MODELLING THE DESIGN OF MODELS: AN EXAMPLE USING CRISP-DM

Kannengiesser, Udo (1);
Gero, John S. (2)

1: Johannes Kepler University Linz;
2: University of North Carolina at Charlotte, USA

ABSTRACT

Design is widely understood as a domain-independent notion, comprising any activity concerned with creating artefacts. This paper shows that models can be viewed as artefacts, and that the design of models resembles the design of artefacts in other domains. The function-behaviour-structure (FBS) ontology of design is applied to models, mapping generic characteristics of models derived from literature on modelling onto basic, design-ontological categories. An example of model design, namely the CRISP-DM model for designing data mining models, is analysed and compared with models of designing in other domains (systems engineering, mechanical engineering, software engineering, and service design). The results show that there are fundamental commonalities but also differences, revealing the need for further research in developing a theory of model design.

Keywords: Design theory, Product modelling / models, Ontologies, Design methodology, Model design

Contact:
Kannengiesser, Udo
Johannes Kepler University Linz
Austria
udo.kannengiesser@jku.at

Cite this article: Kannengiesser, U., Gero, J. S. (2023) 'Modelling the Design of Models: An Example Using CRISP-DM', in *Proceedings of the International Conference on Engineering Design (ICED23),* Bordeaux, France, 24-28 July 2023. DOI:10.1017/pds.2023.271

1 INTRODUCTION

Design is increasingly recognized as a fundamental human activity whose basic processes, methods and principles transcend the various domains in which it takes place. Apart from traditional (mostly engineering and architecture) design domains concerned with developing physical products, systems and software, a number of other domains have adopted an explicit "design" perspective when discussing and reflecting upon their practice of developing their own artefacts. Accordingly, fairly recent types of design have emerged, including organisation design, enterprise design, business process design, service design, communication design, research design, experiment design, and game design.

A number of authors (e.g. Churchill (2013), Hotie and Gordijn (2019), Kropp and Thalheim (2017), and Zott and Amit (2010)) assume that activities of modelling, i.e. creating models for various purposes and in various domains, constitute an additional type of design: "model design". Based on the ubiquitous use of models in many domains including engineering design (McAdams and Dym 2004; Eckert and Hillerband 2018), model design can be viewed as orthogonal to most other design types. For example, while model-based software engineering (MBSE) is concerned with the design of software, it also deals with the design and use of abstract models of the software and the interactions of the software in its environment. Mechanical design using CAx models can be viewed in a similar way, combining the design of a physical product and the design of a set of models representing that product. The two types of design are intertwined: Changes to the product being designed are typically reflected in the model design, and changes to the model representing the product can afford changes in the product design. Model design in these instances has a supportive role and would not be carried out without the design of the artefact being modelled. There are other instances where models are designed independently of the design of other artefacts. For example, data mining models are commonly designed as tools for gaining insights in domains such as business and science.

While there is some literature on model theory (Stachowiak 1973; Thalheim 2011), there is much less work on the process of designing models. There is no general definition of model design, and existing model design approaches are limited to specific domains, such as the design of business models (Zott and Amit 2010), process models (Hotie and Gordijn 2019) and data mining models (Chapman *et al*. 2000). The relationship between these approaches and established insights, models and methods of design research remains largely unclear. The lack of design-theoretical foundations may lead to inadequate understanding of model design and, consequently, ineffective and inefficient models being designed.

This paper contributes to the development of a model design theory by presenting a design-ontological view of models, and by analysing a specific approach of model design − the CRoss-Industry Standard Process for Data Mining (CRISP-DM) (Chapman *et al*. 2000) − and comparing it with design approaches in other domains: the INCOSE (2015) model of systems engineering, Pahl and Beitz' (2007) model of engineering design, Kruchten's (2004) Rational Unified Process (RUP) of software design, and a Design for Six Sigma (DFSS) model of service design (El-Haik and Roy 2005). The differences and commonalities found support positioning model design in the landscape of theories and models from other design disciplines, allowing identifying whether existing design methods may support the specificities of model design.

The paper is structured as follows: Section 2 provides an overview of some of the literature on models, modelling and model design. Section 3 applies the function-behaviour-structure (FBS) ontology (Gero 1990; Gero and Kannengiesser 2004; Gero and Kannengiesser 2014) to models, showing that their general constructs, as provided in the literature, fit with basic ontological categories of design. Section 4 presents how a process of data mining, as an example for model design, is analysed and compared with models of design in other domains. Section 5 presents the results of the analysis, which are discussed in Section 6. Section 7 draws conclusions and proposes future research on model design.

https://doi.org/10.1017/pds.2023.271 Published online by Cambridge University Press

