

Automated Documentation of Research Processes Using RDM

L. Griem^{1,2}, Richard Thelen³, and M. Selzer²

¹ Institute for Applied Materials – Microstructure Modelling and Simulation (IAM-MMS), Karlsruhe Institute of Technology (KIT), Germany

² Institute of Nanotechnology – Microstructure Simulation (INT-MSS), Karlsruhe Institute of Technology (KIT), Germany

³ Institute of Microstructure Technology (IMT), Karlsruhe Institute of Technology (KIT), Germany

Abstract. Published research results usually represent only a fraction of the data generated at a research institute. The unpublished data created in the process of producing the final result, however, often contain valuable information that can be reused. Through research data management, all these data should be stored centrally according to the FAIR principles (Findable, Accessible, Interoperable, Reusable). However, a significant part of knowledge is often not found in the data, but in the processes that led to their generation. It is therefore important to map these processes to archive and document this knowledge in a structured way. Procedures for documenting scientific processes already exist and are actively used at research institutes. However, these are often analogue or paper-based and hence do not meet the requirements for FAIR data management. At the Institute for Microstructure Technology of the KIT, such a paper-based procedure is used to document the production of microstructure components. During their manufacturing, it is essential to adhere to the correct process parameters in order to enable error-free production. Therefore, a so-called job ticket always accompanies the production of components. On this job ticket, the correct process sequence is listed and a detailed description of the respective process step is given. Depending on the component to be produced, a distinction is made between different types of job tickets according to internal conventions. On the one hand, there are so-called green job tickets, which describe a standardised process sequence, and on the other hand, blue job tickets, which are intended to document experimental manufacturing processes. The process sequence on the blue job tickets is initially empty and is filled in during the manufacturing process. Common to both types of job tickets is that they are stored in the institute's archive after completion of the component production. However, since the job tickets are paper-based, the corresponding archive of job tickets cannot be searched quickly and, given the sheer volume of archived job tickets, represents an unmanageable collection of data. The existing system for process documentation is therefore to be implemented with the help of the research data infrastructure Kadi4Mat [1] in accordance with FAIR principles, thereby making the available process knowledge more accessible.

1. Methods

The Kadi4Mat research data infrastructure provides a set of software modules that address different aspects of research data management. These include KadiWeb, which provides a repository for structured data storage, and KadiStudio, a workflow editor that enables reproducible documentation of research processes in the form of automated workflows.

In KadiWeb, research data is always stored in the form of records. These are containers in which research data is stored and provided with descriptive metadata. The design of the metadata is left to the users and can therefore be adapted to all possible use cases and metadata schemes. In order to ensure the FAIR storage of research data, it is necessary that uniform metadata schemas are used within research institutions. For this purpose, metadata templates exist that can be used when creating records and assigning defined metadata to them.

The workflow editor KadiStudio is based on the concept that every scientific process can be divided into a multitude of atomistic, not further divisible, simple work steps [2]. These steps are available in KadiStudio as executable functions and can be freely recombined to model a workflow using an intuitive graphical user interface. The library of available nodes can be extended by the user as desired, allowing KadiStudio to be continuously developed and adapted to specific use cases.

2. Results

The two modules described above, KadiWeb and KadiStudio, are used to model the job ticket documentation. KadiWeb is used to store and archive the job tickets in a machine-readable format, while KadiStudio is used to actually fill in the job ticket with work instructions.

In order to achieve uniform documentation of the job tickets, templates were first created on KadiWeb. Depending on the type of job ticket, green or blue, different approaches were taken. For green job tickets, templates were created in which all the individual steps of the manufacturing process were defined in the metadata as shown in Figure 1.

Typ	Value	Field Type	Action
Typ	null	String	✎
Nummer der Laufkarte	null	Integer	✎
Losnummer	null	String	✎
IFA-Nummer	null	String	✎
Erzeugnisnummer	null	String	✎
Auftraggeber	null	String	✎
Beistellteile, Masken etc.	null	String	✎
Fertigungsgegenstand	null	String	✎

Figure 1. Green Job Ticket Template in KadiWeb.

