The Human Side of Service Engineering, Vol. 62, 2022, 250–255
https://doi.org/10.54941/ahfe1002564

Roles and Competences of Data Science Projects

Claudia Dukino, Damian Kutzias, and Maike Link

Fraunhofer Institute for Industrial Engineering IAO Nobelstrasse 12, 70569 Stuttgart, Germany

ABSTRACT

In recent years, it has become increasingly apparent that data is playing an ever more important role in companies and the need for IT (Information Technology) applications to support people in their activities is growing. The requirements for data-driven projects for the automation and augmentation of processes and tasks are significantly higher than for standard IT projects. An essential requirement is to learn from company data and to use it for new applications. The composition of the project team plays an essential role. It is necessary to recognize which roles and competences are required for the implementation of the project and to recognize how these may change during the project. In the following, due to the interdisciplinary nature of data science projects, new competences and roles for project execution will be identified and discussed. Possible risks that can arise from unstructured project planning and from role and competence planning will be identified. The differences compared to standard projects are highlighted and the challenges compared to them are examined. To support project planning, the use of tools can be helpful. The requirements a tool or method should fulfil in order to add value for a broad spectrum of enterprises are addressed. Exemplary criteria in this context are neutrality, branch independence and free availability of the method. In addition, the usability and areas of application of such tools are discussed.

Keywords: Data science, Data analytics, Project management, Methodology, Project team, Industrial science, Business analytics, Competences

INTRODUCTION

Data science along with its sub-disciplines was often presented as an interdisciplinary field both for practitioners and researchers alike (Asamoah et al., 2020; Saltz & Shamshurin, 2019; Schulz et al., 2020; Zhang & Hu, 2017). Whereas classical projects may have a high predictability for resource requirements and reachability of the defined goals, data-based projects often lack such certainty. The interdisciplinarity and the accompanying complexity of data science projects results in the same for the competence structure: interdisciplinary teams are often required for successful project implementation (Egger & Yu, 2022). The skills that are important and necessary to be able to realise data projects effectively, were examined by Hattingh et al. The authors identify organizational, social, analytical, technical, ethical, regulatory and cognitive capabilities (Hattingh et al., 2019). In data science projects, different project team members fulfil various project- and competence-specific roles which are not easily replaceable. The difficulty here is to identify the necessary roles before and during the project to have them available in the company and to include them in the project process at the right time. In the following, relevant roles of data science projects will be explained alongside methods and tools which are available to identify them at an early stage and to involve them accordingly during the course of the project.

ROLES & COMPETENCES IN DATA SCIENCE PROJECTS

In a previous paper, seven data science process models were presented and examined for their approach (Kutzias et al., 2021). The point "Project team and Competences" was identified as an important component in four of the seven procedure models. While the CRISP-DM (Cross-Industry Standard Process for Data Mining) only states that personnel sources are to be considered (Chapman et al., 2000), some newer models provide detailed descriptions of roles and the assigned competences. For example, the EDDA process model (Engineering Data-Driven Applications) describes the roles of subject-matter expert, software developer, data domain expert, and data scientist (Hesenius et al., 2019) separately from the process. The TDSP model (Team Data Science Process) describes six project roles and assigns them tasks and artifacts: Solution Architect, Project Manager, Data Engineer, Data Scientist, Application Developer, and Project Manager (Microsoft, 2020). The topic is considered most extensively in the DASC-PM (Data Science Process Model). In this process model, detailed descriptions of competence profiles as well as roles and their requirements for the project are specified and a competence and role diagram is provided for each project phase (Schulz et al., 2020).

The need for interdisciplinarity and the creation of new roles and competences described in the models was also perceived in the work of the "SmartAIwork" project (funded by the German Federal Ministry of Education and Research) as an important and changing component in the introduction of artificial intelligence, especially in the field of data science (Tombeil et al., 2021).

