This is the published version of a paper presented at 16th IFAC Symposium on Information Control Problems in Manufacturing - INCOM 18.

Citation for the original published paper:

Bertoni, A. (2018)
Role and Challenges of Data-Driven Design in the Product Innovation Process
In: Proceedings of the 16th IFAC Symposium on Information Control Problems in Manufacturing - INCOM 18

N.B. When citing this work, cite the original published paper.

Permanent link to this version:
http://urn.kb.se/resolve?urn=urn:nbn:se:bth-16444
Role and Challenges of Data-Driven Design in the Product Innovation Process

Alessandro Bertoni*

*Department of Mechanical Engineering, Blekinge Institute of Technology, Karlskrona, 37179, Sweden (Tel: +46 455 385 502; e-mail: alessandro.bertoni@bih.se).

Abstract: The paper discusses the role and the challenges of integrating Data-Driven Design (DDD) models in the product innovation process. Firstly, based on academic literature on the Product Innovation Value Streams and on Model-Based Decision Supports, the paper highlights how and when the use of DDD models can support a more effective product innovation process. Secondly, it highlights a list of challenges to be considered for the development of new DDD models for decision support. Ultimately, those challenges are formalized in an evaluation matrix intended to guide the further development of DDD.

Keywords: Data-Driven Design, Product Innovation, Product Value Stream, Knowledge Value Stream, Decision Making, Product Service Systems Development, Design Methods.

1. INTRODUCTION

Over many decades both research and industry have developed practices, methods and tools for decision support centred on technical and engineering quantifiable aspects such as product performances and safety. The use of simulation-based decision support systems such as Multi-body or Computational Fluid Dynamics and knowledge-based approaches is common in many product development activities in a wide array of industry. However, competitions, regulations, and the “servitization” of manufacturing offers have led to the need of satisfying a larger and heterogeneous set of customer needs, including lifecycle implications, supply chain impact and more high-level global challenges (Meier et al., 2010), that are commonly tacitly and subjectively perceived by engineers (Isaksson et al. 2013). In this context, the established simulation-based models are limited in supporting engineers to evaluate product concepts because of: (1) the limited availability of first-hand data commonly residing also within the product lifecycle (Gautam and Singh, 2008); (2) the lack of established methods to quantify the relations between the design variables and the lifecycle and service performances; (3) the lack of effective approaches to communicate the results of lifecycle and service-related calculations to engineers and designers, usually not accustomed to deal with such type of data (Bertoni et al., 2013). However, while the lack of data has long been a common denominator for such issues, the rapidly lowering cost and technological barriers to data collection and analysis, open unprecedented opportunities for the development of new design practices addressing those challenges. In this direction, Data-Driven Design (DDD) models, including the use of data mining and machine learning, have been increasingly discussed in literature as product innovation enablers, and have been recognized as being able to fit the gap between tools used in decision-making and their linkage to data (Kusiak, 2006). In line with such context, the paper has two objectives. First, it proposes a framework describing the role of DDD as a part of a Model-Based Decision support system. Second, it describes a list of challenges to be considered for the development of DDD support methods, summarizing them in an evaluation matrix with the intent to support the future development of DDD models.

2. RESEARCH APPROACH

Participatory action research (PAR) (Whyte, 1991) has been the main approach adopted for the definition of the industrial challenges. The author has directly participated in research teams composed of researchers and industrial practitioners both in the field of aerospace and construction equipment product development and systems engineering. Data were collected by means of qualitative approaches, such as informal discussion, workshops, semi-structured and open-ended interviews and company conferences. In addition to PAR narrative literature review was run in the field on Data Driven Design, including applications of data mining and machine learning in early product development, and in the field of model-based decision support for design.

