NORDDESIGN CONFERENCE
REYKJAVIK, ICELAND, 12-14 AUGUST 2024

Requirements Analysis of a Research Data Management System in Collaborative Projects

Laura Mueller[1], Andreas Maximilian Schultz[1], Osman Altun[2], Johanna Uhe[2], Oliver Koepler[3], Soeren Auer[3], Iryna Mozgova[1]

[1]Paderborn University, Germany
[2]Leibniz University Hannover, Germany
[3]Leibniz Information Centre for Science and Technology University Library, Germany

Abstract: The collection, analysis, and reuse of data throughout the entire product lifecycle and the resulting potential are establishing in engineering context. Research Data Management (RDM) systems are used in scientific context to enable the storage of research data in a structured way with the aim of reusing research data. Nevertheless, RDM-systems are still a novelty in engineering disciplines and therefore its implementation is accompanied by a number of challenges. In order for RDM-systems to be used extensively, they must fulfil the individual user requirements of the engineering sciences. This paper presents a methodological approach to derive user requirements for a RDM-system and perform a first categorization by applying the Kano model. A survey among researchers is conducted to initially verify and confirm the identified requirements.

Keywords: Requirements Management, Knowledge Management, Research Data Management System, Kano Model, Product Data Management (PDM)

1 Introduction

Data management is important in all areas of engineering, both in the industrial environment and in research. Product Lifecycle Management (PLM) is the management of the product throughout the entire product lifecycle (PLC) (Stark, 2022). This implies the consistent traceability of information and therefore, data across all phases of the entire PLC. Product data management, which is part of PLM, plays an important role during the design phase. Within the engineering areas, the data is characterized by a high degree of diversity, so there are different file formats and challenges regarding interoperable processing (Mozgova et al., 2020). The aim of information systems is to prepare data so users from the various phases of the PLC can access, store, serve and reuse product information (Sudarsan et al., 2005). In this context, many obstacles arise at different levels, and various approaches are currently being researched. Ontology-based solutions that enable semantic interoperability and knowledge management have emerged as an important driver (Fraga et al., 2020).

However, the application of these solutions does not fully reflect the requirements to scientific documentation in conducting research and the organisation of the collection, analysis, and publication of research data. The recognized potential of reusing data and the proposed FAIR data principles (Wilkinson et al., 2016) emphasize the importance of Research Data Management (RDM). With the help of RDM, research data is managed throughout the entire data life cycle with the aim of reusing research data and drawing more benefits out of it. Its importance is confirmed by international and national data strategies for RDM. At international level for example the European Open Science Cloud (EOSC) provides researchers services for RDM and at national level there are many initiatives that have been established during the last few years with the aim of providing RDM.

RDM-Systems are essential in order to exploit the potential of RDM within research projects. With their help, datasets are prepared in such a way that they can be reused by other researchers. Its usage is an important contribution to knowledge generation, but researchers still hesitate to use it due to uncertainty (Wilms et al., 2020). The use of research data involves many challenges due to the lack of common platforms and standardised processes, especially in the context of engineering (Sheveleva et al., 2022). In order for RDM-systems to be used effectively, they must fulfil the requirements of the individual areas and researchers. The engineering sector in particular is characterised by a high degree of diversity, which is why the derivation of requirements needs specific investigation.

This paper shows an example of a RDM-System and deals with the question of user requirements in engineering context. A methodical approach is used to derive requirements by applying the Kano model (Rotar and Kozar, 2017). For this purpose, various interviews were realized, with developers of the system and researchers. With the interviews as a basis and the orientation on the Kano model, an initial categorisation of the requirements for a RDM-system is made. Also, a survey is conducted to carry out an initial check whether the requirements are confirmed.

2 Background

This section provides an overview of the basic content referred to in this paper. First, general functions and a description of two exemplary RDM-systems used in Collaborative Research Centers (CRCs) are presented. Furthermore, the basics of the Kano model are described in order to derive user requirements. The focus here is on categorising user needs and their differences.

2.1 Research Data Management System

RDM-systems face major challenges in terms of their degree of utilisation. On the one hand, it is a new field of application for many researchers, on the other hand, RDM-systems only offer added value if they are used correctly. In this context Tang and Hu emphasize that training opportunities and institutional commitment to resources is crucial (Tang and Hu, 2019). Two RDM-systems used in practice are presented below.

