test costs. Ideally, the process would replace the test lab.

---

[29] Company 2 experimented with research prototypes but there is no validated and verified solution yet. Therefore, this case focuses on designing a new IS.

The focused user profiles are the *test laboratory operator* performing quality tests and the *quality engineer* responsible for addressing failed tests, e.g., adjusting the production process. The lab operator needs information about the **predicted defects** per produced item to, for instance, focus on related tests. Quality engineers need to know the **correlations** between process characteristics and known product defects to adjust production machines' parameters. Such changes minimize the risk that potentially defective products leave the factory, resulting in costly recalls and unsatisfied users.

5.2 Workshops

Before performing the workshops, representatives of the companies received an *online teleconference briefing*. This briefing introduced the workshop's topic, identified relevant participants, and defined the target product. After this briefing, each company participated in *online workshop sessions* to evaluate the method developed in this dissertation and the QualiExplore tool.

The workshop sessions were two hours long, and the participants used an online board similar to a physical whiteboard with cards. This board had six prepared work areas to structure the workshop. Table 18 provides an overview of these areas and the application procedure described in Chapter 4.4.2.

Table 18: Overview of workshop structure and method's application procedure

No.	Online board areas	Related method application procedure steps
1	Application case information	-
2	Contents needed to support the case	Identify the needed information and its characteristics.
3	Information characteristics needed for the case	Identify the needed information and its characteristics.
4	Relevant information quality factors	Identify relevant factors amongst the ones mentioned in the diagrams. The users can define custom categories and factors.
5	Resolution of the top three factors	Identify how the IS design should manage the relevant factors.
6	Workshop feedback	-

Area one asked participants for general information about their use case, such as the company's role, target product, information sources and channels, and the main benefits of usage information. The *second area* detailed the contents needed for the focused work tasks. Chapter 5.1 contains the information from areas one and two. *Areas three* and *four* covered the parts that QualiExplore supports. The former contained prepared cards for the ten quality characteristics identified in Chapter 4.2.1. The *fifth area* specified how participants would address the most relevant quality factors.

Finally, the **workshop's feedback** area asked to evaluate four aspects in an open discussion. First, it was critical to assess if the workshop matched the participants' expectations to identify problems concerning the briefing and its understandability. The following two aspects focused on how helpful working with quality characteristics and factors was. This part identified problems related to the method's approach and specific characteristics or factors described in QualiExplore. Finally, the fourth aspect concerned QualiExplore's practical utility by asking if the participants would use it at work. Its purpose was to identify specific areas for tool improvements and shortcomings.

6 Method evaluation

This chapter describes the evaluation of the demonstration's results. It identifies how effective the method and QualiExplore were during the workshop. *Chapter 6.1* presents the workshop participants' contributions, and *Chapter 6.2* describes my observations during the workshop.

6.1 Workshop results

The following paragraphs summarize the results concerning relevant quality characteristics, quality factors, and the participant's feedback. Table 19 summarizes the results for the identified quality characteristics.

Table 19: Information quality characteristics identified as relevant

No.	Characteristics	Use Case 1	Use Case 2
1	Accessibility	X	X
2	Accuracy	X	X
3	Availability	-	X
4	Completeness	X	X
5	Confidentiality	X	-
6	Credibility	-	X
7	Currentness	X	X
8	Efficiency	X	X
9	Precision	X	-
10	Understandability	X	X

Both workshops considered the majority of suggested quality characteristics as relevant. In most cases, the participants detailed characteristics with at least one card containing information on how the characteristic fits the use case. No participant suggested new information characteristics.

Participants in both workshops selected several filter statements in QualiExplore, as illustrated in Figure 33.

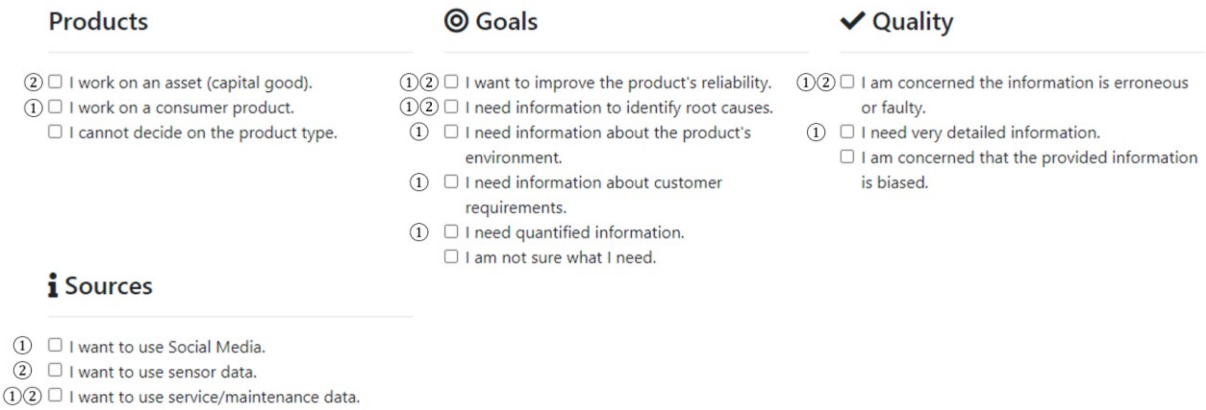

Figure 33: Selected QualiExplore filter statements in use cases one and two

Workshop participants of use *case one* selected 11 out of 15 filter statements. This selection flagged all quality factors in the second step, effectively rendering the filtering mechanism ineffective. The participants had to check all factors quickly to avoid overextending the workshop's time slot for this task. Board area four collected 28 cards indicating relevant factors and, for a few, a short note about measures to address them. The filter selection resulted in 60 flagged factors for *case two*. This number was still significant and resulted in the joint decision to work through the flagged factors until this task consumed the planned time for this slot. The participants collected seven cards with relevant factors and outlined measures for three of them.

Feedback in both workshops was generally positive. Participants of the first one indicated they found the workshop helpful in getting ideas on how to advance their software. For instance, they see sentiment analysis and techniques to address bias in Social Media as relevant. An essential conclusion of the second workshop was that the participants became aware of the critical role of maintenance in achieving high information quality. Maintenance personnel checks, for instance, machine-embedded measurement systems that can degrade until the resulting information is no longer accurate enough.

The participants identified the following shortcomings:

I. In workshop one: working with quality characteristics was **abstract**, and the workshop or QualiExplore should introduce concepts to make them more understandable.
II. In workshop one: finding matches between the quality factors and the use case context was challenging because the participants did not know the **jargon**, e.g., used in metrology.
III. In workshop one: QualiExplore should add **potential measures** to address quality factors. This feature would be an incentive to use the tool.

IV. In workshop two: the statements for the filter criterion “Products” did not fit the use case’s interest in improving process production quality, i.e., using one or more production machines. Adding an option to express interest in a **production process** might address this shortcoming.

6.2 Workshop observations

The most notable observation was that all participants struggled with information characteristics. They found them abstract and challenging to transfer to the use case’s specific context. Introducing examples concerning the use cases addressed this issue.

Several factors depend on specific methods to collect or process PUI. For example, the “self-report design” factor refers to self-report studies. Since the participants were unfamiliar with many of these methods, the factor descriptions were unclear and required explanation.

The filter selection resulted in so many matches that the workshop’s planned timeslot was too short. Participants in the first workshop reacted to this situation by briefly covering all factor descriptions – mainly stating whether a factor is relevant and sometimes why. In the first workshop, many flagged factors turned out to be slightly relevant but not enough to address them with a specific measure. The second workshop kept the planned timeslot but discussed a subset of the flagged factors.

7 Discussion

This discussion critically reflects the workshop results and concludes on how the developed method and QualiExplore fulfill the requirements. *Chapter 7.1* focuses on the workshops' evaluation results, while *Chapter 7.2* revisits the requirements. *Chapter 7.3* discusses the method's limitations from an application-independent point of view.

7.1 Critical reflection on the evaluation results

This chapter analyses the workshop results and related observations. It discusses their implications and identifies improvements for the workshop and QualiExplore. Each of the following paragraphs represents one discussion topic outlined in the evaluation results.

Case-specific understanding of information characteristics. The first finding of the evaluation is that users understand the approach and can work with information characteristics. This finding generally confirms earlier work by Weber (refer to Chapter 3.3.3), but it adds the constraint that users require support in transferring the generic information characteristics to their use case. The supporting person can *provide examples fitting the use case* and thus help overcome situations where the other participants cannot assess the relevance of characteristics. He or she would help the participants transform conscious information needs into formalized ones. The supporting person does not necessarily have to be the IM expert applying the method – it could be, for instance, one of the producer's employees with competencies in IQM. If the support above takes too much time, a complementary approach could be to extend QualiExplore with *context-dependent definitions and examples*. This new function means the static texts depend on the selected filter statements.

More filters and tighter filter scopes. The QualiExplore filters turned out to be weak because they could not sufficiently reduce the number of relevant factors. Extending the workshop could grant participants enough time to discuss all matches. However, this approach does not solve the problem that too many matches turn out to be not or less relevant, as observed in the first case's workshop. One solution could be to add more filters allowing participants to express their interest in PUI precisely, but selecting too many filters would again lead to too many matches. A more effective solution would need to consider tighter filter scopes, i.e., fewer factors related to one filter. This approach also has a downside because many factors refer to two or more filters, and selecting many filters would result in too many matches. A complementary approach is implementing AND and OR logic operators in the matching.[30] For example, the "I need very detailed

[30] The current QualiExplore version uses OR logic only.

information" and "I want to use sensor data" filters would match factors related to sensor data and the precision characteristic.

Add the filter "I work on a process". The second use case is about production process improvements to increase product quality. This process covers machines to produce products, which matches the definition of a capital item (refer to Chapter 2.1.1). Consequently, the "I work on an asset (capital good)" filter generally fits this case. Confusion could arise because the statement refers to two cases. First, the machine producer could improve the machine's design with usage information. This case is similar to the improvement of a consumer product. Second, the organization using the machine could analyze its PUI to improve the production process – for instance, by optimizing machine parameters and environmental conditions. The latter case is harder to explain; therefore, adding an "I work on a process" filter could clarify this distinction. Besides, addressing this process perspective is essential in integrated product development (Ehrlenspiel and Meerkamm, 2017) – therefore, it fits this dissertation's and QualiExplore's scope.

Include measures influencing quality factors. The workshop shortcoming III indicates that adding information on managing quality factors could incentivize using QualiExplore. One challenge of this feature is that its implementation requires a study to identify relevant measures. Its study design could include steps to investigate the literature or conduct expert interviews. In both cases, the study has to take an interdisciplinary perspective to cover, for instance, metrology, user research, and data analytics. Besides, QualiExplore will need to present measures adequately if this study finds many potential measures per factor. Finally, filters to select relevant measures might be necessary to avoid information overload.

Formalism as an application barrier. The developed method and QualiExplore use strict terminology and structures to present PUI quality. This approach was not as practical during the workshops as expected. Since the participants had difficulties transferring the concepts to their situation, further support was necessary. This additional effort is a problem because the method's application becomes more costly – probably so costly that perceived benefits cannot outweigh it. Continuous extensions of QualiExplore's knowledge base could help understand PUI quality better but would amplify this problem. The applied formalism and the much less formal practical understanding of information quality will be difficult to balance. Therefore, formalism will become a barrier to the method's and QualiExplore's application. Overcoming it may require an approach that does not rely on strict terms and structures or can substantially increase the perceived benefits to marginalize the effort.

7.2 Fulfillment of the requirements

The requirements defined in Chapter 4.1.2 focus on the method's capability to support PUI need analysis, while the ones in Chapter 4.5.1 concern the tool and its capability to simplify the method's application. The following sub-chapters revisit these requirements in light of the evaluation results. Table 20 and Table 21 summarize this assessment.

Table 20: Summary of method requirements

Requirements	Findings
Identify relevant quality characteristics systematically	●
Spot the need for compromised information needs	◐
Apply the method in situations with multiple PUI types	●
Require minimum effort to use the method	◐

●: Fulfilled ◐: Partially fulfilled ○: Not fulfilled

Table 21: Summary of tool requirements

Requirements	Findings
R1: The user should be able to access the software easily	●
R2: The user should be able to see the cause-effect relations	●
R3: The user should be able to customize the cause-effect diagrams	◐
R4: The user should be able to receive relevant factors	◐
R5: The user should be able to discover relevant factors by accident	◐
R6: The user should be able to track the method's application progress	◐
R7: The user should be able to receive a description of the factors	◐

●: Fulfilled ◐: Partially fulfilled ○: Not fulfilled

7.2.1 Method requirements

Identify relevant quality characteristics systematically.

This dissertation identified PUI quality characteristics in application case descriptions. Using case descriptions ensures that the characteristics are relevant for development-related tasks. A complete set of these characteristics would allow users to go through each potential item and assess its relevance in the light of the focused use case. This dissertation provides an initial set and proposes a step in the method's application procedure to ask workshop participants for additional quality characteristics. Both measures contribute to systematically identifying characteristics for a specific use case. Therefore, the method **fulfills** this requirement.

Spot constraints for information needs.

Quality-related factors are the leaves in the developed cause-effect diagram and belong to one or more PUI quality characteristics. Method users can see which factors contribute to a characteristic and must assess the necessity and realization of mitigation measures. If necessary measures are infeasible, the employees must expect a compromise between what they need and what the IS can provide. The developed method does not propose trade-offs between quality characteristics because cause-effect diagrams typically miss dependencies between branches and leaves. Given these aspects, the method fulfills this requirement **partially**.

Apply the method in situations with multiple PUI types.

Information from the Internet and maintenance was relevant in the first demonstration use case. The second case focused on measurements and maintenance information. Each case covers at least two PUI types; therefore, the method **fulfills** this requirement.

Require minimum effort to use the method.

The effort during the demonstration included the workshop's preparation, application, and post-processing. The demonstration did not reveal areas where the participants could further reduce the effort – on the contrary, filtering can result in too many matches and higher effort to investigate them. As outlined above, improved filtering could address this issue. However, it could still be possible that method users decide to investigate all factors in detail. Besides, explanations are necessary if participants are unfamiliar with specific methods for PUI collection and processing (refer to Chapter 6.2). In both situations, the effort would be the minimum required to investigate all selected factors. For these reasons, the method fulfills this requirement **partially**.

7.2.2 Tool requirements

R1: The user should be able to access the software easily

QualiExplore is accessible through a web browser and thus usable on mobile devices with an Internet connection. Business notebooks and tablets are standard equipment in companies; therefore, the developed tool **fulfills** this requirement.

R2: The user should be able to see the cause-effect relations

QualiExplore uses a tree structure to illustrate the relationship between quality characteristics and quality-related factors. Branches are collapsible to grant the user control over the branches' visibility. The tool **fulfills** this requirement.

R3: The user should be able to customize the cause-effect diagrams

The tool has two configuration files containing information about the filters and the cause-effect relations. Users can modify these files to customize QualiExplore's contents. The usability for customization is comparably low because the user has to modify the files on the server, which requires programming expertise. In summary, the tool fulfills this requirement **partially**.

R4: The user should be able to receive relevant factors

Users can automatically flag factors for further investigation by selecting filters. The implemented filters were rather broad in the evaluated version. Filter selections in both workshops resulted in at least 60 matches. The participants in the first workshop considered many of them less or not relevant. QualiExplore could benefit from more filters and tighter filter scopes, as argued above. For these reasons, the tool fulfills this requirement **partially**.

