International Journal of Production Management and Engineering

https://doi.org/10.4995/ijpme.2024.19123

Received: 2023-01-16 Accepted: 2024-01-19

A requirement-driven approach for competency-based collaboration in industrial data science projects

Marius Syberg[a1*], Nikolai West[a2], Jörn Schwenken[a3], Rebekka Adams[c], Jochen Deuse[a,b]

[a] Technical University Dortmund, Institute of Production Systems, Leonhard-Euler-Str. 5, Dortmund 44227, Germany.

[b] Centre for Advanced Manufacturing, University of Technology Sydney, 11 Broadway, Ultimo NSW 2007, Australia.

[c] NEOCOSMO GmbH, Science Park 2, 66123 Saarbrücken, Germany.

[a1] *marius.syberg@ips.tu-dortmund.de*

Abstract:

The digitization of learning resources has led to an increase in specialized collaboration platforms across various fields, including the need for manufacturing companies to develop and maintain expertise in Industrial Data Science (IDS). This paper presents an approach to integrating collaborative and competency-based needs specific to industrial data analytics into a functional collaboration platform. We define the unique requirements of IDS projects and translate them into platform features. These features are then implemented and tested in an online platform within a research project, validating their effectiveness in a dynamic value network setting. The platform's primary innovation lies in its tailored design for IDS project practitioners from diverse domains, ensuring sustainable integration of data analytics in industrial settings. The initial version of this collaborative platform is currently accessible online and undergoing validation.

Key words:

Industrial data science, data analytics, industrial production, platform economy, competence developement.

1. Introduction

With the increasing digitization of industrial production, data science becomes the focus of companies that want to remain competitive in the age of globally competing value networks. In principle, every data science project in an industrial environment follows a similar process, with certain processes and workflows recurring throughout the projects (Schulz et al., 2020). This finding is well-known to practitioners of data analytics since the advent of Knowledge Discovery in Databases (KDD) at the latest (Fayyad et al., 1996). Nowadays, a broad range of process models aims at supporting the organization of KDD and data science projects through step-by-step instructions. As a data science process model, the so-called 'Cross Industry Standard Process for Data Mining' (CRISP-DM) is particularly widespread (Chapman et al., 2000). The CRISP-DM will be discussed in the further course of this work in Section 2. In addition to a process model, human competences are of the essence for the successful completion of any data analysis project. For this purpose, Mazarov et al. (2020) define four areas of competences and assigned roles for the different responsibilities in an IDS project:

1. Management
2. Data Scientist
3. Domain Expert
4. Information Technology (IT) Staff

In the same way that the areas of competence for different responsibilities can be divided into four groups. As such, Deuse et al. (2022) advocate to

To cite this article: Syberg, M., West, N., Schwenken, J., Adams, R., Deuse, J. (2024). A requirement-driven approach for competency-based collaboration in industrial data science projects. *International Journal of Production Management and Engineering, 12*(1), 79-90. https://doi.org/10.4995/ijpme.2024.19123

differentiate the segmental objectives of an IDS project into four elements. The result is a process chain that sequences these tasks and defines interaction points. For reference, the proposed model referred to as the 'Process Chain of Industrial Data Science' is shown in Figure 1.

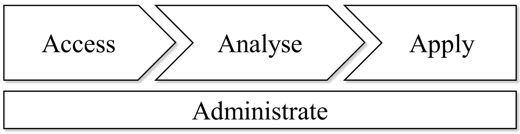

Figure 1. Process Chain of Industrial Data Science for structuring objectives during the analysis process.

The segmentation of IDS projects along the process chain allows collaborators of the aforementioned areas of competence to contribute their respective abilities in a well targeted manner. In addition, the process chain facilitates the development of analysis results in a reusable form by clearly separating individual deliverables. Albeit, to achieve a good reusability of solutions for data access, data analysis, data application and data administration, a suitable platform for collaboration with an appropriate IT infrastructure is required. We provide a tangible case for such a collaboration platform in Section 4.

Another worthwhile approach to fostering the impact of IDS projects is strengthening the ability to work closely together with colleagues of other departments or even other enterprises. Small and medium-sized enterprises (SME) in particular benefit from this approach, which we refer to in the rest of this article as *collaboration*. In contrast, the term *cooperation* describes a group of individuals working together in support of another's goals. Instead, a collaboration follows a shared objective. As such, the result of many partners cannot be directly attributed to a single party, but is the collective result of their joint efforts (Briggs et al., 2003; Henke & Kuhn, 2017). In the same way, joint knowledge discovery is augmented through collaborative work. Collaboration also aids in overcoming the difficulties presented by varied system landscapes (Henke & Kuhn, 2017). It reduces the challenges and financial burdens of substantial investments for a small and medium-sized enterprise (SME) (Zahoor et al., 2020). The only way to achieve effective knowledge management and quick development cycles is through sufficient collaboration between the aforementioned roles and between industrial SMEs acting as both users and system providers. Effective collaboration at these tiers necessitates coordination, communication, and organization.

Collaboration is not the only important factor for SMEs when using IDS; competence development is also crucial. Collaboration fosters the growth of subject-matter knowledge. However, the process of developing competence must also be tailored to each individual (Bauer et al., 2017). Digitization not only necessitates the development of competencies but also serves as a facilitative tool in this developmental process. (Bauer et al., 2017). Due to substantial shifts brought about by digitization, it is essential to enhance employee competency profiles across multiple levels. In the realm of IDS, it becomes crucial to identify, foster, and secure competencies focused on technology and data, as well as those related to processes, customers, and organizational aspects. (acatech, 2016). If they adhere to the situated learning principles, the use of digital technologies is effective for promoting competences (Stegmann et al., 2016). User-activating elements are another essential component of success that furthers knowledge transfer by enhancing the effectiveness of educational activities, personal engagement, and overall drive. (Hamari & Koivisto, 2015). To make the areas of data analysis, collaboration and competence development accessible to many SMEs in a combination, the development via a digital approach is obvious. A platform-based approach allows an easy distribution of services to many users (Schuh et al., 2011). Collaboration platforms are used for different purposes. Lee et al. (2003) are early to describe a web-based platform for e.g. knowledge management, focusing on connecting different systems and simplifying availability to users. Schuh et al. (2011) describe the use of a collaboration platform in the tooling industry. Smith et al. (2017) even present an approach with the FeatureHub, in which a collaboration platform is used for joint work on feature engineering. A platform approach therefore seems to be the right way forward. However, the requirements that a platform must fulfil in order to serve SMEs as an enabler for the successful implementation of IDS projects have not yet been defined and represent a research gap.