The RDM-systems used as a reference in this paper were developed over the last four years for two CRCs that research on current mechanical engineering topics. The goal is to have a modular, adjustable and reusable system that can be applied to different purposes in both CRCs. The core components are a Comprehensive Knowledge Archive Network (CKAN) instance and a semantically enhanced mediawiki, called SemanticMediaWiki (SMW). Both systems are semantically linked and use domain specific vocabulary. CKAN is used as a data repository for storing the actual research raw data which can be text files (measurement results), graphical files like images or plots as well as binary data from proprietary software like Word or Excel. This system offers the possibility to push datasets into a public repository. SMW is used as a knowledge management system to store information about the actual data by generating protocols. This metadata describe the process of data-creation, parameters of the creation process, ownership of data and many more information that help to describe the research data. Both software components are free to use and their code is open-source which facilitates the development of extensions that were needed to fulfil our end-users' requirements. The RDM-teams of both CRCs are staffed overlappingly which enables a close exchange of best practices between both projects. Therefore, both systems provide help-pages with advices for the end users as well as videos with explanations of the systems.

The CRC 1153 aims the development of novel process chains for the production of hybrid solid components (Behrens and Uhe, 2021). The RDM-system in CRC 1153 is used to store information about sample statuses. CKAN is used for storing datasets which belong to samples passing through the tailored-forming process chain. SMW stores information about the protocols and helps the researchers to keep track of the sample status and process history.

The researchers in CRC 1368 collaborate on the development of oxygen-free production processes that are enabled by creating an oxygen-free environment at normal pressure with specific gases (Wegewitz et al., 2023). State of the art is generating oxygen-free environments by applying an extremely high vacuum which requires expensive and difficult apparatuses. The RDM-system is used to protocol the performed process steps that itself are developed in the CRC. Therefore, the SMW needs to provide a flexible way of inventing and creating protocols for newly designed production processes. CKAN is used similarly as a data storage facility. The integration of an electronic laboratory journal is implemented by an interface to the open-source software eLabFTW which is used by several project partners. A more detailed description of the semantic architecture of the RDM is given by Altun et al (Altun et al., 2023). The users of the systems are researchers and project leaders whose satisfaction with the systems is important.

2.2 Kano Model

The analysis and evaluation of customer satisfaction is part of quality management systems referring to requirements (EN ISO 9001). Customer needs and expectations can be determined by using the Kano model (Rotar and Kozar, 2017). Some requirements are articulated by the customer (stated needs) and others are taken for granted as implicit needs (Pfeifer and Schmitt, 2021). These would not be mentioned by the customer because they have become a matter of course. Requirements can be classified into three different categories: Basic, Performance and Excitement (Figure 1).

