

Towards a Guideline Affording Overarching Knowledge Building in Data Analysis Projects

Dorothea Schneider¹[\[https://orcid.org/0000-0002-6633-5052\]](https://orcid.org/0000-0002-6633-5052) and Wibke Kusturica²[\[https://orcid.org/0000-0001-6131-2620\]](https://orcid.org/0000-0001-6131-2620)

¹ Technische Universität Dresden, Institute of Mechatronik Engineering, Dresden, Germany

² University of Applied Sciences Zwickau, Institute for Management and Information, Zwickau, Germany

Abstract. Tight and competitive market situations pose a serious challenge to enterprises in the manufacturing industry domain. Competing in the use of data analytics to enhance products and processes requires additional resources to deal with the complexity. On the contrary, the possibilities afforded by digitization and data analysis-based approaches make for a valuable asset. In this paper we suggest a guideline to a systematic course of action for the data-based creation of holistic insight. Building an overlaying corpus of knowledge accelerates the learning curve within specific projects as well as across projects by exceeding the project-specific view towards an integrated approach.

Keywords: Data Mining, Knowledge Bases, Reference Model, Methodological Knowledge, Domain-Specific Information Systems Development.

1 Introduction

Demand and supply for insights derived from all kinds of accessible data sources in enterprises are higher than ever before as the pressure to keep up with global competitors meets the ever-growing possibilities of data acquisition and exploitation. A plethora of methods and tools is available to deal with and make use of these resources: from sensors to algorithms, from Industrial Internet of Things (IIoT) solutions to programming libraries and software. [1]

While all business sectors face this situation equally and therefore must deal with similar challenges, the complexity of the task is particularly high in the manufacturing industry domain. [2] [3] This holds true especially for tasks within data-driven enhancement projects (EP) in the manufacturing industry domain which require a high level of innovation and are conducted in a project-based manner like one-of-a-kind production, research and development (R&D), customer-specific machinery and plant engineering or the design of cyber-physical production systems. [4]

First and foremost, conducting successful data analysis projects does not only include the activities directly associated with analyzing data but involves the execution of several elaborate steps as well as strategic measures. To systematically align all relevant aspects affecting the analysis outcome in a wider sense will result in distinct quality improvement. [3]

In our research we aim at providing the means to support achieving strategic goals by conducting data analysis projects which systematically connect relevant information fragments on all levels of aggregation from all relevant sources. Therefore, our research is driven by the following research question (RQ):

RQ: How can a reference model be provided for complex tasks in the industrial domain which provides methodological support for the data-driven construction and utilization of an overlaying corpus of knowledge?

To answer this question, we developed an artifact in the form of a reference model to equip the user with a wide range of methodological support for conducting informed data analyses. The goal of the suggested framework is to not only derive insight about the examined topic of an active data mining project but to preserve and build on the findings exceeding project boundaries. The reference model aims to inspire rigorous and holistic investigation, to provide the means for communication, project management and documentation and to build the foundation for future software applications to support this holistic project-exceeding data mining approach thus also paving the way for an analysis and optimization of the activities undertaken within data mining projects themselves.

Following this approach this paper is structured as follows: In Section 2, we describe our motivation, we then sum up foundations and basic concepts in Section 3. Derived from the key activities of the sensemaking approach as described by [5] and more specifically by [6] a set of design principles is suggested, as will be described in Section 4. In fulfillment of the defined design principles a framework is presented in Section 5 to structure necessary methodological measures and to allocate useful activities within five layers of information aggregation. By presenting the reference model we advocate for a systematic course of action aiming at the creation of holistic insight. Finally, we draw a conclusion and give an outlook for further research in Section 6.

2 Motivation

The major purpose of the presented long-term design science research project is to elaborate methodological support for data-driven knowledge extraction projects in the manufacturing industry domain. Therefore, our main objective is to help artifact users gain a sophisticated understanding of the principles by which to conduct data-driven knowledge extraction projects, to reduce the associated hurdles for manufacturing companies and to create a basis to address and solve them in the future in a repeatable manner. The application of the presented reference model enables domain experts to derive cumulative knowledge, rather than re-inventing technical concepts and methodological procedures under new labels in every new project setting. [7]