The challenges of data-driven projects for employees and companies are not only based on the technological requirements. New competences are necessary for successful project realisation requiring a variety of different roles to work in interdisciplinary teams. Employees have to adapt to such team structures. The major differences regard to the communication and workflow: For example, engineers have to dive into the application domain and also think about operational processes. Furthermore cross-cutting competences such as classifying personal performance in the overall system, understanding adjacent disciplines and being able to explain oneself and continuous learning, are required (Ganz et al., 2021). This means that it is precisely the combination of action-, communication- and knowledge-oriented roles and competences that makes a significant contribution to the success of projects (Belbin, 2010). Depending on the relevance of data-based solutions as well as the size and strategy of the company, a change in the competence profiles and their qualification may become necessary, and with it a further development of the corporate strategy.

In addition to the new or expanded competences for working in interdisciplinary teams and for a changed communication and way of working, new roles and expertise are needed for the implementation of a data project.

But which roles in particular make data projects successful? In recent years, various researchers have been discussing the relevant rules required to be able to successfully carry out data projects. To this end, a large-scale study was conducted by Accenture in 2017 with more than 1,000 large companies, which identified three new essential role categories in data science projects. These are trainers - who teach the systems how to work, explainers - who are tasked with bridging the gap between technologists and senior management, and sustainability experts - who are tasked with ensuring that the systems work over the long term and that unintended changes are fixed as quickly as possible. For each of these three categories, three roles were considered in more depth, which can be found in the study (Wilson et al., 2017). Several other researchers addressed the roles and competence profiles of data science projects: (Crisan et al., 2021) provide an extensive overview of different roles mapped to skill requirements by examining 12 studies on data scientists for commonality in role definitions. They were able to identify nine data science roles and assign them to processes (preparation, development & engineering, analysis, communication). In addition, their expertise in the areas of statistics, computer science, domain knowledge, and human-centered design was assessed.

The roles defined in the studies can help to find the roles for own projects and to evaluate them using the methods and tools in the next section and to assign them with tasks and responsibilities.

TOOLS AND METHODS

Depending on the size of the company, it can be meaningful to address competence management across the company instead of locally for a project. Since the nature of project and enterprise management differs, also the tools to address corresponding challenges are different. In this section, three methods and tools are presented as examples that can support competence management and the compilation of necessary roles and tasks for upcoming data science projects both strategically and operationally in the company. The aim is to enable the organization as well as the employees to cope with the upcoming project and work tasks in the best possible way (Dworschak et al., 2020). The following list gives an overview of the discussed tools and briefly presents their application level before diving into more details in the subsections afterwards.

- The “Fraunhofer Competence Compass” deals with the meta-level and thus addresses a rather high strategic altitude to show how the company can be set up to support innovative projects as data science projects.
- The “role and stakeholder analysis” on the operational level examines which actors are required in the project and which tasks they have to fulfil.
- The “RACI matrix” helps to define the responsibilities in the project.

Fraunhofer Competence Compass

(Dworschak et al., 2020) argue that competence management 4.0 is understood to mean competence management that takes account of the requirements of digitization and developments toward Industry 4.0. An appropriately aligned concept enables digitization activities and innovation projects to be supported in this context. The main focus of this is coping with the changes in work tasks, activities and processes brought about by new, previously unknown requirements. One instrument for the adequate implementation of holistic competence management 4.0 is the Fraunhofer Competence Compass (based on DIN PAS 1093 (Stracke, 2009)). This approach comprises six steps of competence management: strategy connection, competence strategy, competence model, competence measurement, competence development and competence evaluation (Dworschak et al., 2020).

In order to effectively and actively support change processes and digitization activities, the distinction between strategic and operational competence management must be overcome. Within the framework of the Competence Compass, this is achieved by deriving strategies at the team, process, and task levels. By identifying existing and required task-related competences, requirements and the resulting necessary role- and competence-related adjustments become visible (Dworschak et al., 2020).