The blue job ticket template, only contains basic metadata and no processing steps. Instead, additional templates were created for all possible processing steps. These can be added to a blue job ticket to document the manufacturing process. All metadata entries were further

equipped with sanity checks to ensure that only meaningful data can be entered. To automate the documentation of job tickets using these templates, a workflow in KadiStudio was used, shown in Figure 2.

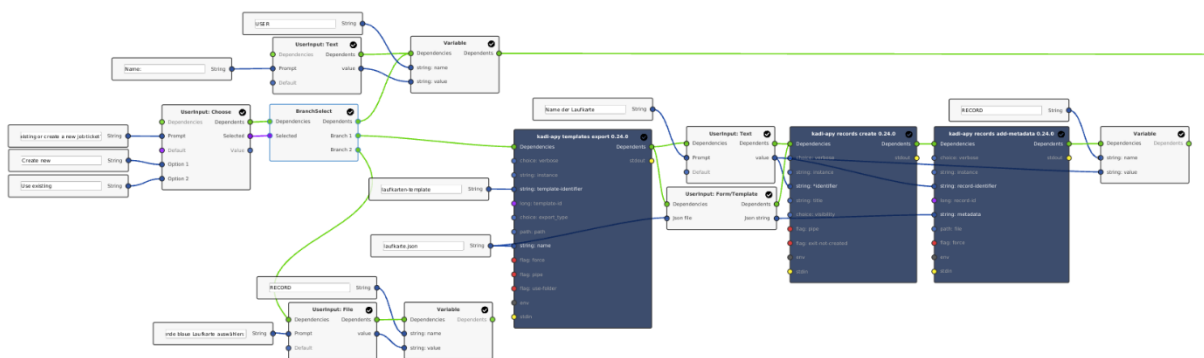


Figure 2. Created job ticket workflow in KadiStudio.

When executing this workflow, the user is first asked whether a green or blue job ticket is to be processed, and is then given the option of creating a new job ticket record in KadiWeb or continuing to work on an existing record. To do this, KadiStudio communicates with KadiWeb in the background and loads all available job ticket records and templates for the user to choose from. The selected job ticket is then used for further processing. In the case of a green job ticket, the next processing step is automatically retrieved from the template. For blue job tickets, the user can interactively select the next step from all available working step templates. In both cases, KadiStudio creates an input mask shown in Figure 3 from the corresponding template and prompts the user to fill it in. Once the mask has been filled in, the metadata is added to the record of the job card in KadiWeb. In this way, the documentation of job tickets is completely digitalised and automated.

Figure 3. User prompt generated from the metadata templates.

3. Conclusion

Integrating the job ticket documentation into the research data infrastructure Kadi4Mat offers many benefits. The job tickets can be quickly searched for specific parameters, equipment or processing. In addition, the user is interactively guided through the manufacturing process in the workflows. This makes it easier for new employees to get started and ensures quality assurance with the help of sanity checks.

Competing interests

The authors declare no competing interests.

Funding

This work was carried out with the support of the Joint Laboratory Model and Data-driven Materials Characterization (JL MDMC), a cross-centre platform of the Helmholtz Association. The authors would also like to thank the Federal Government and the Heads of Government of the Länder, as well as the Joint Science Conference (GWK), for their funding and support within the framework of the NFDI4Ing consortium. Funded by the German Research Foundation (DFG) - project number 442146713. This work was partly carried out with the support of the Karlsruhe Nano Micro Facility (KNMFi, www.knmf.kit.edu), a Helmholtz Research Infrastructure at Karlsruhe Institute of Technology (KIT, www.kit.edu).

Acknowledgement

The authors would like to thank the organizers of the CoRDI conference for providing the possibility to present their research.

References

1. Brandt, Nico et al. Data Science Journal, vol. 20, 2021, p. 8.
2. Griem, L, et al. Data Science Journal, vol. 21, 2022, p. 16.