Specialists dealing with data analysis projects in the industrial domain face the necessity to cover the methodological skillset required in data science as well as a deep understanding of the domain fundamentals to consider relevant causalities and interactions and to purposefully derive and interpret results according to their context. Hence throughout all industrial sectors on the one hand domain experts successfully gain and apply data analytics knowledge while on the other hand data analysts engage in various domain contexts and oftentimes both have to team up with each other and with additional professionals like computer scientists and mathematicians to derive the desired outcome. While tremendous progress is underway in the domain-specific training of and proficient cooperation with data scientists and in the successful realization of data analytics projects the potential for even better outcome is huge. [8] [9] The main hurdles are the intricate communication between domain experts and data scientists, the scarcity of human resources for data analytics projects and the lack of domain-specific standardized procedures which lead to a singular quality of the execution and the use of results of data-driven analyses. These shortfalls especially hold true where a limited number of experts must realize data analytics projects next to rivaling work tasks as is the case in small and medium sized companies (SME), start-ups and R&D or planning departments. [3]

A pre-study in the form of an exploratory study with six qualitative expert interviews aimed to identify the challenges that occur while setting up a data-driven knowledge extraction project confirmed these hurdles. The interviews were designed as partially standardized interviews using open to semi-open questions as initial starting points for the conversation and took between 70 and 180 minutes. The complete listing of the formulated questions and results will be provided by the authors upon request. The answers showed that practitioners tend to rely on traditional procedures and experience-based knowledge.

Their understanding of Data Mining (DM) mainly focused on the core analysis activities like the application of algorithms and often underestimated the effort and importance of peripheral aspects like the determination of target-aimed questions, data preparation to produce structured evaluable data sets, conclusive feature engineering and context-sensitive model building. The interviewees expressed their wish for more structure and guidance in data analytics projects while they found existing standard processes too generic to apply for their domain as well as not sufficiently considering real-life problems like data acquisition, data quality and operational data processing.

3 Foundation

Pursuing a long-term research project in the field of information systems (IS) aiming at the design of an artifact in the form of a reference model we comply with the design science paradigm stated by [10]. We furthermore adopt the three-cycle view of design science research (DSR) presented in [11] to address the relevance, design and rigor of the developed artifact. Additionally we rely on the steps for DSR research recommended by [12] to apply the paradigm to our research as follows: The **problem identification and motivation** for our research is constituted by the experience from numerous research projects and a pre-study in the form of expert interviews as described in Section 2. We then derived theory-based research **goals and objectives** by the definition of design principles as described in Section 4 followed by the **design and development** of the artifact, the outcome of which is presented in Section 5. While applying the findings in practice the derivation of a context-specific model should then be **demonstrated** and **evaluated** within future research. In an iterative manner the insights from an initial implementation within an example scenario should be used to further enhance the artifact and undergo subsequent evaluation phases to then be **transferred to the community**.

When attempting to represent and reduce reality to fulfill a subjective purpose like the understandable formulation of complex facts [13] for a class of similar problems a **reference model** is provided by introducing a model which is of recommendatory and universal character and allows for the derivation of application-specific models.[14] Consequently reference models are a generic type of model representing the essence of a common-practice or best-practice view on a class of similar problems intended for re-use and acting as a blueprint for the derivation of specific models. [15]

The addressed application field of the presented reference model comprises tasks in the industrial domain which require a high level of innovation and are conducted in a project-based manner. When attempting to support such tasks there are various user roles and artifacts to take account of, notwithstanding that more than one user role can be fulfilled by one individual. These roles and artifacts are depicted in figure 1.