2 MODELS, MODELLING, AND THE DESIGN OF MODELS

Compared to the ubiquitous use of models in a variety of domains, there is relatively little research in developing a general theory of modelling (Ritchey 2010). In the area of semiotics, a model is understood as a sign in the sense that it represents another, material or immaterial object (Nöth 2018). Similarly, Stachowiak (1973) states that a key feature of a model is that it has a mapping to an "origin". The original object may be anything in the existing or non-existing (i.e., imagined) world (Guizzardi and Proper 2021). Models vary in terms of scope, precision, granularity and abstraction, and can have various means of representation, including text, symbols, graphs, diagrams, tables and pictures (Thalheim 2011). Some of the typical models used in engineering design include mathematical models, product visualisations, geometric models, physical models, data models, abstract models and numerical models (Eckert and Hillerbrand 2022). An example of a data mining model is shown in Figure 1. It is conceptualised as a process model, consisting of an input, a transformation and an output (Gero and Kannengiesser 2007).

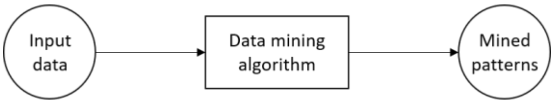


Figure 1. A data mining model as an example

Models typically omit some of the information about the object being modelled. For example, the model depicted in Figure 1 omits various details such as the data types, the data sources, the particular data mining algorithm and any pre- and post-processing steps. Some information captured in a model may also be distorted, idealised, or even extended beyond what was previously known about the object (Thalheim 2011). What is represented in a model ultimately depends on its purpose (Browning 2010). A common distinction is often made between descriptive and prescriptive purposes (Rothenberg 1989). Thalheim (2011, p. 543) distinguishes several types of purposes: "construction of a system, simulation of real-world situations, theory construction, explanation of phenomena, or documentation of an existing system." In addition to the purpose, contextual factors drive the construction of a model, i.e. who uses the model when and with what tools (Stachowiak 1973). The model in Figure 1, for example, has been kept very simple to convey the basic idea of a model to the readers of this paper who are likely to be non-experts in data mining. More details may have been added to the model if it was produced for a specialised data mining audience.

Some researchers make the point that modelling is not designing. For example, Reijers (2021) proposes that process design has a wider scope than process modelling, entailing design decisions such as "[h]ow to organize a process, which technologies to involve, and to whom to assign responsibilities within a process" (*ibid*, p. 3). On the other hand, there is a widespread assumption that modelling is an instance of designing in its own right − in particular, when abstracting away from the particular, domain-specific entities being modelled, resulting in so-called conceptual models (Churchill 2013). However, the connection between modelling and designing is often only loosely made; and there is little research in design-theoretical foundations or principles of model design. Thalheim (2010, p. 3122) states that "conceptual modelling inherits most principles of engineering", providing modularity and top-down vs. bottom-up development strategies as examples. He defines the following activities as constituents of modelling: understanding, conceptualising, abstracting, defining, constructing, refining and evaluating (*ibid*, p. 3117-3118) − which map onto a number of categories in Sim and Duffy's (2003) ontology of generic engineering design activities. Kaschek (2018) points out that modelling belongs to an "analysis−synthesis cycle": Initially, a model is the result of analysing what is being modelled: its "origin". Yet, it is then extended or modified to enhance control or predictability of the origin.

3 A FUNCTION-BEHAVIOUR-STRUCTURE (FBS) ONTOLOGY OF MODELS

We can develop a design view of models based on the function-behaviour-structure (FBS) ontology (Gero 1990; Gero and Kannengiesser 2004; Gero and Kannengiesser 2014) that is a widely used design ontology and defines three categories of properties for objects that have been or can be designed:

https://doi.org/10.1017/pds.2023.271 Published online by Cambridge University Press

- *Structure (S)* is defined as the components of an object and their relationships. According to Thalheim (2011, p. 549), models consist of "constructs" and relationships between them. There are five fundamental types of relationships: aggregation/participation, generalization/specialization, exhibition/characterization, classification/instantiation, introduction/utilisation (Thalheim 2010, p. 3111). For example, the structure of a data mining model can be viewed as shown in Figure 1: three components (i.e., input data, data mining algorithm and mined patterns) interconnected by aggregation/participation relationships (i.e., the arrows between the components).
- *Behaviour (B)* is defined as the attributes derived from the object's structure and its interaction with the environment. For example, literature on modelling often stresses the importance of model quality, which can be syntactic (measuring correspondence between a model and a representation language; e.g. number of syntactic errors), semantic (measuring correspondence between a model and its original; e.g. coverage of concepts) and pragmatic (measuring correspondence between the model and its interpretation by a user; e.g. redundancies in the model) (Krogstie *et al*. 1995). Similarly, Thalheim's (2011) notion of "model value" fits with the definition of behaviour: It is based on the three categories of quality defined in ISO/IEC 9126, namely, internal quality (derived solely from a model's structure; e.g. correctness and complexity), external quality (derived from the interaction of the model's structure with an execution environment; e.g. execution time), and quality of use (derived from the interaction of the model's structure with the user's working environment; e.g. time needed for understanding, creating or modifying the model).
- *Function (F)* is defined as the teleology of an object. Functions of models can be the various purposes associated with the use of a model (Guizzardi and Proper 2021; Thalheim 2011), specialised to the individual circumstances and domains at hand. Examples include the (model-based) design of a system, the discovery of new business insights (e.g., by using data mining models) and a better understanding of a phenomenon (e.g., by using conceptual models).

