




STRUCTURING AND CONTEXTUALIZING HISTORICAL DATA FOR DECISION MAKING IN EARLY DEVELOPMENT

K. V. Sigsgaard , J. K. Agergaard, C. A. Bertram, N. H. Mortensen, I. Soleymani, W. Khalid, K. B. Hansen and G. O. Mueller

DTU-Technical University of Denmark, Denmark

 krvsig@mek.dtu.dk

Abstract

It is important to be able to compare and evaluate different solutions early in development. This paper proposes a method for structuring historical data into a data model that can support the evaluation of new design concepts. The data is contextualized by linking it to a hierarchical decomposition of existing products. Two case studies were conducted to evaluate the value of using historical data when evaluating new concepts. The cases confirm that the proposed method is useful for evaluation of new concepts.

Keywords: big data analysis, design analysis, systematic approach

1. Introduction

Decision making in the early development process is arguably the most important part of a product's life (Saravi et al., 2013). This is mainly due to the majority of impacts on cost, environment, etc. being defined in the early stages of design (Martin et al., 2007; McAloone and Bey, 2009; Saravi et al., 2013; Shehab and Abdalla, 2001; Zrim et al., 2006). It is also here, that changes are cheaper to make (Martin et al., 2007). Decisions made in the early stages are often uncertain and lack information (Saravi et al., 2013). Supplying designers with relevant information can support the decision making and improve the design by improving the estimates of the design impacts. It is important to have correct estimates of the impacts, as over-estimates become self-fulfilling and under-estimates end up being more costly. (Martin et al., 2007).

To support this type of analyses, a more general and flexible data model has to be developed. Such a data model would be applicable in a broader range of projects. This would mean that model development time would be shorter per project in the long run. However, literature only provides examples of application of data to specific impact measures. Therefore, this paper suggests a method for structuring historical data according to a design decomposition. The decomposition helps contextualize the data, making it more flexible to analyses. The proposed method allows the model to be reused in a multitude of analyses for a broader range of impact assessments. The paper tries to answer the following research questions:

1. How is it possible to contextualise historical data by linking the data to a decomposition of a product?
2. How can structuring and contextualisation of data provide flexible data analytics to early development?

This paper presents state of the art literature regarding three topics: decision making in early development, the use of data in early development and how to decompose complex systems and products. The proposed method is then presented, followed by two different case studies. The case studies show application of the method, highlighting how variations in the data influences the flexibility of the output. The result are presented in the discussion and conclusion.

2. Method

The research in this paper is based on a literature review of the existing body of literature and case studies from two companies. As the study expands on the use of data in the development process, the research review focuses on examples of applying data in the decision making process (Heller et al., 2014; McAloone and Pigosso, 2017; Staub-French et al., 2003) and the decomposition of systems as a method to handling complexity from Theory of Technical Systems (Hubka and Eder, 1988).

The case studies are based on companies that have addressed data structuring in two different ways. Both case companies design production systems; the first company designs internal production plants and buys the machinery from suppliers, while the second company designs and produces commercial equipment.

Both case studies are based on the data from the companies' Enterprise Resource Planning (ERP) systems and workshops, meetings and semi-structured interviews with internal experts such as engineers, workers, maintenance responsables, system responsables, etc. The collaboration has allowed for validation of the research. The data from the case studies was structured in a data model using QlikView for the first case company and in SQLiteStudio and PowerBI for the second case company.

3. Literature

The following literature review highlights current research on decision making in development, data driven decision making and decomposition of products. The uncertainties of the early stage decision making and examples of applying data to reduce the uncertainty are examined. The second part of the literature study covers Theory of Technical Systems, which suggest a method for achieving an overview of a technical system that allows the designer to structure and contextualize the data.

3.1. Decision making in the development process

In the development process, decision making is a series of decisions leading to a finished design model. Each decision made leads to more aspects being restricted. All decisions are mutually dependent, e.g. a decision on shape can limit the possible materials. Each decision has an influence on the product impacts such as the final costs, profitability, environmental performance, etc. (Weustink et al., 2000). The mutual dependence of the decisions can lead to an iterative loop of changes affecting other aspects of the design that then need re-designing. This can be a frustrating process, especially for more complex designs. Methods such as the Objectives tree, Quality function deployment, Morphological chart, Weighted objectives method, etc. aid the decision making by quantifying the solution space (Cross, 2008).

To support the decision making, estimates of design impacts are made for the different concepts being decided upon. Decisions are made based on weighing the impact of the different conceptual design features. The more knowledge about the impacts the better, as complexity of a product rises when decisions are made based on a low level of information (Lu and Suh, 2009).