2. Fundamentals

The following section presents the state of the literature in several areas in order to define the requirements that a platform must meet. First, the fundamental approaches and requirements in IDS projects are described. This is followed by a description of how collaboration and competence building lead to success in IDS projects, as well as which aspects must be included in platform development.

2.1. Approaches and requirements of Industrial Data Science projects

As mentioned, projects in the field of IDS follow certain patterns. While data mining is a non-transparent entity for many SMEs, the procedure is widely researched in science. While Fayyad et al. (1996) describe in the pioneer model KDD the technical process of knowledge discovery in databases in a linear way, the CRISP-DM, which is widely used today, transfers the process of a data analysis project into an iterative loop (Chapman et al., 2000; Fayyad et al., 1996). The main focus is on the recurring character of these projects.

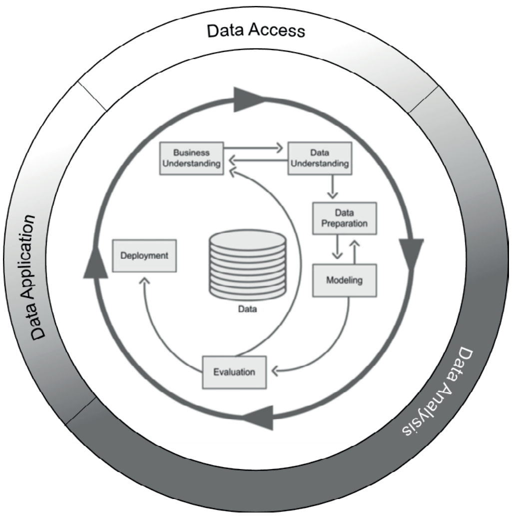

Figure 2. Process Chain of IDS according to the CRISP-DM (cf. Chapman et al., 2000; Deuse et al., 2022).

This results in always the same roles in these projects. Mazarov et al. (2020) have defined specific roles for various tasks within the AKKORD project, aligning them with the shared objectives in Industrial Data Science (IDS). A Data Scientist, an IT Employee, a Domain Expert and the Management share the different tasks. Both Kühn et al. (Kühn et al., 2018) and Deuse et al. (Deuse et al., 2021) emphasize the need of an additional, orchestrating role (Citizen Data Scientist) as necessary in IDS projects. As he is defined as a person using and understanding Data Science but primary working outside of the field he is able to understand and connect the other roles (Mullarkey et al., 2019). Other role concepts exist in the literature, but for SMEs the idea of a role concept as such is both a barrier and an opportunity. First of all, a company must fill these roles in order to be able to carry out projects successfully. In most cases, the necessary competences must first be built up with a certain effort. At the same time, the iterative nature of such projects offers the added value that these competences can be useful in the medium and long term by being used in many further projects. It must therefore be possible to divide the users own data analysis task into elements (modules), which in turn can be reused. These can be built up and divided using the Process chain of IDS also developed in the project (Figure 1) (Deuse et al., 2022).

The different analyses follow the steps of *data access, data analysis* and *data application*. *Data access* involves identifying pertinent data sources for an analytical task, acquiring any missing data through appropriate methods, and making it available for analysis. *Data analysis* includes data preprocessing and the analysis of given data itself. To achieve the companies goals, data analysis should be designed for scalability, versatility, simplicity, and timeliness (Ismail et al., 2019). For the *data application*, suitable visualizations and interfaces are provided in the last phase. There must be a certain technical basis for this. The requirements for a data backend system have already been published by Eiden et al. as part of the project (Eiden et al., 2020). At the same time, it needs to be a goal to ensure uploading and storing data while taking security criteria into account (Yang Liu et al., 2021). Users also need to be able to integrate results directly into their own systems or visualize them with dashboards (Chen et al., 2012; Moore, 2017). Assistance and administration must be included to overcome the existing barriers to enter the topic.

2.2. Collaboration in Industrial Data Science

A key strategy for improving the execution of IDS projects in SMEs is to engage in 'collaboration', which involves working closely with others. 'Collaboration' is defined differently and inconsistently in the literature. Appley and Winder (1977) define collaboration as a relational system in which individuals work in a conceptual framework on equal terms towards a goal, taking into account each other's motives (Appley & Winder, 1977).

This social-philosophical approach is similar to the definition of an acatech study on collaboration in industrial research, which defines collaboration as partners working together to achieve common goals more efficiently (acatech, 2016; Henke & Kuhn, 2017). They especially achieve results that cannot be directly attributed to any partner. Tasks to be executed are then derived from these goals (Briggs et al., 2003).

This intensive form of working with others is recognized as a necessary prerequisite for the implementation of projects in the context of digitization, which brings with it a number of advantages. Primarily, collaborative work aids in the collective development of knowledge. In this context, collaboration is particularly beneficial in addressing the challenges posed by heterogeneous system landscapes, as it facilitates the joint effort in developing interfaces and standards, thereby making these advancements accessible to a wider user base (Aulkemeier et al., 2019). From this point of view, collaboration ensures the holistic nature of solutions for individual companies and value networks. Collaboration secondly shortens development and innovation cycles (Hipp, 2021; Pisano, 1990). Dynamic markets and inconsistent framework conditions require companies to react to changes in an agile and adaptable manner. Market orientation is the relevant factor (Henke & Kuhn, 2017). Improved knowledge management and the pooling of cognitive resources accelerate developments and ensure market proximity. Finally, collaboration equally reduces risks and the barrier of high investment for an SME. As long as companies see the danger of technical lock-in, they risk to become either dependent on suppliers or not to face problems individually enough by using suboptimal applications in the data science area (Zerdick et al., 2001). The exchange with different providers and the cooperation with other providers in the context of a collaboration lowers this barrier.