In the FBS ontology, behaviour is not only connected to structure but also function. This is the result of humans associating purposes with behaviour, in which case the behaviour becomes a desired or "expected" behaviour (Be) that can then be used to assess an object's "actual" behaviour derived from structure (Bs). For models, this connection has been observed by Thalheim (2011, p. 553), who stated that the "assessment of model value [occurs] in dependence on purpose". Divergence between Be and Bs can be seen as the main driver of the process of designing, as it can cause new cycles of synthesis and analysis. The coverage of concepts required for current model purposes can thus be seen as a key expected behaviour for model design. As Kaschek (2018, p. 30) puts it, "[m]odel use is about an information deficit with regard to some item, a so-called model origin[…]: someone wishes to achieve a certain goal with regard to the origin[…]. They, however, turn out to be ignorant of some information (subjectively) required for achieving that goal. The key idea to solve this problem is to obtain the required information with regard to the model and then transfer it to the origin[…]."

The FBS framework (Gero 1990), which describes eight fundamental processes in designing, incorporates the dichotomy of behaviour, together with two additional categories, each of which can be expressed in FBS categories: Requirements (R) are needs, demands, wishes and constraints that are explicitly provided at the outset of designing. In the context of model design, R may include a specified purpose of the model and a specified modelling language such as the entity-relationship notation. Description (D) is any form of external output of the design process. In model design, these may be the various graphs, symbols, diagrams, texts etc. produced. The FBS framework is shown in Figure 2. Comparisons between Be and Bs are covered by process 4 (called "evaluation"); while synthesis and analysis are represented by processes 2 and 3, respectively. R is transformed into F and then Be by process 1, and D is produced by process 5. Additional processes (6−8) are defined to cover the reformulation of S, Be and F after an initial design structure has been specified.

4 ANALYSIS OF MODEL DESIGN: DATA MINING AS AN EXAMPLE

4.1 The CRISP-DM model of designing data mining models

CRISP-DM (CRoss-Industry Standard Process for Data Mining) (Chapman *et al*. 2000) is today the most widely used approach for carrying out data mining and data science projects (Martinez-Plumed *et* *al*. 2021; Saltz 2021). It provides an industry-independent and technology-neutral process model for developing data mining models. CRISP-DM is the result of practitioners synthesizing their data mining experience into a systematic and standard approach. One of the goals associated with the development of CRISP-DM was to "create the impression that data mining is an established engineering practice" (Wirth and Hipp 2000). There are six phases defined in CRISP-DM, as shown in Figure 3:

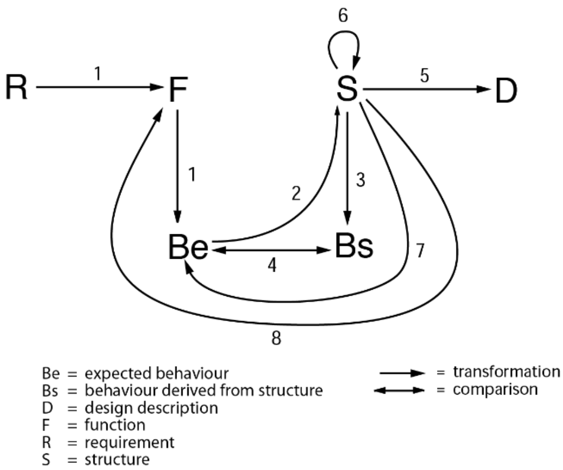


Figure 2. The FBS framework (after Gero (1990) and Gero and Kannengiesser (2014))

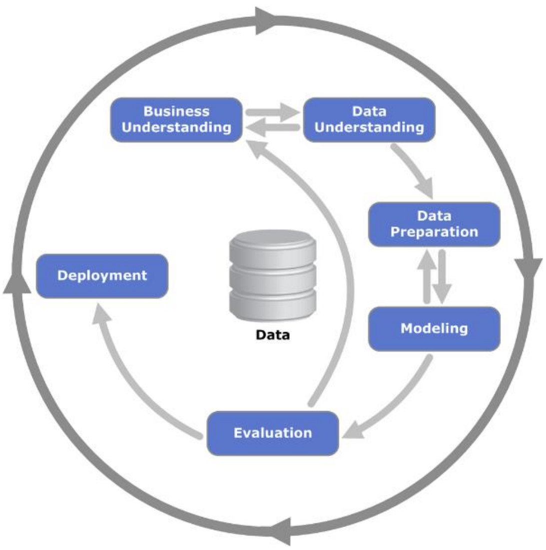


Figure 3. The phases in the CRISP-DM model of data mining (Chapman et al. 2000)

1. *Business Understanding*: is concerned with comprehending the project objectives from a business perspective and deriving a data mining problem definition.
2. *Data Understanding*: is about collecting and familiarising with the data.
3. *Data Preparation*: includes the steps required to produce the final dataset to be input into the modelling tool.
4. *Modeling*: develops and parameterises the transformation techniques to be applied on the dataset.
5. *Evaluation*: assesses the data mining model with respect to the business objectives.
6. *Deployment*: implements and documents the model in the target organisation.