Knowledge is becoming more and more important and is a critical component for a firm's competitive success (Drucker, 2004). The importance of knowledge management is therefore increasing. Wiig (1997) defines the objective in knowledge management as follows:

“To make the enterprise act as intelligently as possible to secure its viability and overall success and to otherwise realize the best value of its knowledge assets”.

Court (1997) has identified three groups of knowledge: 1. *General knowledge*: Knowledge based on everyday experience and are not limited to specific domains; 2. *Domain specific knowledge*: Knowledge based on experience from a specific domain; 3. *Procedural knowledge*: Is a combination

of one and two where the knowledge is based on a task in a specific domain. These three types of knowledge are based on the experiences gained through work. The knowledge generated through experiences is very valuable. Knowledge management identifies multiple challenges, one of which is that knowledge generated from a project might be lost when the project is finished (Grudin, 1996). Ramesh and Tiwana (1999) suggest a list of nine problems identified in knowledge management with a focus on new product development. In the early development process, knowledge can be based on the designer's experience or a collection of data. However, basing requirements and cost estimations solely on an engineer's knowledge tend to lead to less accurate estimates (Staub-French et al., 2003). Therefore, there is a need for using knowledge and data in development. It is believed that historical data can support the knowledge management by providing data to support decision-making and use of historical data makes it easier to collect knowledge between projects and over time.

3.1.1. *Introducing data into decision making*

Introduction of data quantifies estimations and lets the designers take more informed decisions. In literature there are several examples of applying data to different aspects of the development process. Some examples are highlighted in the following.

Staub-French et al. (2003) presented a method for estimating costs of construction projects based on symbolic representation of product features. A generic template is used in every project where engineers input the requirements for the activities, allocation of resources, and execution of resources for each design feature. The generic format of the template allows reuse of the considerations in future projects. Test cases presented in the study showed it helped designers take better cost decisions earlier in the development process.

Life cycle assessment (LCA) is used to gain quantified insights into the environmental impacts of a design and is a regular method from eco-design practices. LCA is based on data about the different aspects of a design. In the early stages when a lot of this data is unknown, similar products can be used to do the LCA (McAloone and Pigozzo, 2017). Similarly, Heller et al. (2014) showed how introduction of data from Life Cycle Management systems can improve early decision making. The cost estimations of a future product are based on requirements for the product. The data used for the estimations comes from the LCA made on similar, existing products.

Knowledge representation works with the problem of developing a sufficiently precise notation for representing knowledge. This notation of knowledge should be a representation scheme (Mylopoulos, 1981). This knowledge should be based on a knowledge base consisting of facts. When using the data as knowledge in development it is important to understand the knowledge base and representation scheme already used by the company. A layer of structuring and contextualization is then built on top of the existing knowledge representation.

The above examples show that data about a specific design impact can be useful to the decision making process. However, it has not been possible to find studies on applying data to look into a broad range of aspects of the design. This might be because of the difficulty of collecting the data, as this type of data is often distributed in separate silos in the company and therefore can be difficult to combined (Das and Mishra, 2011). The process of applying data is a balancing act because it is costly to include a lot of details. On the other hand, too few details in a model will make the model losses its ability to create useful results (Persson, 2002).

3.2. **Decomposition of technical systems**

Theory of Technical Systems is the theory of how technical systems are structured, classified, and decomposed. A technical system is something that fulfils a function for those that use it. It can be a product, a service, a plant, etc. To quantify a technical system it can be split into hierarchical levels. The hierarchy is a way of decomposing the system and quantifying the details of it. One way of decomposing a technical system is by complexity (Figure 1). Each level in the hierarchy consists of groups, and the lower the complexity of the structure level, the fewer the number of parts in a group (Hubka and Eder, 1988).

Groups with lower complexity has a greater breadth of application in other technical systems, e.g. a bolt can be applied in many different connection scenarios, while a specific type of fixture will only be relevant in certain applications. When working at a defined level in the hierarchy, the solutions created at this level must be implemented at the higher level, while the connection to the lower level is only presented on a conceptual level. E.g., the bolt has several detailed features that are important during manufacturing, but at a higher and more complex assembly level, the bolt's feature is seen only as a way to fix two parts to each other. (Hubka and Eder, 1988).