R5: The user should be able to discover relevant factors by accident

QualiExplore presents all quality-related factors and highlights the subset related to the selected filters. Users can decide to investigate factors without flags and discuss their relevance. I did not observe serendipity for quality-related factors in the workshops, but the unconstrained presentation would allow it. The tool fulfills this requirement **partially**.

R6: The user should be able to track the method's application progress

Flag icons highlight factors that belong to the selected filters. The red flag turns green if the user marks a factor with the "Proceed" button. A progress bar indicates the total number of highlighted factors and how many of them are green. The progress bar does not cover factors that the user identified by accident. Therefore, the tool fulfills this requirement **partially**.

R7: The user should be able to receive a description of the factors

Each factor in QualiExplore has a field that contains a description and, if available, links to the related references. Fulfilling this requirement is controversial because QualiExplore has descriptions for all factors, but the workshop participants indicated they are often too abstract. Besides, short descriptions extracted from the application case literature promote ambiguity and cause additional workshop effort to resolve it. In this light, the tool fulfills this requirement **partially**.

7.3 Limitations of the method

I identified several potential weaknesses and limitations during the method's development and evaluation. This chapter outlines and organizes these findings based on the method's scope and development.

7.3.1 Method scope

The method has two main scope limitations. First, it focuses on the usage phase, as illustrated in Chapter 2.2.1, and second, explicit knowledge, as outlined in Chapter 3.1.1. The following paragraphs discuss these limitations.

PUI acquired after the usage phase

The method scope could imply that PUI channels exclusively stem from the usage phase, i.e., when the product is with the user. Wellsandt and Thoben (2016) identified that the end-of-life phase also generates and collects information about product usage. Products are sometimes returned to the producer to settle warranty cases or for refurbishment, recycling, and disposal. During the end-of-life treatment, the producer can analyze the product and create information about its state after usage. This state indicates, for instance, wear and degradation. Producers may also retrieve data storage devices to analyze already collected but, so far, inaccessible information.

PUI acquired during the end-of-life phase can be highly relevant for reliability engineering. It might contain indications of how or why a product failed. This insight, in return, allows employees in the development department to improve the involved components to make them, for instance, more reliable. For products with a usage phase of several years, acquiring PUI at the end of life might be too late for meaningful changes because that product might have become obsolete.

Implicit knowledge about product usage

Implicit knowledge about product usage can reveal opportunities for improvement that measurements and documented information, e.g., from user complaints, do not cover. For instance, measurements may not reveal handling problems because they do not measure the relevant parameters or lack the required precision. If users are unaware that specific handling impacts the product behavior, they do not communicate it to the producer.

Acknowledged product design methods, such as Lead Users (Luthje and Herstatt, 2004) and Design Thinking (Brown and Wyatt, 2010), try to formalize the implicit knowledge of users. These methods involve the users in the development process and allow development-related employees to understand user needs and problems in greater detail. Personal contact between the developers and the users may create a trustworthy working atmosphere that encourages mutual understanding and sharing of relevant implicit knowledge.

7.3.2 Method development

The main weaknesses related to the method's development are the underlying information quality model (Chapter 4.2.1) and the PUI application cases used to identify relevant information characteristics (Chapter 4.2.1.2). This chapter will clarify these weaknesses and potential consequences.

7.3.2.1 Information quality model

The method development in Chapter 4 used an acknowledged information quality model to support the classification of text passages. This model contains descriptions for information quality characteristics, but it *does not detail the relations* between them, despite their importance for identifying trade-offs (refer to Chapter 7.2.1). The following paragraphs will outline and discuss relations I identified during the method development.

Accuracy and precision are information quality characteristics that get confused so often that ISO standards add warnings to their definitions (ISO, 2017). One reason for this confusion is that both characteristics relate to each other. Appendix L summarizes PUI-related standards and guidelines that describe accuracy and precision from the perspective of domains such as metrology, data quality management, software engineering, and systems engineering.

The heterogeneous descriptions in the appendix indicate that **accuracy** has at least two sub-concepts. First, it is a qualitative assessment of the correctness or the absence of errors in a dataset. Second, it means a quantitative measure of the magnitude of the error.

Concerning PUI, Raghupathi et al. (2015), Aldaz et al. (2013), and Finch and Luebbe (1997) used the *qualitative meaning* of accuracy when referring to, for instance, "good accuracy" or "lack of accuracy". Bleda et al. (2012) used the *quantitative meaning* of accuracy when describing measurements, while Abrahams et al. (2015) and Tucker and Kim (2011) used it during their validation of text data analysis methods.

The descriptions for **precision** indicate at least two sub-concepts as well. First is the precision of measurements – this concept stems from metrology. The second meaning focuses on the presence of specific information to reliably distinguish product states, product behavior, and usage situations. This sub-concept differs from the terminology used in metrology.

The *precision of measurements* means how close the values of repeated measurements are. Ideally, the value should be the same for each repetition. This dissertation's application case study did not find this sub-concept.

The application case descriptions typically use precision to refer to the *presence of information* needed for a specific task. This information allows employees to differentiate specific product states, product behaviors, or usage situations. For measurements, application case authors used, for instance, *decimal places* (Table 1 in Abramovici et al., 2011) and *sampling rate* (Ghosh et al., 2016).

In metrology, the concept of resolution covers the meaning of the terms above. **Resolution** is the "*smallest change in a quantity being measured that causes a perceptible change in the corresponding indication*" (ISO, 2007). For human-authored information, the meaning is different. For instance, Abrahams et al. (2015) stated that some of their investigated PUI contained too little information to describe or confirm a product defect with the developed method. Rephrasing this statement indicates one relation between precision and completeness: if an analyst wants to confirm a specific product defect, the PUI has to contain specific information. Other authors are more explicit about the PUI precision, pointing out that it should include information about *product instances* (Bueno and Borsato, 2014) and *sub-components* within a product (Igba et al., 2013). This meaning is again similar to resolution in metrology.

Completeness is an information characteristic several authors referred to in the application case descriptions. The following list summarizes the aspects that I identified during the method development. Completeness refers to the following:

- A *data item* where all expected data attributes exist (ISO, 2008).
- A *measurement dataset* not missing critical data (Vichare et al., 2007).
- A *dataset that describes an entire process,* such as a use situation or maintenance activity (Aldaz et al., 2013), (D'Antonio et al., 2015), (Igba et al., 2013).
- A *dataset that describes all related products* (Bueno and Borsato, 2014).
- A *dataset that describes the entire customer group* (Thiruvenkadam et al., 2008).
- A *consumer information source* without missing data items and with sufficient breadth and depth (Thiruvenkadam et al., 2008).

This dissertation structures the aspects above into three perspectives, as summarized in Table 22.

Table 22: Perspectives on PUI completeness

Perspectives	Meaning based on findings
Data	Data item has all expected attributes
	Dataset has all expected data items
Application	PUI describes a process in detail
	PUI describes all product instances
	PUI describes an entire customer group
Channel	Contains all information needed for a task

- First, the *data perspective* considers data items and datasets. One of its objectives is to ensure that a data item has values for all required attributes, such as timestamps representing a time series. A second objective is that datasets contain all data items required for one or more tasks. Missing data or the presence of data items with unsatisfying quality characteristics constrain the use of this dataset for the given tasks.
- Second, the *application perspective* concerns the completeness required for specific development-related tasks. A root-cause analysis, for instance, could benefit from service information about all product instances with failures. Missing product failure documentation could result in erroneous conclusions or delay root-cause analysis to collect the missing information.
- Third, the *channel perspective* evaluates the communication channels that convey PUI. This evaluation determines if the respective channels contain all the required information to perform a development-related task. If not, the producer's employees should consult supplementary channels.

The following example outlines how the data and application perspectives of completeness relate to accuracy and precision. It focuses on a PUI-supported root-cause analysis because it is acknowledged and frequently used in practice.

A PUI-supported root-cause analysis creates an *analysis result* by investigating an analysis dataset. The *analysis dataset* contains *usage datasets,* and each has item-level PUI, containing usage datasets about the product's components.

The *precision* of the analysis result depends on, for instance, the analyst's capability to identify the specific product components associated with the root cause. Consequently, an analysis dataset should contain item-level PUI to increase this precision. Furthermore, PUI created through measurement is typically a time series with a defined sampling rate. Analysts may recognize unanticipated product states, product behaviors, and usage situations only if they can identify the related data patterns among the measured values (data points).

The *completeness* of the analysis dataset depends on the precision. Specific causes may become visible only if the analysis dataset includes the product instances and the data points with an interesting pattern. The more complete the dataset is in terms of instances, and the more complete the time series in the usage datasets are, the more likely it is to find anomalies. These anomalies can hint at root causes. Furthermore, if the dataset is complete and does not contain anomalies, there is a more robust justification to search for the cause elsewhere.

As indicated above, investigating the relations between quality characteristics is complex. This manuscript's outlook will revisit it as an area for improvement.

7.3.2.2 Application cases

I developed the method using a selection of application case descriptions, which likely introduced a **sampling bias**. This bias could have resulted in missing important information quality characteristics and quality-related factors. The impact of this sampling bias on the results is comparably small. Method users should adjust the identified characteristics and quality-related factors to meet the needs of their specific PUI case. The incompleteness of the tool's content due to a sampling bias, therefore, should not be harmful to method users. Furthermore, ISO 25012 allows custom information quality characteristics, provided there is a mapping with the characteristics suggested in this ISO standard.

The second source of bias is that I performed the qualitative content analysis alone. The results are more **subjective**; consequently, the method could miss relevant information characteristics and quality-related factors. Furthermore, it could contain erroneous content created during the classifications. Literature suggests involving multiple analysts to reduce subjectivity during content analysis. I decided against this practice because the content analysis took around 50 hours, which is significant enough to discourage potential participants. The impact of subjectivity on the developed method is small. The method's adjustability mentioned above addresses the problem of missing out on relevant characteristics and factors. Errors are a problem and could give the method users wrong information, especially about the relevance of specific quality-related factors. An effective approach to address this problem is continuously improving the method based on application experiences.

Finally, the **maximum and minimum age** of the application cases could have rendered the method outdated. Older application cases, such as Finch and Luebbe (1997), are timely because they focus on the fundamental aspects of – at that time – new technologies and their application. For instance, statements about the quality of user-generated content were relevant for Usenet, but they are also relevant for Social Media, which emerged several years later. The youngest investigated application cases are from 2016. Since then, academia has published several articles covering similar technologies and work tasks. For instance, Meyer et al.

(2021) provide a recent overview of PUI applications for product planning. Most likely, PUI application case literature will evolve with the technologies that collect, organize, and present PUI. Therefore, users of the developed method must interpret quality characteristics and factors in light of the available technologies and adjust them where necessary. Since the method's users should adjust the content to their needs, the age of the investigated application cases does not render the method outdated.

8 Summary and Outlook

This chapter summarizes the approach and findings of this dissertation and outlines future work based on the discussion and current research challenges.

8.1 Summary

This dissertation aimed to develop a method that helps organizations describe precise PUI needs for product improvement. Since there was no acknowledged conceptual framework for PUI in literature, the **first task** was to *identify key concepts and terminology* necessary to understand PUI and its application in product development. An overview of products and services identified shopping goods and capital items as the most suitable focus areas. Products belonging to these classes can have use phases of several years, are subject to maintenance services, and contain embedded electronics and software. These factors benefit the creation of PUI. Taylor's information need typology explains how information needs emerge from a feeling of uncertainty experienced while employees perform work tasks for a producer. Information from the usage phase – i.e., the middle of a product's life cycle – is one of three product-related information types that employees use to satisfy their information needs. Working on the first task revealed synonyms for PUI, such as in-service data, field data, and installed-base information – a sign of literature fragmentation. A custom communication model summarized the PUI domain's information sources, destinations, and communication channels. The last step of the first task specified concepts and terminology in PRIM, which manages the application, information systems, and processing infrastructure of product-related information. This part differentiated PRIM from the related management disciplines of PDM, PIM, and PLM. The first task indicated that the key challenges of using PUI in PRIM are: a) creating item-level information, b) combining PUI from different channels, and c) information processing. Literature covers these challenges well from a technical perspective. The application perspective is not a focus, though it includes identifying the need for PUI.

The **second task** was identifying the State-of-the-Art for information need analysis concerning PUI and development-related work tasks. Since this scope was too narrow, I had to extend it to work tasks in general. The first step was to refine the related concepts and terminology. Most notably, this included acknowledged definitions for "information" and "information need" – concepts intensely discussed in the literature about *information behavior*. Information is meaningful data and is often referred to as explicit knowledge, while an information need ranges from an unexpressed feeling to a formal question passed on to an IS. The next step clarified that, in general, employees need product-related information to understand and solve *complex problems* at work. More specifically, international *quality management* standards from the ISO 9000 and ISO 10000 series named complaints management and measuring customer satisfaction as relevant tasks that need PUI. These standards state that product-related information should consider sample sizes, errors, completeness, and accuracy. Such statements were the first indicators of PUI quality concerning product improvement. Most specific was the literature about *reliability engineering*, which concerns failure measurement, prediction, accommodation, and prevention. Employees who perform reliability engineering tasks benefit from information about a product's load conditions. The last part of the State-of-the-Art identified one general and four IM-specific methods to analyze information needs. These methods involve information users and have multiple steps and iterations. Information quality is a crucial concept for these methods used to express requirements for needed information regarding, for instance, relevance, timeliness, and reliability. I concluded from the State-of-the-Art that information need analysis should concretize needs through information characteristics. This approach is helpful because IM experts can design systems to provide information that meets these characteristics. Information characteristics should originate from specific work tasks because the task's context determines the characteristic's exact meaning. The *gap* in the State-of-the-Art was the lack of support in identifying the characteristics that affect the PUI quality. Besides, development-related employees struggle to formalize the specific PUI they need – partially because this field is comparably new.

The **third task** was to develop the method for PUI need analysis. It followed the "method engineering" approach that detailed the method's concepts and motivation, theoretical foundation, procedure and syntax, and application. A framework of concepts organized the relations between work tasks, employees, information needs, interaction, communication channels, and information characteristics. Method engineering started with the deduction of four basic requirements. I used an inductive, qualitative content analysis of 22 application case descriptions to create the method's theoretical foundation. Its first part identified *13 relevant information quality characteristics* using the content analysis and consultation of the ISO 25012:2008 information quality model – an acknowledged international standard. The second part focused on the factors that influence these quality characteristics. Liu and Chi's *evolutional data quality concept* provided the theory of how information quality changes over time. They structured the data life cycle into data collection, organization, presentation, and application. Each step has one data quality concept and several factors that change the quality characteristics. The data quality of an earlier step contributes to the quality of the following ones. Liu and Chi used a cause-effect diagram to visualize how these four qualities relate to factors. I identified factors in the application cases that refer to various biases and data creation, access, processing, and application. The method's procedure uses Liu and Chi's *cause-effect diagram* combined with the identified PUI-related information characteristics and factors. I differentiated factors that may change the characteristics and factors that change the requirements. The developed cause-effect illustrations focused on the former because I could not distinguish the identified factors further without introducing additional subjectivity.

With the theoretical foundation and the adapted cause-effect diagrams, the last part of method development was to design an *application procedure*. The designed procedure suggests performing a workshop with development-related employees. Its principal activities include identifying the needed information and related characteristics, identifying relevant factors, defining additional factors, and identifying how the IS design should manage relevant factors. Since the developed cause-effect diagrams are complex and may cause information overload, the final step was to develop a software tool that supports the method's application. Seven requirements specified what the tool should do – this manuscript called the realization *QualiExplore*, a website with a filter function, and a visualization of the PUI quality characteristics and the related factors.