The authors of CRISP-DM stress that iterations between different phases are required (Chapman *et al*. 2000). Yet, they do not provide any detailed description of how the iterations occur or which phases they involve. The only indication provided are the loops in the graphical model shown in Figure 3.

https://doi.org/10.1017/pds.2023.271 Published online by Cambridge University Press

4.2 Method of analysis

The method used to compare CRISP-DM with the four other models of designing (INCOSE, Pahl and Beitz, RUP and DFSS-ICOV) follows the two-step approach developed by Kannengiesser and Gero (2022). In the first step, the different design approaches are brought into a uniform FBS representation. For space reasons, the detailed mappings onto FBS issues (i.e., instances of one of the six FBS categories: R, F, Be, Bs, S and D) are not elaborated in this paper. Readers may refer to Kannengiesser and Gero (2022) for further details. The output of this step are sequences of FBS issues, where every sequence corresponds to a simulated execution of individual steps described in each approach. For CRISP-DM, a number of alternative FBS sequences are considered, to account for the different iterations possible shown in Figure 3. Specifically, the variants shown in Figure 4 have been defined.

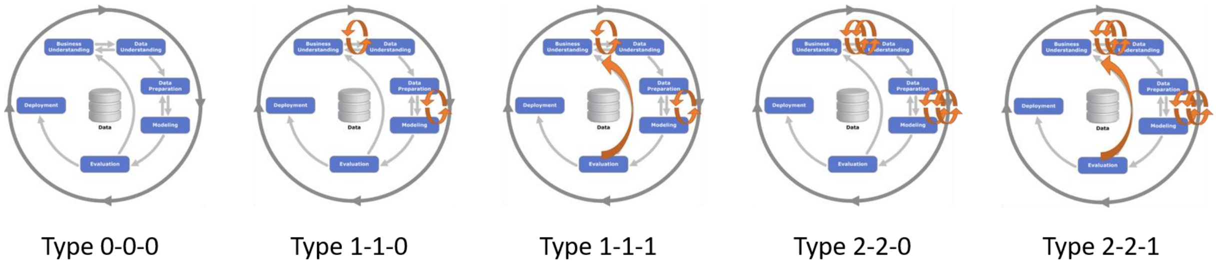


Figure 4. Iteration types in CRISP-DM considered for analysis (Types are labelled using an "X-Y-Z" schema, indicating the number of iterations at the first (X), second (Y) and third (Z) location in the CRISP-DM model), based on Figure 3 in iconised forms

In the second step, three statistical analyses are run on these representations:

- Correspondence analysis: reduces the dimensionality of data to visualise it on a two-dimensional plot. In the context of this study, two categories of data are considered: the design approaches (and the phases defined within them) and the six types of FBS design issues.
- Cumulative occurrence analysis: aggregates the occurrences of each FBS issue over all design steps, allowing the following characterisations of the resulting graphs: regression lines, slopes and first occurrences at start.
- Markov model analysis: calculates the probabilities of moving from one FBS issue to another during a design approach.

