

Bausteine Forschungsdatenmanagement
Empfehlungen und Erfahrungsberichte für die Praxis von
Forschungsdatenmanager*innen

Boosting data value

Part 1: Data publication practices in the institutional repository
TUDOdata

Patrick Kibiesⁱ Wibke Kleinaⁱⁱ Olaf Kletkeⁱⁱⁱ
Lars Schumann^{iv} Julia Stapels^v Bernd Zey^{vi}

2026

Zitiervorschlag

Kibies, Patrick, Kleina, Wibke, Kletke, Olaf, Schumann, Lars, Stapels, Julia and Zey, Bernd.
2026. Boosting data value. Part 1: Data publication practices in the institutional repository
TUDOdata. *Bausteine Forschungsdatenmanagement. Empfehlungen und Erfahrungsberichte
für die Praxis von Forschungsdatenmanager*innen* Nr. 1/2026: S. 2-10. DOI:
[10.17192/bfdm.2026.1.8994](https://doi.org/10.17192/bfdm.2026.1.8994).

Dieser Beitrag steht unter einer
[Creative Commons Namensnennung 4.0 International Lizenz \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/).

ⁱTechnische Universität Dortmund, IT & Medien Centrum und Fakultät für Chemie und
Chemische Biologie. ORCID: [0000-0002-9358-0600](https://orcid.org/0000-0002-9358-0600)

ⁱⁱTechnische Universität Dortmund, Universitätsbibliothek. ORCID: [0000-0003-2978-9909](https://orcid.org/0000-0003-2978-9909)

ⁱⁱⁱTechnische Universität Dortmund, Universitätsbibliothek. ORCID: [0000-0002-7825-1278](https://orcid.org/0000-0002-7825-1278)

^{iv}Ruhr-Universität Bochum, Cluster of Excellence RESOLV, ZEMOS. ORCID: [0009-0006-0664-3885](https://orcid.org/0009-0006-0664-3885)

^vTechnische Universität Dortmund, Universitätsbibliothek. ORCID: [0000-0003-1347-9607](https://orcid.org/0000-0003-1347-9607)

^{vi}Technische Universität Dortmund, Universitätsbibliothek. ORCID: [0000-0003-2551-9653](https://orcid.org/0000-0003-2551-9653)

Abstract

In mediating the continuous successful integration of research data management (RDM) practices and data publications into scientists' occupational routine, institutions face the challenge to transfer abstract RDM concepts and practices into concrete infrastructures, processes, and support offers. As data publications are becoming increasingly relevant, institutional RDM efforts meet a growing interest to publish research data as proof of scientific merit.

This work is a practice report on how TU Dortmund University established the publication process with data curation as core element for its institutional research data repository "TUDOdata". The devised publication process is discussed with respect to the usefulness for researchers and the institutional resources required. Research data management professionals are provided with practical insights based on experiences made during conceptualization and implementation of the publication procedure in an institutionally hosted repository utilizing the software Dataverse.

1 Introduction: Curation of research data as a means of ensuring data quality

As research data publications become more common, data quality and rich meta-data increasingly gain importance (Azeroual, 2021). However, across disciplines, most researchers find it difficult to understand and work with data created by other researchers and even with their own after a time (Baker, 2016). Possible solutions for this are detailed descriptions, documentation, and metadata that are provided with the data, so that no prior knowledge about the respective dataset is necessary in order to be able to work with it. Descriptive metadata ensure the FAIR (Findable, Accessible, Interoperable, Re-usable) principles (Wilkinson et al., 2016) not only for using foreign research data, but even to find and (re-)use one's own data. The application of FAIR principles offers a structured framework for improving data quality, particularly by enhancing representational and accessibility-related aspects. Integrating these principles into the management practices of data repositories serves as an effective procedural mechanism to foster systematic improvements (RfII – Rat für Informationsinfrastrukturen, 2019). Furthermore, adherence to globally recognized standards increases the visibility of datasets and strengthens their perceived reliability and trustworthiness among users. Consistent implementation of FAIR principles ensures that datasets achieve high scores in external assessments of data quality. The current standard approach of publishing data is to provide it as supplementary material alongside a text publication, which usually does not fulfill the criteria for data accessibility requested for works supported with funding by e.g., the Deutsche Forschungsgemeinschaft (DFG) (Deutsche Forschungsgemeinschaft, 2015) or the European Commission