In summary, the study suggests that dividing tasks among different roles within a company is essential for IDS projects. Simultaneously, effective collaboration between these roles and between small to medium-sized industrial firms as end-users and system providers is crucial. Successful collaboration at these levels hinges on key factors such as coordination, clear communication, and organized planning.

2.3. Competence Development in Industrial Data Science

Alongside collaboration, competency development plays a pivotal role for SMEs when dealing with IDS. Collaboration inherently contributes to knowledge enhancement within a specific domain. However, a personalized and focused approach to competency development is equally crucial (Bauer et al., 2017). Competence itself is defined as the ability to further develop knowledge and skills (Bergmann, 2000; Rauner, 2005). Digitization itself serves as both the catalyst for competency development and a valuable tool to facilitate this process (Bauer et al., 2017). SMEs have a particularly high demand for data science. According to different studies, SMEs and experts are well aware of the great potential and future significance of data science in this field. A lack of knowledge and problems in the organization are cited ahead of costs or IT-related issues (Dobler et al., 2020; Moeuf et al., 2020). Today, many companies use computer-based technologies for educational purposes (Lalic et al., 2017). Accordingly, environments from the e-learning area are particularly suitable for data science applications (Gorecky et al., 2017). In addition to the direct application of the learned content, interaction with other participants and the possibility to work on topics with his/her own pace promotes learning success (Derouin et al., 2005; Kong, 2011; Lalic et al., 2017). With the profound changes in work processes brought by digitalization, it will be necessary to further develop the competence profiles of employees at various levels. The different roles in a Data Science project already mentioned fulfill different tasks, which in turn require different competences. In research, the division of the employees competences is discussed (Dietzen et al., 2016), a separation into person-related and technical competences dominates (Erpenbeck et al., 2021; Ştefănică et al., 2017). The competences required by employees in the respective roles are closely linked to the activities they perform and the work processes in which they are involved (Becker, 2010; Fischer, 2000). Given the significant transformations brought about by digitization, it becomes imperative to enhance the competency profiles of employees across different levels (Arnold et al., 2016; Spöttl et al., 2016). In the context of IDS, it is essential to highlight the importance of recognizing, fostering, and securing competencies in technology, data, processes, customers, and organizational aspects (acatech, 2016). Depending on the role, technical competences for data analysis are just as relevant as personal competences for coordinating, collaborating and informing all stakeholders. The efficient and effective use of data analysis technologies in particular places special demands on both personal and technical competences in those roles. Against this background, situated learning modules for application-related, individual competence development are necessary. On the other hand, structures and practice-oriented recommendations for action are to be developed, which make it possible to record competences and to network them across the users. For a target-oriented competence development, the detection of the actual state at the beginning and during the process is essential. In general, objective performance tests are preferable to self-assessments (Dietzen et al., 2016; Zinn et al., 2015). This is only possible to a limited extent for person-related competences (intelligence, interest...), so self-assessments must be used here. The use of digital technologies is particularly effective for promoting competences if they follow the principles of situated learning (Stegmann et al., 2016). These principles are authenticity through the availability of real problems, multiple perspectives, social and collaborative learning, and guidance and support (Mandl & Kopp, 2006; Rosen & Salomon, 2007). Gamification is a method that supports knowledge transfer by using playful elements (Deterding et al., 2011). Thus, the goal of gamified elements is no longer just personal entertainment, but focuses on increasing the effectiveness of learning activities, participation, motivation of the learners. (Hamari & Koivisto, 2015; Sailer & Homner, 2020; Wood & Reiners, 2015)

In summary, in the AKKORD research project, a digital knowledge service with individual learning paths and data science related content regarding technical and personal competences is to be developed. This should follow the principles of blended learning.

Like collaboration, this not only shortens innovation cycles, but also anchors knowledge in the long term.

3. Requirements for an IDS Collaboration Platform

In the triad of people, technology and organization, the AKKORD project addresses the sub-goals of value-creating collaborations, integrated and networked analysis and development of action competences and recommendations. The basis is the creation of a complete, networked and integrated database to implement all in a modular, data-driven reference kit in the form of a collaborative service platform. In the previous section, three sub-areas were described. The general approach in IDS as well as the collaboration *in* these projects and the competence development *for* these projects were examined in more detail to define requirements for the platform solution.

Meeting these requirements transforms a platform into a facilitator for prosperous data science projects within SMEs. Nevertheless, the literature lacks requirements models tailored to collaboration platforms in the IDS domain. Hence, we have drawn upon models from other domains or with distinct focuses, such as learning or collaboration platforms, and amalgamated the requirements outlined in the following section. We use requirements from the procedure in IDS projects as well as collaboration and competence development with regard to these projects developed from the fundamentals in Section 2.

Additionally several models from distantly related approaches were taken into account. Spath et al. (2007) present a web-based open source software for enterprise collaboration. For this, necessary functionalities are defined, which in turn are to be understood as requirements for a solution. These are divided into four areas. In the ‘communication support’ section, functionalities such as e-mail, contact management and forum options are presented. ‘Project management’ covers the management of resources, project progress and tasks. ‘Information and data management’ describes a wiki and suitable documentation. Finally, the ‘administrative functions’ section includes the management of the platform itself and a troubleshooting system. (Spath, 2007)

Ismail et al. (2019) characterize process data analysis pipelines to assist data engineers in their design. For this purpose, they build on literature and existing platform requirements for such a pipeline. These are also related to the approach taken in the AKKORD project. The authors build their requirements on the FURPS model, which has been merged into ISO/IEC 25000 today (Grady & Caswell, 1987). A data analysis support platform must be designed with a focus on functionality, user-friendliness, dependability, performance, and maintainability (Ismail et al., 2019).