5 RESULTS

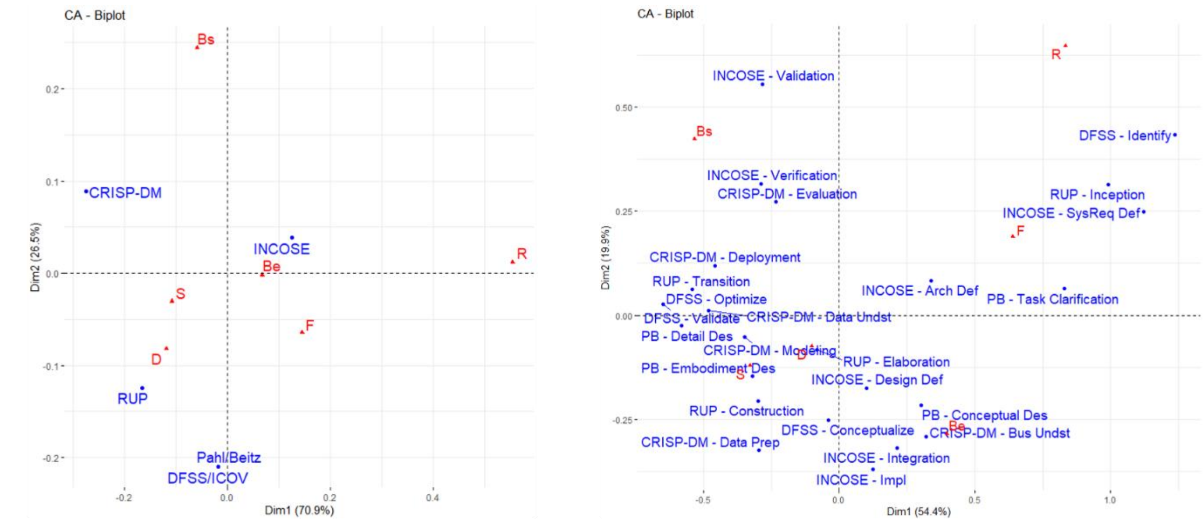
5.1 Correspondence analysis

Two correspondence analyses were carried out; one considering the number of each of the six types of FBS issues in the different approaches, and one considering their number in the different phases of each approach. The results are shown in Figures 5(a) and (b).

It can be seen in Figure 5(a) that CRISP-DM is the only approach located in the top-left quadrant, indicating that it is quite different to the other design approaches. It is positively associated with Bs issues (and, to a lesser extent, S issues), and negatively associated with R and F issues (and, to a lesser extent, Be issues). This indicates that CRISP-DM is strongly oriented towards design solutions as opposed to design problems.

The plot in Figure 5(b) shows that no phase in CRISP-DM is located in the top-right quadrant. In this quadrant the phases concerned with requirements analysis in the other design approaches are located, highly associated with R and F issues. The most problem-oriented phase in CRISP-DM is *Business Understanding* (shortened to *Bus Undst* in Figure 5), based on its close connection to Be issues. The phase of *Evaluation* in CRISP-DM is closely connected to Bs issues, similar to the INCOSE phases of *Verification* and *Validation*. These results confirm the highly solution-focused character of CRISP-DM.

https://doi.org/10.1017/pds.2023.271 Published online by Cambridge University Press

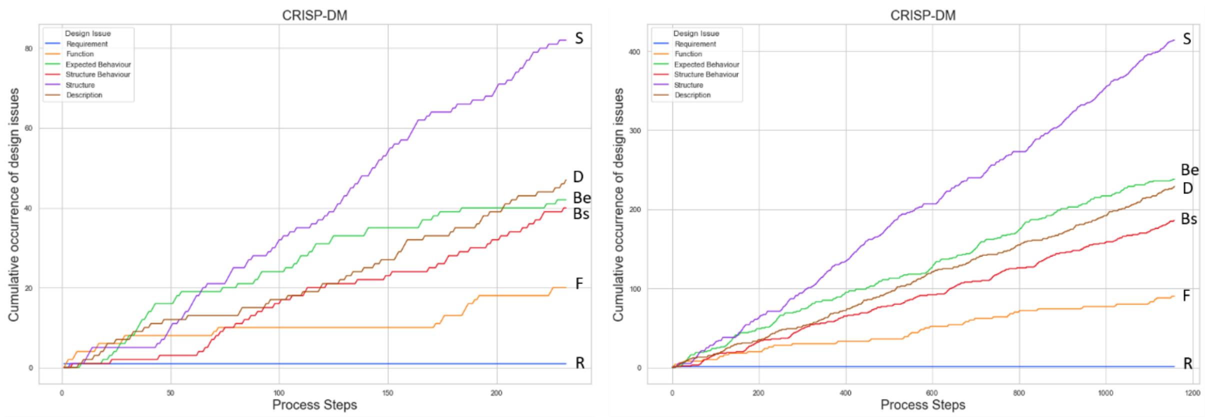


a) b)

Figure 5. Results of correspondence analysis: (a) across the five design approaches, and (b) across the different phases in the design approaches

5.2 Cumulative occurrence analysis

The cumulative occurrence of FBS issues in CRISP-DM is shown for iteration types 0-0-0 and 2-2-1 in Figure 6.



a) Type 0-0-0 b) Type 2-2-1

Figure 6. Cumulative occurrence of FBS issues in CRISP-DM: (a) for type 0-0-0 (i.e., no iterations), and (b) for type 2-2-1

The graphs for Bs, S and D issues are linear in all iteration types. The graphs for F and Be issues flatten towards the end when no iteration occurs (see Figure 6(a)). These graph behaviours are consistent with those observed for the other four design approaches (Kannengiesser and Gero 2022).

As the number of iterations increases in CRISP-DM, the graphs for F and Be issues become linear (see Figure 6(b) for type 2-2-1). This is different from the other design approaches, which can be explained based on the unique style of iteration considered for CRISP-DM: It includes a large loop from *Evaluation* back to *Business Understanding*, having the effect that any non-linearities over the course of the design process are "ironed out" as the graphs scale up with increasing numbers of iteration.