(European Commission - Directorate-General for Research & Innovation, 2017). Individually published datasets are typically harder to find with limited reach, the datasets are documented insufficiently, and machine-readable file formats are chosen rarely. To overcome those issues a first step is to publishing research data in specialized repositories to facilitate metadata annotation. Since researchers usually are no experts in the field of RDM, the quality of data preparation and metadata annotation in non-curated open research data platforms tends to be poor – oftentimes metadata is missing and data annotation is very heterogeneous (Umbrich et al., 2015). In addition, datasets are unknowingly published under disadvantageous or inappropriate licenses. An effective measure to make data FAIR and to exploit its full potential is curation.

In this practice report, we present the publication and curation process introduced at TU Dortmund University on an institutional level, annotated with practical experiences, aiming to enhance data quality standards across all positions and disciplines of researchers.

2 Institutional integration of data publication procedures at TU Dortmund University

Providing a curated repository on an institutional level can be advantageous, albeit costly. Services and infrastructure can be offered to researchers without surcharge, servers and storage are maintained by the institution's central services allowing for a high degree of control and data sovereignty. Especially in politically uncertain and highly dynamic times, where scientific databases that are central for whole disciplines are politically forced to charge for their formerly free services or even prevent access by international collaborators, data that was once considered securely published and archived falls under an emerging risk of deaccessioning (Elbing, 2025). In this context, providing access to research data as a publicly-funded institution such as a university via e.g., a repository can reduce the risk of such data loss as publicly funded institutions are required to handle their infrastructure with responsibility towards society, thus increasing trust in such a platform.

Concerning ease-of-use, an institutional repository can facilitate the practice of data publication as a trusted platform flanked by dedicated support staff who offer tailored assistance and advice around the publication process. On the cost-side, providing a central service is expensive both in technology costs and administrative overhead as technical solutions need to be implemented on-site to ensure provision of data storage and long-term storage perspectives (Rothfritz et al., 2025).

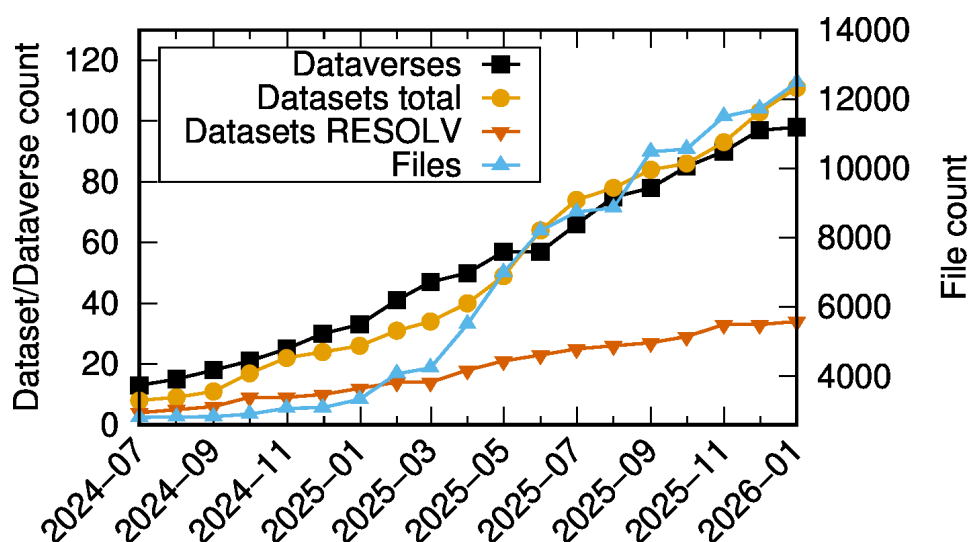


Figure 1: Count of published Datasets (persistently identifiable and addressable entities containing meta- and research data) and Dataverses (collections of Datasets, means of repository structuring) (left y-axis) and number of published Files (right y-axis) of the research data repository TUDodata plotted against time. Black: number of Dataverses, orange: number of Datasets, dark orange: number of Datasets published by the Cluster of Excellence RESOLV, blue: number of files.