The **fourth task** was the method's demonstration and evaluation. Two use cases with companies served as demonstration settings. One case focused on a software provider aiming to improve an existing PUI-based IS about automobile components. The second case featured a home appliances producer and product improvements through PUI from production machines. Workshops with experts from the two companies tested the method and the support tool. The participants' main feedback was that working with quality characteristics is too abstract, and method improvements should introduce concepts to make them more understandable. Besides, presenting potential measures to control factors could incentivize using the software tool. Observations during the workshop confirmed that participants had difficulties working with information characteristics. Contextualized examples helped address this issue. Furthermore, the participants did not know several specific methods to collect or process PUI, which required additional explanation of the related factors. The most critical observation of the tool was that filters were ineffective due to their broad scope.

The **final task** was to critically discuss the fulfillment of all defined requirements and the method's scope, development, and application. The method fulfills two requirements and partially fulfills the other two. Shortcomings included missing information about trade-offs between quality characteristics and ineffective filters in the tool that increased the workshop's effort. QualiExplore fulfills two requirements and five in part. Shortcomings included too few filters, too broad filters, and abstract factor descriptions. Furthermore, the discussion outlined limitations, such as the relevance of acquiring PUI after the usage phase and the importance of formalizing implicit knowledge through methods such as Lead Users and Design Thinking. The method development did not cover relations between information quality characteristics and might have introduced a sampling bias during the application case selection from the literature.

8.2 Outlook

The outlook of this dissertation has four parts. First is the systematic development of an acknowledged PUI domain respecting inter-disciplinary concepts, and second is a potential improvement for the developed method to increase its utility. The third part covers changes to improve QualiExplore's user experience beyond the shortcomings identified in the discussion, and the final part outlines future research topics related to PUI and its application in product development.

8.2.1 Developing an acknowledged theoretical foundation

One of this dissertation's main challenges was managing the fragmented research about PUI and its application in product improvement. An interdisciplinary effort involving product development, information behavior, and information quality could systematically research specific PUI aspects.

One area to investigate further is to *identify relationships* between PUI characteristics. Follow-up research could revisit the relationships mentioned in this dissertation and continue with the literature. For instance, Eppler (2006) outlines five possible goal conflicts between information quality characteristics, but it is unclear to what extent his findings are relevant for PUI. Finch (1999) mentioned that PUI users might have to accept certain forms of bias and should consult complementary information acquired with other methods. Results in this area could lead to more accurate PUI quality models or justify employee training programs to increase their awareness of information quality.

A second area is the investigation of additional literature to refine the results of Chapter 4.2. Besides other application cases, *general articles* covering PUI-related technologies and methods are relevant. Karkouch et al. (2016), for instance, studied information quality concerning the IoT. They adopted the information characteristics of Wang and Strong (1996). Besides, they identified factors that change information quality. Rämänen et al. (2013) and Mahlamäki and Rämänen (2015) investigated the factors that influence the quality of manually collected asset data. They identified factors related to data collection tools, employee motivation, work tasks and employee competencies, and organization and culture. Madhikermi et al. (2016) developed a quality assessment for maintenance reporting procedures. They adopted the information characteristics from Krogstie et al. (1995) and identified quality measures for three characteristics. Leopold (2012) developed a data collection approach to improve product reliability analysis for utility vehicles. He defined nine information characteristics and used them to assess the quality of reliability-related information channels. Schulte (2006) developed a methodology to integrate customer feedback into a product development task. He defined five characteristics to describe the quality of feedback information. Follow-up research in this area could confirm the characteristics used in this dissertation and reveal new characteristics and quality-related factors.

Finally, *customer knowledge* is a concept related to PUI. It consists of the knowledge for, about, and from customers (Gebert et al., 2002). PUI analysis can lead to knowledge about the latter two (He et al., 2018). I mention this topic because the related literature connects this dissertation to the knowledge management domain for potential follow-up research.

8.2.2 Potential method improvements

Chapter 7.3.2.1 clarified that this dissertation used an information quality model that does not include the relationships among quality characteristics. Integrating them could lead to new views on quality-related factors. The integration could focus on two steps. First, follow-up research should *integrate methods suited to represent relationships* because cause-effect diagrams inherently lack this feature. One potential method is the *knowledge graph* (Hogan et al., 2020). It contains nodes and labeled edges that could represent information quality characteristics, categories, quality-related factors, and relationships among them. The relationships include cause-effect labels inherent to an Ishikawa diagram and others describing, for instance, how characteristics affect each other. Second, adding relationships would result in many labeled arrows between PUI characteristics and quality-related factors. Consequently, the method's user would likely experience information overload. A tool, such as QualiExplore, would help use the method effectively.

8.2.3 Potential tool improvements

One QualiExplore improvement is *tracking the method's application progress* when users accidentally discover relevant factors. An easy solution for this issue could allow users to add factors to the progress calculation. A more complex solution could use a *supervised machine learning approach*. The user could tag factors as relevant, and the tool would learn which filter-factor combinations occur frequently. With this extension, the tool would incorporate the user's knowledge about the factors.

The second improvement is to *extend the tool's scope to all steps* of the developed method. Its goal would be to reduce the effort to apply the method. QualiExplore could support the user at each step – including steps where interactions with other users are essential. This improvement could minimize the need for a workshop and increase the users' acceptance of applying the method.

A third improvement concerns the *graphical user interface* (GUI). The current interface does not allow the user to modify the files containing the filters, characteristics, categories, and factors. A GUI would lower the effort to change these contents and make the tool easier to transfer to other domains.

Finally, the fourth improvement could focus on improving filters by *adding personal factors* (e.g., education, experience, and emotions) *and situational factors* (e.g., time available to perform the task). This extension would align the tool with information seeking models, such as Byström and Järvelin's model presented in Chapter 3.1.3. Consequently, users could express their interests and constraints more precisely.

8.2.4 Complementary research areas

Several societal and legal topics emerged during this dissertation that would affect PUI's quality and application. This final part of the outlook will cover them briefly to outline points of interest for practitioners and researchers.

Data privacy protection. Since May 2018, the General Data Protection Regulation (GDPR) has regulated personal data processing (European Commission, 2018). Consequences for PUI quality may result from the following GDPR parts.

Article 15 covers the right of access. It regulates that an organization must inform an individual upon request which data categories it processes, the data processing, the data receivers, and how the organization created it. Furthermore, the organization must hand over a copy of the personal data if requested. This right reduces the effort to access PUI for individuals. If a competitor's employees rightfully request their personal data, they will learn about data creation, processing, and its usage for product improvement. This situation could *affect the competitive advantages* of the producer and its competitors.

Article 17 grants individuals the right to request the erasure of personal data. This right could imply that organizations must remove personal data from the training data for machine learning, the trained models, and information gained through applying the model. One potential outcome is an *intensifying sampling bias* when the training data no longer includes specific user groups. A second outcome could be the *increased effort to maintain the machine learning models* trained with frequently changing PUI.

Data ownership and license agreements. There is no data ownership right in the European Union (Dutch-Brown et al., 2017; Martens et al., 2020). End-user license agreements are an approach to address this complex legal situation. For instance, Toyota Motor Sales states that they use data collected via their Connected Services program for quality assurance, analysis, research, and product development (N.n., 2020b). Their online privacy notice is an informative example of an agreement where a producer asks for consent to use, store, and share PUI[31]. It is

[31] Toyota Connected Services collects and uses, amongst others, personal information, location data, vehicle status data, driving data, vehicle health data, multimedia screen data,

not clear, though, who should provide this consent. Potential candidates are the vehicle's owner, an intermediary (rental car), the driver, or passengers. A similar situation exists for information publicly available on the Internet. For a producer that wants to use this PUI, it may not be clear whom it should ask for consent. Potential candidates include the author, users mentioned by the author, and the channel provider. The providers range from global corporations, such as Meta or Google, to a private person providing an online discussion forum. With this spectrum of potentially involved stakeholders, utilizing PUI without significant consent management effort is challenging. A data ownership right might simplify the challenges above and stimulate efficient PUI integration into product development.

Inaccuracy through more analysis. An important mechanism for inaccurate PUI resides in the evolutional data quality concept itself. Data life cycles with many processing steps will likely introduce additional quality-related factors. The more steps the producer adds, the more factors the IM must manage. This procedure would continue until the management costs exceed the benefits of the PUI. Important cost drivers are the *number of unknown harmful factors* and the *number of factors the producer cannot influence*.

The first driver is difficult to avoid in general. Still, the systematic investigation of changes regarding the processing steps could increase the chances of identifying harmful factors. The second driver depends on the number of processing steps that integrate information third parties provide, such as weather, climate, geography, traffic, or demographic data. Typically[32], the producer cannot influence the accuracy of externally acquired information. Therefore, assessing how introducing more data analysis affects PUI quality is crucial. This precaution is even more relevant if the analysis relies on models generated with machine learning. In such cases, producers should request information about the training data and the machine learning mechanism before applying additional data analysis.

Disinformation. Kumar and Shah (2018) define disinformation as information on the Internet that a bad actor spreads intending to deceive. Policy and academia acknowledge disinformation as a problem, especially on digital platforms like Amazon's online shop and Facebook's social networking service (Kumar and Shah, 2018; European Commission, 2020). User-generated content on the Internet is susceptible to disinformation because it is cheap to create and difficult to identify. Bad actors can pay humans or use bots and sockpuppets to create fake accounts that spread disinformation (Kumar and Shah, 2018). For measurements, the mechanisms to create disinformation are different. Actors could manipulate

voice recordings, voice recognition recordings, safety system data, navigation system data, and traffic and weather data.

[32] Some producers may have financial incentives or other leverage to influence factors.

the embedded systems within products to deliberately produce or transmit false sensor readings (Hou et al., 2019).

Concerning product development, bad actors could create false PUI to affect, for instance, the product's appearance and functionality to facilitate bad user experiences that reduce a producer's reputation. Intentions with more severe consequences for users are the deliberate introduction of product safety and security risks. Influencing product design through disinformation could be a subtle method for bad actors to hurt individual companies or an entire industry. The literature analysis did not indicate cyberattacks of this kind, but governments' increasing interest in cyber defense (N.n., 2020a) indicates that bad actors potentially possess wide-ranging capabilities already. An approach to address disinformation is using services, such as ReviewMeta (N.n., 2020c), that assess the quality of user-generated product reviews. However, the most effective approach is to consider PUI as complementary information that employees should compare with other available information.

My conclusion for the outlook is that applying PUI for product improvement has a problematic future. While the underlying technologies are mature enough to show PUI's application potential, non-technical barriers will likely slow down or prohibit wider adoption in praxis.

References

1. Abrahams, A.S., Fan, W., Wang, G.A., Zhang, Z.J., Jiao, J., 2015. An Integrated Text Analytic Framework for Product Defect Discovery. Production and Operations Management 24, 975–990.
2. Abrahams, A.S., Jiao, J., Fan, W., Wang, G.A., Zhang, Z., 2013. What's buzzing in the blizzard of buzz?: Automotive component isolation in social media postings. Decision Support Systems 55, 871–882.
3. Abramovici, M., Gebus, P., Göbel, J.C., Savarino, P., 2018. Utilizing unstructured feedback data from MRO reports for the continuous improvement of standard products. In: Maier, A., Kim, H., Oehmen, J., Salustri, F., Škec, S., Kokkolaras, M. (Eds.) Design information and knowledge. Curran Associates Inc, Red Hook, NY, pp. 327–336.
4. Abramovici, M., Lindner, A., 2011. Providing product use knowledge for the design of improved product generations. CIRP Annals - Manufacturing Technology 60, 211–214.
5. Abramovici, M., Lindner, A., Walde, F., Fathi, M., Dienst, S., 2011. Decision support for improving the design of hydraulic systems by leading feedback into product development, in: Proceedings of the 18th International Conference on Engineering Design (ICED11). Impacting Society through Engineering Design. The 18th International Conference on Engineering Design (ICED11), Copenhagen, Denmark. 15-19 August 2011. Design Society, Design Methods and Tools Part 1.
6. Abramovici, M., Neubach, M., Fathi, M., Holland, A., 2009. Knowledge-based feedback of product use information into product development, in: DS 58-8: Proceedings of ICED 09, the 17th International Conference on Engineering Design. International Conference on Engineering Design, Palo Alto, USA. 24-27.08, pp. 227–238.
7. Abramovici, M., Schulte, S., 2007. Optimising customer satisfaction by integrating the customer's voice into product development, in: Proceedings of the 16th International Conference on Engineering Design (ICED07). The 16th International Conference on Engineering Design (ICED07), Paris, France. 28-31 August, 2007. Design Society, Paper no. 147.
8. Aldaz, G., Steinert, M., Leifer, L.J., 2013. Instrumenting the user: Needfinding by combining data logging and immediate self-reporting, in: Proceedings of the 19th International Conference on Engineering Design (ICED13). Design for harmonies: Human Behavior in Design. The 19th International Conference on Engineering Design (ICED13), Seoul, Korea. 19-22 August, 2013. Design Society, Castle Cary, Somerset, pp. 417–426.
9. Anderson, E.W., Fornell, C., Lehmann, D.R., 1994. Customer Satisfaction, Market Share, and Profitability: Findings from Sweden. Journal of Marketing 58, 53.
10. Armstrong, J.S., Overton, T.S., 1977. Estimating Nonresponse Bias in Mail Surveys. Journal of Marketing Research 14, 396.
11. Augustin, S., 1990. Information als Wettbewerbsfaktor: Informationslogistik - Herausforderung an das Management. Verl. Industrielle Organisation, Zürich, 207 pp.
12. Belkin, N.J., Vickery, A., 1985. Interaction in information systems: A review of research from document retrieval to knowledge-based systems. British Library, United Kingdom.
13. Bernstein, P.A., Haas, L.M., 2008. Information integration in the enterprise. Commun. ACM 51, 72.
14. Betsch, T., Funke, J., Plessner, H., 2011. Denken – Urteilen, Entscheiden, Problemlösen. Springer Berlin Heidelberg, Berlin, Heidelberg.
15. Bleda, A.L., Maestre, R., Santa, G., Jara, A.J., Skarmeta, A.G., 2012. Web of Things as a Product Improvement tool: Furniture as Case Study, in: The Sixth International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS-

2012). Sixth International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing, Palermo, Italy. 4-6 July 2012, pp. 846–851.
16. Brinkkemper, S., 1996. Method engineering: Engineering of information systems development methods and tools. Information and Software Technology 38.
17. Brown, T., Wyatt, J., 2010. Design Thinking for Social Innovation. Development Outreach 12, 29–43.
18. Bueno, M.R., Borsato, M., 2014. A Method for Identifying Product Improvement Opportunities through Warranty Data, in: Moving integrated product development to service clouds in the global economy. Proceedings of the 21st ISPE Inc. International Conference on Concurrent Engineering. International Conference on Concurrent Engineering (CE-2014). 8-11 September 2014. IOS Press, Amsterdam, pp. 122–131.
19. Byström, K., Järvelin, K., 1995. Task complexity affects information seeking and use. Information Processing & Management 31, 191–213.
20. Case, D.O., Given, L.M. (Eds.), 2016. Looking for information: A survey of research on information seeking, needs, and behavior. Emerald Group Publishing Limited, Bingley, 507 pp.
21. Cato, W.W., Mobley, R.K., 2002. Computer-managed maintenance systems: A step-by-step guide to effective management of maintenance, labor, and inventory, 2nd ed. Butterworth-Heinemann, Boston, 180 pp.
22. Choo, C.W., 2002. Information management for the intelligent organization: The art of scanning the environment, 3rd ed. Information Today, Medford, NJ, 325 pp.
23. Condra, L.W., 2001. Reliability improvement with design of experiments, 2nd ed. Marcel Dekker, New York, NY, 398 pp.
24. Cooper, T., 1994. Beyond recycling: The longer life option. New Economics Foundation, London.
25. Daft, R.L., Lengel, R.H., 1984. Information richness: A new approach to managerial behavior and organizational design. Research in Organizational Behavior 6, 191–233.
26. D'Antonio, G., Sauza Bedolla, J., Chiabert, P., Lombardi, F., 2015. PLM-MES integration to support collaborative design, in: Proceedings of the 20th International Conference on Engineering Design (ICED15). The 20th International Conference on Engineering Design (ICED15), Milan, Italy. 27-30 July, 2015. Design Society, Glasgow, Scotland, pp. 81–90.
27. Deng, Q., Franke, M., Hribernik, K., Thoben, K.-D., 2018. Exploring the integration of social media feedback for user-oriented product development. In: Maier, A., Kim, H., Oehmen, J., Salustri, F., Škec, S., Kokkolaras, M. (Eds.) Design methods and tools, vol. 4. Curran Associates Inc, Red Hook, NY, pp. 453–462.
28. Deng, Q., Wellsandt, S., Hribernik, K., Thoben, K.-D., 2021. Understanding Users and Products in Product Development: The Application of Product Usage Information and Its Challenges. Proc. Des. Soc. 1, 3299–3308.
29. Deng, Q., Wellsandt, S., Hribernik, K.A., Thoben, K.-D., 2019. Towards Understanding the Role of Product Usage Information in Product Design Improvement. In: Fortin, C., Rivest, L., Bernard, A., Bouras, A. (Eds.) Product Lifecycle Management in the Digital Twin Era, vol. 565. Springer International Publishing, Cham, pp. 369–378.
30. Detlor, B., 2010. Information management. International Journal of Information Management 30, 103–108.
31. Dorner, D.G., Gorman, G.E., Calvert, P.J., 2015. Information needs analysis: Principles and practice in information organizations. Facet Publishing, London, 269 pp.
32. Dörner, D., 1974. Die kognitive Organisation beim Problemlösen: Versuche zu einer kybernetischen Theorie der elementaren Informationsverarbeitungsprozesse beim Denken. Huber, Bern, 370 pp.