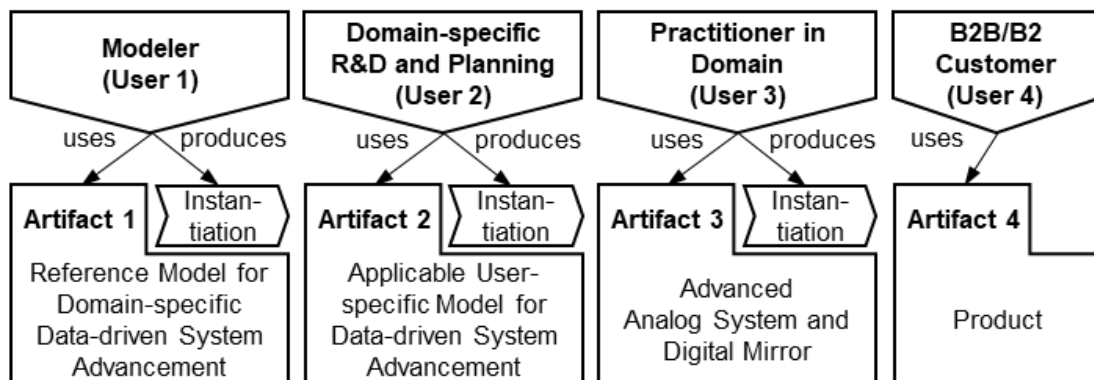


Figure 1. Addressed users and artifacts

As drawing conclusions by the statistical or algorithms-based study of large amounts of data today is widely established throughout all disciplines, numerous attempts have been made to **standardize the data mining process** especially in the field of computer science and economic analyses. Such procedure models generally consist of generic steps to structure and guide the planning and execution of DM projects.[20] Prominent standard operating models are subsequently named. Knowledge discovery in databases (KDD) is a description of the central building blocks of the overall multi-step procedure for complex real-world analysis tasks aiming at the discovery of knowledge in large amounts of data.[17] [18] Subsequent approaches like SEMMA and CRISP-DM emerged from the basic concept of KDD. The cross-industry standard process for DM (CRISP-DM) comprises the steps business understanding, data understanding, data preparation, modeling, evaluation and deployment, thus adding a more strategic perspective to the KDD core concept [19] [20]. The sample, explore, modify, model, and assess (SEMMA) methodology was developed by the SAS Institute to methodically organize the functions of its statistical and business intelligence software SAS Enterprise Miner, its constituent phases naming the concept in the form of an acronym. The analytics solutions unified method (ASUM) draws on a combination of agile and traditional implementation principles to achieve set solution goals and therefore complements the defined analysis phases by an additional project management stream to support the organizational realization. [21]

4 Design Principles

The concept of **sensemaking** originated in social psychology and was set in an organizational context by [5]. The approach describes how human beings in a social setting derive understanding of their surroundings by combining various information, creating connections and finally adding their own reasoning to it. The concept is described extensively in [22]. [6] sums up relevant literature and derives five key activities found in previous work as listed in table 1 which constitute the making of sense and thereby act as design goals for the developed reference model.

As the developed framework is supposed to not only support the understanding of facts and the creation of insight but also its utilization for the in-project and project-exceeding enhancement of the target-system, one more key activity is needed to complement the sensemaking key activities. By including the creation and utilization of a knowledge base we want to create a linkage to the field of knowledge management and thereby create the concept of **knowledge making**. By coining the term, we want to emphasize a creative, intuitive and iterative character of the approach, orienting on human behavior and the cognitive and social processes it originates in.

In DSR the concept of **design principles** (DP) provides the means to specify prescriptive design knowledge in a way that allows for a precise formulation to describe how the mechanisms of a technology or approach help to achieve particular aims.[23] According to [24] design principles should describe which actions are made possible through the use of an artifact and explain the material properties which make that action possible while naming the boundary conditions under which this description holds true. More precisely [24] suggests the formulation of a DP in the following form: "Provide the system with [material property—in terms of form and function] in order for users to [activity of user/group of users—in terms of action], given that [boundary conditions—user group's characteristics or implementation settings]." Following this suggestion, we formulated design principles for the presented reference model based on the derived knowledge making key activities as shown in table 1.

Table 1. Sensemaking [6] and knowledge making key activities with derived DPs

Sensemaking key activities	Knowledge making key activities	Design Principles: provide the reference model with features
(S1) Triggered by disruptive ambiguity	(K1) Open up new possibilities: Provoke action	...providing input to provoke action, in order for users to proactively advance their course of action in EPs, according to their given resources.
(S2) Acts of noticing and bracketing, create initial sense	(K2) Inform and classify: Provide a structure and give information impulses	...for structuring the user's course of reasoning and action, in order for users to gain a comprehensive understanding of relevant aspects to perform EPs, regarding the given domain and project context.
(S3) Requires labelling and categorizing, find common ground	(K3) Label and categorize: Assign methods and prior selections to structure	...to provide a catalogue of relevant action methods and tools, previously chosen alternatives or standard configurations and suitable search terms, in order for users to gain awareness of actionable alternatives.
(S4) Involves presumption to guide action, connect the abstract with the concrete	(K4) Assume and iterate: Intuitive selection of solution options, heuristic procedure	...allowing for heuristic solution approaches, in order for users to initialize solution finding in an intuitive way and vary solution configuration easily, considering established solution approaches and insight of previous data analysis projects.
(S5) Involves communication, draws on the resources of language	(K5) Provide communication basis: Provide means to discuss solution alternatives	...to visualize structure and solution alternatives, in order for users to apply the visual and textual representation of the reference model as means to debate and cooperatively decide on practicality, options and implications of solution configurations, considering the user's heterogeneous professional background
-	(K6) Capture and utilize: Build overarching body of knowledge	...the means to include a knowledge base, in order for users to connect, preserve, document and utilize information fragments and their relations, considering the user's preconditions to obtain a balance of a potent solution and manageable effort.