3. THEORETICAL FRAMEWORK

3.1. Data-driven approaches in design

Data-driven models are used in engineering both with the predictive goal of forecasting the value of a variable and with the descriptive goal of understanding and discovering patterns in the available data (Anand and Büchner, 1998). The need to capture and analyze data capable of overcoming the limit of human analysis is broadly discussed in literature (Fayyad and Stolorz, 1997), but a limited number of examples of data mining applications in engineering design are presented. Tsent and Hiao (1997) were among the first to
propose the use of data mining for the recognition of functional requirements patterns in design. Later Vale and Shea (2003) focused on accelerating the design synthesis by proposing an approach observing and analyzing the effects of sequences of modifications on the design objectives and rating the modifications accordingly with a figure of merit. With a similar purpose, Geng et al. (2012) used historical data to extract parameter-translation rules to transform customer requirements into design requirements and module characteristics. Design synthesis was also in the focus of Wickel and Lindemann (2005) whose recommendations for engineering changes are derived from a database of engineering changes. Quintana Amate et al. (2015) described a case study of a data-driven design optimization of wing covers to estimate how the time to remanufacture a product is affected by one change in a value parameter. Furthermore, Lützenberger et al. (2016) used data to improve the definition of the design parameters. The identification of product families has been in focus of many authors: Agard and Kusiak (2004) to define new product families, Ma and Kim (2016) to derive market value predictions and define clusters of families, Song and Kusiak (2009) to identify sub-assembly patterns from customers’ orders. The theme of analyzing and forecasting the demand trend has also been addressed by Woon et al. (2003) to discover hidden knowledge about trends in product development. More recently Pajo et al. (2015) have proposed a method to mine social media to gather information about actual and future customers’ needs.

Ma et al. (2014) used data trend mining technique allowing the identification of patterns of design attributes to create predictive models allowing the optimization of the design features. Other authors have applied Apriori algorithms with different goals: to describe association relationships between product map knowledge and customer data (Liao et al., 2008), to mine the customer knowledge to improve the development process and the customer relationship management (Liao et al. 2010), to identify the most relevant attribute that influence the purchase of a product (Bae and Kim 2011) or to identify the target market segment for a new design with the intent to support managerial decision-making (Lei and Moon, 2013). In summary, the application of data-driven model in design has captured increased attention in recent years. The main research contributions here identified focus on three main areas of interest: design synthesis, product family and customer needs. Figure 1 visualizes such areas clustering the references described in this section.

### 3.2. Model-based support for early engineering design decision-making

The use of computer-based decision support systems to overcome humans’ cognitive and information processing limitation is field of study since the 70s (Weiss et al., 1978). Literature highlights three dominating aspects of concern on decision making, namely: 1) information about the current situation and history; (2) the relation between basic processes and actions or decisions; and (3) the decision process (Wierzbicki et al., 2000). The reality in complex decisions environment shows the demand of decision support systems that integrate those three aspects, however normally people tend to isolate only one of such aspects, or even a sub-aspect, and develop a specific decision support for that (Wierzbicki et al., 2000). This is done by the rational decision of not studying an overly complicated environment with a high level or cross-influencing factors. Literature highlights that in large-scale product development programs individual intuition plays a relevant role in the decision-making, building on the tacit knowledge and experience of the design team members (Flanagan et al., 2007). In a new product development process, the complexity and the speed of change of information does not grant that ad hoc intuitive decisions are better than well prepared analytical decisions; rather the benefit of the use of computerized support for decision is not disputed (Wessel and Wierzbicki, 2000) (Flanagan et al., 2007). On the one hand group decisions can be supported by models capturing the priorities and the relevant criteria for decisions; on the other hand, decisions can also be supported by models providing a means for discussing and exchanging evaluation on possible alternatives according to chosen criteria. The first type of models is generally referred in literature as “Substantive models”, representing a part of the reality relevant for the decision situation; while the second types of models are generally referred as “Preferential models”, representing the preference of the decision makers and other aspects of the decision process (Wierzbicki et al., 2000). Wierzbicki et al., (2000) have analysed the role and the timely development of substantive and preferential models in a decision-making process. According to their description three stages can be identified.

- **Early Stages of decision process (Model Building):** here the problem is recognized and preliminary data are gathered in order to define alternative hypothesis about the problem, the model type and the data structure. The model design is created either through the modification of existing models of through the integration of possible modification of existing model blocks. In this stage substantive model parameters are estimated through statistical analysis or parametric optimization procedures.