Süptitz et al. (2013) define requirements for 'Virtual Research Environments'. Although this is thematically far from IDS in SMEs, a consideration of the ideas is obvious. Interdisciplinary collaboration and targeted competence development as enablers for the success of these environments are similar. In addition to the models of Spath et al. and Ismail et al., a focus is placed here on data management, communication and collaboration as well as an interdisciplinary and modular orientation (Süptitz et al., 2013). This is aligned with the conclusions from the general fundamentals mentioned in Section 2.

3.1. General requirements

Broken down, it must be possible to perform data analytics, develop competences in a targeted manner, and interact with one another on the related platform solution. To achieve this, it is imperative to establish the *technical infrastructure* for data analysis, as well as for learning and collaboration on the platform. This also aligns with the necessity for *individualization*. Although the process of data analysis projects may follow a recurring pattern, it is vital for the success of SME projects that individual issues are effectively resolved. Moreover, ensuring *accessibility* and security serves as a foundational requirement. *Accessibility* entails that the platform should be user-friendly for a wide range of SMEs, allowing for easy integration, utilization of analytical results in other software, and data storage capabilities. Security encompasses safeguarding personal and company-specific data throughout all stages of analysis and learning processes, including up-/downloads and storage. (see Table 1). The general requirements have an effect and also apply to the three areas. The extent to which they are prerequisites is explained in the respective section.

Table 1. General requirements for a competency-based collaboration plattform.

General requirements:
Individualization: User individual choice of contents; Company individual choice of contents; Individual assistance systems
Technical Infrastructure: Powerful backend; Intuitive user interface; Execution of analysis directly in the system
Accessability: Access for many users with different technical prerequisites; general possibility to integrate in IT infrastructure or use as open cloud service, Upload, download and store data
Security: Safety of personal and company specific data; Following international legal requirements

3.2. Industrial Data Science specific requirements

The requirements for a collaboration platform for IDS are based on the FURPS Model and extended in several details (Grady & Caswell, 1987). This secures a stable data analytics pipeline. We interpret the *functionality* requirement of FURPS as encompassing the initial breakdown of the analysis process into reusable modules that can be configured in various ways. Additionally, it should be feasible to utilize analysis results through interfaces or visualization options. For *usability*, modules need to be developed for all phases of an IDS project, with a focus on industrial production, aligned with one of the process models mentioned in Section 2 to ensure project success. Users should have the ability to select or create modules as needed, with proper documentation to facilitate accessibility for new users. In this context, *reliability* entails handling a certain level of data pre-processing and analysis process preparation to achieve the necessary data quality. *Performance* relies on a high-capacity server environment within a cloud-based solution, closely tied to the technical infrastructure. Regarding content, the platform's performance potential is fully realized by effectively integrating practical data analytics with competency development and learning. *Supportability* is upheld through the collaborative nature of the platform.

Table 2. IDS specific requirements, based on the FURPS Model (Grady & Caswell, 1987).

IDS specific requirements
Functionality: Fragmentation of analysis process; Configurability; Use of results in other environments
Usability: Creating modules for all phases of an IDS Project; Selection and creation of modules by users; Documentation
Reliability: Preprocessing data and Preparation of analysis to ensure Data Quality (West et al., 2021)
Performance: High performance analytic environment (cloud-based); Linkage to competence development
Supportability: Secured by platform approach itself; Structured database, up- and download of data and results; technical problem support

Technically, support is guaranteed appropriate *administration* and *technical infrastructure* need to be guaranteed. Structured databases, the functions for uploading and downloading, and technical maintenance must be ensured, as well as the organization and support within the processes. The requirement of *individualization* in this area means above all individual adaptation possibilities of the modules. For example, a time series analysis may make sense for one company in the area of sales, but for another user in the area of maintenance. Accordingly, the module can be reused if it is configurable. At the same time, assistance systems must be implemented that suggest useful modules to users according to their company and individual profile.

3.3. Collaboration specific requirements

The collaboration-specific requirements are a direct outcome of the literature review presented in Section 2.2. Consequently, the primary emphasis lies on leveraging the network effect. A crucial requirement is the development of modules with universal applicability. Simultaneously, the platform's value is enhanced when users share their individually created and customized modules with others. Another essential requirement is to facilitate intuitive *communication*. The platform should include features such as forums, comment options, and chat functions to promote interaction among users. Furthermore, *information* accessibility is of paramount importance. To keep users informed about the constantly evolving field of data science, incorporating features like news feeds or implementing a Wiki can be beneficial. These features would provide users with timely and transparent updates on general news and modifications to modules or learning content that are relevant to their interests.

Table 3. Collaboration specific requirements.

Collaboration specific requirements
Communication: Features to encourage an exchange between users (Forum, comment options, chat)
Information: Keep users on current state of Data Science and their projects/learning (e.g. information feed, Wiki)

Individualization with regard to collaboration can be achieved in particular through suitable account and profile design, which should be coupled with contact management. This enables contact to be made between different users. Additionally, this is important company-internal if several users work on the platform. With regard to role concepts, this makes it possible to identify missing roles or to expand teams. With regard to *administration*, according to Spath et al. (2006), project, resource and task management is an important function here in order to work collaboratively. The *technical infrastructure* meets the requirements already mentioned.

3.4. Competence development specific requirements

Furthermore, in the realm of multidisciplinary competence development for IDS, there are corresponding requirements in addition to the establishment of a knowledge service, such as a Wiki. The first requirement pertains to *content efficiency*, which is ensured by closely aligning it with the developed role concept (refer to Section 2). Learning modules related to content must be created or externally integrated to impart data science knowledge in an industry-oriented manner. This entails teaching not only the methods themselves but also the procedures based on procedural models (e.g., CRISP-DM, Process Chain of IDS). Moreover, instruction on utilizing the platform and its modules efficiently is essential. The second requirement concerns *learning support*. According to literature, various concepts can enhance learning outcomes. Diverse approaches, including gamification, the utilization of various media (text, video), as well as progress tracking, tests, and certificates, should be implemented to ensure successful learning results.