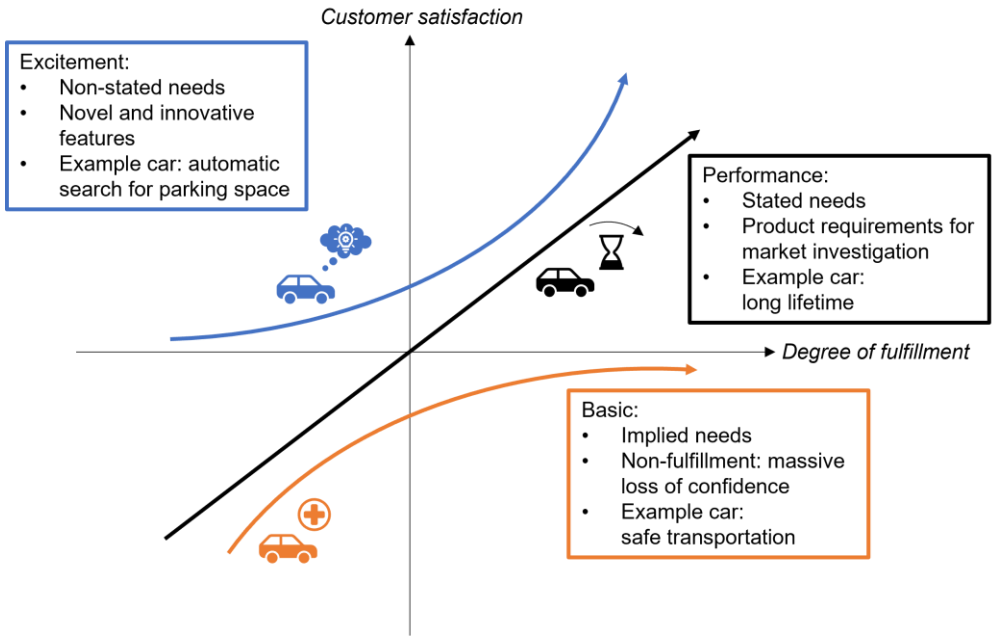

Figure 1. Kano model

Depending on the fulfilment of requirements from the categories mentioned, a different degree of customer satisfaction will be achieved (Pfeifer and Schmitt, 2021). The fulfilment of basic requirements would not satisfy the customers. However, non-fulfilment leads to a massive loss of confidence. Requirements which belong to Performance are explicitly expressed and compared to other products. The excitement requirements are not expected by customers and therefore represent a delight factor. In comparison, there will be no feeling of dissatisfaction if these are not fulfilled (Sauerwein et al., 1996).

3 Methodology

To enable researchers to recognize the added value of RDM-systems and use them sustainably, user requirements must be met depending on the research area. This paper shows a methodological approach to derive and categorize user requirements.

Currently, there are no open modular adaptive solutions for data management systems in collaborative projects. Therefore, for both CRCs, a solution has been developed that meets the projects' request and requirements (Mozgova et al., 2020). With the help of the Kano model products can be designed in a more customer-centred way as stated in 2.2. By applying the Kano model we aim to categorize the different requirements that were obtained during the initial project phase and during subsequent interactions with the researchers. These requirements were complemented by interviews with the expert developers.

The basic requirements, as stated in 2.2, are requirements that are mandatory for a user of a specific system and rarely expressed directly. Users mostly anticipate a functionality or quality of a product which is common and from their point of view not necessarily expressible. These requirements are difficult to obtain by interviews or workshops from end-users. Therefore, we asked experienced developers, which basic requirements should be fulfilled by a RDM-system.

The same holds true for the excitement requirements. These are hardly expressed by the end-users since they are either not conscious about what is possible with a technology, or they deem a requirement impossible to be fulfilled. The result is the same for both reasons: excitement requirements are difficult to be obtained from end-users and have to be retrieved by other means. We interviewed the developers and asked them to imagine what features could be excitement requirements for end-users.

The performance requirements are typically stated by end-users and express their need and desire towards a specific product. Still, requirements engineering is a challenging discipline on its own to have precise and clear requirements, but among the three Kano-categories these are the easiest to obtain.

Finally, we created a survey that was issued to the end-users where we asked them about their satisfaction towards the RDM-system. The survey is used as an initial validation concerning requirements of all three categories as we asked questions about the end-users' satisfaction with the system which comprises a wholistic view. The research steps are presented in the following Chapter.

4 Results

This chapter presents the results of determined requirements according to the presented method and are shown in Figure 2. Moreover, an initial evaluation will be presented by conducting a first survey with researchers. The survey refers to selective requirements and the use of the RDM-system.

4.1 Determined Requirements

In a first step, the basic requirements are determined. For this we interviewed the developers of the systems and asked them which requirements are crucial for an RDM-system to be accepted. In general, it can be emphasized that IT safety and security must be considered. The systems should also be accessible and usable independently of the device. Within the data repository and the knowledge management system, it must be possible to generate protocols and upload datasets in accordance with the research processes. Without the ability to store research results in a structured way, the usefulness of a RDM-system is not fulfilled. Here it is important to avoid redundancies when entering data. It should also be possible to assign different authorizations to users and to search the systems or use filters to narrow down the information. In order to reuse research data, a RDM-system should make it possible to e.g., search for specific datasets or metadata.

The researchers were surveyed to derive the performance requirements. Several individual interviews were held to determine individual requirements in relation to the respective research areas. Not all wishes could be considered when implementing the systems. However, overarching requirements can be derived that affect all CRC researchers. It was expressed generally that the systems offer a user-friendly overview and trainings for better usability and understanding of the RDM-system.

Specifically for the knowledge management system, the following performance requirements were stated. Each sub-project carries out different procedures and experiments. It must therefore be possible to adapt the protocols. Furthermore, it should be possible to add remarks. Another point mentioned is the compliance with standardized entries. A given example is the list of materials. By using a standardized international number which is unique for every material a misunderstanding concerning the material can be reduced. Another requirement is to be informed via e-mail when other researchers add new information to the protocol.

Specifically for the data repository, the following performance requirements were stated. When datasets are uploaded, it should be possible to link them to the information from the knowledge management system. In addition, the possibility of collaborative work was requested if several subprojects work together. Therefore, editing authorizations must be flexibly assignable by admins which refers to the basic requirements as a precondition. For the reuse of research results, a further requirement is that stored datasets can be pushed into a public repository to reference them in a publication.