- **Middle Stages of decision process (Model Analysis):** In this stage, the simulation of simple or multi-objective substantive models is run together with sensitivity and post-optimal analysis. Substantive models are used to generate and pre-select a number of design options or decision scenarios.
possibly with the help of a partial preferential model if the number of decision options is large.

- **Late Stages of decision process (Choice and Implementation):** In this stage decision makers interact with preferential models that suggest solution to a decision problem. The preferential models generate options with special tools designed for the purpose to guide to the final decisions. Models can also be used for monitoring the effect that a decision can generate with the purpose of checking model adequacy and viability.

3.3. The Product Innovation Process and the Value Streams

The starting point for the development of the framework is the model proposed by Kennedy et al. (2008), who describes the innovation process as divided into a knowledge value stream and a product value stream. Kennedy’s model is increasingly proposed in literature as a lean enabler for systems engineering (see for instance Stenholm et al., 2015; Isaksson et al., 2015). The first stream represents the capture and reuse of knowledge about markets, customers, technologies, products and manufacturing capabilities, which is general across projects and organizations. The product value stream is specific for each project and consists of the flow of tasks, people and equipment needed for creating, for example, drawings, bill of materials and manufacturing systems. Isaksson et al. (2015), provided a “framework for model-based decision support for value and sustainability” further contextualizing such model from the perspective of a manufacturer of complex systems, highlighting the main development activities linked to the value streams. According to Isaksson et al., (2015) progressing along the two streams, different needs are observable when considering decision support. Four main activities are identified along the streams and can be summarized as follows:

- **Concept Screening/Scoping:** In this phase possible solutions need to be screened quickly and with limited effort and time, typically in order of hours. The information available is limited and immature and the solution space is very open.

- **Concept trade-off:** here a set of the most promising solutions is selected for further analysis. The solution space is more limited but the trade-off is still driven by simple simulation models with low maturity and dependent by variable input.

- **Emerging Design (product commitment):** here decisions are made to enable the design team to confine the design space and down select a limited number of concepts from the previous set. The model development and execution can take within one or few days and a range of models for the design space becomes available.

- **Concept development:** here the knowledge value stream is abandoned to commit to a specific product value stream. Product and process definition are refined to minimize risks and costs. The time frame for decision support tools is still time-constrained, yet studies may now expand to weeks.

With more clear product definitions follows and higher quality of data available.

4. THE ROLE OF DATA-DRIVEN DESIGN IN THE PRODUCT INNOVATION PROCESS

Data in early stages embed context-dependent aspects, recalling tacit or experience-driven knowledge of the decision makers, which are often very difficult to be formalized into mathematical or statistical rules. In respect to Kennedy’s model, the use of DDD models in the initial concept screening stage shall deal mainly with the acquisition of preliminary data about the problem to be solved, in order to define consistent hypotheses and investigate the eventual correlation of variables. This corresponds to the “problem recognition” phase and deals with the formalization of substantive models to capture the priorities in the evaluation. Moving into the concept trade-off phase DDD models shall encompass a data availability and allow the creation of substantive models capable of better representing the portion of reality relevant for decision making. At this stage, parametric analysis and statistics can be applied to the “model design” phase where one or more new models or derivatives are developed. These activities are however not enough to complete the concept trade-off. This is because the decision process enters into the Model Analysis phase, where simple multi-objectives optimization shall be run. Here DDD models shall encompass more data-intensive calculations, new data do not necessarily need to be collected, rather sensitivity analysis and design of experiments (DOE) can support decisions making by simulating the effect of input variations. At this point the DDD models can be still seen as substantive, i.e. describing the situation as-is with no reference to decision makers preference. Entering the Emerging Design phase, partial preferential models start to be needed: solution options are generated and eventually pre-selected if the number of solutions is large. This is now possible thanks to a clearer AS-IS description generated by the previous steps allowing a scenario-based assessment of the product performances based on decision-makers’ preferences. This enables the decisions to converge toward the final choice of committing to a concept. DDD models at this stage encompass the comparison of the early simulation results with real experimental data, so to develop advanced simulation models on request. Model users need to be capable of interacting with the DDD models to estimate the value of a solution based on different preferences used as input. In other words, in this stage the DDD models are no longer playing the role of describing the situation AS-IS, rather they shall allow estimating the effect of a decision to predict how preferable a solution is in comparison to others. At last the product innovation process enters the concept development stage in the product value stream. Here the implementation of previous decisions is in focus and the DDD models no longer have the objective of describing the as-is situation or predicting the future, rather the can be seen as limited to the continuous monitoring of the process, eventually leading to model prediction updates. Figure 2 summarizes the framework for DDD application in the product innovation process, highlighting the parallelism between data-driven
activities, product innovation stages and phases of the decision process.