33. Dörner, D., 1979. Problemlösen als Informationsverarbeitung, 2nd ed. Kohlhammer, Stuttgart, 151 pp.
34. Duncker, K., 1974. Zur Psychologie produktiven Denkens, 3rd ed. Springer, Berlin, 135 pp.
35. Duriau, V.J., Reger, R.K., Pfarrer, M.D., 2007. A Content Analysis of the Content Analysis Literature in Organization Studies: Research Themes, Data Sources, and Methodological Refinements. Organizational Research Methods 10, 5–34.
36. Dutch-Brown, N., Martens, B., Mueller-Langer, F., 2017. The economics of ownership, access and trade in digital data. JRC Technical Reports 2017-01. European Commission.
37. Edler, A., 2001. Nutzung von Felddaten in der qualitätsgetriebenen Produktentwicklung und im Service. Dissertation, Berlin, Germany.
38. Ehrlenspiel, K., Meerkamm, H., 2017. Integrierte Produktentwicklung: Denkabläufe, Methodeneinsatz, Zusammenarbeit, 6th ed. Hanser, München, XXVIII, 991 str.
39. Elo, S., Kyngäs, H., 2008. The qualitative content analysis process. Journal of advanced nursing 62, 107–115.
40. Eppler, M.J., 2006. Managing Information Quality: Increasing the Value of Information in Knowledge-intensive Products and Processes, 2nd ed. Springer, Berlin, Heidelberg.
41. Eppler, M.J., Mengis, J., 2004. The Concept of Information Overload: A Review of Literature from Organization Science, Accounting, Marketing, MIS, and Related Disciplines. The Information Society 20, 325–344.
42. European Commission, 2018. The GDPR: new opportunities, new obligations DS-01-18-082-EN-N. European Union, Luxembourg.
43. European Commission, 2020. Tackling online disinformation. European Commission. https://ec.europa.eu/digital-single-market/en/tackling-online-disinformation. Accessed 25 October 2020.
44. Ferrara, E., Varol, O., Davis, C., Menczer, F., Flammini, A., 2016. The rise of social bots. Communications of the ACM 59, 96–104.
45. Finch, B.J., 1999. Internet discussions as a source for consumer product customer involvement and quality information: An exploratory study. Journal of Operations Management 17, 535–556.
46. Finch, B.J., Luebbe, R.L., 1997. Using Internet conversations to improve product quality: An exploratory study. Int J Qual & Reliability Mgmt 14, 849–865.
47. Frensch, P.A., Funke, J. (Eds.), 1995a. Complex problem solving: The European perspective. Lawrence Erlbaum Associates, Hillsdale, NJ.
48. Frensch, P.A., Funke, J., 1995b. Definitions, traditions, and a general framework for understanding complex problem solving. In: Frensch, P.A., Funke, J. (Eds.) Complex problem solving: The European perspective. Lawrence Erlbaum Associates, Hillsdale, NJ, pp. 3–25.
49. Freund, L., 2015. Contextualizing the information-seeking behavior of software engineers. J Assn Inf Sci Tec 66, 1594–1605.
50. Funke, J., 2003. Problemlösendes Denken, 1st ed. Kohlhammer, Stuttgart, 309 pp.
51. Gebert, H., Geib, M., Kolbe, L., Riempp, G., 2002. Towards customer knowledge management: Integrating customer relationship management and knowledge management concepts, in: The Second International Conference on Electronic Business (ICEB 2002). International Conference on Electronic Business, Taiwan. 10-13.12, pp. 296–298.
52. Ghosh, D.D., Kim, J., Olewnik, A., Lakshmanan, A., Lewis, K.E., 2016. Cyber-Empathic Design: A Data Driven Framework for Product Design, in: Proceedings of the ASME 2016 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference. International Design Engineering Tech-

nical Conferences and Computers and Information in Engineering Conference, Charlotte, North Carolina, USA. 21-24 August, 2016. American Society of Mechanical Engineers, New York, N.Y., V02AT03A012.
53. Graves, A., 2013. Six Sigma terms: Six M's. Six Sigma Daily. https://www.sixsigmadaily.com/six-ms/. Accessed 20 July 2018.
54. Hajian, S., Bonchi, F., Castillo, C., 2016. Algorithmic Bias: From Discrimination Discovery to Fairness-aware Data Mining, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 22nd ACM SIGKDD International Conference, San Francisco, California, USA. ACM, New York, NY, pp. 2125–2126.
55. Hans, C., Hribernik, K.A., Thoben, K.D., 2010. Improving reverse logistics processes using item-level product life cycle management. IJPLM 4, 338.
56. Harter, S.P., 1992. Psychological relevance and information science. J. Am. Soc. Inf. Sci. 43, 602–615.
57. He, W., Zhang, W., Tian, X., Tao, R., Akula, V., 2018. Identifying customer knowledge on social media through data analytics. Journal of Ent Info Management 8, 17.
58. Heckman, J., 1977. Sample Selection Bias As a Specification Error (with an Application to the Estimation of Labor Supply Functions). National Bureau of Economic Research, Cambridge, MA.
59. Herring, J.P., 1999. Key intelligence topics: A process to identify and define intelligence needs. Comp. Int. Rev. 10, 4–14.
60. Higgins, J.P.T., Altman, D.G. Assessing Risk of Bias in Included Studies. In: Higgins, J.P., Green, S. (Eds.) Cochrane Handbook for Systematic Reviews of Interventions. Wiley-Blackwell, pp. 187–241.
61. Hogan, A., Blomqvist, E., Cochez, M., d'Amato, C., Melo, G. de, Gutierrez, C., Gayo, J.E.L., Kirrane, S., Neumaier, S., Polleres, A., Navigli, R., Ngomo, A.-C.N., Rashid, S.M., Rula, A., Schmelzeisen, L., Sequeda, J., Staab, S., Zimmermann, A., 2020. Knowledge Graphs. https://arxiv.org/pdf/2003.02320.
62. Hou, J., Qu, L., Shi, W., 2019. A survey on internet of things security from data perspectives. Computer Networks 148, 295–306.
63. Hou, L., Jiao, R.J., 2020. Data-informed inverse design by product usage information: a review, framework and outlook. J Intell Manuf 31, 529–552.
64. Hull, E., Jackson, K., Dick, J., 2011. Requirements Engineering. Springer London, London.
65. Igba, J., Alemzadeh, K., Anyanwu-Ebo, I., Gibbons, P., Friis, J., 2013. A Systems Approach Towards Reliability-Centred Maintenance (RCM) of Wind Turbines. Procedia Computer Science 16, 814–823.
66. Igba, J., Alemzadeh, K., Gibbons, P.M., Henningsen, K., 2015. A framework for optimising product performance through feedback and reuse of in-service experience. Robotics and Computer-Integrated Manufacturing 36, 2–12.
67. ISO, 2007. International vocabulary of metrology - Basic and general concepts and associated terms (VIM). Beuth Verlag GmbH. https://www.iso.org/obp/ui/#iso:std:iso-iec:guide:99:ed-1:v2:en.
68. ISO, 2008. Software product Quality Requirements and Evaluation (SQuaRE) - Data quality model. Beuth Verlag GmbH, Berlin 35.080.
69. ISO, 2012. Data quality - Part 2: Vocabulary. Beuth Verlag GmbH, Berlin 01.040.25; 01.040.25.
70. ISO, 2015a. Quality management systems - Fundamentals and vocabulary. Beuth Verlag GmbH, Berlin 01.040.03; 03.120.10.
71. ISO, 2015b. Quality management systems - Requirements. Beuth Verlag GmbH, Berlin 03.120.10.

72. ISO, 2016. Data quality management: Process reference model. Beuth Verlag GmbH.
73. ISO, 2017. Systems and software engineering - Vocabulary, 2nd ed. Beuth Verlag GmbH, New York, USA.
74. Kapur, K.C., Pecht, M., 2014. Reliability engineering. Wiley, Hoboken NJ, XIX, 489 S.
75. Kari, J., 2007. Conceptualizing the personal outcomes of information. Information Research 12, paper 292.
76. Karkouch, A., Mousannif, H., Al Moatassime, H., Noel, T., 2016. Data quality in internet of things: A state-of-the-art survey. Journal of Network and Computer Applications 73, 57–81.
77. Kim, N.-H., An, D., Choi, J.-H., 2017. Prognostics and Health Management of Engineering Systems. Springer International Publishing, Cham.
78. King, S., 2014. Big Data. Springer Fachmedien Wiesbaden, Wiesbaden.
79. Kiritsis, D., 2011. Closed-loop PLM for intelligent products in the era of the Internet of things. Computer-Aided Design 43, 479–501.
80. Kiritsis, D., Bufardi, A., Xirouchakis, P., 2003. Research issues on product lifecycle management and information tracking using smart embedded systems. Advanced Engineering Informatics 17, 189–202.
81. Klesse, M., Herrmann, C., Braendli, P., Muegeli, T., Maier, D., 2004. Information Quality and the Raising Demands of Regulators: Reengineering the Customer Investigation Process at Credit Suisse, in: Proceedings of the 9th International Conference on Information Quality. International Conference on Information Quality, Cambridge, Massachusetts, USA. 5-7 November, pp. 418–432.
82. Kotler, P., Keller, K.L., 2012. Marketing management, 14th ed. Pearson, Boston.
83. Krcmar, H., 2015. Informationsmanagement, 6th ed. Springer Gabler, Wiesbaden, 793 pp.
84. Krogstie, J., Lindland, O.I., Sindre, G., 1995. Defining quality aspects for conceptual models. In: Falkenberg, E.D., Hesse, W., Olivé, A. (Eds.) Information System Concepts. Springer US, Boston, MA, pp. 216–231.
85. Kubat, M., 2015. An Introduction to Machine Learning. Springer International Publishing, Cham.
86. Kuestner, C., Wartzack, S., 2015. The realization of an engineering assistance system for the development of noise-reduced rotating machines, in: Proceedings of the 20th International Conference on Engineering Design (ICED15). The 20th International Conference on Engineering Design (ICED15), Milan, Italy. 27-30 July, 2015. Design Society, Glasgow, Scotland, pp. 71–80.
87. Kuhlthau, C., 1993. A Principle of Uncertainty for Information Seeking. Journal of Documentation 49, 339–355.
88. Kumar, S., Shah, N., 2018. False Information on Web and Social Media: A Survey. https://arxiv.org/pdf/1804.08559.
89. Lazar, J., Feng, J.H., Hochheiser, H., 2017. Research Methods in Human-Computer Interaction, 2nd ed. Morgan Kaufmann.
90. Leckie, G.J., Pettigrew, K.E., Sylvain, C., 1996. Modeling the Information Seeking of Professionals: A General Model Derived from Research on Engineers, Health Care Professionals, and Lawyers. The Library Quarterly: Information, Community, Policy 66, 161–193.
91. Lenzerini, M., 2002. Data integration: A theoretical perspective, in: Proceedings of the twenty-first ACM SIGMOD-SIGACT-SIGART symposium on principles of database systems. the twenty-first ACM SIGMOD-SIGACT-SIGART symposium, Madison, Wisconsin. 3/6/2002 - 5/6/2002. ACM, New York, NY, pp. 233–246.
92. Leopold, T., 2012. Ganzheitliche Datenerfassung für verbesserte Zuverlässigkeitsanalysen. Dissertation. IMA, Stuttgart, 127 pp.