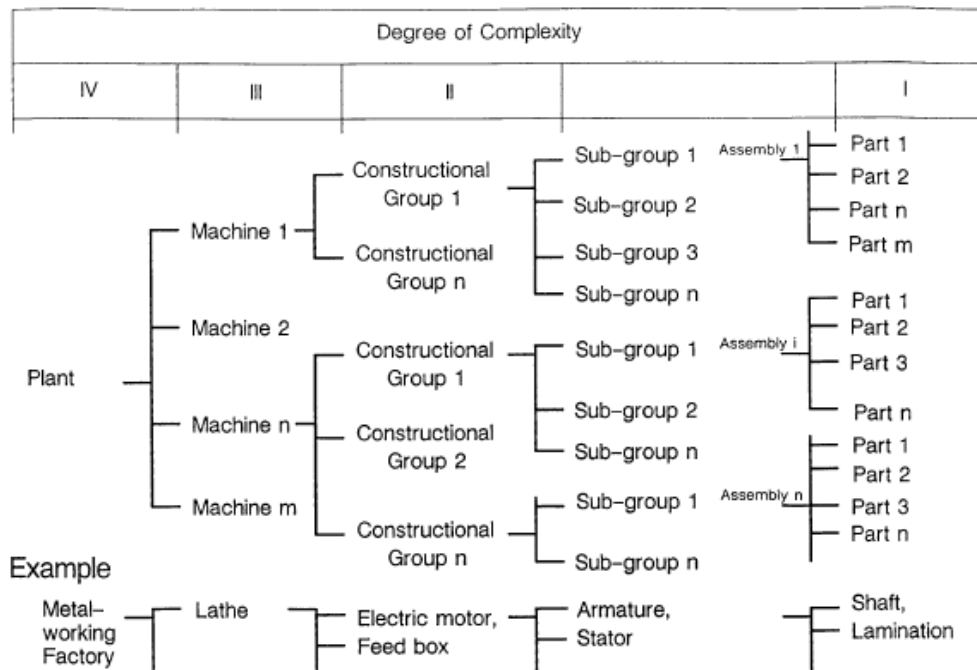


Figure 1. Deconstruction of a technical system.
Figure from Hubka and Eder (1988)

4. A method to structure and contextualize data

The proposed model uses the theory from decomposition of technical systems to link and contextualize historical data. By using the model, it is possible to create a flexible data model that can be used to evaluate concepts early in development. Flexibility makes it possible to evaluate products in all levels of the deconstruction of the technical system, thereby enabling comparison of different systems or subsystems all the way down to individual components. This paper focuses on how to structure the historical data so it is possible to quickly identify the relevant data and use it for analysis in a specific case. The process of the data analyses are out of the scope for this paper.

The main idea behind the model is to deconstruct the product to contextualize the data. By deconstructing the product it is possible to break down the product into sub systems, modules or even individual parts. By using hierarchies to break down the product, it is possible to see which systems the product contains and which components are in the systems. The key is then to link the data at the lowest level possible in the hierarchy.

To better understand the proposed model, a simple example with a bike is presented. The goal in the bike example is to identify material costs and customer complaints for a bike. First, the bike(Product) is decomposed into three systems: the frame, the drivetrain and the wheel. Each systems is then decomposed into components(Figure 2). By linking the material cost and customer complaints to each component in the hierarchy, it is possible to get the material cost for the frame by summing the material costs of the parts related to the frame. In the same way it is possible to get the customer complaints for any desired aggregate level. This is possible because the hierarchy is used to contextualize the data: The context makes it easy to aggregate all parts and systems in a product.

Cost and customer complaints are two different types of data and are often collected in separate departments and even in different systems. Linking and contextualising the data is the key for using the data. In the case of the bikes there are two tables of data and a table for the decomposition. The goal is to link the data to the lowest level in the decomposition, as this gives the highest flexibility. In this case, the lowest possible link is the component level. Figure 3 shows the concept of linking the data. There are two data tables: the cost and the customer complaints, where each component is linked to the corresponding component in the decomposition table. This paper is not focusing on how to link data, there are several tools that can do this and in each case study the specific program used is specified.

