Information Quality in PLM: A Product Design Perspective

Stefan Wellsandt, Thorsten Wuest, Karl Hribernik, Klaus-Dieter Thoben

Information Quality in PLM: A product design perspective

Stefan Wellsandt[1,3], Thorsten Wuest[2], Karl Hribernik[1], Klaus-Dieter Thoben[3,1]

[1]BIBA - Bremer Institut für Produktion und Logistik GmbH, Bremen, Germany

[2]Industrial and Management Systems Engineering, Benjamin M. Statler College of Engineering and Mineral Resources, West Virginia University, Morgantown, USA
thwuest@mail.wvu.edu

[3]Faculty of Production Engineering, University of Bremen, Germany
{wel, hri, tho}@biba.uni-bremen.de

Abstract.
Recent approaches for Product Lifecycle Management (PLM) aim for the efficient utilization of the available product information. A reason for this is that the amount of information is growing, due to the increasing complexity of products, and concurrent, collaborative processes along the lifecycle. Additional information flows are continuously explored by industry and academia – a recent example is the backflow of information from the usage phase. The large amount of information that has to be handled by companies nowadays and even more in the future, makes it important to separate "fitting" from "unfitting" information. A way to distinguish both is to explore the characteristics of the information, in order to find those information that are "fit for purpose" (information quality). Since the amount of information is so large and the processes along the lifecycle are diverse in terms of their expectations about the information, the problem is similar to finding a needle in a hay stack.
This paper is one of two papers aiming to address this problem by giving examples why information quality matters in PLM. It focuses on one particular lifecycle process, in this case product design. An existing approach, describing information quality by 15 dimensions, is applied to the selected design process.

Keywords: Product Lifecycle Management; quality management; usage data; product design; data quality; usage information

1 Introduction and problem description

Closing the information loops along the product lifecycle is a recent effort undertaken by research projects [1], [2]. One of the reasons why closing information loops is so important is the expectation that designers and manufacturers will be able to create products (and services) of higher quality. This expected increase in product quality is largely based on the assumption that information about the products' in-use behavior

('usage information') will lead to better decisions in processes like product design. Usage information can substantiate decisions and thus increase their transparency within collaborative working environments. Recent research on information about product usage typically focuses on the capabilities and general appropriateness of different approaches, methodologies or solutions that make usage information available to certain decision-makers (e.g. [3] and [16]). Important for the actual integration of the information is the technical capability and a use/business case, as well as the adequacy of the information for the given case. Due to the heterogeneity of usage information, it is difficult to decide what information is actually relevant for a certain decision process – currently, the quality dimensions for usage information are largely unknown.

This paper will discuss the importance of information quality in PLM. For reasons of complexity the scope of the paper has to be significantly limited. This is done by focusing on one exemplary information loop (i.e. from usage to design). Furthermore, only one decision-process is selected for the following discussion.

2 Related work

2.1 Information flows in PLM

Handling product data and information along the complete product lifecycle is stated as PLM [4]. A product's lifecycle can be structured into three subsequent phases stated as 'beginning of life' (BOL), 'middle of life' (MOL) and 'end of life' (EOL). The concept of PLM was further extended during the EU-funded large-scale research project PROMISE – it demonstrated the possibilities of closing information loops among different processes of the lifecycle [5]. The recent concept of PLM is illustrated in Figure 1. Internal information flows within the phases are not covered in the illustration.

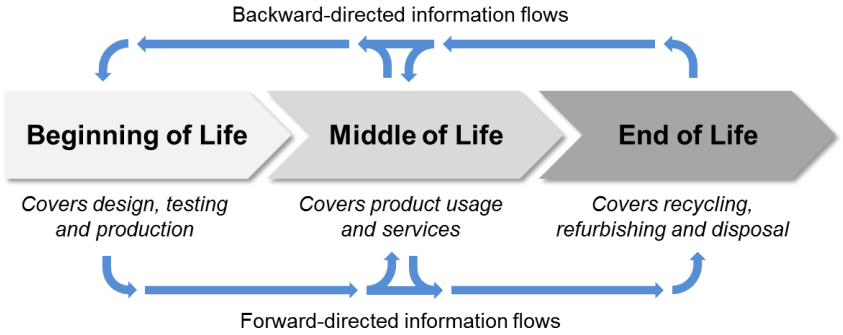

Fig. 1. A product lifecycle model and its major information flows [6]

Among the three lifecycle phases, two types of information flows can be established. The forward-directed flows are the ones that are typically mandatory to design, produce, service and dismiss the product. Backwards-directed flows are typically optional and allow optimization and control of processes. One recent example for optimization

is the improvement of product design through the integration of usage information from the MOL phase – this approach is sometimes called 'fact-based design' [7].