93. Lim, S., Tucker, C.S., 2016. A Bayesian Sampling Method for Product Feature Extraction From Large-Scale Textual Data. J. Mech. Des 138, 61403.
94. Liu, L., Chi, L., 2002. Evolutional Data Quality: A Theory-Specific View, in: Proceedings of the Seventh International Conference on Information Quality. International Conference on Information Quality, MIT Sloan School of Management, Cambridge, MA, USA. November 8-10, 2002, pp. 292–304.
95. Lundqvist, M., Sandkuhl, K., Seigerroth, U., Holmquist, E., 2011. Method for Information Demand Analysis: IDA User Guide and Handbook. Jönköping University.
96. Luthje, C., Herstatt, C., 2004. The Lead User method: an outline of empirical findings and issues for future research. R&D Management 34, 553–568.
97. Lützenberger, J., Klein, P., Hribernik, K., Thoben, K.-D., 2016. Improving Product-Service Systems by Exploiting Information From The Usage Phase. A Case Study. Procedia CIRP 47, 376–381.
98. Madhikermi, M., Kubler, S., Robert, J., Buda, A., Främling, K., 2016. Data quality assessment of maintenance reporting procedures. Expert Systems with Applications 63, 145–164.
99. Magniez, C., Brombacher, A.C., Schouten, J., 2009. The use of reliability-oriented field feedback information for product design improvement: A case study. Qual. Reliab. Engng. Int. 25, 355–364.
100. Mahlamäki, K., Rämänen, J., 2015. Factors Influencing the Quality of Manually Acquired Asset Data. In: Amadi-Echendu, J., Hoohlo, C., Mathew, J. (Eds.) 9th WCEAM Research Papers, vol. 75. Springer International Publishing, Cham, pp. 273–284.
101. Markeset, T., Kumar, U., 2005. Product support strategy: Conventional versus functional products. J of Qual in Maintenance Eng 11, 53–67.
102. Martens, B., Streel, A. de, Graef, I., Tombal, T., Duch-Brown, N., 2020. Business-to-Business data sharing: An economic and legal analysis. JRC Technical Reports 2020-05. European Commission.
103. Mayer, R.J., Crump, J.W., IV, Fernandes, R., Keen, A., Painter, M.K., 1995. Information Integration for Concurrent Engineering (IICE) Compendum of Methods Report: Interim Technical Paper for Period February 1991 to March 1995 F33615-90-C-0012. Knowledge Based Systems Inc.
104. Mayring, P., 2000. Qualitative Content Analysis. Forum Qualitative Sozialforschung / Forum: Qualitative Social Research, Vol 1, No 2 (2000): Qualitative Methods in Various Disciplines I: Psychology. https://doi.org/10.17169/fqs-1.2.1089.
105. Mentzas, G., Hribernik, K., Thoben, K.-D., Kiritsis, D., Mousavi, A., 2018. Part 6 Summary: Predictive Maintenance in Industry 4.0: Methodologies, Tools and Interoperable Applications. In: Zelm, M., Jaekel, F.-W., Doumeingts, G., Wollschlaeger, M. (Eds.) Enterprise Interoperability. John Wiley & Sons, Inc, Hoboken, NJ, USA, pp. 233–237.
106. Messias, J., Schmidt, L., Oliveira, R., Benevenuto, F., 2013. You followed my bot!: Transforming robots into influential users in Twitter. First Monday - Peer-reviewed journal on the Internet 18.
107. Mey Goh, Y., McMahon, C., 2009. Improving reuse of in-service information capture and feedback. Jnl of Manu Tech Mnagmnt 20, 626–639.
108. Meyer, M., Wiederkehr, I., Koldewey, C., Dumitrescu, R., 2021. Understanding Usage Data-Driven Product Planning: A Systematic Literature Review. Proc. Des. Soc. 1, 3289–3298.
109. Mitchell, M.L., Jolley, J.M., 2013. Research design explained, 8th ed. Wadsworth, Cengage Learning, Belmont, CA, xxiii, 723.

110. Moody, D., Walsh, P., 1999. Measuring The Value Of Information: An Asset Valuation Approach, in: Proceedings of the 7th European conference on information systems. European Conference on Information Systems, Copenhagen, Denmark. June 23-25, 1999, pp. 1–17.
111. N.n., 2016a. ADAC Untersuchung: Datenkrake Pkw. ADAC e.V. https://www.adac.de/infotestrat/adac-im-einsatz/motorwelt/datenkrake_auto.aspx. Accessed 3 January 2019.
112. N.n., 2016b. ADAC Untersuchung an vier Fahrzeugen: Welche Daten erzeugt ein modernes Auto? Allgemeiner Deutscher Automobil-Club. https://www.adac.de/infotestrat/technik-und-zubehoer/fahrerassistenzsysteme/daten_im_auto/default.aspx. Accessed 4 February 2018.
113. N.n., 2016c. Welche Daten erzeugt ein modernes Auto?: ADAC Untersuchung an vier Fahrzeugen. ADAC e.V. https://www.adac.de/infotestrat/technik-und-zubehoer/fahrerassistenzsysteme/daten_im_auto/default.aspx. Accessed 3 January 2019.
114. N.n., 2017a. Feedback Mechanisms Across the Lifecycle for Customer-driven Optimization of Innovative Product-Service Design (Falcon). Falcon Consortium. https://cordis.europa.eu/project/id/636868. Accessed 16 October 2022.
115. N.n., 2017b. Project Deliverable 3.2: The FALCON Ontology. Falcon Consortium. http://www.falcon-h2020.eu/wp-content/uploads/2017/05/FALCON_Deliverable_D3-2_Final.pdf. Accessed 19 January 2022.
116. N.n., 2018. Product Service Design and Manufacturing Intelligence Engineering Platform Manutelligence (Manutelligence). Manutelligence Consortium. https://cordis.europa.eu/project/id/636951. Accessed 16 October 2022.
117. N.n., 2019. Upgrade yourself – the Toolbox App. Robert Bosch GmbH. https://www.bosch-professional.com/static/specials/upgrade/gb/en/. Accessed 16 September 2022.
118. N.n., 2020a. Cyber defence. North Atlantic Treaty Organization. https://www.nato.int/cps/en/natohq/topics_78170.htm. Accessed 25 October 2020.
119. N.n., 2020b. Privacy Notice: Connected Services. Toyota Motor Sales, U.S.A., Inc. https://www.toyota.com/privacyvts/. Accessed 4 October 2020.
120. N.n., 2020c. ReviewMeta. Review Honesty Inc. https://reviewmeta.com/. Accessed 25 October 2020.
121. N.n., 2022a. Automotive Sensors. Bourns. https://www.bourns.com/products/automotive/automotive-sensors. Accessed 17 January 2022.
122. N.n., 2022b. BOB Community. Robert Bosch Power Tools GmbH. https://www.bosch-professional.com/de/de/community/. Accessed 10 September 2022.
123. N.n., 2022c. Commercial Vehicle Sensors. Bourns. https://www.bourns.com/products/automotive/commercial-vehicle-sensors. Accessed 10 September 2022.
124. N.n., 2022d. Customer Privacy Notice: Information from or about your Tesla vehicle. Tesla Motors Inc. https://www.tesla.com/legal/privacy#vehicle-data. Accessed 10 September 2022.
125. N.n., 2022e. Product Information Management. SDZeCOM GmbH & Co. KG. https://www.sdzecom.de/loesungen/product-information-management/. Accessed 16 September 2022.
126. Ni, J., Li, J., McAuley, J., 2019. Justifying Recommendations using Distantly-Labeled Reviews and Fine-Grained Aspects, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Conference on Empirical Methods in Natural Language Processing (EMNLP), Hong Kong, China. Association for Computational Linguistics, Stroudsburg, PA, USA, pp. 188–197.

127. O'Connor, P.D.T., Kleyner, A., 2012. Practical reliability engineering, 5th ed. Wiley-Blackwell, Oxford, 484 pp.
128. Oehlmann, R., Thoben, K.-D., Weber, F., 1997. Capturing and assessing formal interaction at task level in product development. Computers in Industry 33, 179–189.
129. Oriol, M., Stade, M., Fotrousi, F., Nadal, S., Varga, J., Seyff, N., Abello, A., Franch, X., Marco, J., Schmidt, O., 2018. FAME: Supporting Continuous Requirements Elicitation by Combining User Feedback and Monitoring, in: 2018 IEEE 26th International Requirements Engineering Conference (RE). 2018 IEEE 26th International Requirements Engineering Conference (RE), Banff, AB. 20.08.2018 - 24.08.2018. IEEE, pp. 217–227.
130. Ortegon, K., Nies, L., Sutherland, J.W., 2016. Remanufacturing. In: Produ, T.I.A.f., Laperrière, L., Reinhart, G. (Eds.) CIRP Encyclopedia of Production Engineering. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 1–4.
131. Oxford University Press, 2017. Oxford dictionary in english language. Oxford University Press. https://en.oxforddictionaries.com/.
132. Pahl, G., Beitz, W., Blessing, L., Feldhusen, J., Grote, K.-H., Wallace, K. (Eds.), 2007. Engineering Design: A Systematic Approach. Springer-Verlag London Limited, London.
133. Peffers, K., Tuunanen, T., Rothenberger, M.A., Chatterjee, S., 2007. A Design Science Research Methodology for Information Systems Research. Journal of Management Information Systems 24, 45–77.
134. Picot, A., Reichwald, R., Wigand, R., 2001. Die grenzenlose Unternehmung, 4th ed. Gabler Verlag, Wiesbaden.
135. Picot, A., Reichwald, R., Wigand, R., 2008. Information, Organization and Management. Springer-Verlag Berlin Heidelberg, Berlin, Heidelberg.
136. Plungis, J., 2018. Who Owns the Data Your Car Collects? Consumer Reports, Inc. https://www.consumerreports.org/automotive-technology/who-owns-the-data-your-car-collects/.
137. Podsakoff, P.M., MacKenzie, S.B., Lee, J.-Y., Podsakoff, N.P., 2003. Common method biases in behavioral research: a critical review of the literature and recommended remedies. The Journal of applied psychology 88, 879–903.
138. Podsakoff, P.M., MacKenzie, S.B., Podsakoff, N.P., 2012. Sources of method bias in social science research and recommendations on how to control it. Annual review of psychology 63, 539–569.
139. Podsakoff, P.M., Organ, D.W., 1986. Self-Reports in Organizational Research: Problems and Prospects. Journal of Management 12, 531–544.
140. Polanyi, M., 1958. Personal knowledge: Towards a post-critical philosophy, 1962nd ed. Routledge.
141. Porter, M.E., 1985. Technology and Competitive Advantage. Journal of Business Strategy 5, 60–78.
142. Porter, M.E., 1991. Towards a dynamic theory of strategy. Strat. Mgmt. J. 12, 95–117.
143. Porter, M.E., 1998. Competitive advantage: Creating and sustaining superior performance, 1st ed. Free Press, New York, 557 pp.
144. Porter, M.E., Heppelmann, J.E., 2014. How Smart, Connected Products Are Transforming Competition. Harward Business Review, 1–23.
145. 2008. Proceedings of the 2008 IEEE International Professional Communication Conference (IPCC 2008). IEEE, Piscataway, NJ.
146. Pumacy GmbH, 2021. Recalls, Investigations and Complaints regarding vehicles worldwide. Pumacy GmbH. https://carwatch.net/. Accessed 23 July 2021.
147. Raghupathi, D., Yannou, B., Farel, R., Poirson, E., 2015. Customer sentiment appraisal from user-generated product reviews: A domain independent heuristic algorithm. Int J Interact Des Manuf 9, 201–211.

148. Rämänen, J., Mahlamäki, K., Borgman, J., Nieminen, M., 2013. Human Role in Industrial Installed Base Information Gathering. Procedia CIRP 11, 406–411.
149. Rockart, J.F., 1979. Chief executives define their own data needs. Harvard Business Review 57, 81–93.
150. Rowley, J., 2007. The wisdom hierarchy: Representations of the DIKW hierarchy. Journal of Information Science 33, 163–180.
151. Rozin, P., Royzman, E.B., 2001. Negativity Bias, Negativity Dominance, and Contagion. Personality and Social Psychology Review 5, 296–320.
152. Saaksvuori, A., Immonen, A., 2008. Product Lifecycle Management. Springer Berlin Heidelberg, Berlin, Heidelberg.
153. Sampson, S.E., 1996. Ramifications of Monitoring Service Quality Through Passively Solicited Customer Feedback. Decision Sciences 27, 601–622.
154. Sánchez López, T., Ranasinghe, D.C., Harrison, M., McFarlane, D., 2012. Adding sense to the Internet of Things. Pers Ubiquit Comput 16, 291–308.
155. Sandkuhl, K., 2011. Information Demand Patterns: Capturing Organizational Knowledge about Information Flow. In: Laux, F. (Ed.) PATTERNS 2011. The Third International Conferences on Pervasive Patterns and Applications : September 25-30, 2011, Rome, Italy. IARIA, Wilmington, DE, USA, pp. 1–6.
156. Sandkuhl, K., Seigerroth, U., 2018. Method engineering in information systems analysis and design: a balanced scorecard approach for method improvement. Softw Syst Model 2, 1–25.
157. Saracevic, T., 2016. The Notion of Relevance in Information Science: Everybody knows what relevance is. But, what is it really? Synthesis Lectures on Information Concepts, Retrieval, and Services 8, i-109.
158. Schulte, S., 2006. Integration von Kundenfeedback in die Produktentwicklung zur Optimierung der Kundenzufriedenheit. Dissertation. Shaker, Bochum, Germany, VIII, 230 Seiten.
159. Sendler, U., 2009. Das PLM-Kompendium: Referenzbuch des Produkt-Lebenszyklus-Managements. Springer Berlin Heidelberg, Berlin, Heidelberg.
160. Shin, J.-H., Jun, H.-B., Catteneo, C., Kiritsis, D., Xirouchakis, P., 2015. Degradation mode and criticality analysis based on product usage data. Int J Adv Manuf Technol 78, 1727–1742.
161. Stone, T.M., Choi, S.-K., 2013. Consumer preference estimation from Twitter classification: Validation and uncertainty analysis, in: Proceedings of the 19th International Conference on Engineering Design (ICED13). Design for harmonies: Human Behavior in Design. The 19th International Conference on Engineering Design (ICED13), Seoul, Korea. 19-22 August, 2013. Design Society, Castle Cary, Somerset, pp. 457–466.
162. Strong, D.M., Lee, Y.W., Wang, R.Y., 1997. Data quality in context. Commun. ACM 40, 103–110.
163. Tao, F., Sui, F., Liu, A., Qi, Q., Zhang, M., Song, B., Guo, Z., Lu, S.C.-Y., Nee, A.Y.C., 2019. Digital twin-driven product design framework. International Journal of Production Research 57, 3935–3953.
164. Taylor, R.S., 1968. Question-Negotiation and Information Seeking in Libraries. CRL 29, 178–194.
165. Thiruvenkadam, G., Brombacher, A.C., Lu, Y., den Ouden, E., 2008. Usability of consumer-related information sources for design improvement, in: Proceedings of the 2008 IEEE International Professional Communication Conference (IPCC 2008). 2008 IEEE International Professional Communication Conference (IPCC 2008), Montreal, QC. 13-16 July, 2008. IEEE, Piscataway, NJ, pp. 1–7.

166. Tucker, C., Kim, H., 2011. Predicting emerging product design trend by mining publicly available customer review data, in: Proceedings of the 18th International Conference on Engineering Design (ICED11). Impacting Society through Engineering Design. The 18th International Conference on Engineering Design (ICED11), Copenhagen, Denmark. 15-19 August 2011. Design Society, pp. 43–52.
167. Tukker, A., 2004. Eight types of product–service system: Eight ways to sustainability? Experiences from SusProNet. Bus. Strat. Env. 13, 246–260.
168. Tukker, A., van den Berg, C., Tischner, U., 2006. Product-services: a specific value proposition. In: Tukker, A., Tischner, U. (Eds.) New business for old Europe. Product-service development, competitiveness and sustainability. Greenleaf Pub, Sheffield, South Yorkshire, England, pp. 22–34.
169. U.S. Consumer Product Safety Commission, 2021. Duty to Report to CPSC: Rights and Responsibilities of Businesses. https://www.cpsc.gov/Business--Manufacturing/Recall-Guidance/Duty-to-Report-to-the-CPSC-Your-Rights-and-Responsibilities. Accessed 26 July 2021.
170. U.S. Department of Transportation, 2017. Motor Vehicle Safety Defects And Recalls: What Every Vehicle Owner Should Know DOT HS 808 795. National Highway Traffic Safety Administration. https://www.nhtsa.gov/sites/nhtsa.gov/files/documents/14218-mvsdefectsandrecalls_041619-v2-tag.pdf.
171. Uckelmann, D., Harrison, M., Michahelles, F., 2011. Architecting the Internet of Things. Springer Berlin Heidelberg, Berlin, Heidelberg.
172. Vakkari, P., 2003. Task-based information searching. Ann. Rev. Info. Sci. Tech. 37, 413–464.
173. Verein Deutscher Ingenieure, 2019. Design of technical products and systems - Model of product design, 2019th ed. Beuth Verlag GmbH, Berlin 03.100.40.
174. Vichare, N., Rodgers, P., Eveloy, V., Pecht, M., 2007. Environment and Usage Monitoring of Electronic Products for Health Assessment and Product Design. Quality Technology & Quantitative Management 4, 235–250.
175. Voß, S., Gutenschwager, K., 2001. Informationsmanagement. Springer Berlin Heidelberg, Berlin, Heidelberg.
176. Vuori, V., 2006. Methods of Defining Business Information Needs, in: Proceedings of the International Conference on Electronic Business (ICEB). International Conference on Electronic Business, Tampere, Finland. 2006.
177. Wang, R.Y., Strong, D.M., 1996. Beyond Accuracy: What Data Quality Means to Data Consumers. Journal of Management Information Systems 12, 5–33.
178. Watkins, R., West Meiers, M., Visser, Y.L., 2012. A Guide to Assessing Needs: Essential Tools for Collecting Information, Making Decisions, and Achieving Development Results.
179. Weber, F., 2005. Formale Interaktionsanalyse: Ein Beitrag zur systematischen Gestaltung von Informations- und Kommunikationsstrukturen im Concurrent Enterprise durch die Berücksichtigung von Informationseigenschaften. Dissertation, Bremen.
180. Weiser, M., 1999. The computer for the 21 st century. SIGMOBILE Mob. Comput. Commun. Rev. 3, 3–11.
181. Wellsandt, S., Hribernik, K., Thoben, K.-D., 2015a. Content analysis of product usage information from embedded sensors and web 2.0 sources: A first analysis of practical examples, in: Proceedings of the 21st ICE/IEEE International Technology Management Conference. IEEE International Conference on Engineering, Technology and Innovation/ International Technology Management Conference (ICE/ITMC), Belfast, Ireland. 22-24 June, pp. 1–9.
182. Wellsandt, S., Hribernik, K., Thoben, K.-D., 2015b. Sources and Characteristics of Information about Product Use. Procedia CIRP 36, 242–247.