Role and Stakeholder Analysis

There are various overviews of roles and stakeholders in the literature and on the Internet. With new competences and roles for data science projects, it is important to know which roles to consider for the upcoming project. The overviews of roles and competencies by Wilson (Wilson et al., 2017) and Crisan (Crisan et al., 2021) can be helpful here. In order to visualize these, the following method has proven to be practicable for us in business workshops: metaplan with cards. In our view, the following aspects should be elaborated for each role/stakeholder when creating the overview:

- Designation of the role/stakeholder; e.g. business data expert.
- Definition of the role/stakeholder; e.g. business expert from the application domain with a focus on data
- Tasks of the role/stakeholder; e.g., generation and interpretation of data
- Responsibility of the role/stakeholder: internal/external

After placement, it becomes apparent which roles can be filled or not yet filled. Accordingly, missing roles and stakeholders may still have to be qualified, hired or assigned externally.

Responsibility Assignment Matrix (RACI)

The RACI matrix is a widely used method in project management (Howard, 2012; O'Neill, 2010) for the distribution of responsibilities and their overview in the project. RACI stands as an acronym and is defined as follows:

R - Responsible - Actors who are responsible for the execution of the project, e.g., data scientist, business unit employee, IT architect, etc.

A - Accountable - Persons who are responsible for the project and its substeps and make decisions regarding them, e.g., project manager, IT manager, etc.

C - Consultant - Can be internal and external experts who contribute information to the decision within the project.

I - Informed - Form the group of people who are informed about the results but cannot influence the decision, e.g., IT users.

Usually, the tasks or activities are plotted on the Y-axis and the involved actors on the X-axis. In the connection of the two axes a letter assignment or a color coding corresponding to the letters takes place. The advantages of this matrix are that it strengthens the understanding among each other as to who bears what responsibility, thus also simplifying communication. It can help to define interfaces in the area of business process modeling. And it is very easy to apply, since paper and pencil can be sufficient and numerous free templates are available on the Internet.

CONCLUSION

This paper is focussed on roles and competences in and around data science projects. At first, different levels at which such roles and competences are relevant and have differences to other types of projects are discussed alongside with corresponding challenges and tasks. Two different overviews from existing research are presented, one focussed on new tasks coming with a categorisation of those new roles and one with a focus on competence requirements for presented roles coming with a mapping of relevant processes for the presented roles. In order to be able to utilise the known new role and competence requirements for data science projects, methods are presented which can provide a structured approach for planning. These tools deal with different levels ranging from project planning to strategic competence management on enterprise level.

REFERENCES

Asamoah, D. A., Doran, D., & Schiller, S. (2020). Interdisciplinarity in Data Science Pedagogy: A Foundational Design. *Journal of Computer Information Systems*, *60*(4), 370–377. https://doi.org/10.1080/08874417.2018.1496803

Belbin, R. M. (2010). *Management Teams: Why they succeed or fail* (3rd ed.). Routledge. https://doi.org/10.4324/9780080963594

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

Crisan, A., Fiore-Gartland, B., & Tory, M. (2021). Passing the Data Baton: A Retrospective Analysis on Data Science Work and Workers. In N. Elmqvist, B. Fisher, P. Londstorm, R. Maciejewski, M. Meyer, S. Miksch, L. G. Nonanto, N. Riche, H.-W. Shen, R. Westermann, J. Wood, & J. Yang (Eds.), *IEEE Transactions on Visualization and Computer Graphics* (pp. 1860–1870). IEEE. https://doi.org/10.1109/VDS51726.2020.00005

Dworschak, B., Karapidis, A., Zaiser, H., & Weisbecker, A. (2020). Kompetenzmanagement 4.0 - Kompetenz und Kompetenzentwicklung in einer digitalisierten Arbeitswelt. In R. Knackstedt, K. Kutzner, M. Sitter, & I. Truschkat (Eds.), *Kompetenzmanagement in Organisationen. Grenzüberschreitungen im Kompetenzmanagement: Trends und Entwicklungsperspektiven* (pp. 145–157). Springer-Verlag.