Following the working-definition argued by Wellsandt et al., usage information is *"[…] any product-related information that is created after the product is sold to the end customer and before the product is no longer useful for a user"* [6]. Usage information can originate from sources like product embedded sensors, maintenance reports, shopping websites, social networks, product reviews and discussion forums [8]. Information from these heterogeneous sources feature very different characteristics concerning their format (e.g. structured vs. unstructured data), scope (e.g. plain data vs. multimedia) and the lifecycle activities covered in the content (e.g. use, maintenance and repair).

2.2 Information Quality (IQ)

The topic of IQ has been intensely discussed for at least two decades; several sophisticated definitions for 'information quality' exist. Since the purpose of this paper is not to discuss these fundamental concepts, a thoroughly discussed definition is selected for this paper. From a general perspective, the quality of information can be defined as the degree that the characteristics of specific information meet the requirements of the information user (derived from ISO 9000:2005 [9]). Based on this understanding, Rohweder et al. propose a framework for information quality that is an extension of the work conducted by Wang and Strong [10] – it contains 15 information quality dimensions that are assigned to four categories as summarized in Table 1.

Table 1. Dimensions of information quality [11]

Quality category	Scope	Quality dimensions
Inherent	Content	Reputation; free of error; objectivity; believability
Representation	Appearance	Understandability; interpretability; concise representation; consistent representation
Purpose-dependent	Use	Timeliness; value-added; completeness; appropriate amount of data; relevancy
System support	System	Accessibility; Ease of manipulation

The selected definition of information quality is split into four categories, i.e. inherent, representation, purpose-dependent and system support. Each category has dimensions that characterize information by two to five dimensions. A description of some definitions of these quality dimensions is provided in Table 2.

Table 2. Selection of quality dimensions and their description (based on [11])

Quality dimension	Description
Reputation	Credibility of information from the information user's perspective
Free of error	Not erroneous; consistent with reality
Objectivity	Based on fact; without judgement
Believability	Follows quality standards; significant effort for collection and processing

Understandability	Meaning can be derived easily by information user
Interpretability	No ambiguity concerning the actual meaning; wording and terminology
Concise representation	Clear representation; only relevant information; suitable format
Consistent representation	Same way of representing different information items
Accessibility	Simple tools and methods to reach information
Ease of manipulation	Easy to modify; reusable in other contexts

In order to receive a specific statement about the actual quality of an information item, the as-is characteristics of the item must be compared with the required characteristics. The better the matching is, the higher the information quality is considered.

3 Methodology and Scope

In order to substantiate the framework proposed by Rohweder et al., the requirements of the information users (decision-makers) must be identified and compared with the proposed IQ dimensions (see Table 1). For this purpose, the information flow from MOL to BOL is targeted in this paper – the main subject is usage information. The targeted decision-making process is 'requirements elicitation', an information-intensive decision-making processes conducted early in product design.

Process description. Requirements elicitation (REL) is a systematic, and oftentimes iterative, process aiming to retrieve information from users (and other stakeholders) – the main result of this process is a list of explicit user requirements [12]. Techniques for information retrieval include surveys, questionnaires and observation. Recent approaches, like fact-based-design, aim for the retrieval of actual product usage information, in order to improve, for instance, the requirements list (see section 2.1).

Required characteristics. In literature, quality dimensions for the results of the elicitation activity, i.e. documented requirements, are readily available (e.g. IEEE 830 standard and [13]). A non-comprehensive list of requirements quality (RQ) dimensions is provided in Table 3. It is valid both for individual requirements and whole lists of requirements. RQ dimensions and IQ dimensions have overlapping in some areas.

Table 3. Quality dimensions for requirements according to IEEE 830 and [13]

Criterion	Description
Valuable	Has a specific economic value or benefit for the business
Correct	Free from errors and in accordance with facts
Clear	Only one way to understand the requirements; no ambiguity
Understandable	Target audience can perceive the intended meaning
Complete	All relevant requirements are captured
Consistent	Not containing any logical contradictions
Assessable	Importance or relevance can be estimated
Verifiable	Fulfillment is testable in limited time
Modifiable	Variable components and the influences in case of changes are clearly identifiable
Traceable	Identifiable; connected to other requirements and documents

Relevancy	Fulfills a specific stakeholder need or goal (e.g. needed for realization)
Realizable	Can be realized in practice; resource efforts are estimated

Additional, but more general, information requirements result from decision-making in business environments, e.g. cost-efficiency of collection and use of information.