The process for determining the excitement requirements was similar to the basic needs. With the help of the expert opinion of our developers, the following excitement requirements could be identified. The use of Artificial Intelligence (AI) represents a great enrichment in terms of gathering knowledge. It can be used for an optimized search and automatic analysis. In this meaning the use of AI can be interpreted as a proactive assistant with its help researchers will get support when using the RDM-system. A consistency check and an automatic content check of entries are further excitement requirements. In this context, users receive automatic notifications from the system if, for example, protocols are not completed or assigned parameters do not match the machine. Further excitement factors are the possibility of integrating additional systems which are automatically synchronized. Given the example of the CRC 1368 electronic lab notebooks represent this case.

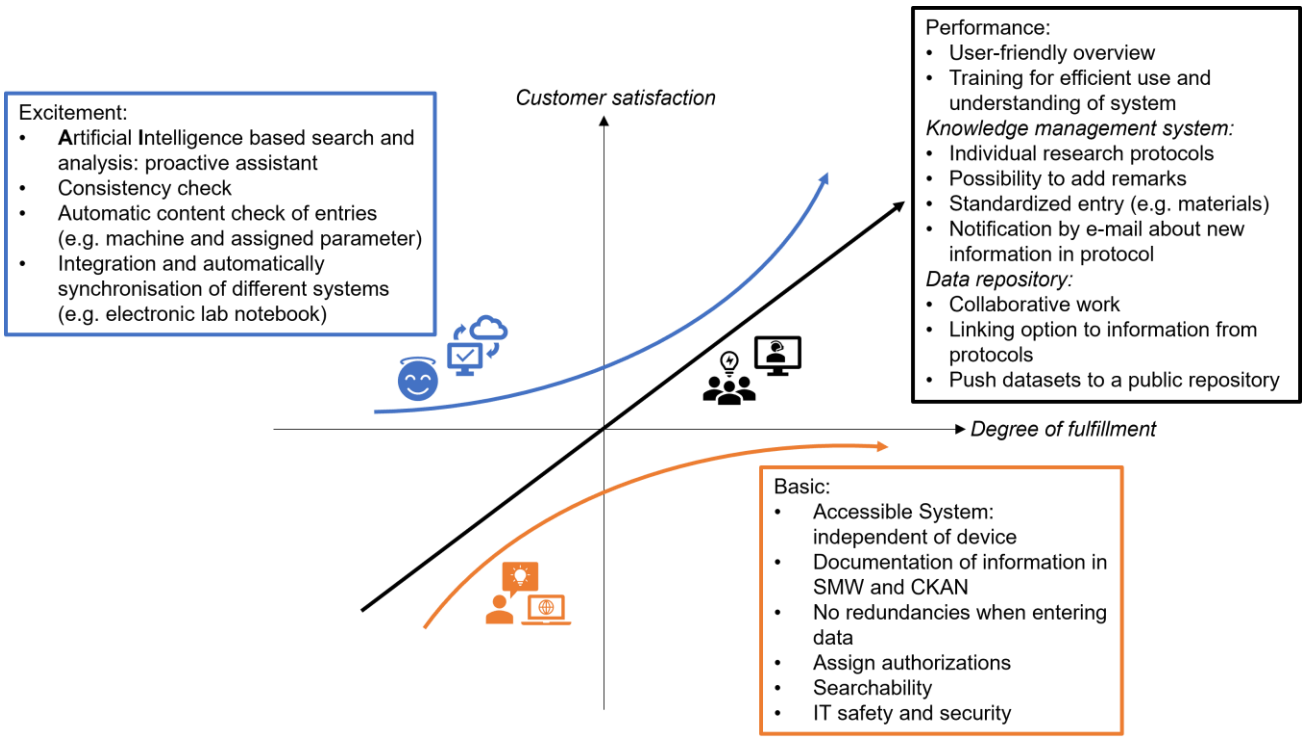

Figure 2. Requirements mapped with Kano model

4.2 Survey

An initial survey was conducted so that a first check of the identified requirements could be carried out. It was realised within the CRC 1368 and a total of 18 researchers of the subprojects participated. This target group represents those who use the RDM-system for their research activities. The purpose of the survey was to gain an initial insight into the views of the researchers. Some of them use the RDM-system more than others which might be only 9 Persons have participated completely in the meaning of answering all questions. Only their responses are taken into account to make the answers comparable and analysable. The main questions concerning the RDM-system, the different categories for answers and the percentage response can be seen in Table 1. Further questions such as the belonging to subprojects are not included in this paper as they do not contribute to a deeper understanding.