Egger, R., & Yu, J. (2022). Data Science and Interdisciplinarity. In R. Egger (Ed.), *Tourism on the Verge. Applied Data Science in Tourism* (pp. 35–49). Springer International Publishing. https://doi.org/10.1007/978-3-030-88389-8_3

Ganz, W., Friedrich, M., Hornung, T., Schneider, B., & Tombeil, A.-S. (2021). *Arbeit mit Künstlicher Intelligenz: Fallbeispiele aus Produktion, Sacharbeit und Dienstleistungen*. Stuttgart. Fraunhofer-Institut für Arbeitswirtschaft und Organisation (IAO). http://publica.fraunhofer.de/dokumente/N-636254.html

Hattingh, M., Marshall, L., Holmner, M., & Naidoo, R. (2019). Data Science Competency in Organisations: A Systematic Review and Unified Model. In C. de Villiers & H. Smuts (Eds.), *Proceedings of the South African Institute of Computer Scientists and Information Technologists conference* (pp. 1–8). ACM Press. https://doi.org/10.1145/3351108.3351110

Hesenius, M., Schwenzfeier, N., Meyer, O., Koop, W., & Gruhn, V. (2019). Towards a Software Engineering Process for Developing Data-Driven Applications. In *2019 IEEE/ACM 7th International Workshop on Realizing Artificial Intelligence Synergies in Software Engineering (RAISE)* (pp. 35–41). IEEE. https://doi.org/10.1109/RAISE.2019.00014

Howard, D. (2012). *It Release Management: A Hands-on Guide*. CRC Press.

Kutzias, D., Dukino, C., & Kett, H. (2021). Towards a Continuous Process Model for Data Science Projects. In C. Leitner, W. Ganz, D. Satterfield, & C. Bassano (Eds.), *Lecture Notes in Networks and Systems. Advances in the Human Side of Service Engineering* (Vol. 266, pp. 204–210). Springer International Publishing. https://doi.org/10.1007/978-3-030-80840-2_23

Microsoft. (2020). *Team Data Science Process Documentation*. https://docs.microsoft.com/en-us/azure/machine-learning/team-data-science-process/lifecycle

O'Neill, M. G. (2010). *Green IT for Sustainable Business Practice: An ISEB Foundation Guide*. British Informatics Society Limited.

Saltz, J. S., & Shamshurin, I. (2019). Achieving Agile Big Data Science: The Evolution of a Team's Agile Process Methodology. In C. Baru (Chair), *2019 IEEE International Conference on Big Data*. Symposium conducted at the meeting of Institute of Electrical and Electronics Engineers; IEEE Computer Society.

Schulz, M., Neuhaus, U., Kaufmann, J., Badura, D., Kerzel, U., Welter, F., Prothmann, M., Kühnel, S., Passlick, J., Rissler, R., Badewitz, W., Dann, D., Gröschel, A., Kloker, S., Alekozai, E. M., Felderer, M., Lanquillon, C., Brauner, D., Gölzer, P., . . . Gehrke, N. (2020). *DASC-PM v1.0 - Ein Vorgehensmodell für Data-Science-Projekte*. Hamburg, Elmshorn. https://doi.org/10.25673/32872

Stracke, C. M. (2009). *Personalentwicklung unter besonderer Berücksichtigung von Aus- und Weiterbildung: Kompetenzmodellierung in der Personalentwicklung*. Beuth Verlag GmbH.

Tombeil, A.-S., Dukino, C., Zaiser, H., & Ganz, W. (2021). *KI-Ambition als Treiber für die Realisierung von Digitalisierung: Wann ist weniger mehr? Automatisierung und Unterstützung in der Sachbearbeitung mit Künstlicher Intelligenz: Vol. 8*. Fraunhofer Verlag.

Wilson, H. J., Daugherty, P. R., & Morini-Bianzino, N. (2017). The Jobs That Artificial Intelligence Will Create: A global study finds several new categories of human jobs emerging, requiring skills and training that will take many companies by suprise. *MITSloan Management Review*, *58*(4), 13–16. https://sloanreview.mit.edu/article/will-ai-create-as-many-jobs-as-it-eliminates/

Zhang, Y., & Hu, J. (2017). Discovering the interdisciplinary nature of big data research. In J.-Y. Nie, Z. Obradovic, T. Suzumura, R. Ghosh, R. Nambiar, C. Wand, . . . M. Toyoda (Chairs), *2017 IEEE International Conference on Big Data (Big Data),* Boston, MA.