2.1 Data publication in context of the institutional RDM support, exemplified in the service “TUDodata”

For TU Dortmund University, the scientific benefits of hosting an institutional repository outweighed the anticipated costs (Zey et al., 2025). Incidentally, hosting a federated research entity such as the cluster of excellence RESOLV (Ruhr Explores Solvation, a cluster investigating the field of solvation science, with most research groups having a background in chemistry)¹ has led to a high number of early adopters which aided greatly in developing an adequate curation process. The high research output of such a cluster (Figure 1) underlines the necessity to provide such a service and the pull effect successful data publication practices can have on the scientific community of an institution. As such, a formalized publication workflow was devised (Section 2.2) that offers close collaboration between researchers and curators, but at the same time runs along a pre-defined set of checklist questions that are evaluated for each dataset (Kibies et al., 2025). The checklist is designed to ensure the FAIR principles (spanning topics from rich metadata to open file types and permissive licensing) in each dataset by assisting researchers in creating a FAIR dataset from the beginning and allowing curators to employ a standardized review process.

¹<https://www.solvation.de/>

2.2 Curation as an integrated element of the publication process in TUDOdata

TUDOdata is faced with a broad variety of disciplines storing and publishing data. TU Dortmund University has 17 faculties spanning over natural sciences, engineering, social sciences and humanities. The repository is intended to publish datasets authored by individual researchers up to joint research collaborations. To account for this granularity, its structure replicates the university's organizational chart by being structured into faculties, research groups and consortia. Since its inception in July 2024, 111 individual datasets have been published, amounting to roughly 1.4 datasets per week as of 2026-01-30 (see Figure 1). Initially, the amount of Dataverses was rising comparably. As Dataverses are collections used to structure the repository itself, fewer new Dataverses than datasets are published once the structure of faculties/workgroups/projects is adequately depicted within the repository. The development shown suggests that the initial adoption phase is slowly developing towards routine publication practices.

In order to ensure high quality with respect to the FAIR principles (Wilkinson et al., 2016), a curation process was implemented as a quality management tool according to the Principles of RDM at TU Dortmund University ("Principles of Research Data Management at TU Dortmund University", 2019) and TU Dortmund University's rules for good scientific practice ("Rules of Good Scientific Practice at TU Dortmund University", 2019).

Currently, for all incoming submissions, TU Dortmund University allocates one person as a data curator and the cluster of excellence RESOLV dedicates one person to equivalent tasks for the cluster's own researchers and publications. However, both are also engaged in data stewardship, technical support and development, and training besides the curation activities.

To illustrate how data curation is part of the whole publishing process, we will give a short overview of the workflow from initial contact to a published dataset, which is compactly shown in Figure 2.

For first-time users, an initial login to TUDOdata is necessary to create a user account. Logging into the system is possible via institutional login of research facilities being member of the authentication and authorization infrastructure of the Deutsches Forschungsnetz (DFN-AAI), supplied via a Shibboleth Identity Provider (IdP), with an ORCID account, or for users not having access to either via local e-mail address-based user accounts.