Table 4. Requirements for competence development of IDS projects [1, 4, 10].

Competence Development specific requirements
Content Efficiency: Develop knowledge in Data Science Methods and Algorithms, Data Science Procedure Models and Using the Platform properly
Learning Support: Ensure targeted learning by determining users initial level, using Gamification, different media, awarding and certification

Individualization is also required in Competence Developement. First of all, a functionality has to be implemented that determines the user's initial level. Based on the user's intended role and their initial skill level, a personalized learning path comprising specific learning modules should be recommended. With regard to the *technical infrastructure*, it must be possible to include further learning content in line with the constant new developments. The *administration* refers in particular to the learning progress and eventual bug fixes.

4. Development of a Platform concept for Collaborative Industrial Data Science

In the research project AKKORD we transferred the requirements into a platform solution as a part of the AKKORD reference toolkit. The following part describes the solution in detail.

4.1. Adaption of requirements into findings

In response to the specific needs outlined in four distinct categories, the AKKORD research project has formulated a concept that equips users with the necessary skills in Industrial Data Science (IDS). This is achieved through collaborative, modular data analysis techniques. A key component of this initiative, showcased in Figure 2, is the ‘Learning and Collaboration Platform,’ integral to the AKKORD reference toolkit.

The ‘AI Toolbox’ facilitates active engagement in modular data analysis, whereas the ‘Work & Learn Platform’ focuses on collaboration and the development of competencies. A strategic linkage exists between these two segments, with their separation aimed at maximizing the practical application and dissemination of research findings, particularly for SMEs and the industrial manufacturing sector. Both segments, while connected, also function as standalone outcomes of the project and can be utilized independently. The platform’s innovation is primarily reflected in its design, which is meticulously tailored to meet the needs of industrial users. The ‘Work & Learn Platform’ is adept at imparting a broad spectrum of IDS skills, ranging from Database Technology to Python Programming, even within the dynamic and demanding context of a corporate setting. It enables seamless collaboration on data science topics, both within an organization and across different companies.The ‘AI Toolbox’ stands out for its ability to facilitate modular data analyses with minimal programming knowledge, offering an accessible, browser-based interface. Additionally, it supports custom implementations to cater to unique business use cases. This approach not only makes data analysis more approachable for a wider audience but also provides flexibility for tailored business applications. Within the scope of the AKKORD project, the ‘Work & Learn Platform’ is being constructed as a digital workplace, utilizing the NEOCOSMO platform PIIPE to meet these specific requirements. Its homepage features a variety of topic channels and magazine articles, customizable through user subscriptions to ensure relevance. The course section lists both internal and external courses and includes an interface for monitoring progress and selecting suitable learning materials. These courses, developed in collaboration with consortium partners and other platforms, cover specialized Data Science subjects and practical project management. Adapted from university curricula and modified for real-world application with industry partners, the courses are enriched with diverse learning aids like images, YouTube videos, slideshows, and quizzes. These elements are designed to make the learning journey more engaging and diverse for the user. A progress bar is visible for courses already underway. Beyond courses, the platform offers a collaborative wiki, a magazine section, and forum-like features to facilitate knowledge sharing and direct discussion. Upon registration, users undergo a self-evaluation to create a tailored profile, which then guides the suggestion of learning paths and courses. As an example, an experienced project manager with no prior experience in Data Science would follow a different learning path compared to a Quality Management expert who has been working with data throughout their career. We ensure that, based on the initial survey responses, both individuals receive the courses that cater to their specific needs and backgrounds. We Service providers also have a dedicated space to offer their services, linking them with other platform content. The ‘AI-Toolbox,’ implemented using RapidMiner software, is a visually-oriented data science workflow tool that simplifies analysis process construction. Managed by project partners VPE, PDtec, CONTACT, and ARENDAR, this backend is modular in design. Within AKKORD, RapidMiner is employed to develop various modules aligned with IDS process chain phases, customizable as per user needs. Each module is equipped with a data schema, clarifying the data format and attributes required for effective utilization. The ongoing implementation of collaborative analysis within the platform exemplifies this innovative approach.

5. Validation of the AKKORD platform

The description of the platform solution developed in AKKORD demonstrates that the requirements have been largely met. The fundamental criteria for ensuring platform accessibility and security have been addressed. Ongoing validation will

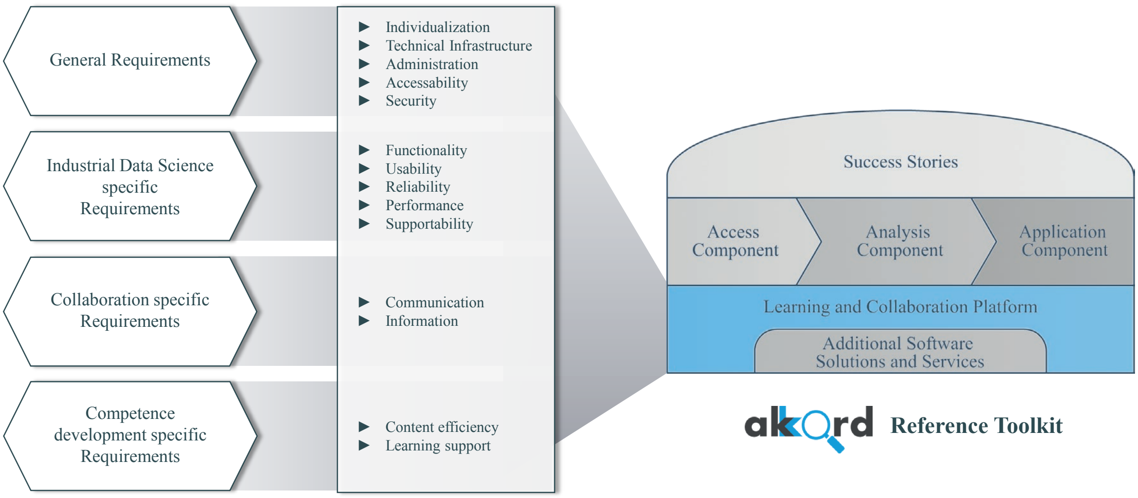

Figure 3. Adaption of requirements into functions of the Platform in the AKKORD reference toolkit.