5 Reference Model

We want to motivate a highly strategic and integrated practice in data-driven enhancement projects [EP] in the manufacturing industry domain and to support this mindset by suggesting a framework to guide the efforts. The development of this reference model is driven by the needs identified in industrial practice and numerous research projects and realized by employing well-researched approaches grounded in established theory. We set up a grid-like structure to assign relevant methodologies to the respective analysis project phases and thereby fulfill the design principles formulated in Section 4. We based our approach on three widely established concepts: standard procedure models, the concept of data aggregation and the field of knowledge management. We attempt to provide the means for the effective combination and domain-specific adaption of these concepts while additionally overcoming their shortcomings as described in section 1 and further elaborated in [25] and [3].

We especially want to emphasize the importance of considering the various **aggregation levels** as described in table 2 in which information fragments can occur in, calling attention in particular to the intense interaction of all five levels of aggregation implying the necessity to expand awareness to each of them and their interrelations within each step of action. More specifically speaking an integrated consideration and

operationalization is needed throughout all project phases as the strong focus on DM core analysis activities was one of the main hurdles found in the pre-study described in Section 2. The reference model supports practitioners in the inclusion of all aspects, from aggregation level 1, being the least connected state of raw data and the physical system realization and data acquisition up to level 5, comprising the overarching management of highly connected complex information constructs.

Table 2. Aggregation levels

Aggregation level	Description
AL 1: Analogous level	Tangible components of the real-life system
AL 2: Representation level	Data objects representing and accompanying the real-life system
AL 3: Transfer level	Measures derived from representing and accompanying data
AL 4: Implementation level	Implementations derived from feature sizes of the transfer level
AL 5: Information level	Highly connected information comprising facts and interrelations, decision support

Data aggregation is often depicted in a form similar to the *traditional knowledge pyramid*, although revised and refined approaches can be found superseding this strictly hierarchical view. [26] Within the scope of our research we adopt the view that information fragments can exist in various states of aggregation, starting from incrementally small pieces of data like a single binary number, but also forming states of light aggregation as in protocols or logfiles or of higher aggregation like in the form of data sets, tables, charts or reports, where data is set into context and provides declarations exceeding its alpha-numerical value. We therefore deem it valid to speak of information when referring to aggregated data. Data aggregation states then stretch to strongly aggregated forms of where aggregated chunks of information further connect to complex constructs representing relations comprising formal logic thus resembling the processing of insight and thought in the human mind. We therefore argue that the term information is suitable to describe aggregated forms of data and highly aggregated information equals knowledge in the daily use of language. In table 2 we convey this understanding to the manufacturing industry domain introducing an additional level of *analogous real-life objects* which the relevant data relates to and originates in.

Relevant objects within AL 1 can be controllers, motors, GPS trackers and sensors or transport systems, accompanied by the respective digital counterparts in AL 2 like output data of controllers, performance data of motors, GPS data and other sensor data. Furthermore AL 2 addresses additional descriptions of the target-system as e.g. conceptual models. Within AL 3 a suitable concept must be chosen to gather, process and contain any relevant information fragments to transfer them to higher levels of aggregation and derive and utilize insight. A suitable concept can be an enterprise-specific analysis framework, an individual adoption of the DM standard processes described in section 3 or domain-specific adoptions like the "DMME: Data mining methodology for engineering applications" as presented in [3]. Within AL 3 and the central analysis project phase of the chosen concept resides the core activity constituting the success of the EP: Proceeding in an intensely iterative character and closely observing the relation to any other grid point highly context-sensitive feature engineering is made possible. Within AL 4 the found facts and interrelations are implemented by integrating the derived insight within physical instantiations, instantiations of digital shadows or digital twins, simulation models or visualizations.