Question 1) to 4) refer to the basic requirements. If the handling is simple, it implies that the systems basically work and research results can be structured, stored, and reused. Regarding the first question 88,89 % consider uploading datasets in CKAN easy. Concerning the degree of simplicity when generating a protocol in SMW given in question 2) the majority states a neutral handling at 33,33 %. Question 3) and 4) involve the concrete degree of satisfaction in relation to two common use cases of the RDM-system. The process of entering machine information in the SMW is rated as satisfying to very satisfying at 66,67 % and the process of linking datasets with publications is rated as satisfying to very satisfying at 77,78 %. In summary, the basic functions of the knowledge management system and the data repository are rated as satisfied to very satisfied by the majority on the basis of individual examples. Thus, some derived basic requirements can be confirmed.

Question 5) and 6) regard to the performance requirements. Participants were asked if they would like more training or tutorials to help them use the RDM-system more effectively. 55,55 % state yes including maybe and definitely. When asked specifically which format they would like for such training courses or tutorials, 33,33 % prefer online videos and 11,11 % webinars. No other formats were mentioned by the participants in an additional comment field. Regarding to Figure 2 one performance requirement is confirmed in the meaning of establishing more trainings for better usability and understanding of RDM-system.

Question 7) refers to excitement requirements which delight end-users. This is a free-text question in which the participants were not given an answer option. They were asked which additional functions or tools should be added to the RDM-systems. There are only two responses: firstly, more user-friendly and clearly arranged systems and secondly integration of other systems. Last mentioned overlaps with the requirements derived from the developers so that this requirement is confirmed. User-friendly and clearly arranged systems are expressed as a new excitement requirement.

Question 8) generally refers to the understanding of RDM-systems. It is asked if the researchers see added value in using the RDM-system for research activities and those of the CRC as whole. The answers here are very scattered. 44,44 % state

yes but it is divided into three categories: only CRC, only research activities and for both. The answer hardly and no is rated at 55,55 %. This result emphasizes the general need for training with regard to the actual purpose and usefulness of RDM-systems in engineering.

Table 1. Survey within CRC concerning RDM-system

Question	**1) How easy is it for you to upload datasets to CKAN?**					
Answer	Very easy	easy	neutral	difficult	Very difficult	
Percent	0 %	55,56 %	33,33 %	0 %	11,11 %	

Question	**2) How easy is it for you to generate protocols in SMW?**					
Answer	Very easy	easy	neutral	difficult	Very difficult	Never used it before
Percent	11,11 %	11,11 %	33,33 %	22,22 %	0 %	22,22 %

Question	**3) How do you evaluate the process of entering machine information in the SMW?**					
Answer	Very satisfying	satisfying	neutral	unsatisfying	Very unsatisfying	
Percent	11,11 %	55,56 %	22,22 %	11,11 %	0 %	

Question	**4) How do you evaluate the process of linking datasets with publications?**					
Answer	Very satisfying	satisfying	neutral	unsatisfying	Very unsatisfying	Never used it before
Percent	11,11 %	66,67 %	11,11 %	0 %	0 %	11,11 %

Question	**5) Would you like more training or tutorials to help you make better use of the RDM-system?**					
Answer	Yes, definitely	Yes, maybe	Neutral	No, not necessary	No answer	
Percent	22,22 %	33,33 %	11,11 %	0 %	33,33 %	

Question	**6) If yes, which format would you prefer for such training courses or tutorials?**					
Answer	Online videos	Written instructions	webinars	Presence workshop	other	No answer
Percent	33,33 %	0 %	11,11 %	0 %	11,11 %	44,44 %

Question	**7) Which additional functions or tools would you like to add to the RDM-system?**	
Answer	More user-friendly and clearly arranged	Integration of other systems

Question	**8) Do you see added value in using the RDM-system for your research activities and those of the CRC as a whole?**					
Answer	Yes, for my research activities and for the CRC	Yes, only for CRC	Yes, only for my research activities	hardly	no	
Percent	11,11 %	22,22 %	11,11 %	22,22 %	33,33 %	

5 Conclusion and Outlook

The integration of RDM-systems for documentation of research results in engineering context is categorised as a novelty in many areas. Therefore, its implementation is accompanied by a number of challenges. The research results in engineering are characterised by very heterogeneous data, so that there are individual requirements for the use of RDM-systems depending on the field of application. Its usage is related to the fulfilment of end-user requirements. A possible concept for requirements identification and categorisation has been shown.