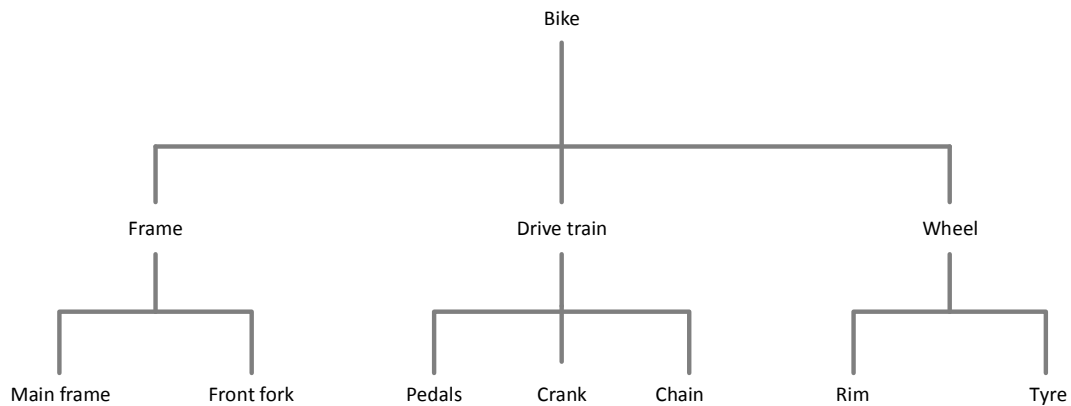


Figure 2. Simple decomposition of a bike

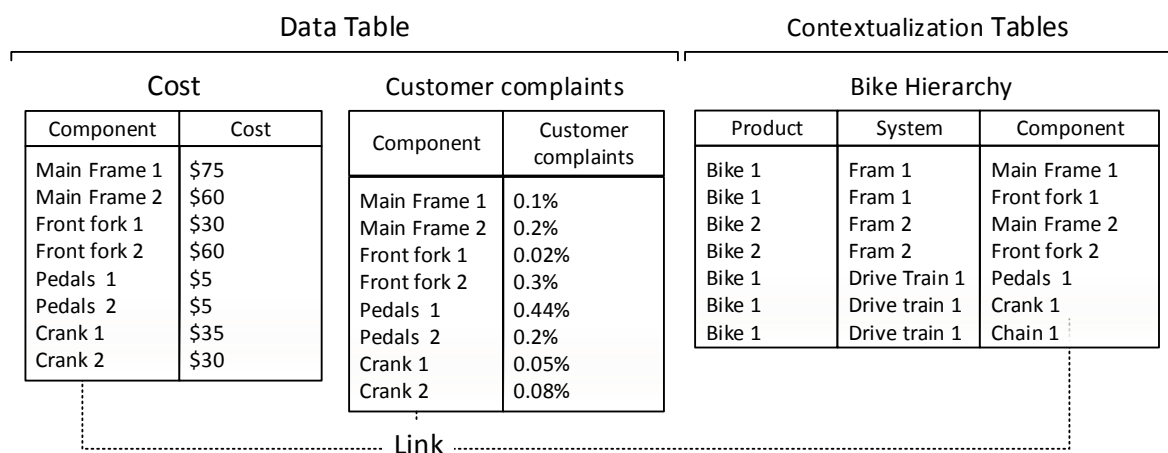


Figure 3. Linking data with decomposition of the bike

Expanding the bike example to two bikes, it is possible to compare the material price between the two cranks but also to compare the price between the two frames or the overall price of the bikes (Figure 4).

When designing more complex systems with thousands of pieces, it might not be feasible to link the data on the lowest level because of time constraints, limitation in the data, etc. However, understanding the link between the data and the hierarchy will give an understanding of the limitations of the data model (Figure 5).

Based on the two case studies general requirements for building the model were identified. The first requirement is that relevant data has to be available. Secondly, there has to be a similar decomposition for the different products and lastly there needs to be a clear link between the data and the decomposition. However further studies should be performed to fully specify the requirements and differentiate them for different use cases.

To understand the limitations and possibilities of the model, it is important to understand at what level the data can be linked to the decomposition. The level of the link is paramount to the comparisons that

can be made with the final model. It is also important to understand the decomposition to make sure that only parts or systems that are at the same level are compared.