183. Wellsandt, S., Nabati, E., Wuest, T., Hribernik, K.A., Thoben, K.D., 2016. A survey of product lifecycle models: Towards complex products and service offers. IJPLM 9, 353.
184. Wellsandt, S., Stietencron, M. von, Hribernik, K., Henriksen, B., Røstad, C.C., Thoben, K.-D., 2015c. Fact-based Design for Leisure Boats: The HighSea-experiment Setup. Procedia CIRP 38, 74–77.
185. Wellsandt, S., Thoben, K.-D., 2016. Approach to Describe Knowledge Sharing between Producer and User. Procedia CIRP 50, 20–25.
186. Wellsandt, S., Thoben, K.-D., Klein, P., 2018. Information feedback in product development: Analysing practical cases, in: DS 92: Proceedings of the DESIGN 2018 15th International Design Conference. 15th International Design Conference. May, 21-24, 2018. Faculty of Mechanical Engineering and Naval Architecture, University of Zagreb, Croatia; The Design Society, Glasgow, UK, pp. 1745–1756.
187. Wilberg, J.M.M., 2019. From data to value: Facilitating strategy development for connected products. Dissertation, Germany.
188. Wilson, T.D., 2000. Human Information Behavior. Informing Science 3, 49–56.
189. Wilson, T.D., 2002. The nonsense of 'knowledge management'. Information Research 8, paper no. 144.
190. Wollnik, M., 1988. Ein Referenzmodell des Informationsmanagements. Information Management 3, 34–43.
191. Wynn, D.C., Eckert, C.M., 2017. Perspectives on iteration in design and development. Res Eng Design 28, 153–184.
192. Yarmey, A.D., 1979. The psychology of eyewitness testimony. Free Press, New York, N.Y.
193. Zins, C., 2007. Conceptual approaches for defining data, information, and knowledge. J. Am. Soc. Inf. Sci. 58, 479–493.

Appreciation of student contributions

This dissertation contains results that were obtained in the course of supervising the following student projects:

Master's theses

Miao Guo (2013): *Ermittlung der Anwendungspotentiale produktintegrierter Sensoren im Bereich der Anforderungsermittlung*.

Kotschei Borgy (2013): *Auswirkungen von Qualitätsunterschieden von Produktdaten auf betriebliche IT-Systeme am Beispiel eines Produktkonfigurators*.

David Trzebiatowski (2013): *Ansatz zur Reduzierung von Fehlern im Produkt-Anforderungsmanagement bei modularen Baukastensystemen*.

Matthias Pantleon (2017): *Mitarbeitermotivation als Einflussfaktor beim Umgang mit Nutzungsinformationen in der Produktentwicklung*.

André Gerdes (2017): *Ansatz zur Planung und Umsetzung eines Produkt-Avatars für die Produktentwicklung*.

Fatemeh Talati Alishah (2018): *Informationen in der Produktion – Nutzungssituationen der Produkte in sozialen Medien*.

Waldemar Gellert (2018): *Künstliche Intelligenz in der Produktion – Entwicklung eines Konzeptes zum Umgang mit Qualitätsunterschieden bei Produktionsdaten*.

Onur Gezen (2018): *Entwicklung einer Methode zur Identifikation von Kommunikationsproblemen im Kontext der Produktentwicklung*.

Steven Brümmerhoff (2019): *Entwicklung einer Methode zur Analyse von Datenqualitätsproblemen am Beispiel eines virtuellen B2B-Marktplatzes*.

Nikolai Becker (2019): *Engineering 4.0 – Chancen, Herausforderungen und Rahmenbedingungen von Nutzungsinformationen in der Produktentwicklung*.

Bachelor's theses

Johannes Finder (2013): *Klassifizierung von Methoden und Techniken zur Beschreibung von Produktnutzungsprozessen*.

Student projects

Regina Bektev, Marian Göbel, Marie-Louise Heidbrink, Philipp Maron, Fatemah Talati Alishah (2017). *Nutzungsinformationen in der Produktentwicklung*.

List of figures

List of tables

Appendix

This Appendix provides details that allow readers to investigate my arguments and decisions. Sections A-C focus on Chapters 2-3, and the following ones contain the text passages I used to identify the PUI quality characteristics and quality-related factors. The final section L summarizes relations between information characteristics – an aspect relevant for this dissertation's outlook.

A. Conditioning variables of Freund's information seeking model

Table 23: Conditioning variables for information needs of software engineers (Freund, 2015)

Types	Variables	Descriptions
Requirements	Level of detail	Conceptual level, depth and scope of information
	Sanction	Distinction between approved information and information from unofficial sources. The latter is considered more accurate and reliable
	Situatedness	Degree to which an information item is gained through the firsthand experience of the author versus information based on theoretical understanding
	Specificity	Degree to which information relates to a particular situation compared as opposed to generic information applicable to various situations
	Purpose	Refers to what the information was created for
Constraints	Awareness	Seeker's awareness of information
	Findability	How easy it is to find the needed information
	Abundance	Quantity of information available
	Cooperation	Level of cooperation between people sharing information
	Permission	Accessibility of the information
	Time	Available amount of time to seek out and use the information

B. Information quality models

Table 24 provides the proposed characteristics of data and information from (ISO, 2008), (Strong et al., 1997) and (Eppler, 2006). Each reference uses a different term to refer to characteristics: the ISO standard uses "quality model characteristics", Strong et al. prefer "quality dimension", and Eppler suggests "quality criteria". This dissertation adopts the ISO 9000 standard series terminology that uses the general term "characteristics".

Table 24: Examples of data and information quality characteristics

Similarity	ISO 25012:2008	Strong et al.	Eppler
Similar	Accessibility	Accessibility	Accessibility
	Accuracy	Accuracy	Accuracy
	Confidentiality	Access security	Security
	Consistency	Consistent representation	Consistency
	Completeness	Completeness	Comprehensiveness
	Credibility	Reputation	Correctness
	Currentness	Timeliness	Timeliness
	Understandability	Ease of understanding	Clarity
	Traceability	-	Traceability
	-	Concise representation	Conciseness
Different	Availability	Amount of data	Applicability
	Compliance	Believability	Convenience
	Efficiency	Interpretability	Currency
	Portability	Objectivity	Interactivity
	Precision	Relevancy	Maintainability
	Recoverability	Value-Added	Speed

The first column of the table indicates the two similarity categories. Eight items are similar in their wording and meaning among the three models. Traceability and conciseness are present in two of the models.

C. Overview of the application cases

Table 25 summarizes the application cases investigated in this dissertation's state-of-the-art analysis. The table contains a case identifier, the identified PD objective, targeted products, communication channels, and the reference.

Table 25: Application case descriptions covered in the state-of-the-art analysis

No.	Objective	Products	Channels	Reference
1	Product defect discovery	Cars and car components	Online discussion forum	(Abrahams et al., 2015)
2	Vehicle component isolation within a text	Cars and car components	Online discussion forum	(Abrahams et al., 2013)
3	Support of product failure analysis	Manufacturing machine components	Embedded sensors, service reports	(Abramovici and Lindner, 2011)
4	Support of product failure analysis	Hydraulic systems	Embedded sensors, service reports	(Abramovici et al., 2011)
5	Identification of user needs	Hearing aids	Embedded sensors, self-reports	(Aldaz et al., 2013)

No.	Objective	Products	Channels	Reference
6	Degradation and wear analysis	Furniture	Embedded sensors	(Bleda et al., 2012)
7	Support of failure analysis	Fuel injection of a car	Warranty claims	(Bueno and Borsato, 2014)
8	Redesign of manufacturing operations	Manufacturing machines	Embedded sensors	(D'Antonio et al., 2015)
9	Support of quality improvement methods	Fly fishing equipment	Mailing lists	(Finch and Luebbe, 1997)
10	Support of quality improvement methods	Power tools	Usenet*	(Finch, 1999)
11	Identify consumer perceptions and usage intent	Shoes	Embedded sensors, customer surveys	(Ghosh et al., 2016)
12	Optimize product performance	Wind turbine gearbox	Embedded sensors, service reports	(Igba et al., 2013)
13	Optimize product performance	Gearbox	"codified" and "personalized" knowledge	(Igba et al., 2015)
14	Noise reduction	Wind turbine	Embedded sensors	(Kuestner and Wartzack, 2015)
15	Accurate product feature extraction from text	Smartphones	Micro weblog	(Lim and Tucker, 2016)
16	Improve reliability	Not mentioned	Embedded sensors, service reports	(Magniez et al., 2009)
17	Predict sentiment from online product reviews	Home theater system	Online product reviews	(Raghupathi et al., 2015)
18	Improve product design or maintenance through degradation analysis	Lift arm of heavy construction equipment	Embedded sensors	(Shin et al., 2015)
19	Estimate consumer preference for particular product characteristics	Smartphone	Micro weblog	(Stone and Choi, 2013)
20	Understand root causes of customer dissatisfaction	Wireless audio product and an ironing system	Help Desk, Online discussion forums	(Thiruvenkadam et al., 2008)
21	Identify customer preference trends	Smartphone	Product reviews	(Tucker and Kim, 2011)
22	Improve thermal design	Notebooks	Embedded sensors	(Vichare et al., 2007)

**Usenet is an independent service of the Internet. The World Wide Web service has online discussion forums with similar functionality.*

Some references relate to the same or a similar application case. Table 26 summarizes the unique application cases and the related references.

Table 26: Unique application cases identified in the literature

Application case	**References**	**Relation of references**
Product defect discovery for vehicles	(Abrahams et al., 2013; Abrahams et al., 2015)	The papers contain consecutive research for the same application case.
Support of root-cause analysis for machine components	(Abramovici et al., 2011; Abramovici and Lindner, 2011)	The papers apply one approach for feedback integration to different products.
Identification of user needs	(Aldaz et al., 2013)	No relation.
Measurement of furniture use and deterioration	(Bleda et al., 2012)	No relation.
Detect and evaluate failure patterns	(Bueno and Borsato, 2014)	No relation.
Production process improvement	(D'Antonio et al., 2015)	No relation.
Support of quality improvement methods	(Finch and Luebbe, 1997; Finch, 1999)	The papers contain consecutive research for the same application case.
Understand and model user perceptions	(Ghosh et al., 2016)	No relation.
Optimize product performance	(Igba et al., 2013; Igba et al., 2015)	Both papers describe a framework to utilize in-service data for wind turbine improvement.
Noise-reduced wind turbine design	(Kuestner and Wartzack, 2015)	No relation.
Customer preference modeling	(Tucker and Kim, 2011; Stone and Choi, 2013; Lim and Tucker, 2016)	All papers focus on cell phones and extracting product features from the Internet. Two contain consecutive research.
Reliability engineering	(Magniez et al., 2009)	No relation.
User appraisal analysis	(Raghupathi et al., 2015)	No relation.
Degradation mode and criticality analysis	(Shin et al., 2015)	No relation.
Understand root causes of customer dissatisfaction	(Thiruvenkadam et al., 2008)	No relation.
Reliability engineering	(Vichare et al., 2007)	No relation.

D. Categorization according to quality characteristics

Table 27 summarizes the categorization of text passages according to the data quality model provided in ISO 25012:2008.

Table 27: Categorization of text passages according to ISO 25012:2008

No.	Characteristics	Text passages from application case descriptions	References
1	Completeness	"Periodic sensing can reduce power consumption relative to continuous monitoring, but presents a risk of loosing (sic) critical data."	(Vichare et al., 2007)
2		"Completeness of a consumer information source as a whole. This refers to the extent to which data is not missing and is of sufficient breadth and depth for the task at hand."	(Thiruvenkadam et al., 2008)
		"den Ouden et al […] observed that the data in these forums give a more in-depth view of how consumers experience their products. They note that although the data are very rich in content with highly specific information about problem situations, they do not provide statistical information since they may not necessarily represent the total group of consumers that bought the product."	(Thiruvenkadam et al., 2008)
3		"Good feedback to design can be achieved through correct documentation and reporting of all field O&M activities, especially those that have to do with RCM."	(Igba et al., 2013)
4		"Measuring devices should be supported by mathematical techniques able to real-time integrate and analyze data collected from several heterogeneous sources, to provide a complete picture of the current state of the process and make available useful indications to improve the process itself."	(D'Antonio et al., 2015)
5		"Often there is lack of information in warranty databases, where the exact number of units that had the problem or the correct mileage of the vehicle that failed are unknown, or information about certain failure are recorded only for the specific units that presented problems and not for all related units […]."	(Bueno and Borsato, 2014)
6		"Usually, a study will aim to get an accurate and complete representation of a participant's life as it pertains to the issue being investigated."	(Aldaz et al., 2013)
7	Credibility	"The unstructured nature of online customer review data relieves respondents from the traditional predefined structure of a survey type approach and enables respondents to provide an unbounded assessment of their product preferences."	(Tucker and Kim, 2011)
8		"The resulting preference model must be valid. This means that the model must produce predictions that are close to the real preferences."	(Stone and Choi, 2013)

No.	Characteristics	Text passages from application case descriptions	References
9	Credibility	"We propose to validate our model by asking 38 humans to do exactly the same task that our model, i.e. to perform 15 rate reviews on a scale of 0-5."	(Raghupathi et al., 2015)
10		"With ground truth data, the performance of the proposed methodology can be compared to existing methods for accuracy to achieve a solution."[33]	(Lim and Tucker, 2016)
11		"Good feedback to design can be achieved through correct documentation and reporting of all field O&M activities, especially those that have to do with RCM."	(Igba et al., 2013)
12		"An increase in the popularity of using this type of information to aid in quality management practices would undoubtedly result in attempts to contaminate the data and provide disinformation."	(Finch, 1999)
13		"In the aeronautics market, traceability is a highly important requirement: the manufacturer should guarantee at any time accurate information about the product, processes, materials, stocks, etc. "	(D'Antonio et al., 2015)
14	Currentness	"In many applications, it is not always known a priori what parameters need to be measured, nor with what frequency or precision."	(Vichare et al., 2007)
15		"Internet forums, Frequently Asked Questions (FAQs) at service, helpdesk and many other related information sources can potentially provide valuable decision support information very quickly [...]."	(Thiruvenkadam et al., 2008)
16		"The information related to the reliability of products on the market should be provided as quickly as needed."	(Magniez et al., 2009)
17		"Designers aiming to utilize online product-related data will therefore be able to acquire and mine relevant product-feature related data, while minimizing noise and time challenges."	(Lim and Tucker, 2016)
18		"Moreover, collecting real-time sensor data can help the designers in analyzing trends and extracting patterns over an extended period to make more effective decisions."	(Ghosh et al., 2016)
19		"The fact that it can be obtained at any time, not just when customer interview processes are going on, could create a competitive advantage by giving the manufacturer an early jump on the development of the improvement."	(Finch and Luebbe, 1997)
20		"Timeliness of information is especially critical in environments characterized by shortened product life cycles. It can result in quicker identification and action for conformance problems as well as reduced time-to-market."	(Finch, 1999)

[33] Accuracy refers to how many of the ground truth data are correctly classified by the algorithm.