With the help of the Kano model, basic and excitement requirements were derived through interviews with the developers of the systems. Performance requirements were derived by interviewing the researchers. Subsequently, an initial survey was conducted within the CRC. Overall, some requirements could be confirmed and in one case supplemented. However, further research may be required in order to make more concrete statements. Nevertheless, the chosen method shows an initial derivation of requirements and how these can be categorised in relation to RDM-systems in the engineering sciences. In order for further requirements to be checked, more surveys must be carried out for a holistic validation. It is also planned to carry out these surveys in CRC 1153 in the future in order to analyse and compare the systems. A methodical evaluation of the requirements, for example by the House of Quality and User Experience Design, will follow as a further outlook. In this way, requirements can be evaluated and provide a decision-making aid which requirements should be implemented in order to increase the quality of research processes documentation with RDM-systems.

Acknowledgements

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - SFB 1153 – 252662854 and SFB 1368 – 94563137.

References

Altun, O., Hinterthaner, M., Barienti, K., Nürnberger, F., Lachmayer, R., Mozgova, I., Koepler, O., Auer, S., 2023. Contextualization for Generation FAIR Data: a Dynamic Model for Documenting Research Activities. International Conference on Product Lifecycle Management, IFIP, 20, Montreal. (in press)

Behrens, A., Uhe, J., 2021. Introduction to tailored forming. Production Engineering 15, pp. 133-136.

EN ISO 9001:2015. Quality managment systems – Requirements.

Fraga, A.L., Vegetti, M., Leone, H.P., 2020. Ontology-based solutions for interoperability among product lifecycle management systems: A systematic literature review. Journal of Industrial Information Integration 20, 100176. https://doi.org/10.1016/j.jii.2020.100176

Mozgova, I., Koepler, O., Kraft, A., Lachmayer, R., Auer, S., 2020. Research Data Management System for a large Collaborative Project. Balancing Innovation and operation: The Design Society, NordDesign. https://doi.org/10.35199/NORDDESIGN2020

Pfeifer, T., Schmitt, R., 2021. Masing Handbuch Qualitätsmanagement, volume 7. Carl Hanser, München.

Rotar, L.J., Kozar, M., 2017. The Use of the Kano Model to Enhance Customer Satisfaction. Organizacija, Kranj 50,4, 339-351.

Sauerwein, E., Bailom, F., Matzler, K., Hinterhuber, H.H., 1996. The Kano Model: How To Delight Your Customers. International Seminar on Production Economics 1, 4, pp. 313-327.

Sheveleva, T., Wawer, M.L., Oladazimi, P., Koepler, O., Nürnberger, F., Lachmayer, R., Auer, S., Mozgova, I., 2020. Creation of a Knowledge Space by Semantically Linking Data Repository and Knowledge Management System – a Use Case from Production Engineering. IFAC-PapersOnLine 55, 10, 2030-2035.

Stark, J., 2022. Product Lifecycle Management (Volume 1). Springer Nature Switzerland AG, Cham. https://doi.org/10.1007/978-3-030-98578-3

Sudarsan, R., Fenves, S.J., Siriam, R.D., Wang, F., 2005. A product information modeling framework for product lifecycle management. Computer-Aided Design 37, 1399-1411. https://doi.org/10.1016/j.cad.2005.02.010

Tang, R., Hu, Z., 2019. Providing Research Data Management (RDM) Services in Libraries: Preparedness, Roles, Challenges, and Training for RDM Practice. Data and information management 3, 2, pp. 84-101.

Wegewitz, L., Maus-Friedrichs, W., Gustus R., Maier, H.J., Herbst, S., 2023. Oxygen-Free Production – From Vision to Application. Advanced engineering materials 25, 12, 2201819.

Wilkinson, M.D., Dumontier, M., Aalsber, I.J.J. et al., 2016. The FAIR Guiding Principles for scientific data management and stewardship. Scientific data, 3, 160018.

Wilms, K.L., Stieglitz, S., Ross, B., Meske, C., 2020. A value-based perspective on supporting and hindering factors for research data management. International Journal of Information Management 54, 102174.

Contact: Laura Mueller, Paderborn University, Data Management in Mechanical Engineering, Warburger Str. 100, 33098, Paderborn, Germany, +49 5251 60-2227, laura.mueller@upb.de