The knowledge base constituting AL 5 can take many forms, from the incorporation by an individual, classical SQL databases or ontologies to intelligent agents. Lastly the successful utilization of the concept will depend on what the respective knowledge base affords. Despite AL 5 constituting the bottleneck of the implementation, the more suitable its chosen way of instantiation is for the occasion the more intense the usage in practice will be. Highly formalized approaches and machine-readable implementations allow for complex and potent operations but require high effort to set up and maintain. Depending on the application

situation the manageable effort of a lightweight solution can advance implementation success. We suggest orienting on existing solutions like for example extensively elaborated for the application of ontologies in the manufacturing domain in [27].

Two more aspects are vital to exploit the full potential of data analytics in the industrial domain: to take into account the **dimorphic system character** of the target system consisting of analogous and digital components and to focus on **context-sensitive engineering of conclusive features** as this step constitutes the heart of the project and is complemented by the choice and application of fitting tools and methods, only rendered possible by the utilization of aforementioned concepts providing the necessary context. [29]

As pointed out by [29] and further elaborated by [30], the concepts described in Section 3 share the common essence of a stepwise description of the data mining project phases along with similar core principles of the activities performed during the respective steps. Attempting to capture the essence of the various data mining procedure models we derived a generalized version of data mining project phases as can be seen in figure 2. Based on the specification of the analysis project goal in phase 1 (P1) a conceptualization phase follows in phase 2 (P2). The data analysis core activities are performed in phase 3 (P3) and 4 (P4). First data is collected by setting up the necessary physical infrastructure and accumulating all accessible and presumably relevant information fragments, growing and extending the data pool. Then feature engineering, model building and extraction of relations follow, reducing the data build-up to a set of connected information which can then be deployed. Phase 5 (P5) draws on the preceding phases and can and should be conducted in parallel from the start as it preserves and makes available the methodological and meta-information of the data analysis project as well as comprises the supervision of its execution during and after the project.

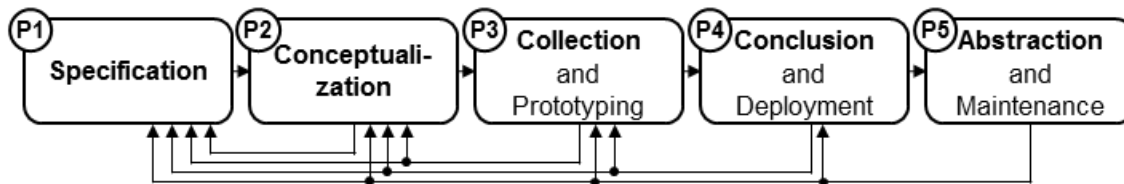


Figure 2. Generalized procedure model

The phases described above provide the reference model with a basic sequence of actions to perform in a data analysis project and can be replaced by any adequate alternative during instantiation, e.g. a standard process or an enterprise specific procedure. Concurrently the necessity to consider various aggregation levels of available and derived information fragments pertains for all project steps. The aggregation level view in combination with the project phases forms a grid as presented in figure 3 to address the methodological repertory of each combination of layer and phase allowing for the mapping of relevant methods accompanied by respective meta-information.

First application experience in the support of research projects showed the value of the reference model to promote a more integrated method of operation, but also made obvious how providing the means for intuitive applicability is crucial for the successful implementation of the approach. [4] [36]

Future work will be devoted to the demonstration, evaluation and revision of the concept in practice. Additionally, a thorough analysis of existing and common methodological elements will be conducted by the analysis of research publications within leading journals and by assessment of accessible information on their application in practice to develop an appropriate classification and identify any additional elements that should be included. Moreover, having provided the means to document the usage of methods and their specification as well as having examined their classification allows for the construction of a formalized body of knowledge addressing the creation of knowledge itself. Future work will comprise the development of a taxonomy of methodological principles at hand to then be conveyed to an ontology defining logical relations, rules and principles allowing for decision support by typecasting similar EPs and deriving suitable solution approaches. While this paper focused on the motivation and the theoretical grounding of the concept, some consideration should also be given to its compliance with existing standards and tools to accelerate interoperability. The integration with standardized approaches like the *Reference Architecture Model Industrie 4.0* (RAMI4.0) or with data management aspects like the data lifecycle approach can create synergies and add a helpful dimension to support the organizational implementation of the suggested method within enterprises. [37]

Acknowledgements

Parts of this research were funded by the German Federal Ministry of Education and Research, grant number 03ZZ0212G "Zwanzig20 Agent-3D Verbundvorhaben: QUALIPRO".