The graphs show that all six FBS issues first occur at the beginning of the CRISP-DM process. This is consistent with the other design approaches for R, F, Be and D issues. However, in all of the other approaches the first occurrence of Bs and S issues was observed only later in the process. This is because CRISP-DM is the only approach where the suitability of existing design solutions is assessed in the first phase. Other approaches are based on the assumption of greenfield design, where existing solutions do not exist or are intentionally ignored for the new design.

5.3 Markov model analysis

First-order Markov models produced for all iteration types show the same patterns in terms of so-called dominant state transitions − transitions between FBS issues that are the most likely to occur, based on the highest probability of occurrence in the model for that transition. They are shown in Figure 7, together with those identified for the other design approaches (see Kannengiesser and Gero (2022)). The results show that CRISP-DM shares two dominant state transitions with the other approaches: R to F, Bs to S, and D to S. The dominance of the F-to-F transition and of the S-to-S transition in CRISP-DM is shared only by RUP and DFSS/ICOV, respectively. The dominant Be-to-Be transition in CRISP-DM is unique among the approaches. Therefore, a rather unique feature emerges from the CRISP-DM Markov model: There is an apparent lack of dominant transitions transforming function to expected behaviour and then structure − which are generally viewed as essential activities in generating a design, based on their abductive nature (Roozenburg 1993). Since only dominant transitions are shown, this does not imply that there are no such transitions, only that they are not dominant.

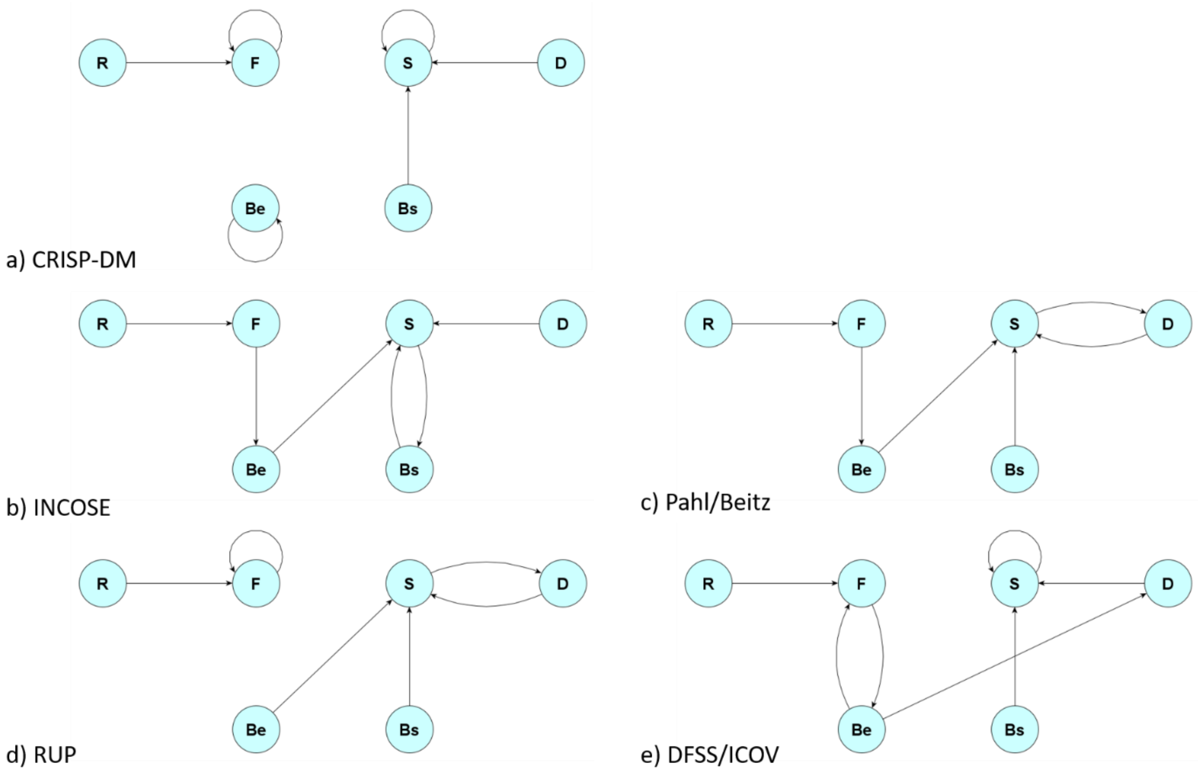


Figure 7. Dominant state transitions in the Markov models of (a) CRISP-DM, (b) INCOSE, (c) Pahl and Beitz, (d) RUP and (e) DFSS/ICOV

6 DISCUSSION

Overall, the analysis shows that CRISP-DM can be studied as a design approach, in the same way as INCOSE, Pahl and Beitz, RUP and DFSS-ICOV. Its individual steps can be mapped onto the six types of design issues, resulting in sequences of FBS issues that can be characterised quantitatively and compared to the other approaches. The results of the comparison show similarities in multiple aspects, but also differences:

- CRISP-DM is more solution-oriented in its emphasis on Bs and S and its lack of emphasis on R, F and Be issues (see Section 5.1).
- The rate at which F and Be issues are generated is constant the more iterations occur (see Section 5.2).

https://doi.org/10.1017/pds.2023.271 Published online by Cambridge University Press

- Bs and S issues first occur at the beginning, as existing solutions are analysed in the first phase (see Section 5.2).
- There is a lack of dominant abductive transformations from F to Be and from Be to S issues (see Section 5.3).