Figure 2. The framework describing the role of Data-Driven Design in the Product Innovation Process.

5. THE CHALLENGES OF DATA-DRIVEN DESIGN IN THE PRODUCT INNOVATION PROCESS

Introducing DDD models to enhance decision making in the product innovation process it is not a straightforward process but it needs to account for number of challenges emerged both from literature and empirical studies. Those are summarized here:

5.1 Context dependency

Different theories for decision making, are based on the “context independence” axiom (Holt, 1986). Such axiom states that the preferential choice between two items does not depend on other choices that may be available. This means that, intuitively, if the intent of a new the design is to develop a joint with the highest resistance to stress, the decision makers will prefer the concept proving the highest resistance to stress during the simulations. However, research shows the inconsistency of the axiom in case of uncertain data (as in early product development), since decision-makers often use models expressing the relative value of a concept based on the background and local context of a decision (Tversky and Simonson, 1993). This is particularly relevant where decisions are made trading off sets of data, whose relevancy is dependent on the context in which the trade-off is run. In case of “traditional” computer-based simulations, such as finite element analysis, the uncertainty, and thus the risk, is mitigated by the high reliability of the output, while the same situation is not met in case of models based on more immature data. This generates two effects in the decision, namely, trade-off contrast and extremeness aversion. The first makes a product to appear more or less attractive based on the other products that are already present in the background. The second relates to the fact that, in the presence of risk or uncertainty, extreme values are relatively less attractive than intermediate values (Tversky and Simonson, 1993). The challenge with DDD models is that decision makers will most likely judge a concept biased by their previous experience of products, tending to avoid the choice of concepts that appear to be “extreme” and not in line with what was the “expected value” before running the model itself. These challenges bring up two questions to be considered when developing DDD models in the presence of contextual dependency:

- Is the model mitigating the phenomenon of trade-off contrast?
- Is the model mitigating the bias of extremeness aversion?

5.2 Data interpretation

The value of a concept when trading off design alternatives is relatively perceived by different stakeholders interpreting it from their particular standpoint. What are the criteria considered as relevant and by who, and thus which data shall be collected and analysed, is a question that can highly be debated in a design team. A model shall address the challenge of not taking for granted how the designers individually interpret and “weight” the information obtained from the data. Models shall not be perceived as “black boxes” rather they need to be understandable and transparent for the users that shall be able to understand the priority and the relevance given to the data. Such challenge is similar to the one of contextual dependency, but it differentiates in the way that it considers different individual interpretation, not derived by contextual dependency, but from the personal understanding of the model by each individual. This challenge raises two questions:

- Is the model clearly defining what data have been considered and how?
- Are the figures/data/descriptions used in the model easy to understand irrespectively from the individual role/background in the design team?

5.3 Information completeness

Earlier in the design process the decision are taken, lower is the availability of information that can be used to base such decision upon (Ullman, 1992). The challenge for DDD methods is to provide consistent support for decision while coping with the intrinsic incompleteness of information. This issue highlights a necessary trade-off between the range of the data collected and the necessity to provide an effective analysis. In other words, the data considered need to encompass multidimensional aspects, but they need to be manageable by the users, not providing false sense of accuracy in the evaluation when this is not present. The challenge of the information completeness raises two questions:

- Is the model capable of highlighting all the aspects relevant to make an informed decision?
- Is the model implementing measures to mitigate the risk to provide false sense of accuracy?
5.4 Nature of the data

The fourth challenge deals with the nature of data. The data available during design can both have a numerical and a nominal nature; i.e. they can deal with computed performances but also with specific product features. The capability to merge in a single model both nominal and numerical values (or quantitative and qualitative assessment) is seen as a challenge for the development of DDD models. For this reason, the question formulated for evaluating the capability of a method to cope with this challenge is:

- Does the model allow the consideration of both nominal and numerical data simultaneously?