determine the extent of required adaptations and interface extensions. Notably, security concerns have emerged in German companies, particularly regarding the transmission of sensitive data to the cloud. Nevertheless, the browser-based approach holds promise for accessibility. The technical infrastructure is built upon the successful existing system (PIIPE), offering versatile administration functions and supporting competence development. This encompasses user profile management, course management, and the wiki. The open approach fosters user-driven growth. While the ability of the administration and infrastructure to effectively support the use of analysis modules cannot be definitively assessed at present, the structural requirements have been met.

The platform guarantees accessibility through its browser-based design, ensuring usability for all users. Simultaneously, security is upheld through robust access management. Additionally, the possibility of implementing the platform within an organization's own intranet offers an alternative security measure.

6. Conclusion

In conclusion, this paper highlights the pivotal role of competence development and collaboration in addressing IDS challenges for SMEs. The ongoing digitization necessitates these efforts while also providing the means to implement digital solutions. Managing IDS projects requires a platform that meets specific requirements, particularly in facilitating modular data analysis, competence development, and collaboration both internally and externally. The AKKORD research project has successfully met these requirements through the development of an online platform featuring two key elements: the 'AI Toolbox' for modular data analysis, designed as a user-friendly, no-code solution, and the 'Work & Learn Platform' tailored to fostering collaboration and competence development in an industrial context. The project's final phase will involve integration and validation with additional users, ensuring any necessary adjustments are made. Promisingly, this platform has the potential to offer comprehensive support to all companies, particularly SMEs, engaged in IDS projects. It enables companies to establish networks, build competences, and conduct targeted data analyses with ease, empowering SMEs to nurture their employees into Citizen Data Scientists. Furthermore, there is potential for innovative business models to emerge, leveraging the principles of the platform economy, provided network effects are effectively harnessed.

Data Availability

The data that support the findings of this study are available on request from the corresponding author, Marius Syberg, upon reasonable request.

Acknowledgements

The work on this paper has been supported by the German Federal Ministry of Education and Research (BMBF) as part of the funding program 'Industry 4.0 - Collaborations in Dynamic Value Networks (InKoWe)' in the project AKKORD (02P17D210).

Authors contribution

Marius Syberg developed the idea presented and led the research and elicitation of requirements for the platform.

Nikolai West led the research project and conceptualised the requirements into functionalities of the platform.

Jörn Schwenken worked on the role concept as part of the project and influenced the development of the platform with regard to the qualification of users in relation to the roles.

Rebekka Adams worked on the development of the 'Work & Learn Platform' in terms of technology, methodology and content.

Prof. Dr.-Ing. Jochen Deuse developed the conceptual project idea and, as a professor in the field of Industrial Data Science, brings the data analytical requirements to the contribution.

References

Acatech (Ed.). (2016). *acatech Position. Kompetenzen für Industrie 4.0: Qualifizierungsbedarfe und Lösungsansätze (Competencies for Industry 4.0. Requirements for Qualitifcations and Solutions)*. Herbert Utz Verlag GmbH.

Appley, D. G., & Winder, A. E. (1977). An Evolving Definition of Collaboration and Some Implications for the World of Work. *The Journal of Applied Behavioral Science*, *13*(3), 279–291. https://doi.org/10.1177/002188637701300304

Arnold, D., Butschek, S., Steffes, S., & Müller, D. (2016). *Digitalisierung am Arbeitsplatz: Bericht*, *FB468*.

Aulkemeier, F., Iacob, M.-E., & van Hillegersberg, J. (2019). Platform-based collaboration in digital ecosystems. *Electronic Markets*, *29*(4), 597–608. https://doi.org/10.1007/s12525-019-00341-2

Bauer, W., Dworschak, B., & Zaiser, H. (2017). Weiterbildung und Kompetenzentwicklung für die Industrie 4.0 (Education and skills development for Industry 4.0). In B. Vogel-Heuser, T. Bauernhansl, & M. ten Hompel (Eds.), *Handbuch Industrie 4.0 Bd.1* (pp. 125–138). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-45279-0_36

Becker, M. (Ed.). (2010). *Berufliche Bildung in Forschung, Schule und Arbeitswelt: Vol. 5. Von der Arbeitsanalyse zur Diagnose beruflicher Kompetenzen: Methoden und methodologische Beiträge aus der Berufsbildungsforschung*. Lang. https://doi.org/10.3726/978-3-653-00477-9

Bergmann, B. (2000). *Kompetenzentwicklung und Berufsarbeit*. Waxmann.

Briggs, R., Vreede, G.-J. de, & Jr, J. (2003). Collaboration Engineering with ThinkLets to Pursue Sustained Success with Group Support Systems. *J. Of Management Information Systems*, *19*, 31–64. https://doi.org/10.1080/07421222.2003.11045743

Chapman et al. (Ed.) (2000). *CRISP-DM 1.0: Step-by-step data mining guide*.