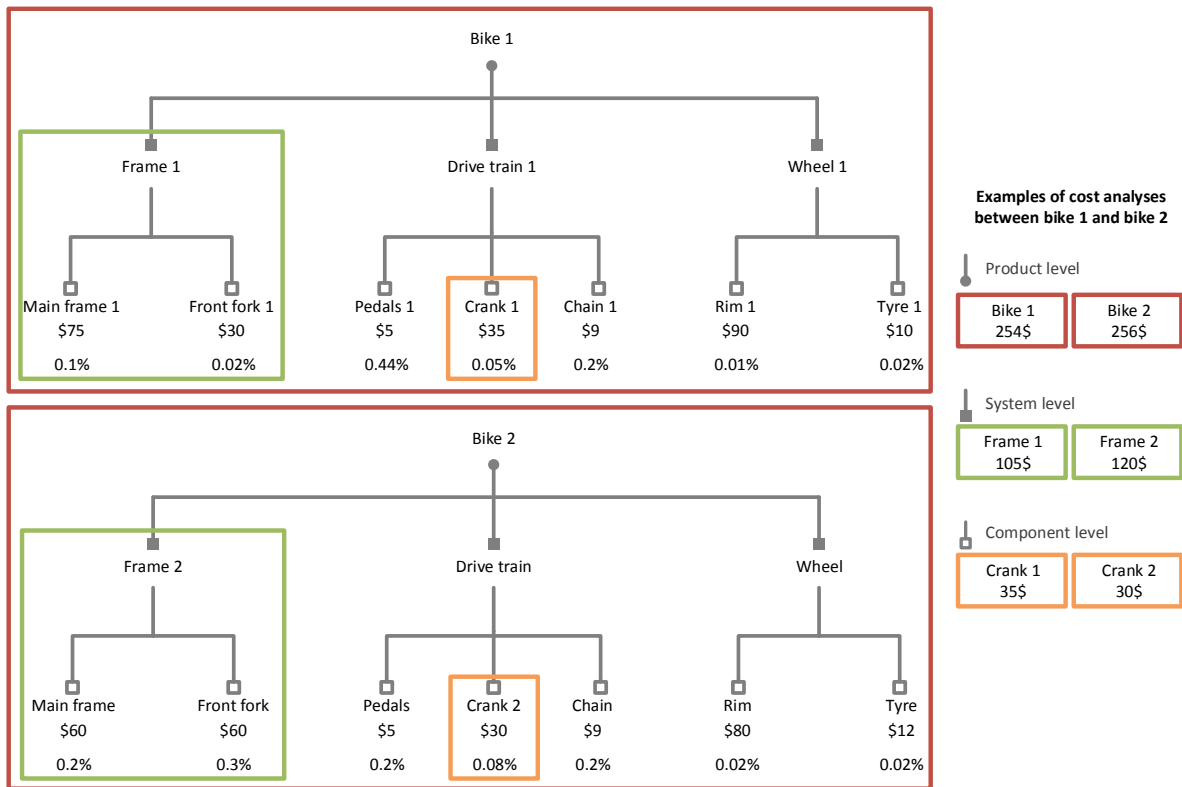


Figure 4. Comparison between two bikes

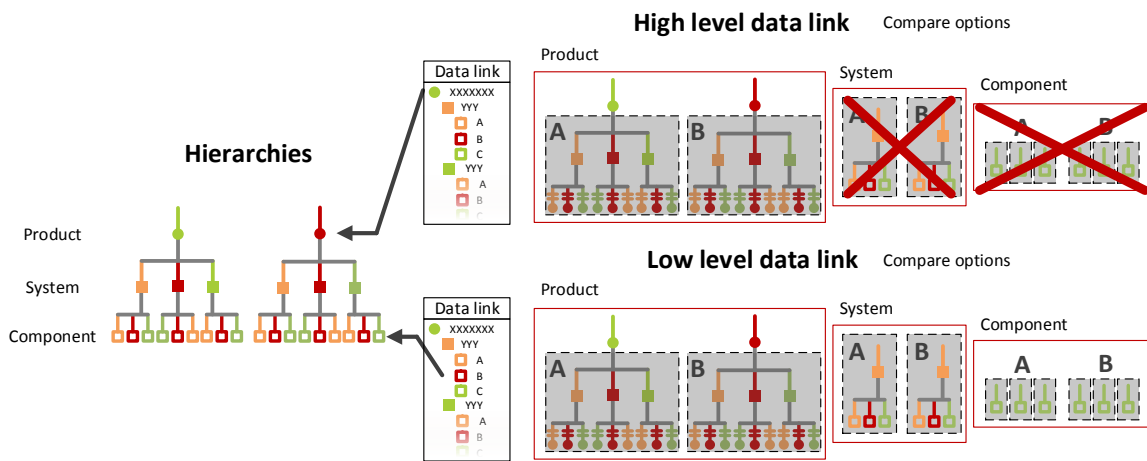


Figure 5. Limitation in high link data and low linked data

5. Case studies

Two case studies are presented in this paper. Both studies focus on the use of historical data to take data-driven decisions in the development of production facilities. In both cases, the main focus is contextualizing the data by linking it to the decomposed hierarchy of the production facilities. The difference in the two cases is that the first case study links the data at a low level in the hierarchy, and the second case study links the data at a higher level in the hierarchy. The knowledge gained from the two case studies is discussed in the discussion section.