5.5 Individual cognitive limitation

The last challenge is related to information design and to the limited cognitive capabilities of individuals to rationally analyse and manage a high amount of information in a short time. The issue of cognitive limitation is addressed by the research in information design, which comprises analysis, planning, presentation and understanding of a message (Petterson, 2010). The use of the right information design strategy helps to inform, to simplify and to make information accessible while improving the clarity of communication. In this sense, an effective information design allows individuals to capture patterns and relationships not so easily deduced otherwise (Meirelles, 2013). This challenge leads to the formulation of the question:

- Is the model representing data and results so to deal with the limitation of the individual cognitive capability of the designers?

Table 1 summarizes the challenges presented in this section into an evaluation matrix describing the dimensions to consider for the development of DDD models.

<table>
<thead>
<tr>
<th>Challenges</th>
<th>Is the challenge addressed?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No</td>
</tr>
<tr>
<td>Context dependency</td>
<td>Trade-off contrast</td>
</tr>
<tr>
<td></td>
<td>Extremeness aversion</td>
</tr>
<tr>
<td>Data Interpretation</td>
<td>Clear definition</td>
</tr>
<tr>
<td></td>
<td>Cross-functional understanding</td>
</tr>
<tr>
<td>Information completeness</td>
<td>All relevant aspects</td>
</tr>
<tr>
<td>Nature of data</td>
<td>Both nominal and numerical data</td>
</tr>
<tr>
<td>Cognitive limitation</td>
<td>Data and results representation</td>
</tr>
</tbody>
</table>

The matrix shall be seen as a subjective qualitative tool for model assessment. The distinction between “partially” and “fully” addressing a specific challenge resides in the scientific empirical validation of the desired effect. In other words, a challenge shall be considered as fully addressed, if established and scientifically proven methods are adopted (e.g. colours have been proven to be effective in addressing cognitive limitations in data representation). Alternatively, when the model adopts approaches yet to be fully verified in practice, the challenge shall be considered “partially” addressed, thus in need of further verification.

5. CONCLUSION

Technology for information communication and the Internet of Things have created unprecedented opportunities to collect, interpret and analyse data in a variety of contexts. The use of machine learning and data mining algorithms allows the discovery of correlations in data and the prediction of results in a relatively short timeframe. The application of those algorithms in the design of new solutions is increasingly referred in literature as the development of Data-Driven Design models. The development of DDD models for the design of new product innovation is still in its infancy and no established approaches are applied in industry. Only a few applications are nowadays presented in literature and there is a limited discussion about the role of such models in respect to the established design process. The framework for DDD presented in this paper is intended to clarify the role that DDD models play in the product innovation process, highlighting what types of models are needed at what stages of the development, focusing at the same time to what types of decisions can be supported and what is the impact of an increase data availability in the decision process. The evaluation matrix summarizing the identified challenges shall be seen as a reference to verify theoretically how a new approach fits with the decision-making issues highlighted in literature, so to plan corrective actions before testing the approach in the real world. The framework and the evaluation matrix are the results of extensive literature review in the domain of data-driven support for model-based decision making, this knowledge was combined with qualitative empirical data collected by the author in a number of research projects dealing with early-stage decision making in complex development processes. The results can, therefore, be seen a formalization of the empirical evidence about the use of data-intense models in early design, framed into the current literature on product innovation and decision making. The paper does not claim to provide complete picture of all the possible application of DDD models in product innovation, but it is meant to provide a reference upon which to further investigate the benefits of such models from both a product innovation and a decision-making perspective, to ultimately push forward the state of the art and the state of the practice of data-driven design.

ACKNOWLEDGEMENT

The research has received financial support from the Swedish Knowledge and Competence Development Foundation through the Model Driven Development and Decision Support research profile at Blekinge Institute of Technology.

REFERENCES