After creating an account the user has the same level of permissions as a visitor without an account by default. Therefore, the identity provider is used for authentication only. Authorization, in terms of granting access permissions to unpublished content and to create as well as publish content, is a manual administrative task performed by the TUDOdata team. Because of this, users are required to contact the research

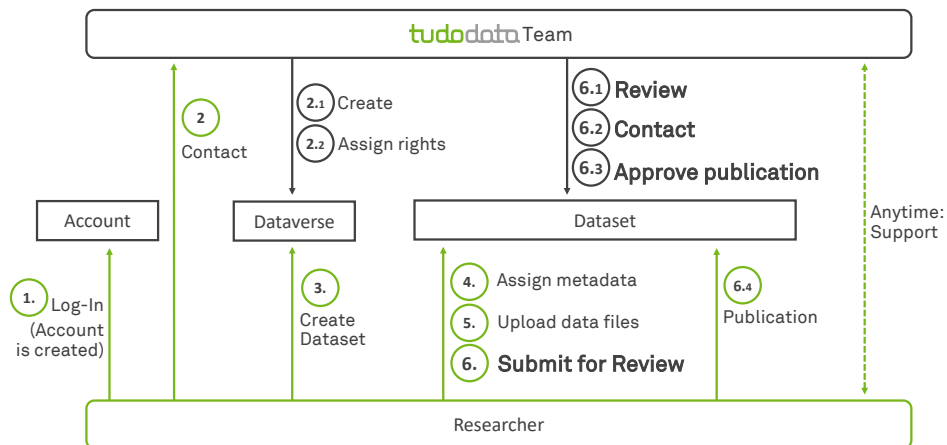


Figure 2: Graphical representation of the publication workflow in TUDodata. The time axis is from left to right and the numbers indicate the usual steps over time. The bottom party represents the researchers, the top party is the TUDodata team, thus arrows pointing upwards represent actions by researchers, while arrows pointing downwards are actions by the TUDodata team. Core steps of the curation process are highlighted in bold and larger font. The figure is a modification of the work licensed under CC-BY 4.0 by Bernd Zey, 2024, <https://doi.org/10.17877/TUDODATA-2024-LYPSTGCJ>

data service of TU Dortmund University. In the following communication between researchers and curators, necessary access and the placement of the planned data publication within the TUDodata structure will be discussed, adequate data collections (named “Dataverse” in the Dataverse software) will be created by the research data service and permissions to add datasets to the Dataverses are granted.

Communication can take place via e-mail or in a face-to-face meeting. Such a consultation is offered to all researchers and is usually accepted. During such a meeting, the publication workflow is discussed, the data is described (file types, folder structures, naming), the repository and its functions are looked into together, and, if necessary, the Dataverses are created. In this context, it is important to talk about the most important legal issues: does the dataset inadvertently contain personal data and are there any licensing issues (e.g., reuse of third-party data; software) to consider?

The before-mentioned structure of Dataverses depicts the organizational structure of the university, with top-level Dataverses for institutions like TU Dortmund University, but also for large institution-spanning research units like clusters of excellence or other joint collaboration initiatives, followed by the level of faculties, institutes, and research groups. An example of a collaborative research project is the cluster of excellence RESOLV, which was the first large-scale facility to obtain a top-level data collection named RESOLVdata with its own DOI provider and an independent research data management support structure. This structure, together with the fine granularity

of permissions in TUDOdata will allow for federated curation together with research data specialists on smaller-scale facilities distributing the curation load and integrating a higher degree of discipline-specific competence into the data management process.

In the process of preparing a dataset, the researchers are encouraged to always contact the RDM service personnel if they have any questions. Moreover, documents are available to support the researchers: (1) a handout and a tutorial for TUDOdata and RESOLVdata “How do I publish my data?” (Schumann & Zey, 2024; Zey, 2024) and (2) a website containing the most frequently asked questions and answers (Research Data Service of TU Dortmund University, 2025).

After the researchers prepared a dataset, they can submit it to review by the research data service through the user interface, starting the reviewing process which is formalized by a curation checklist. For organizing the curation process different labels are internally assigned to datasets (like “initial review in progress”, or “ready for publication”). A first review can usually take from a few hours to a few days, depending on how well the dataset is prepared before submission, dataset complexity, and curator availability. A detailed curation report is compiled and shared with the submitting party via e-mail. One or multiple iterations follow to reach publication readiness of the dataset. At any point during this process, additional assistance can be provided in a format that is most helpful for the researchers, either in a meeting or via e-mail. Depending on the number of iterations and response time, this dataset improvement process can take up to multiple weeks. Once the dataset is ready for publication, publication permission is granted to the user and will be revoked once the dataset is published, to ensure all future dataset versions undergo the curation process.