5.1. Low-level data link - first case study

5.1.1. Company, context and goal

The first case is from an oil and gas company that designs and operates offshore oil and gas production facilities. The company wants to reduce the operational costs for future offshore facilities and focuses on maintenance operations as one of the main influence factor. The overall approach is to use historical data from similar platforms or systems to estimate maintenance costs for new facilities by selecting the most cost efficient solutions in relation to the life cycle cost.

5.1.2. How

The main part of the work has been to gather maintenance data from approximately 40 oil and gas installations to understand how to decompose the facilities into systems and equipment and, most importantly, to distribute maintenance hours to every piece of equipment. This has resulted in a data model containing the following:

- Overall information on installed equipment
- A hierarchy of the platforms, systems and equipment
- Amount of hours used for maintenance throughout the hierarchy
- Classification of platforms with information on weight, type and facilities

All data is linked down to equipment level (e.g. pumps, valves, etc.) and the hierarchy and classification is used to contextualize the data. Figure 6 shows the structure of the data and how it was contextualized in the first case. The first case uses one table with data on all the maintenance jobs performed on each piece of equipment. The maintenance jobs are linked through equipment to two contextualization tables. The first table is “Hierarchy of the platforms”, which contains information on the location of the equipment and the platform. The second table is the “Equipment table”, which is a classification of the equipment.

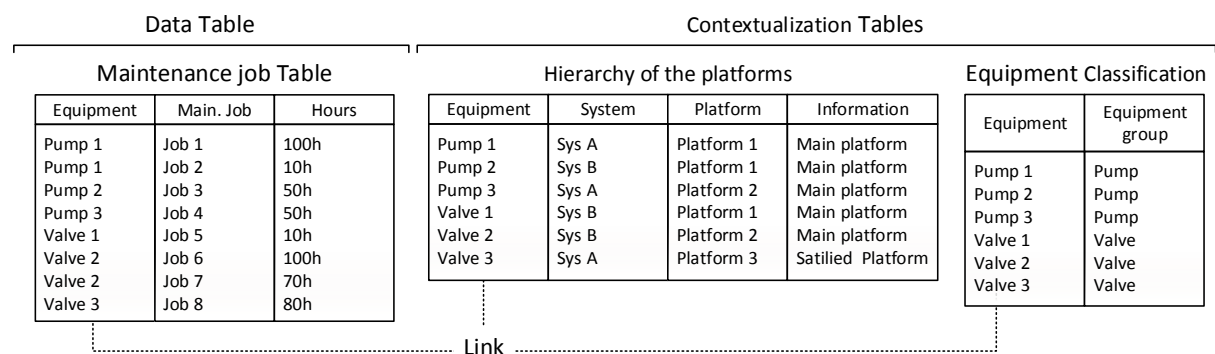


Figure 6. Simplified model for the first case study

5.1.3. Findings

With the data model built, it has been possible to compare and evaluate different concepts by comparing them to similar, existing solutions. More surprisingly, the model also help to estimate maintenance for new systems by modifying the composition of equipment the system will contain. The data model also facilitates an overview of how the maintenance hours are distributed which helps prioritization of the critical parts.

5.1.4. Limitations

The model has two main limitations. Firstly, the user of the model has to have a good understanding of the existing installation to fully understand the data behind the model. Secondly, it is time consuming to gather the data at this level of detail and other companies might not have data at the same level of detail and coherence. It worked well in the case company, as they had already implemented a comprehensive ERP system and a decomposition of the facilities.

5.2. High-level data link - second case study

5.2.1. Company, context and goal

The second company is an engineer-to-order production plant manufacturer. They design, engineer and build processing plants based on specifications negotiated with the customer. The company wants to rationalize their product portfolio, by looking into historical data from past projects, revealing best and worst practices. This could potentially lead to a suite of highly cost-efficient standard plants and merits for handling and pricing customizations. The goal is to better utilize past experiences to take better decisions in the present.

5.2.2. How

Efforts have been concentrated on gathering documentation, associate data from past projects and consolidating these within a data model. The resulting data model contains the following information:

- At what scope, situation, location and circumstances the plant was built
- What equipment was chosen for the plant and how each piece of equipment was designed
- How budgets progressed for each project and how cost developed
- What process calculations were done for each plant

The company develops plants on a project-by-project basis and all the findings are consequently related to each specific project. All individual data is therefore linked to the specifications of each project and not the equipment. There are only a few exceptions on costing information where there is a link to specific equipment. Figure 7 shows the main structure of the data and how it was contextualized in the second case. The model consists of two data tables and two tables for contextualization. In this case the data link is made on a project level which is a higher level, but the lowest available level. The first table contains budget and related cost, and time consumption is included in the second table. The contextualization's first table contains the overall project info and the second table contains an equipment list for each project.