7 CONCLUSION

Our design-ontological perspective of models, using the FBS ontology, and our analysis of an example of model design indicate that models can be designed in a similar way as other artefacts. The process of model design covers all six fundamental FBS design issues, while there are differences in the emphasis on specific issues overall or over the course of designing. These insights were derived from only one instance of model design, namely, CRISP-DM. More research is needed, which may be based on similar analyses of other model design approaches, in order to enhance validity. The design-ontological perspective of models lends itself to a range of model development that includes domains not traditionally considered as design activities. For example, in presenting a research project it is often necessary to include a research plan, which needs to be developed for the specific research questions to be explored or hypotheses to be tested (Harrington, 2020). Military planning can be viewed through this design-ontological perspective of models and as a consequence can be modelled as a design activity (US Army, 2010).

A better understanding of modelling as a design activity can potentially lead to improved methods in this area. Here adopting a design view may be helpful again: Prescriptive models of designing, such as the process model defining the CRISP-DM approach that was visualised in Figure 3, can themselves be viewed as artefacts that can be designed and re-designed for enhanced performance. We can view the design of models of designing as meta-design. In such as meta-design context, the results of our analysis (e.g., CRISP-DM's emphasis on solutions at the expense of problems) represent the behaviour (Bs) derived from its process model structure. The characteristics of the other design approaches studied (INCOSE, RUP, etc.) could serve as benchmarks for expected behaviour (Be), which may then be used to re-design CRISP-DM. For example, this may lead to modifications of CRISP-DM strengthening the focus on the requirements and functions that the data mining model is to address, thus making this approach more problem-oriented.

ACKNOWLEDGEMENTS

This research is supported in part by a grant from the US National Science Foundation to J. G., grant number CMMI-1762415. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

Browning, T.R. (2010), "On the alignment of the purposes and views of process models in project management", *Journal of Operations Management*, Vol. 28 No. 4, pp. 316–332. https://doi.org/10.1016/j.jom.2009.11.007

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C. and Wirth, R. (2000), *CRISP-DM 1.0: Step-by-step data mining guide*, CRISP-DM Consortium.

Churchill, D. (2013), "Conceptual model design and learning uses", *Interactive Learning Environments*, Vol. 21 No. 1, pp. 54–67. http://dx.doi.org/10.1080/10494820.2010.547203

Eckert, C. and Hillerbrand, R. (2018), Models in Engineering Design: Generative and Epistemic Function of Product Models, In: Vermaas, P. and Vial, S. (Eds), *Advancements in the Philosophy of Design*, Springer, Cham. https://doi.org/10.1007/978-3-319-73302-9_11

Eckert, C. and Hillerbrand, R. (2022), "Models in engineering design as decision-making aids", *Engineering Studies*, Vol. 14 No. 2, pp. 134–157. https://doi.org/10.1080/19378629.2022.2129061

El-Haik, B. and Roy, D.M. (2005), *Service Design for Six Sigma: A Roadmap for Excellence*, John Wiley & Sons, Hoboken, NJ.

Gero, J.S. (1990), "Design prototypes: a knowledge representation schema for design", *AI Magazine*, Vol. 11 No. 4, pp. 26–36. https://doi.org/10.1609/aimag.v11i4.854

Gero, J.S. and Kannengiesser, U. (2007), "A function–behavior–structure ontology of processes, *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, Vol. 21 No. 4, pp. 379–391. https://doi.org/10.1017/S0890060407000340