Competing interests

The authors declare no competing interests.

References

1. Schäffer, T., Leyh, C.: Master Data Quality in the Era of Digitization - Toward Inter-organizational Master Data Quality in Value Networks: A Problem Identification. In: Piazzolo, F., Geist, V., Brehm, L., and Schmidt, R. (eds.) *Innovations in Enterprise Information Systems Management and Engineering*. pp. 99–113. Springer International Publishing, Cham (2017).
2. Zschech, P., Heinrich, K., Pfitzner, M., Hilbert, A.: Are you up for the challenge? Towards the development of a big data capability assessment model. *Proc. 25th Eur. Conf. Inf. Syst. ECIS*. 2613–2624 (2017).
3. Huber, S., Wiemer, H., Schneider, D., Ihlenfeldt, S.: DMME: Data Mining Methodology for Engineering Applications – A Holistic Extension to the CRISP-DM Model. In: *Procedia CIRP* (2018). pp. 403–408., Gulf of Naples, Italy (2018).
4. D. Gliem, C. Laroque., U. Jessen, J. Stolipin, S. Wenzel, W. Kusturica: *SimCast – Simulationsgestützte Prognose der Dauer von Logistikprozessen*. Abschlussbericht. Universität Kassel, Fachgebiet Produktionsorganisation und Fabrikplanung; Westsächsische Hochschule Zwickau, Fachgebiet Wirtschaftsinformatik (2019).
5. Weick, K.E.: *The social psychology of organizing*. McGraw-Hill, New York (2006).

6. Seidel, S., Chandra Kruse, L., Székely, N., Gau, M., Stieger, D.: Design principles for sensemaking support systems in environmental sustainability transformations. *Eur. J. Inf. Syst.* 27, 221–247 (2018). <https://doi.org/10.1057/s41303-017-0039-0>.
7. Gregor, S., Jones, D.: The anatomy of a design theory. *J. Assoc. Inf. Syst.* 8, (2007).
8. Zschech, P., Fleißner, V., Baumgärtel, N., Hilbert, A.: Data Science Skills and Enabling Enterprise Systems: Eine Erhebung von Kompetenzanforderungen und Weiterbildungsangeboten. *HMD Prax. Wirtsch.* 55, 163–181 (2018). <https://doi.org/10.1365/s40702-017-0376-4>.
9. Debortoli, S., Müller, O., Brocke, J. vom: Vergleich von Kompetenzanforderungen an Business-Intelligence- und Big-Data-Spezialisten: Eine Text-Mining-Studie auf Basis von Stellenausschreibungen. *Wirtschaftsinformatik.* 56, 315–328 (2014). <https://doi.org/10.1007/s11576-014-0432-4>.
10. Hevner, A.R., March, S.T., Ram, S., Park, J.: Design Science in Information Systems Research. *MIS Q.* 28, 75–105 (2004).
11. Hevner, A.R.: A three cycle view of design science research. *Scand. J. Inf. Syst.* 19, 87–92 (2007).
12. Peffers, K., Tuunanen, T., Rothenberger, M.A., Chatterjee, S.: A Design Science Research Methodology for Information Systems Research. *J. Manag. Inf. Syst.* 24, 45–77 (2007). <https://doi.org/10.2753/MIS0742-1222240302>.
13. Stachowiak, H.: Allgemeine Modelltheorie. Springer, Wien [u.a.] (1973).
14. Vom Brocke, J., Grob, H.L.: Referenzmodellierung: Gestaltung und Verteilung von Konstruktionsprozessen. Logos Verlag, Berlin (2015).
15. Schlieter, H., Esswein, W.: Reference Modelling in Health Care - State of the Art and Proposal for the Construction of a Reference Model. *Enterp. Model. Inf. Syst. Archit.* 6, 36–49 (2011). <https://doi.org/10.18417/emisa.6.3.3>.
16. Kurgan, L.A., Musilek, P.: A survey of Knowledge Discovery and Data Mining process models. *Knowl. Eng. Rev.* 21, 1 (2006). <https://doi.org/10.1017/S0269888906000737>.
17. Brachman, R.J., Anand, T.: The Process of Knowledge Discovery in A First Sketch. *AAAI Tech. Rep. WS-94-03* (1994).
18. Fayyad, U., Piatetsky-Shapiro, G., Smyth, P.: From data mining to knowledge discovery in databases. *AI Mag.* 17, 37–37 (1996).
19. Xu, L.D., He, W., Li, S.: Internet of Things in Industries: A Survey. *IEEE Trans. Ind. Inform.* 10, 2233–2243 (2014). <https://doi.org/10.1109/TII.2014.2300753>.
20. Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C.R., Wirth, R.: CRISP-DM 1.0: Step-by-step data mining guide, (2000).
21. IBM Corporation 2016: Analytics Solutions Unified Method. Implementations with Agile principles, <ftp://ftp.software.ibm.com/software/data/sw-library/services/ASUM.pdf>, (2016).
22. Weick, K.E., Sutcliffe, K.M., Obstfeld, D.: Organizing and the Process of Sensemaking. *Organ. Sci.* 16, 409–421 (2005). <https://doi.org/10.1287/orsc.1050.0133>.
23. Gregor, S., Kruse, L.C., Seidel, S.: The Anatomy of a Design Principle. *J. Assoc. Inf. Syst.* (2020).
24. Chandra, L., Seidel, S., Gregor, S.: Prescriptive Knowledge in IS Research: Conceptualizing Design Principles in Terms of Materiality, Action, and Boundary Conditions. In: 2015 48th Hawaii International Conference on System Sciences. pp. 4039–4048. IEEE, HI, USA (2015). <https://doi.org/10.1109/HICSS.2015.485>.