Chen, Chiang, & Storey (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *MIS Quarterly*, *36*(4), 1165. https://doi.org/10.2307/41703503

Derouin, R. E., Fritzsche, B. A., & Salas, E. (2005). E-learning in organizations. *Journal of Management*, *31*(6), 920–940. https://doi.org/10.1177/0149206305279815

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness. In A. Lugmayr, H. Franssila, C. Safran, & I. Hammouda (Eds.), *Proceedings of the 15th International Academic MindTrek Conference on Envisioning Future Media Environments - MindTrek '11* (p. 9). ACM Press. https://doi.org/10.1145/2181037.2181040

Deuse, J., West, N., & Syberg, M. (2022). Rediscovering scientific management. The evolution from industrial engineering to industrial data science. *International Journal of Production Management and Engineering*, *10*(1), 1–12. https://doi.org/10.4995/ijpme.2022.16617

Deuse, J., Wöstmann, R., Schulte, L., & Panusch, T. (2021). Transdisciplinary competence development for role models in data-driven value creation: The Citizen Data Scientist in the Centre of Industrial Data Science Teams. In W. Sihn & S. Schlund (Eds.), *Competence development and learning assistance systems for the data-driven future* (pp. 37–58). Goto Verlag. https://doi.org/10.30844/wgab_2021_3

Dietzen, A., Nickolaus, R., Rammstedt, B., & Weiß, R. (Eds.). (2016). *Berichte zur beruflichen Bildung. Kompetenzorientierung: Berufliche Kompetenzen entwickeln, messen und anerkennen*. W. Bertelsmann Verlag GmbH & Co. KG.

Dobler, M., Etschmann, R., Kugler, P., Meierhofer, J., Olbert-Bock, S., Redzepi, A., Schumacher, J., Thiel, C., & Tietz, R. (2020). *Data Science für KMU leicht gemacht.* Internationale Bodenseehochschule Labs. https://doi.org/10.25924/opus-3505

Eiden, A., Gries, J., Eickhoff, T., & Göbel, J. C. (2020, December 16). Anforderungen an ein Daten-Backend-System zur Unterstützung industrieller Datenanalyse-Anwendungen in digitalen Engineering-Prozessen dynamischer Wertschöpfungsnetzwerke. In *Proceedings of the 31st Symposium Design for X (DFX2020)* (pp. 81–90). The Design Society. https://doi.org/10.35199/dfx2020.9

Erpenbeck, J., Heyse, V., Meynhardt, T., & Weinberg, J. (2021). *Die Kompetenzbiographie: Wege der Kompetenzentwicklung* (3. durchgesehene Auflage). Waxmann.

Fayyad, U. M., Piatetsky-Shapiro, G., & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *AI Mag*, *17*, 37–54.

Fischer, M. (2000). *Von der Arbeitserfahrung zum Arbeitsprozeßwissen: Rechnergestützte Facharbeit im Kontext beruflichen Lernens*. Leske + Budrich. https://doi.org/10.1007/978-3-663-11783-4

Gorecky, D., Khamis, M., & Mura, K. (2017). Introduction and establishment of virtual training in the factory of the future. *International Journal of Computer Integrated Manufacturing*, *30*(1), 182–190. https://doi.org/10.1080/0951192X.2015.1067918

Grady, R. B., & Caswell, D. L. (1987). *Software metrics: Establishing a company-wide program*. Prentice-Hall.

Hamari, J., & Koivisto, J. (2015). Why do people use gamification services? *International Journal of Information Management*, *35*(4), 419–431. https://doi.org/10.1016/j.ijinfomgt.2015.04.006

Henke, M., & Kuhn, A. (Eds.). (2017). *acatech STUDIE. Kollaboration als Schlüssel zum erfolgreichen Transfer von Innovationen: Analyse von Treibern und Hemmnissen in der Automobillogistik (Collaboration as to the Successful Transfer of Innovations. Analyses of Drivers and Barriers in automotive logistics)*. Herbert Utz Verlag GmbH. http://www.acatech.de/wp-content/uploads/2018/03/Innokey_acatech_STUDIE_Web.pdf

Hipp, A. (2021). R&D collaborations along the industry life cycle: the case of German photovoltaics manufacturer. *Industrial and Corporate Change*, *30*(3), 564–586. https://doi.org/10.1093/icc/dtaa054

Ismail, A., Truong, H.-L., & Kastner, W. (2019). Manufacturing process data analysis pipelines: a requirements analysis and survey. *Journal of Big Data*, *6*(1). https://doi.org/10.1186/s40537-018-0162-3

Kong, S. C. (2011). An evaluation study of the use of a cognitive tool in a one-to-one classroom for promoting classroom-based dialogic interaction. *Computers & Education*, *57*(3), 1851–1864. https://doi.org/10.1016/j.compedu.2011.04.008

Kühn, A., Joppen, R., Reinhart, F., Röltgen, D., Enzberg, S. von, & Dumitrescu, R. (2018). Analytics Canvas – A Framework for the Design and Specification of Data Analytics Projects. *Procedia CIRP*, *70*, 162–167. https://doi.org/10.1016/j.procir.2018.02.031

Lalic, B., Majstorovic, V., Marjanovic, U., Delić, M., & Tasic, N. (2017). The Effect of Industry 4.0 Concepts and E-learning on Manufacturing Firm Performance: Evidence from Transitional Economy. In H. Lödding, R. Riedel, K.-D. Thoben, G. von Cieminski, & D. Kiritsis (Eds.), *IFIP advances in information and communication technology. Advances in Production Management Systems. The Path to Intelligent, Collaborative and Sustainable Manufacturing* (Vol. 513, pp. 298–305). Springer International Publishing. https://doi.org/10.1007/978-3-319-66923-6_35

Lee, W. B., Cheung, C. F., Lau, H., & Choy, K. L. (2003). Development of a Web-based enterprise collaborative platform for networked enterprises. *Business Process Management Journal*, *9*(1), 46–59. https://doi.org/10.1108/14637150310461396

Mandl, H., & Kopp, B. (2006). *Blended Learning: Forschungsfragen und Perspektiven.* https://doi.org/10.5282/ubm/epub.905

Mazarov, J., Schmitt, J., Deuse, J [J.], Richter, R., Kühnast-Benedikt, R., & Biedermann, H. (2020). Visualisation in Industrial Data Science projects (Translation): Visualisierung in Industrial Data-Science-Projekten (Original title). *Industrie 4.0 Management*, *36*(6), 63–66.