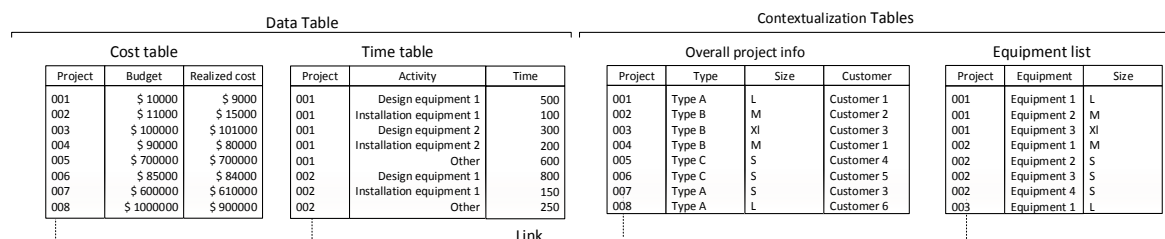


Figure 7. Simplified model for the second case study

5.2.3. Findings

The model has proven useful for evaluating the progress and outcomes of past projects. It has been possible to reveal past best practices and likewise worst practices for plant designs and project execution and utilizing this information for the planning of future projects.

5.2.4. Limitations

For this case study, conclusions can only be drawn on the best practices on a full plant level, not on more specific, less complex levels of decomposition. This stems from the plant projects, only being decomposed by slightly similar system definitions, making them more difficult to compare to each other. Furthermore, all of the gathered data is encapsulated as information about the overall plant project and not as individual solutions such as sub-systems or pieces of equipment.

6. Discussion

Both case studies show that use of historical data can remove some of the uncertainties that occur when developing new production facilities. The two cases also show that by structuring and contextualizing

the data, it is possible to create a tool that can be used to evaluate multiple concepts without changing the created data model.

The main dissimilarity between the two cases is the level at which the data is linked. In the first case, the data is linked at a low level and is therefore more flexible. This is also seen from Theory of Technical Systems, where solutions at lower complexity levels find a greater breadth of application. The flexibility makes it possible to compare groups at the same level of complexity with each other, e.g. platform with platform, system with system, or equipment with equipment. In the second case the data was linked at a higher level, which only allowed comparison at the project level.

The differences between the cases show that there are clear benefits to link at a lower level. However, choosing the level of the link will always be a trade-off between usability and the number of resources used to develop the model. High level linking requires less data collection and preparation, making it a faster and less demanding option. However, as the literature shows, having access to information improves the decision making process. Therefore, creating a data model, which can be used across a variety of projects, can be of great benefit.

7. Conclusion

In the literature, it is widely accepted that most of the impacts are determined in early development. This makes it important to estimate the impacts when the decisions on the design is made. Different models and methods for this purpose are proposed in the current literature. Several studies have also successfully implemented data into the process. However, this study focuses on utilizing data in one data model to analysing the impact in multiple areas instead of having one data model for each analysis. A method for linking historical data to the decomposition of a technical system from Theory of Technical Systems is presented. Linking the data to the decomposition contextualizes the data and allows a wider range of analyses.

The cases presented in this paper show that it has been possible to build a generic and flexible data model in the two case companies. It is shown, that the model can assist evaluation of a variety of impacts in early development. Specifically, it is shown that having data linked at a lower complexity level in the system hierarchy supports a greater breadth of applications.

There are, however, some limitations with these two case studies that require further research. The tool has only been evaluated by experts in the company and there are no studies on the accuracy of the predictions made by the tool. The study does not indicate that the cases are unique, but with only two cases the conclusions are only an indication of the potential of the model. The proposed model also requires that the company already has been storing the detailed data for some time and that they are willing to set aside resources to collect and model it. Further research should focus on determining how accurate the estimations are. More case studies in other industries could also help verify the flexibility and usability of the historical data for general development.

