OTTO VON GUERICKE UNIVERSITÄT MAGDEBURG

www.mrcc.eu
MAGDEBURG RESEARCH AND COMPETENCE CLUSTER VLBA

AG WI | SAP UCC | VLBA Lab

A Methodology for Constructing Patterns for the Management of Data Science Projects

Christian Haertel, Sarah Schramm, Matthias Pohl, Sascha Bosse, Daniel Staegemann, Christian Daase, and Klaus Turowski
Institute of Technical and Business Information Systems, Otto-von-Guericke University Magdeburg, Magdeburg, Germany
{christian.haertel, sarah.schramm, matthias.pohl, sascha.bosse, daniel.staegemann, christian.daase, klaus.turowski}@ovgu.de

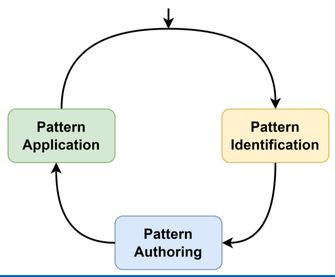
Introduction

As the amount of generated data is steadily increasing, businesses of various domains aim to derive potential advantages and enhance their competitive positioning (de Medeiros et al., 2020). Data Science (DS) aims to extract knowledge and insights from data using various methods and techniques (Chang and Grady, 2019) and thus, has gained increasing significance (Cao, 2017). Organizations often encounter challenges in implementing these projects (Martinez et al., 2021) and there is no widely accepted and applied approach. This is reflected in the low DS project success rate (VentureBeat, 2019), demanding improvements for DS project management (Saltz and Krasteva, 2022). As the literature identifies common problems in the execution of DS projects (Martinez et al., 2021a), the adaptation of the *pattern* concept to DS appears promising.

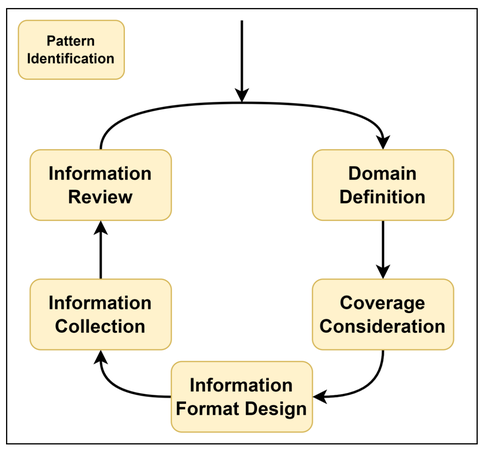
Term	Definition
Data Science	*„Data science is the methodology for the synthesis of useful knowledge directly from data through a process of discovery or of hypothesis formulation and hypothesis testing." (Chang and Grady 2019)*
Patterns	*Patterns capture solutions to recurring problems in a domain in a simple and straightforward form (Fehling et al., 2014). An overview and structured presentation within patterns could ensure alleviated and methodology-independent access to common problems and solutions and, thus, contribute to the improvement of DS project management activities.*

RQ: How can a methodology for the construction of patterns for DS project management be designed and applied?

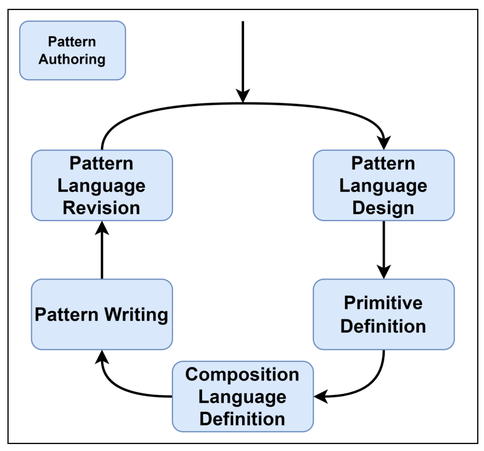
Pattern Construction Process for Data Science Project Management



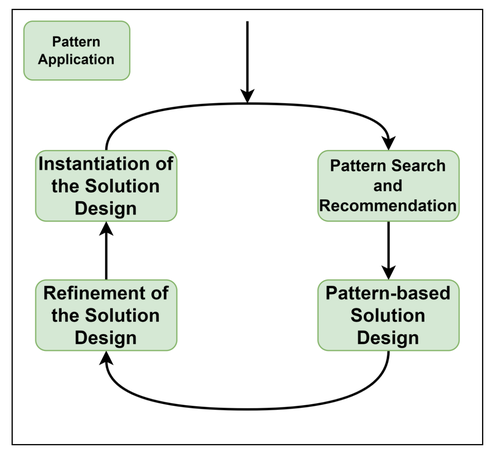
In the literature, there are different approaches to the development of patterns. The process according to (Fehling et al., 2014) was selected and adapted to the field of DS project management since its applicability to various domains has already been demonstrated. The method consists of three phases: pattern identification, pattern authoring, and pattern application, as illustrated. Each stage involves several iteratively traversed activities to continuously improve and adapt the developed results (Fehling et al., 2014). In particular, the first two phases are repeated multiple times to discover and form patterns. Finally, the third phase involves refining the patterns for specific use cases or application environments.