4 Discussion

The quality dimensions in Table 3 describe desired characteristics of documented requirements, i.e. an output of the REL activity. In a PLM scenario, like fact-based-design, the requirements can be derived from usage information effectively serving as an input of the REL activity. Deriving requirements from usage information requires some kind of analysis and interpretation of the information, in order to get valuable design knowledge. Since usage information and requirements are involved in the elicitation activity, their quality characteristics might be related to each other. The relation between the two sets of quality dimensions will be substantiated in the following. For reasons of complexity, each IQ dimension will be put into the context of one RQ dimension at most. Therefore, the given examples for relations among RQ dimensions and IQ dimensions are not meant to be comprehensive – there are, most likely, other influences that are not covered in this paper. Since this paper lacks a specific use case for REL, the 'use'-scope of IQ dimensions is not further covered in this paper. The discussion is structured according to the four categories summarized in Table 1.

4.1 Content scope

Reputation (rep). Quality dimensions, like the ones in Table 1, can be difficult to instantiate for a company, for instance when expertise or resources are lacking. In these cases, previous positive experience with usage information sources can outweigh the lack of precise estimations for the IQ. The reputation is relevant for decision-making in general, thus also relevant for REL.

Free of error (foe). An error is something produced by mistake [14]. Concerning usage information, errors can occur in at least two areas, i.e. measurements and human-authored contents. In case of measurements, errors can be caused by, for instance inappropriate calibration of sensors, poorly placed sensors and measuring wrong events. In case of human-authored information, errors can be a result of, e.g. unskilled authors (e.g. typos and wording) and limited knowledge of authors (e.g. wrong statements and conclusions). When deriving requirements from erroneous usage information, the resulting requirements might be corrupted (e.g. reflecting a non-existent user need) – therefore, the correctness of requirements benefits from correct usage information.

Objectivity (obj). A characteristic of human-authored feedback information is its subjectivity – also stated as response-bias [15]. When dealing with user responses (e.g. in online discussions) information users generally need to take response-bias into account. Measurements, on the other hand, are more 'objective', since they do not have a response-bias [16]. Due to influences like the response-bias, REL decision-makers may

not be able to derive requirements that fulfill the 'complete'-dimension – the available usage information (e.g. from weblogs or forums) is limited/scoped by the perceptions of its author.

Believability (bel). Information that is used to elicit requirements can be extracted from product reviews. These reviews may be authored by renowned professionals that are familiar with a product and terminology to describe it (higher bel) or by common users with unknown identity, knowledge and skills (lower bel). In addition, the reviews can be based on a structured, transparent testing process (higher bel) or an unstructured/unknown process (lower bel). The dimension might influence the "correct"-dimension of the REL process, since a higher believability of usage information might be associated with less errors effectively reducing potential corruption of requirements. Further the dimension is related to the "objectivity"-dimension.

4.2 Appearance scope

Understandability (und). Human authored usage information, e.g. user feedback in discussion forums, is created by users with different backgrounds (e.g. writing skills, language and expertise on topic). The language, for instance, is an important factor in REL as it limits the understandability of usage information. In a similar way, raw measurement data from sensors (e.g. without graph plotting) are barely understandable by decision-makers. Not being able to take these kinds of usage information into account for REL may lead to an incomplete requirements list (e.g. missing key requirements).

Interpretability (int). Extracting the meaning of usage information can be difficult in case that important context information is lost or originally not provided. Missing context may cause ambiguity of usage information. Measurement protocols, for instance, require context information about the sensor that was used to collect the data. In case this context is not provided, tolerances of measurement remain unclear. This IQ-dimension affects the RQ-dimension for "correctness", as ambiguous usage information may lead to erroneous assumptions and finally to flawed requirements.

Concise representation (ccr). Usage information that is based on human-authored contents is not necessarily uniform. Product reviews may contain a mixture of media, like text, pictures and videos, or different languages. When dealing with these information in the REL process, an in concise representation makes analysis more time consuming.

Consistent representation (csr). Contents generated in the Internet (e.g. weblogs, discussion forums) do not follow standardized procedures. Text can be created freely following limited formal structures, such as templates in 'WordPress' and form fields of forum posts. Content can further contain media formats like pictures and videos. Multimedia formats of usage information require more elaborate tools and in general more effort for analysis. Therefore, consistent representation benefits cost-efficient collection and use of information during the REL process.

4.3 Use scope

The five IQ dimensions of the 'use' scope are not covered further in this paper, since at this stage, no specific use case has been chosen. Without a use case, the range of possible requirements from REL is too large to provide value to this paper.

4.4 System scope

Accessibility (acc). Getting access to usage information (in a technical sense) is challenging for several reasons. Usage information is, for instance:

- distributed across different sources (e.g. weblogs and databases);
- heterogeneous concerning its format, i.e. representation;
- potentially copyrighted or otherwise restricted (e.g. forum with registration).