Moeuf, A., Lamouri, S., Pellerin, R., Tamayo-Giraldo, S., Tobon-Valencia, E., & Eburdy, R. (2020). Identification of critical success factors, risks and opportunities of Industry 4.0 in SMEs. *International Journal of Production Research*, *58*(5), 1384–1400. https://doi.org/10.1080/00207543.2019.1636323

Moore, J. (2017). Data Visualization in Support of Executive Decision Making. *Interdisciplinary Journal of Information, Knowledge, and Management*, *12*, 125–138. https://doi.org/10.28945/3687

Mullarkey, M. T., Hevner, A. R., Grandon Gill, T., & Dutta, K. (2019). Citizen Data Scientist: A Design Science Research Method for the Conduct of Data Science Projects. In B. Tulu, S. Djamasbi, & G. Leroy (Eds.), *Lecture Notes in Computer Science. Extending the Boundaries of Design Science Theory and Practice* (Vol. 11491, pp. 191–205). Springer International Publishing. https://doi.org/10.1007/978-3-030-19504-5_13

Pisano, G. P. (1990). The R&D Boundaries of the Firm: An Empirical Analysis. *Administrative Science Quarterly*, *35*(1), 153. https://doi.org/10.2307/2393554

Rauner, F. (Ed.). (2005). *Handbuch Berufsbildungsforschung*. Bertelsmann. http://www.socialnet.de/rezensionen/isbn.php?isbn=978-3-7639-3167-5

Rosen, Y., & Salomon, G. (2007). The Differential Learning Achievements of Constructivist Technology-Intensive Learning Environments as Compared with Traditional Ones: A Meta-Analysis. *Journal of Educational Computing Research*, *36*, 1–14. https://doi.org/10.2190/R8M4-7762-282U-554J

Sailer, M., & Homner, L. (2020). The Gamification of Learning: a Meta-analysis. *Educational Psychology Review*, *32*(1), 77–112. https://doi.org/10.1007/s10648-019-09498-w

Schuh, G., Boos, W., & Völker, M. (2011). Collaboration platforms to enable global service provision in the tooling industry. *Production Engineering*, *5*(1), 9–16. https://doi.org/10.1007/s11740-010-0274-x

Schulz, M., Neuhaus, U., Kaufmann, J., Badura, D., Kuehnel, S., Badwitz, W., Dann, D., Kloker, S., Alekozai, E. M., & Lanquillon, C. Introducing DASC-PM: A Data Science Process Model. In *ACIS 2020 Proceedings* (Vol. 45). (Original work published 2020)

Smith, M. J., Wedge, R., & Veeramachaneni, K. (2017). FeatureHub: Towards Collaborative Data Science. In *2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA)* (pp. 590–600). IEEE. https://doi.org/10.1109/DSAA.2017.66

Spath, D. (Ed.). (2007). *Webbasierte Open Source-Kollaborationsplattformen: Eine Studie der Fraunhofer Gesellschaft*. Fraunhofer IRB Verlag.

Spöttl, G., Gorldt, C., Windelbrandt, L., Grantz, T., & Richter, T. (2016). *Industrie 4.0 – Auswirkungen auf Aus- und Weiterbildung in der M+E Industrie*. Eine bayme vbm Studie, erstellt von der Universität Bremen.

Ştefănică, F., Abele, S., Walker, F., & Nickolaus, R. (2017). Modeling, Measurement, and Development of Professional Competence in Industrial-Technical Professions. In M. Mulder (Ed.), *Technical and Vocational Education and Training: Issues, Concerns and Prospects. Competence-based Vocational and Professional Education* (Vol. 23, pp. 843–861). Springer International Publishing. https://doi.org/10.1007/978-3-319-41713-4_39

Stegmann, K., Wecker, C., Mandl, H., & Fischer, F. (2016). Lehren und Lernen mit digitalen Medien (Teaching and Learning with Digital Media). In R. Tippelt & B. Schmidt-Hertha (Eds.), *Handbuch Bildungsforschung* (pp. 1–22). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-531-20002-6_42-1

Süptitz, T., Weis, S., & Eymann, T. (2013). *Was müssen Virtual Research Environments leisten? – Ein Literaturreview zu den funktionalen und nichtfunktionalen Anforderungen.*

West, N., Gries, J., Brockmeier, C., Göbel, J. C., & Deuse, J [Jochen] (2021). Towards integrated Data Analysis Quality: Criteria for the application of Industrial Data Science. In *2021 IEEE 22nd International Conference on Information Reuse and Integration for Data Science (IRI)*. https://doi.org/10.1109/IRI51335.2021.00024

Wood, L. C., & Reiners, T. (2015). Gamification. In D. Khosrow-Pour (Ed.), *Encyclopedia of Information Science and Technology, Third Edition* (pp. 3039–3047). IGI Global. https://doi.org/10.4018/978-1-4666-5888-2.ch297

Yang Liu, Tao Fan, Tianjian Chen, Qian Xu, & Qiang Yang (2021). FATE: An Industrial Grade Platform for Collaborative Learning With Data Protection. *J. Mach. Learn. Res.*, *22*, 226:1-226:6.

Zahoor, N., Al-Tabbaa, O., Khan, Z., & Wood, G. (2020). Collaboration and Internationalization of SMEs: Insights and Recommendations from a Systematic Review. *International Journal of Management Reviews*, *22*(4), 427–456. https://doi.org/10.1111/ijmr.12238

Zerdick, A., Picot, A., Schrape, K., Artopé, A., Goldhammer, K., Heger, D. K., Lange, U. T., Vierkant, E., López-Escobar, E., & Silverstone, R. (2001). *Die Internet-Ökonomie: Strategien für die digitale Wirtschaft* (3., erweiterte und überarbeitete Auflage). *European Communication Council Report*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-56418-5

Zinn, B., Güzel, E., Walker, F., Nickolaus, R., Sari, D., & Hedrich, M. (2015). ServiceLernLab - Ein Lern- und Transferkonzept für (angehende) Servicetechniker im Maschinen- und Anlagenbau. Advance online publication. *Journal of Technical Education (JOTED), 3*(2). https://doi.org/10.48513/joted.v3i2.61