Gero, J.S. and Kannengiesser, U. (2004), " The situated function-behaviour-structure framework", *Design Studies*, Vol. 25 No. 4, pp. 373–391. https://doi.org/10.1016/j.destud.2003.10.010
Gero, J.S. and Kannengiesser, U. (2014), "The function-behaviour-structure ontology of design", In: Chakrabarti, A. and Blessing, L.T.M. (Eds), *An Anthology of Theories and Models of Design*, Springer, pp. 263–283.
Guizzardi, G. and Proper, H.A. (2021), "On understanding the value of domain modeling", *Proceedings of the International Workshop on Value Modelling and Business Ontologies*, Bolzano, Italy, pp. 51–62.
Harrington, M.E. (2020), *The Design of Experiments in Neuroscience*, Cambridge University Press, Cambridge.
Hotie, F. and Gordijn, J. (2019), "Value-based process model design", *Business & Information Systems Engineering*, Vol. 61, pp. 163–180. https://doi.org/10.1007/s12599-017-0496-y
INCOSE (2015), *Systems Engineering Handbook: A Guide for System Life Cycle Processes and Activities*, Fourth Edition, INCOSE-TP-2003-002-04, International Council on Systems Engineering (INCOSE), San Diego, CA.
Kannengiesser, U. and Gero, J.S. (2022), "What distinguishes a model of systems engineering from other models of designing? An ontological, data-driven analysis", *Research in Engineering Design*, Vol. 33, pp. 129–159. https://doi.org/10.1007/s00163-021-00382-9
Kaschek, R. (2018), "20 years after: What in fact is a model?", *Enterprise Modelling and Information Systems Architectures (EMISAJ)*, Vol. 13, pp. 28–34. https://doi.org/10.18417/emisa.si.hcm.2
Krogstie, J., Lindland, O.I. and Sindre, G. (1995), "Defining quality aspects for conceptual models", In: Falkenberg, E.D., Hesse, W. and Olivé, A. (Eds), *Information System Concepts*, IFIP Advances in Information and Communication Technology, Springer, Boston, MA, pp. 216–231. https://doi.org/10.1007/978-0-387-34870-4_22
Kropp, Y. and Thalheim, B. (2017), "Data mining design and systematic modelling", In: *XIX International Conference Data Analytics and Management in Data Intensive Domains (DAMDID/RCDL'2017)*, Moscow, pp. 273–280.
Kruchten, P. (2004), *The Rational Unified Process: An Introduction*, Addison-Wesley, Upper Saddle River, NJ.
Martinez-Plumed, F., Contreras-Ochando, L., Ferri, C., Hernandez-Orallo, J., Kull, M., Lachiche, N., Ramirez-Quintana, M.J. and Flach, P. (2021), "CRISP-DM twenty years later: From data mining processes to data science trajectories", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 33 No. 8, pp 3048–3061. https://doi.org/10.1109/TKDE.2019.2962680
McAdams, D.A. and Dym, C.L. (2004), "Modeling and information in the design process", *Proceedings of the ASME 2004 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Volume 3a: 16th International Conference on Design Theory and Methodology*, Salt Lake City, Utah, pp. 21–30. https://doi.org/10.1115/DETC2004-57101
Nöth, W. (2018), "The semiotics of models", *Sign Systems Studies*, Vol. 46 No. 1, pp. 7–43. https://doi.org/10.12697/SSS.2018.46.1.01
Ritchey, T. (2010), "Outline for a morphology of modelling methods: Contribution to a general theory of modelling", *Acta Morphologica Generalis*, Vol. 1 No. 1.
Pahl, G. and Beitz, W. (2007), *Engineering Design: A Systematic Approach*, Springer, Berlin.
Reijers, H.A. (2021), "Business process management: The evolution of a discipline", *Computers in Industry*, Vol. 126, pp. 1–5. https://doi.org/10.1016/j.compind.2021.103404
Roozenburg, N.F.M. (1993), "On the pattern of reasoning in innovative design", *Design Studies*, Vol. 14 No. 1, pp. 4−18. https://doi.org/10.1016/S0142-694X(05)80002-X
Rothenberg, J. (1989), "The Nature of Modeling", In: Widman, L.E., Loparo, K.A. and Nielson, N. (Eds), *Artificial Intelligence, Simulation & Modeling*, John Wiley & Sons, New York, pp. 75–92.
Saltz, J.S. (2021), CRISP-DM for data science: Strengths, weaknesses and potential next steps, 2021 *IEEE International Conference on Big Data (Big Data)*, Orlando, FL, pp. 2337-2344. https://doi.org/10.1109/BigData52589.2021.9671634
Stachowiak, H. (1973), *Allgemeine Modelltheorie*, Springer-Verlag, Wien, New York.
Thalheim, B. (2010), "Towards a theory of conceptual modelling", *Journal of Universal Computer Science*, Vol. 16 No. 20, pp. 3102–3137. https://doi.org/10.3217/jucs-016-20-3102
Thalheim, B. (2011), "The theory of conceptual models, the theory of conceptual modelling and foundations of conceptual modelling", In: Embley, D. and Thalheim, B. (Eds), *Handbook of Conceptual Modeling*, Springer, Berlin, Heidelberg, pp. 543–577. https://doi.org/10.1007/978-3-642-15865-0_17
US Army, (2010), *Field Manual 5.0 The Operations Process*, Headquarters of the U. S. Army.
Wirth, R. and Hipp, J. (2000), "CRISP-DM: Towards a standard process model for data mining", *Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining*, Manchester, UK, pp. 29–39.
Zott, C. and Amit, R. (2010), "Business model design: An activity system perspective", *Long Range Planning*, Vol. 43 No. 2–3, pp. 216–226. https://doi.org/10.1016/j.lrp.2009.07.004