Furthermore, the collection of usage information may require special skills and/or knowledge (e.g. data or text mining). Barriers for easy accessibility affect the 'complete'-dimension of requirements, since restricted or too costly access to information might result in missing requirements (that could be derived from the information).

Ease of manipulation (eas). Usage information, like product reviews and posts in discussion forums, may contain pictures and videos. These contents are provided in formats that can be difficult to manipulate (e.g. video stream). The 'eas'-dimension is also ambiguous in relation to REL, since manipulation might not be desired by decision makers. The requirements should be framed in a way that they reflect the user's expectations and needs. Ease of manipulation might affect the 'correctness'-dimension of requirements when manipulation of usage information leads to wrong conclusions. An example concerns losing context information during a copy and paste procedure – in consequence, decision-makers might take wrong assumptions about product or user behavior.

5 Conclusion and Outlook

While the availability of new information sources, such as usage information in design, provides new opportunities to improve products (see [3], [7]), newly created information flows and new kinds of information can introduce problems along the lifecycle. Feeding usage information into the BOL phase, for instance, can cause issues in related decision-making processes. These issues can affect product quality in a negative way, e.g. when incorrect or incomplete requirements are elicited based on flawed usage information. The impact of flawed information may further affect later stages of the lifecycle such as maintenance, disassembly and disposal of products. Therefore, the example provide in this paper helps to justify why IQ has to be considered more thoroughly in PLM. In future research, the following aspects should be considered, in order to extend the understanding of IQ in PLM:

- Collection of additional cases from all lifecycle phases (e.g. production, sales, maintenance and EOL scenarios).

- The adequacy of IQ dimensions for each case has to be argued. This requires an analysis of decision-making processes.
- Interdependencies of IQ dimensions need to be detailed. This should be substantiated by practical examples from use cases.

Acknowledgement

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 636951 (Manutelligence) and grant agreement no. 636868 (Falcon).

References

1. Falcon Project: Feedback mechanisms Across the Lifecycle for Customer-driven Optimization of iNnovative product-services. (2015)
2. Manutelligence Project: Product Service Design and Manufacturing Intelligence Engineering Platform. Available: http://www.manutelligence.eu/. Accessed: 21.04.2015
3. Abramovici, M., Lindner, A.: Knowledge-based decision support for the improvement of standard products. CIRP Annals – Manufacturing Technology, *62*(1), 159–162, (2013)
4. Kiritsis, D.: Closed-loop PLM for intelligent products in the era of the Internet of things. Computer-Aided Design, *43*(5), 479–501, (2011)
5. Promise Project: promise-innovation.com. Available: http://promise-innovation.com/components. Accessed: 26.03.2015.
6. Wellsandt, S., Hribernik, K., Thoben, K.-D.: Sources and characteristics of information about product use. Procedia CIRP, 25th CIRP Design Conference, Haifa, Israel, (2015)
7. Røstad, C., Henriksen, B.: ECO-Boat MOL – Capturing Data from Real Use of the Product. In: Rivest, L., Bouras, A., Louhichi, B. (Eds.): PLM 2012, IFIP AICT 388, 99–110, (2012)
8. Wellsandt, S., Hribernik, K., Thoben, K.-D.: Content analysis of product usage information from embedded sensors and web 2.0 sources. ICE Conference 2015, Belfast, Ireland, in press
9. International Organization for Standardization: ISO 9000:2005. Quality management systems – Fundamentals and vocabulary.
10. Wang, R. Y., Strong, D. M.: Beyond Accuracy: What Data Quality Means to Data Consumers. Journal of Management Information Systems, *12*(4), 5–33, (1996)
11. Rohweder, J., Kasten, G., Malzahn, D., Piro, A., Schmid, J.: Informationsqualität – Definitionen, Dimensionen und Begriffe. In: Hildebrand, K., Gebauer, M., Hinrichs, H., Mielke, M.: Daten und Informationsqualität. Springer, 25–45, (2008)
12. Jiao, J. R., Chen, C.-H.: Customer requirement management in product development: a review of research issues. Concurr. Eng., *14*(3), 173–185, (2006)
13. Ebert, C.: Systematisches Requirements Engineering, 4th ed., (2012)
14. Merriam Webster dictionary: Available: http://www.merriam-webster.com/dictionary/error. Accessed: 12.06.2015
15. Sampson, S. E.: Ramifications of Monitoring Service Quality Through Passively Solicited Customer Feedback. Decision Sciences, *27*, 601–622, (1996)
16. Dornhöfer, M., Fathi, M., Holland, A.: Applying Rules for Representing and Reasoning of Objective Product Use Information Exemplarily for Injection Molding Machines. Informatics Engineering and Information Science, Springer, 616–623, (2011)