Appendix

No.	Characteristics	Text passages from application case descriptions	References
21	Currentness	"These measuring devices should be supported by mathematical techniques able to real-time integrate and analyze data collected from several heterogeneous sources, to provide a complete picture of the current state of the process and make available useful indications to improve the process itself."	(D'Antonio et al., 2015)
22		"Automated component classification […] is helpful as it allows new social media postings to be rapidly and accurately classified."	(Abrahams et al., 2013)
23		"In turn, the time-to-discovery for each defect can be significantly reduced; fewer defective products will reach consumer's hands; and large savings can consequently accrue to manufacturers."	(Abrahams et al., 2015)
24	Accessibility	"The main reason is the publicly (sic) of their data, available with Perl script API's."	(Raghupathi et al., 2015)
25		"In-service data is usually being managed in stand-alone databases […] making accessibility of data difficult for design engineers, in some cases engineers were not aware such data existed."	(Igba et al., 2015)
26		"[...] as the design team has access to the field databases only, and do not create or feed them, it generally cannot verify the accuracy of the information included in the databases, which may lead to difficulties in the assessment or interpretation of the information."	(Magniez et al., 2009)
27		"The only way to access the conversations among the members is to become a list member."	(Finch, 1999)
28	Confidentiality	"Longitudinal studies involving wearable computing may also pose serious privacy risks. […]. Studies must be designed to safeguard this sensitive data, both during the study and afterwards."	(Aldaz et al., 2013)
29	Efficiency	"The pre-partitioned format of the customer feedback platform greatly reduces the complexities that would have resulted from trying to mine the raw data for negative customer opinions."	(Tucker and Kim, 2011)
30	Accuracy	"In other words, this study aims for automatically compute or predict the overall sentiment rating from online reviews, with a good accuracy and without a tedious customization to a product domain or customer polarities."	(Raghupathi et al., 2015)
31		"Since all data related to the acoustical behaviour of a product need to be stored in the product data model to be analysed or exchanged, an accuracy classification before the import of the data in the EAS ALARM is crucial to guarantee correct analysis results. If the data is used as an input for analysis, the outputs precision is influenced by inputs quality."	(Kuestner and Wartzack, 2015)
32		"The data collected from Internet conversations are gathered in an unobtrusive way. This does not necessarily mean, however, that they are any more accurate than those gathered from a survey or that the person posting them is motivated to be more honest and unbiased."	(Finch and Luebbe, 1997)

No.	Characteristics	Text passages from application case descriptions	References
33	Accuracy	"The well-documented drawback of self-reporting is the relatively poor accuracy of the data, either by recall error or deliberate falsification."	(Aldaz et al., 2013)
34		"[...] it generally cannot verify the accuracy of the information included in the databases, which may lead to difficulties in the assessment or interpretation of the information."	(Magniez et al., 2009)
35		"The lack of meaningful information for the root-cause analysis using these databases results either from an oversimplification of the customer complaint (use of fault code that is very often a symptom description instead of a cause) or a lack of accuracy (use of free text leading to a difficult appreciation of the information)."	(Magniez et al., 2009)
36	Precision	"Threads containing less than 50 words or less than two postings were excluded from further analysis as they were found to contain too little information to describe and confirm a defect."	(Abrahams et al., 2015)
37		"Often there is lack of information in warranty databases, where the exact number of units that had the problem or the correct mileage of the vehicle that failed are unknown, or information about certain failure are recorded only for the specific units that presented problems and not for all related units […]."	(Bueno and Borsato, 2014)
38		"The Internet messages that do not mention a specific company are not likely to provide the specificity necessary to guide improvement efforts on existing products. However, as mentioned in Table I, many product-related messages are quite specific and not only mention a company, but also frequently provide model-specific information."	(Finch and Luebbe, 1997)
39		"This can be achieved if a computerised maintenance management system (CMMS) is used, where all failures, safety incidents and maintenance tasks are recorded in detail up to the sub-component level."	(Igba et al., 2013)
40		"To estimate the value of performance indicator in a precise manner, it is important to identify the most relevant parameters from available product data."	(Shin et al., 2015)
41		"den Ouden et al […] observed that the data in these forums give a more in-depth view of how consumers experience their products. They note that although the data are very rich in content with highly specific information about problem situations, they do not provide statistical information since they may not necessarily represent the total group of consumers that bought the product."	(Thiruvenkadam et al., 2008)
42		"In many applications, it is not always known a priori what parameters need to be measured, nor with what frequency or precision."	(Vichare et al., 2007)
43		"They noted that the emphasis of the call centre agents is to settle as many cases satisfactorily as possible and not so much on the process of documenting these cases in a precise manner."	(Thiruvenkadam et al., 2008)

No.	Characteristics	Text passages from application case descriptions	References
44	Understandability	"[...] it generally cannot verify the accuracy of the information included in the databases, which may lead to difficulties in the assessment or interpretation of the information."	(Magniez et al., 2009)
45		"The lack of meaningful information for the root-cause analysis using these databases results either from an oversimplification of the customer complaint (use of fault code that is very often a symptom description instead of a cause) or a lack of accuracy (use of free text leading to a difficult apprecia-tion of the information)."	
46		"Messages which contain query terms for more than one attribute level are omitted, since the subject of those messages is more ambiguous."	(Stone and Choi, 2013)
47		"Hence, when service personnel take unconventional decisions during maintenance, designers might be unable to reuse such information if they do not know the rationale behind."	(Igba et al., 2015)
48		"Unusable data, consisting of messages that were in a language other than English, messages that were garbled and unintelligible, and messages that were posted in duplicate, were removed from the data set."	(Finch, 1999)
49		"While studies have shown that Support Vector Machines and Neural Nets offer high accuracy […], support still exists for more easily interpretable and explanatory approaches like Logistic Regression […]."	(Abrahams et al., 2015)
50	Availability	"By integrating free, publicly available customer review data, the proposed design methodology can have wide applicability to many areas of product design."	(Tucker and Kim, 2011)
51		"[...] potentially much more information could be available for use in the development process pro-vided that it arrives in the right format, properly condensed and prioritized, at the desk of the devel-oper."	(Thiruvenkadam et al., 2008)
52		"To estimate the value of performance indicator in a precise manner, it is important to identify the most relevant parameters from available product data."	(Shin et al., 2015)
53		"The current state of the art is that product developers try to use data of pervious (sic) platform variants, but they are often not aware of which data of pervious (sic) platform variants is available and how it can be used to avoid tonalities."	(Kuestner and Wartzack, 2015)
54		"Since the use of the Internet is equally available to anyone, a long-term competitive advantage based on the utilization of this technology is not likely to result. However, since it is available, not using it to its potential may result in a competitive disadvantage."	(Finch and Luebbe, 1997)
55		"One way to support such inclusive approach is the deployment of information systems to efficiently manage design and production operations, and to maximize the efficacy of the available data."	(D'Antonio et al., 2015)

E. Quality characteristics not covered by ISO 25012:2008

Table 28 summarizes the quality characteristics identified in the application case descriptions but are not part of the ISO 25012:2008 data quality model. The names of characteristics represent terminology widely used in literature. Moody and Walsh (1999) mentioned value. Case and Given (2016) and Saracevic (2016) define relevance. Richness is a concept described by Daft and Lengel (1984). It is the cornerstone of the media richness theory.

Table 28: Identified quality characteristics not covered in ISO 25012:2008

No.	Characteristics	Text passages from application case descriptions	References
1	Value	"With all of this electronic "talking" going on, is any of it worth listening to? If so, what is valuable and what is not? And how can valuable information be used?"	(Finch and Luebbe, 1997)
2		"Uncovering value from social media comments involves a systematic process of gathering social media comments, text preprocessing, feature extraction, and advanced text modeling and analytics."	(Abrahams et al., 2015)
3		"Internet forums, Frequently Asked Questions (FAQs) at service, helpdesk and many other related information sources can potentially provide valuable decision support information very quickly [...]."	(Thiruvenkadam et al., 2008)
4		"Comparisons may offer a special value to a manufacturer when included in evaluations. They provide additional information for the person requesting an evaluation because the purchase decision is typically a choice among alternatives."	(Finch, 1999)
5	Relevance	"They concluded, based on the quantity, richness and relevance of data collected from an internet mailing list discussion forum (over a certain period of time about a hardware product) that the information gathered through internet discussions has significant potential for use by consumer product manufacturers."	(Thiruvenkadam et al., 2008)
6		"Increasingly, the design sector employs the weblogs and product review to target relevant information for designer [...]."	(Raghupathi et al., 2015)
7		"Designers aiming to utilize online product-related data will therefore be able to acquire and mine relevant product-feature related data, while minimizing noise and time challenges."	(Lim and Tucker, 2016)
8		"Preliminary field tests may be justified, to understand what data is actually relevant in a fully operating system."	(Vichare et al., 2007)
9		"Interesting information tends to be embedded in a mass of irrelevant data, and it requires patience and skill to extract it."	(Aldaz et al., 2013)

No.	Characteristics	Text passages from application case descriptions	References
10	Richness	"Without the sensor information, it would be virtually impossible for a designer to gather and understand this type of rich information about the product, its use, and its impact on the user."	(Ghosh et al., 2016)
11		"Either conversation in person or expressed in online text form, subjectivity and sentiment add richness to the shared information."	(Raghupathi et al., 2015)
12		"Traditional sources of feedback are not likely to contain comparisons with detailed evaluations of strengths and weaknesses. This is an attribute of the richness associated with this data that probably can not be obtained through traditional means "	(Finch, 1999)
13		"Communication media that can clarify ambiguities or change someone's understanding of an issue are considered the most rich. Numeric documents rank the lowest in richness. Face-to-face conversations rank the highest. Because of the necessary brevity of the instrument, passively solicited feedback might be considered less rich than actively solicited feedback. Internet conversations, although not face-to-face, may be considered more rich, and therefore attractive."	

F. Biases identified in the application case descriptions

Application case authors mention some bias explicitly. In other cases, they refer to complementary references. This dissertation relates all of the identified biases to the information collection activity. In practice, they may occur during other activities, such as a common method bias during the application of information.

Table 29: Identified biases

No.	Biases	Mentioned in references	Original references cited by authors
1	Self-selection	(Aldaz et al., 2013, p. 4; Abrahams et al., 2015, p. 984)	(Heckman, 1977)
2	Non-response	(Abrahams et al., 2015, p. 984)	(Armstrong and Overton, 1977; Duriau et al., 2007)
3	Negativity	(Abrahams et al., 2015, p. 984)	(Rozin and Royzman, 2001)
4	Single source	(Abrahams et al., 2015, p. 984)	(Podsakoff and Organ, 1986)
5	Common method	(Abrahams et al., 2015, p. 984)	(Podsakoff et al., 2003)
6	Selection/exclusion	(Abrahams et al., 2015, p. 984)	-
7	Product popularity	(Abrahams et al., 2015, p. 984)	-
8	Recall error	(Aldaz et al., 2013, p. 3)	(Yarmey, 1979)
9	Response	(Finch, 1999, p. 537)	(Sampson, 1996)
10	Sampling	(Finch and Luebbe, 1997, p. 862)	-
11	Bradley effect	(Raghupathi et al., 2015, p. 203)	*Original reference does not describe the Bradley effect.*
12	Algorithmic bias	(Raghupathi et al., 2015)	-

G. Factors related to collection quality

Table 30 summarizes the identified factors that relate to collection quality.

Table 30: Identified text passages and the derived factors related to collection quality

No.	Factors	Groups	Text passages
1	Customer sophistication	-	"[…] a segment of the customer base which is **more sophisticated** and product-aware **than the average customer** […]" (Finch and Luebbe, 1997, pp. 861–862)
2	Availability of context information	-	"For further analysis, each data set requires detailed **metadata** to describe under which **circumstances the data has been elicited**." (Kuestner and Wartzack, 2015, p. 6) "[...] another aspect of importance to in-service knowledge capture and feedback, which aids data reuse, is the capturing of **context and rationale** alongside the data itself." (Igba et al., 2015, p. 11)
3	Allowed message length	Channel	"The limitation of the use of tweet is its **shortness**. A consumer quickly limits his/her message to the binary answer of satisfaction or dissatisfaction." (Raghupathi et al., 2015, p. 2)
4	Control over medium	Channel	"The fact that Internet information is not being directed to a particular company's inquiries might result in the accuracy of Internet information being higher, however. Extending the issue of accuracy to that of controlling who information is received from presents the potential for a greater problem. With a survey, the **source of the information is controlled**." (Finch and Luebbe, 1997, p. 863)
5	Directedness	Channel	"Complaints are **directed**: service requests come in at the dealership, complaint emails are sent to the manufacturer, and complaint reports are filed for specific vehicle models at the National Highway Traffic Safety Administration (NHTSA). In contrast, defect reports in social media may be **haphazardly interspersed** among other content, such as reviews, comparisons, questions, and social discussions, and debates." (Abrahams et al., 2015, p. 979)
6	Free text	Channel	"[...] **free text** leading to a difficult appreciation of failure descriptions." (Magniez et al., 2009, p. 357)
7	Self-report design	Channel	"Lastly, since sensor data alone cannot provide important information such as user intention and perception, the immediate **self-reports must be carefully designed** with clarity and ease of use as a paramount objective." (Aldaz et al., 2013, p. 4)
8	Depth of discussion	Content	"It does not provide **depth** and diversity of discussion on a company or product that can be obtained from a company-based focus." (Finch, 1999, p. 540)