Domain Definition: creating joint understanding of key terminology and concepts of DS
Coverage Consideration: assessing and narrowing down the scope of DS project management
Information Format Design: establishing a unified structure for information capture and processing (e.g., tools and templates)
Information Collection: use of literature review to acquire relevant material in a structured way
Information Review: result set is further narrowed down in different filtering stages until a feasible set related to the specific problems is identified



Pattern Language Design: determine the pattern format (e.g., name, problem, context, challenges, solution, results, and links)
Primitive Definition: further elaboration of definitions from Information Format Design
Composition Language Definition: determine guidelines and formal specifications of the pattern language
Pattern Writing: documentation of the patterns based on the previously defined structure
Pattern Language Revision: evaluation and revision of pattern language (e.g., through DS practitioners)



Pattern Search and Recommendation: accompany DS patterns with a summary of the problem and solution of each of the patterns to facilitate navigation and identify suitable patterns
Pattern-based Solution Design: expand patterns by the section "Examples" to provide a reference solution of the DS problem within the given pattern.
Refinement of the Solution Design: patterns are constrained and adapted to a specific environment where they should be applied
Instantiation of the Solution Design: determine the means to manage, configure, and deploy the patterns

Name	Alignment of Expectations
Problem	The explorative nature of DS projects increases the difficulty of establishing goals and timelines that confirm with expectations of management and domain users.
Context	DS project expectations are frequently not realized. Oftentimes, possibilities and results strongly depend on available resources, data access, and quality.
Challenges	The availability of resources impact the project outcome. A significant challenge in DS undertakings is data access. Additionally, the data exploration might reveal the unsuitability of the available data for achieving the business objectives. Hence, because of the inherent risks and uncertainties in DS projects, flexibility regarding the modification of the expectations might be necessary. This also applies to other resources like computing infrastructure or personnel.
Solution	The project team, management, domain users, and other stakeholders perform an alignment regarding the potential and limitations of the envisioned DS application. A situation assessment evaluates the feasibility of the set objectives and their added value for the organization. Based on detected similar problems and the corresponding solutions, the relevant resources (e.g., data, budget, competencies) and their availability are discussed.
Result	Development of a joint understanding of appropriate expectations and the approximately required resources and timelines. Based on the situation assessment, confidence is established regarding the feasibility of the DS project and its added value for the organization.
Links	Project expectations result from the Strategic Alignment of the Project and Involvement of Senior Management. Objectives have to be aligned with requirements to the project execution, including the IT Infrastructure and Team Composition to determine a realistic Scope. Moreover, expectations are defined regarding Project Team Competencies to complete the project tasks. During project execution, based on Project Performance Monitoring new or revised requirements and goals can arise.
Example	This pattern can be assigned to Business Understanding, which is a common phase in various DS process models. Here, the project circumstances are communicated with involved stakeholder groups to elaborate opportunities, requirements, and functionalities of the DS application (Schulz et al., 2020). A feasibility study can be used to evaluate the likelihood of fulfilling project requirements and objectives (Schulz et al., 2020).

Notation of the pattern "Alignment of Expectations"

References

Cao, L. (2017). Data Science: Challenges and Directions. Communications of the ACM, 60(8):59–68.
Chang, W. and Grady, N. (2019). NIST Big Data Interoperability Framework: Volume 1, Definitions.
de Medeiros, M. M., Hoppen, N., and Macada, A. C. G. (2020). Data science for business: benefits, challenges and opportunities. The Bottom Line.
Fehling, C., Barzen, J., Breitenbücher, U., and Leymann, F. (2014). A process for pattern identification, authoring, and application. In Eloranta, V.-P. and van Heesch, U., editors, Proceedings of the 19th European Conference on Pattern Languages of Programs, pages 1–9, New York, NY, USA. ACM.
Martinez, I., Viles, E., and Olaizola, I. G. (2021). Data Science Methodologies: Current Challenges and Future Approaches. Big Data Research 24.
Saltz, J. S. and Krasteva, I. (2022). Current approaches for executing big data science projects - a systematic literature review. PeerJ Computer Science, 8(e862).
Schulz, M., Neuhaus, U., Kaufmann, J., Badura, D., Kuehnel, S., Badewitz, W., Dann, D., Kloker, S., Alekozai, E. M., and Lanquillon, C. (2020). Introducing DASC-PM: A Data Science Process Model. ACIS 2020.
VentureBeat (2019). Why do 87% of data science projects never make it into production?

Contact: Christian Haertel, Magdeburg Research and Competence Cluster, christian.haertel@ovgu.de, www.mrcc.eu