References

- Court, A.W. (1997), "The relationship between information and personal knowledge in new product development", *International Journal of Information Management*, Vol. 17 No. 2, Elsevier Ltd, pp. 123-138. [https://doi.org/10.1016/s0268-4012\(96\)00054-0](https://doi.org/10.1016/s0268-4012(96)00054-0)
- Cross, N. (2008), *Engineering Design Methods : Strategies for Product Design*. J. Wiley.
- Das, T.K. and Mishra M.R. (2011), "A Study on Challenges and Opportunities in Master Data Management", *International Journal of Database Management Systems*, Vol. 3 No. 2, pp. 129-139. <https://doi.org/10.5121/ijdms.2011.3209>
- Drucker, P.F. (2004), *Post-Capitalist Society*, Butterworth-Heinemann.
- Grudin, J. (1996), "Evaluating opportunities for design capture", *Design Rationale: Concepts, Techniques and Use*, pp. 453-470.
- Heller, J.E., Löwer, M. and Feldhusen, J. (2014), "Requirement based future product cost estimation using lifecycle assessment data", *Procedia CIRP*, Vol. 15, Elsevier B.V.: pp. 520-525. <https://doi.org/10.1016/j.procir.2014.06.069>
- Hubka, V. and Eder, W.E. (1988), "Theory of Technical Systems A Total Concept Theory for Engineering Design".

- Lu, S.C.Y. and Suh, N.P. (2009), "Complexity in design of technical systems", *CIRP Annals - Manufacturing Technology*, Vol. 58 No. 1, pp. 157-160. <https://doi.org/10.1016/j.cirp.2009.03.067>
- Martin, P., Dantan, J.-Y. and Siadat, A. (2007), "Cost Estimation and Conceptual Process Planning", *Digital Enterprise Technology*, pp. 243-250. https://doi.org/10.1007/978-0-387-49864-5_28
- McAloone, T.C. and Bey, N. (2009), "Environmental Improvement through Product Development - A Guide", Available at: www.ecodesign.dtu.dk
- McAloone, T.C. and Pigosso, D.C.A. (2017), "Ecodesign implementation and LCA", In: *Life Cycle Assessment: Theory and Practice*, pp. 545-576. https://doi.org/10.1007/978-3-319-56475-3_23
- Mylopoulos, J. (1981), "An overview of knowledge representation", *ACM SIGPLAN Notices*, Vol. 16 No. 1, pp. 5-12. <https://doi.org/10.1145/960124.806869>
- Persson, J.F. (2002), "The impact of different levels of detail in manufacturing systems simulation models", *Robotics and Computer-Integrated Manufacturing*, Vol. 18 No. 3-4, pp. 319-325. [https://doi.org/10.1016/S0736-5845\(02\)00024-8](https://doi.org/10.1016/S0736-5845(02)00024-8)
- Ramesh, B. and Tiwana, A. (1999), "Supporting collaborative process knowledge management in new product development teams", *Decision Support Systems*, Vol. 27 No. 1, Elsevier Science Publishers B.V., pp. 213-235. [https://doi.org/10.1016/S0167-9236\(99\)00045-7](https://doi.org/10.1016/S0167-9236(99)00045-7)
- Saravi, M., Newnes, L. and Mileham, T. (2013), "Optimising Performance and Cost at the Early Design Stages", Vol. 3 No. 3, pp. 214-228.
- Shehab, E.M. and Abdalla H.S. (2001), "Manufacturing cost modelling for concurrent product development", *Robotics and Computer-Integrated Manufacturing*, Vol. 17 No. 4, pp. 341-353. [https://doi.org/10.1016/S0736-5845\(01\)00009-6](https://doi.org/10.1016/S0736-5845(01)00009-6)
- Staub-French, S. et al. (2003), "A generic feature-driven activity-based cost estimation process", *Advanced Engineering Informatics*, Vol. 17 No. 1, pp. 23-39. [https://doi.org/10.1016/S1474-0346\(03\)00017-X](https://doi.org/10.1016/S1474-0346(03)00017-X)
- Weustink, I.F. et al. (2000), "A generic framework for cost estimation and cost control in product design", *Journal of Materials Processing Technology*, Vol. 103 No. 1, pp. 141-148. [https://doi.org/10.1016/S0924-0136\(00\)00405-2](https://doi.org/10.1016/S0924-0136(00)00405-2)
- Wiig, K.M. (1997), "Knowledge management: Where did it come from and where will it go?", *Expert Systems with Applications*, Vol. 13 No. 1, Elsevier Ltd: 1-14. [https://doi.org/10.1016/S0957-4174\(97\)00018-3](https://doi.org/10.1016/S0957-4174(97)00018-3)
- Zrim, G., Maletz, M. and Lossack, R. (2006), "Experience based cost management in the early stages of product development", *9th International Design Conference, DESIGN 2006*, pp. 1569-1574.