25. Michalczyk, S., Scheu, S.: Designing an analytical Information Systems Engineering method. In: ECIS (2020).
26. Jennex, M.E.: Re-Visiting the Knowledge Pyramid. In: 2009 42nd Hawaii International Conference on System Sciences. pp. 1–7. IEEE, Waikoloa, Hawaii, USA (2009). <https://doi.org/10.1109/HICSS.2009.361>.
27. Sanfilippo, E.M., Belkadi, F., Bernard, A.: Ontology-based knowledge representation for additive manufacturing. *Comput. Ind.* 109, 182–194 (2019). <https://doi.org/10.1016/j.compind.2019.03.006>.
28. Gulcehre, C., Bengio, Y.: Knowledge Matters: Importance of Prior Information for Optimization. *J Mach Learn Res.* 17, (2013).
29. Mariscal, G., Marbán, Ó., Fernández, C.: A survey of data mining and knowledge discovery process models and methodologies. *Knowl. Eng. Rev.* 25, 137–166 (2010). <https://doi.org/10.1017/S0269888910000032>.
30. Ferstl, O.K., Sinz, E.J.: *Grundlagen der Wirtschaftsinformatik*. Oldenbourg, München (2013).
31. Kritzinger, W., Karner, M., Traar, G., Henjes, J., Sihn, W.: Digital Twin in manufacturing: A categorical literature review and classification. *IFAC-Pap.* 51, 1016–1022 (2018). <https://doi.org/https://doi.org/10.1016/j.ifacol.2018.08.474>.
32. Hall, D., Paradise, D., Courtney, J.F.: Building a theoretical foundation for a learning-oriented knowledge management system. *J. Inf. Technol. Theory Appl. JITTA.* 5, 7 (2003).
33. QualiPro. <https://agent3d.de/qualipro>. Accessed 10.05.2021.
34. DIN SPEC 91345:2016-04. Deutsches Institut für Normung, DIN SPEC 91345:2016-04. Beuth-Verlag, 2016, <https://www.beuth.de/de/technische-regel/din-spec-91345/250940128>; Accessed 10.05.2021.
35. Simon, H.A.: *The sciences of the artificial*. MIT Press, Cambridge, Mass. (2008).
36. Klaus, G., Buhr, M.: *Philosophisches Wörterbuch*. VEB Enzyklopädie, Leipzig (1971).
37. Angée, S., Lozano-Argel, S.I., Montoya-Munera, E.N., Ospina-Arango, J.-D., Tabares-Betancur, M.S.: Towards an Improved ASUM-DM Process Methodology for Cross-Disciplinary Multi-organization Big Data & Analytics Projects. In: Uden, L., Hadzima, B., and Ting, I.-H. (eds.) *Knowledge Management in Organizations*. pp. 613–624. Springer International Publishing, Cham (2018), doi: https://doi.org/10.1007/978-3-319-95204-8_51