No.	Factors	Groups	Text passages
9	Diversity of discussion	Content	"It does not provide depth and **diversity** of discussion on a company or product that can be obtained from a company-based focus." (Finch, 1999, p. 540)
10	Language	Content	"Unusable data, consisting of messages that were in a **language** other than English, […], were removed from the data set." (Finch, 1999, p. 542)
11	Mentioning of a specific product	Content	"It would be critical for a firm monitoring Internet conversations for quality information because **messages that mention specific products** have the greatest potential to enhance product improvement efforts." (Finch, 1999, p. 542)
12	Oversimplification	Content	"The lack of meaningful information for the root-cause analysis using these databases results either from an **oversimplification of the customer complaint** (use of fault code that is very often a symptom description instead of a cause) […]." (Magniez et al., 2009, p. 357)
13	Presence of acronyms	Content	"The future improvements in SENTRAL will be on […], as well as its robustness in terms of the quality of input data: **presence of acronyms**, typographical errors, ironic and sarcastic expressions." (Raghupathi et al., 2015, p. 14)
14	Reporting errors	Content	"Although potential mistakes may stem from **errors during the reporting** in the field database (e.g. wrong part number, wrong fault code, bad diagnosis), changes are traced so that there is quantitative evidence of field problems." (Magniez et al., 2009, p. 358)
15	Semantic defects	Content	"This information is filtered and cleansed of **semantic** and syntactic **defects** […]." (Abramovici and Lindner, 2011, p. 212)
16	Syntactic defects	Content	"This information is filtered and cleansed of semantic and **syntactic defects** […]." (Abramovici and Lindner, 2011, p. 212)
17	Typographical errors	Content	"The future improvements in SENTRAL will be on […], as well as its robustness in terms of the quality of input data: presence of acronyms, **typographical errors**, ironic and sarcastic expressions." (Raghupathi et al., 2015, p. 14)
18	Duplicates	-	"Over 7,000 Twitter messages were collected […]—not counting **duplicate** messages which were omitted." (Stone and Choi, 2013, p. 4) "[…] messages that were posted in **duplicate**, were removed from the data set." (Finch, 1999, p. 542)
19	Accuracy of sensors	Information creation context - technical	"Multicomp moisture sensor: this device is of small dimensions with accuracy between -5 +5 %." (Bleda et al., 2012, p. 848)

No.	Factors	Groups	Text passages
20	Placement of sensors	Information creation context - technical	"When furniture is occupied, its temperature rises and depends on body temperature, so it is necessary to use a **very well integrated temperature sensor**. So the integration of temperature sensors in furniture structure in a ubiquitous way allows contacting directly with the user and obtaining realistic measurements about furniture temperature." (Bleda et al., 2012, p. 848)
21	Unobtrusive monitoring	Information creation context - technical	"If possible the **monitoring process should not be known to user** to avoid any kind of bias." (Vichare et al., 2007, p. 241)
22	Measurement frequency	-	"From the **two periods monitored** in this study, it seems clear that even though the information might be useful, it might not always be of sufficient quantity to direct change." (Finch, 1999, p. 862) "**Periodic sensing** can reduce power consumption relative to **continuous monitoring**, but presents a risk of loosing [sic] critical data." (Vichare et al., 2007, p. 240)
23	Defect detection lags	Measurement frequency	"Timeliness of defect discovery is limited by a number of significant lags, including the **lag between when the consumer acquires the product and when they observe the defect**, and the lag between observing the defect and reporting the defect to the organization or to the user community." (Abrahams et al., 2015, p. 985)
24	Reporting lag	Measurement frequency	"Timeliness of defect discovery is limited by a number of significant lags, including the lag between when the consumer acquires the product and when they observe the defect, and the **lag between observing the defect and reporting the defect** to the organization or to the user community." (Abrahams et al., 2015, p. 985)
25	Speed of data collection	Measurement frequency	"Equally as important as the feedback is the magnitude and speed (algorithm speed as well as **speed of customer review input)** at which this feedback can be acquired using the proposed online customer review process." (Tucker and Kim, 2011, p. 48)
26	Time frame of data	Measurement frequency	"It is of common sense within the organization, that **timeframes** of less than one year show no significant occurrences of faults that could be relevant for the study." (Bueno and Borsato, 2014, p. 126)
27	Providing disinformation	-	"The Internet's power is rooted in free access, but free access is also its weakness. An increase in the popularity of using this type of information to aid in quality management practices would undoubtedly result in attempts to contaminate the data and **provide disinformation**." (Finch, 1999, p. 554) "The well-documented drawback of self-reporting is the relatively poor accuracy of the data, either by recall error or **deliberate falsification**." (Aldaz et al., 2013, p. 3)

No.	Factors	Groups	Text passages
28	Hardware problem	Missing data	"A **technical problem** resulting from a change of Internet server hardware resulted in receiving no messages during the 11-day period from May 13 to May 24. Messages were accumulated for the entire year before the analysis took place." (Finch, 1999, p. 541)
29	Software bug	Missing data	"A **bug** limited the function of the test setup so that one environment was never detected. There is no data about this event." (Aldaz et al., 2013, p. 9)
30	Validity of information	-	"If we upload a condition monitoring system (CMS) measurement of a WT it is only **valid** for this single WT (assembly)." (Kuestner and Wartzack, 2015, p. 7)

H. Factors related to organization quality

Table 31 summarizes the identified factors that relate to organization quality.

Table 31: Identified text passages and the derived factors related to organization quality

No.	Factors	Groups	Text passages
1	Data format	-	"Customer review data in **html format** is stored on an SQL database on the web site which the authors developed. […]. This is a **non-trivial problem** as customer reviews are presented in an **unstructured, unpredictable style**." (Tucker and Kim, 2011, p. 46)
2	Data reduction	-	"Hence for further analysis of the acquired data, it is essential to **simplify the raw sensor data** to a form compatible with the input requirements of the selected models. Data reduction is often the first step in preprocessing and is important for **reducing both data storage space and calculation time**." (Vichare et al., 2007, p. 244) "Subsequently the data may have to be processed to **extract the relevant load parameters** (e.g., cyclic mean, amplitudes, ramp rates, hold periods, power spectral densities, etc.) for PoF model input." (Vichare et al., 2007, pp. 244–245)

J. Factors related to presentation quality

Table 32 summarizes the identified factors that relate to presentation quality.

Table 32: Identified text passages and the derived factors related to presentation quality

No.	Factors	Groups	Text passages
1	Condensation of information	-	“[...] potentially much more information could be available for use in the development process provided that it arrives in the right format, **properly condensed** and prioritized, at the desk of the developer.” (Thiruvenkadam et al., 2008, p. 1)
2	Presentation format	-	“[...] design engineers can get overwhelmed by complex statistics so it is better to keep it simple, using **simple charts (bar, pie, Pareto) or dashboards with green, amber, and red indicators** as straight forward reporting tools.” (Igba et al., 2015, p. 10)
3	Prioritization	-	“[...] potentially much more information could be available for use in the development process provided that it arrives in the right format, properly condensed and **prioritized**, at the desk of the developer.” (Thiruvenkadam et al., 2008, p. 1)

K. Factors related to application quality

Table 33 summarizes the identified factors that relate to application quality. Application quality includes data processing with machine learning. This assignment can be ambiguous because data organization may cover data processing as well.

Table 33: Identified text passages and the derived factors related to application quality

No.	Factors	Groups	Text passages
1	Access during service event	-	"The principal limitation of the information flow for the TTL is that usage information can be **collected only when maintenance service** triggered by a failure is performed." (Shin et al., 2015, p. 1734)
2	Availability of context information	-	"For further analysis, each data set requires detailed **metadata** to describe under which **circumstances the data has been elicited**." (Kuestner and Wartzack, 2015, p. 76) "[...] another aspect of importance to in-service knowledge capture and feedback, which aids data reuse, is the capturing of **context and rationale** alongside the data itself." (Igba et al., 2015, p. 11)
3	Customer context information	Content	"Designer should be able to request information about **customer environment** and **customer profile** to verify field failure information." (Magniez et al., 2009, p. 357)
4	Discussion length	Content	"Threads containing **less than 50 words or less than two postings** were excluded from further analysis as they were found to contain too little information to describe and confirm a defect." (Abrahams et al., 2015, p. 980)
5	Information with privacy-related constraints	Content	"Longitudinal studies involving wearable computing may also pose serious **privacy risks**. Depending on the level of instrumentation, these systems may record the **actions, locations, preferences, and vital signs** of participants." (Aldaz et al., 2013, p. 420)
6	Access permission	Information access context	"We obtained the **permission** of the forum owners and crawled all threads available." (Abrahams et al., 2013, p. 874)
7	Commercial factors	Information access context	"[…] some of the hindering factors which prevent such companies from getting access to or reusing data are a complex combination of cultural, political and **commercial**, and are specific to individual organisations, […]" (Igba et al., 2015, p. 12)

No.	Factors	Groups	Text passages
8	Community membership	Information access context	"Mailing lists allow a list member to send an e-mail message to an entire group of individuals. Any one on the list can respond, and the response is sent to all list members. The only way to **access the conversations** among the members is to **become a list member**." (Finch, 1999, p. 536)
9	Cultural factors	Information access context	"[…] some of the hindering factors which prevent such companies from getting access to or reusing data are a complex combination of **cultural**, political and commercial, and are specific to individual organisations, […]" (Igba et al., 2015, p. 12)
10	Political factors	Information access context	"[…] some of the hindering factors which prevent such companies from getting access to or reusing data are a complex combination of cultural, **political** and commercial, and are specific to individual organisations, […]" (Igba et al., 2015, p. 12)
11	Willingness to share data	Information access context	"Attempts to create a database for wind turbine gearbox component failures have already begun by the gearbox reliability collaborative […]. This will however depend on **how much information turbine operators & manufacturers** are **willing to share**." (Igba et al., 2013, p. 821)
12	Employees' awareness of data existence	-	"In-service data is usually being managed in stand-alone databases […] making accessibility of data difficult for design engineers, in some cases engineers were **not aware** such **data existed**." (Igba et al., 2015, p. 6)
13	False negative error	Method	"A keyword recognition problem is synonymous to a **false negative** or a type II error." (Lim and Tucker, 2016, p. 1)
14	False positive error	Method	"A term disambiguation problem is synonymous to a **false positive** or a type I error." (Lim and Tucker, 2016, p. 1)
15	Right format for use	-	"[...] potentially much more information could be **available for use** in the development process provided that it **arrives in the right format**, properly condensed and prioritized, at the desk of the developer." (Thiruvenkadam et al., 2008, p. 1)
16	Language structure	Method	"Data mining algorithms focused on mining short messages differ from traditional mining algorithms due to their ability to handle online media data containing lower data dimensions and higher variability in **language structure**." (Lim and Tucker, 2016, p. 2)

No.	Factors	Groups	Text passages
17	Domain expert	Method - machine learning	"The list of attributes and levels are **manually created based on the designer's knowledge** of the smartphone market. A more automated process is desired, but the current process involves iterative improvements to the attributes/levels based on the designer's assessment." (Stone and Choi, 2013, p. 459) "However, these two methods are limited in their application to data-driven product design using online media data due to the fact that (1) keyword-based methods use predetermined keywords to return product related data, hereby assuming that the **domain expert always knows the entire set of words** that accurately describe a product/product feature." (Lim and Tucker, 2016, p. 2) "The sample was tagged and labeled by three mechanical engineering **domain experts**." (Abrahams et al., 2015, p. 981)
18	Number of classes	Method - machine learning	"It is well-known that the more classes a multiclass classifier handles, the more likely it is that the classifier's prediction accuracy will degrade [53]." (Abrahams et al., 2013, p. 876)
19	Tagging rules	Method - machine learning	"Threads were **flagged as "defect" if** the discussion could be characterized as pertaining to any of Safety Integrity Levels (SILs) 0 through 4, these being a set of standard motor industry controllability categories for classifying operability hazards […]" (Abrahams et al., 2015, p. 981)
20	Manual classification by Internet users	Method	"Starting at the top right of Fig. 1, users of an online automobile discussion forum manually choose a relevant sub-forum for their new postings. This is effectively **manual component classification**." (Abrahams et al., 2013, p. 873)
21	Method sensitivity	Method	"Finally, 400–600 features produces the optimal (or very nearly the optimal) classification performance in all cases, though in many (6 of 9) cases the relatively flat line (F1-measure score) for the best method […] indicates that the **best method is relatively insensitive** to the number of features used." (Abrahams et al., 2013, p. 877)
22	Robustness of algorithm	Method	"The future improvements in SENTRAL will be on proving the **robustness** of this domain-independent heuristic algorithm for other categories of products and services, as well as its **robustness** in terms of the quality of input data: […]" (Raghupathi et al., 2015, p. 210)
23	Sample size	Method	"1500 was chosen as the **sample size** that would produce an **acceptable margin-of-error** of ±2.5% at the 95% confidence level." (Abrahams et al., 2013, p. 874)

No.	Factors	Groups	Text passages
24	Speed of algorithm	Method	“Equally as important as the feedback is the magnitude and speed (**algorithm speed** as well as speed of customer review input) at which this feedback can be acquired using the proposed online customer review process.” (Tucker and Kim, 2011, p. 48)
25	Unit of analysis	Method	“Thread, rather than posting, was chosen as our **unit of analysis** as each posting within the thread can often only be understood within the context of the entire thread. As a posting is frequently unintelligible when removed from the context of a thread, and classifying every individual posting within a thread results in an order of magnitude increase in the number of items requiring review, the full thread (comprising all postings in response to the original posting, and typically dealing with a single topic or issue) represents a **sensible and efficient unit of analysis** for defect discovery.” (Abrahams et al., 2015, p. 981)
26	Validation	Method	“**Validation** is perhaps the most critical feature of this research field moving forward. There is an abundance of online content and a virtually infinite number of ways to extract information and interpret data. Validation ensures that a preference model accurately predicts consumer decisions. If the framework produces a **valid preference model**, then the designer has access to a continuously updated design tool for as long as the online content continues to grow and the framework remains **valid for that content**.” (Stone and Choi, 2013, p. 464) “We propose to **validate** our **model** by asking 38 humans to do exactly the same task that our model, i.e. to perform 15 rate reviews on a scale of 0-5.” (Raghupathi et al., 2015, p. 207)
27	Noise	-	“Data-driven product design methodologies based on mining online data may have high identification errors of product feature-related information **due to noise** resulting from the differences between writing formats or because of constraints placed by online media platforms.” (Lim and Tucker, 2016, p. 1)
28	Use of secondary sources	-	“The iPhone 4 and iPhone 4S technical specification manuals (technical data sources) from the manufacturer (Apple) and Wikipedia articles for the iPhone 4 and the iPhone 4S (nontechnical data sources) are used as **secondary data sources**.” (Lim and Tucker, 2016, p. 7)
29	Source/author exposure	-	“One must weigh the **risks of identifying a source against the value added by publishing its identity**.” (Finch, 1999, p. 541)

L. Relations between information quality characteristics

This section highlights the heterogeneous meaning of related information quality characteristics. Table 34 summarizes descriptions for accuracy and precision because both concepts are often confused.

Table 34: Descriptions of accuracy and precision in standards and guidelines

Concept	Standard	Description
Accuracy of measurement	ISO Guide 99:2007	Closeness of agreement between a measured quantity value and a true quantity value of a measurand.
Accuracy*	ISO 25012	The degree to which data has attributes that correctly represent the true value of the intended attributes of a concept or event in a specific context of use.
Data accuracy	ISO 8000-2:2012	Closeness of agreement between a property value and the true value.
Accuracy	ISO/IEC/IEEE 24765:2017	1. A qualitative assessment of correctness, or freedom from error. 2. A quantitative measure of the magnitude of error. 3. Within a quality management system, accuracy is an assessment of correctness.
Accuracy of measurement	ISO/IEC/IEEE 24765:2017	The closeness of the agreement between the result of a measurement and the true value of the measurand.
Accuracy	ISO 15939:2017	Accuracy is the extent to which the procedure implementing a base measure conforms to the intended measurement method.
Precision of measurement	ISO Guide 99:2007	Closeness of agreement between indications or measured quantity values obtained by replicate measurements on the same or similar objects under specified conditions.
Precision	ISO 25012	The degree to which data has attributes that are exact or that provide discrimination in a specific context of use.
Precision	ISO/IEC/IEEE 24765:2017	1. The degree of exactness or discrimination with which a quantity is stated. 2. Within a quality management system, precision is a measure of exactness.

* Differentiates syntactic and semantic accuracy.