The publication by the user has two advantages: The exact timing of publication is decided by the dataset author(s), and the legal act of actually publishing a dataset remains their responsibility as well.

Successive versions of a published dataset undergo a shortened version of this curated process, focusing on introduced differences to the already-published version(s). Each modification to the dataset, metadata or data files, implies a new version while the old version remains in the system. Moreover, Dataverse offers the possibility to describe the differences in the metadata field “version notes”.

3 Outlook

Looking into the future, further professionalization and optimization of the institutional research data publication workflow can be mediated by investing into the curators’ professional skills, and by steady reevaluation of the workflow including user feedback.

RDM support on institutional scale also aims at training local data stewards to address discipline-specific topics and act as first-level support regarding RDM since they are thoroughly trained, and knowledgeable about RDM resources and services within the institution. As such, they can easily assist in preparing datasets and are familiar with the curation process during the publication procedure in the repository. Furthermore, federation of the research data repository is currently in progress for the University Alliance Ruhr, consisting of three universities, with potential for further expansion. This means a diversification in publication procedures, depending on the institution in question. At the same time, a convergence process of procedures towards best practices is anticipated, as increasing acceptance of a platform leads to standardization, professionalization and further allotment of resources. The example of TU-DOdata shows in practice how a curated institutional repository boosts data value by improving data quality, adhering to the FAIR principles and providing sustainable infrastructure.

Acknowledgements

We are grateful to Deutsche Forschungsgemeinschaft for funding LS and partially PK in context of the cluster of excellence EXC 2033 (RESOLV), (Projektnummer 390677874).

Contributions

PK: Conceptualization (equal), visualization (equal), writing - original draft preparation (supporting), writing - review & editing (equal)

WK: Conceptualization (supporting), writing - original draft preparation (supporting), writing - review & editing (supporting)

OK: Conceptualization (supporting), writing - review & editing (supporting)

LS: Conceptualization (equal), data curation (equal), methodology (equal), project administration (lead), visualization (equal), writing - original draft preparation (lead), writing - review & editing (equal)

JS: Conceptualization (supporting), writing - original draft preparation (supporting), writing - review & editing (supporting)

BZ: Conceptualization (equal), data curation (equal), methodology (equal), visualization (equal), writing - review & editing (equal)

References

- Azeroual, O. (2021). Datenqualität und -kuratierung als Voraussetzung für Open Research Data. *Information - Wissenschaft & Praxis*, 72(4), 204–211.
<https://doi.org/10.1515/iwp-2021-2158>

- Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature*, 533, 452–454.
<https://doi.org/https://doi.org/10.1038/533452a>
- Deutsche Forschungsgemeinschaft. (2015). DFG Guidelines on the Handling of Research Data. Retrieved September 10, 2025, from
<https://www.dfg.de/resource/blob/172098/4ababf7a149da4247d018931587d76d6/guidelines-research-data-data.pdf>
- Elbing, K. (2025). Informationsdienst Wissenschaft - Zugang zu Forschungsdaten in Gefahr. Retrieved June 6, 2025, from <https://idw-online.de/de/news852540>
- European Commission - Directorate-General for Research & Innovation. (2017). Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020. Retrieved September 3, 2025, from
https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf
- Kibies, P. J., Schumann, L. E., & Zey, B. (2025). Curation checklist for datasets published in TUDodata.
<https://doi.org/10.17877/TUDODATA-2025-MDCU78ZD>
- Principles of Research Data Management at TU Dortmund University. (2019). Retrieved April 16, 2025, from <https://www.tu-dortmund.de/en/research/research-data-management/principles-of-rdm>
- Research Data Service of TU Dortmund University. (2025). The data repository TUDodata. Retrieved July 1, 2025, from <https://fdm.tu-dortmund.de/en/rdm-at-tu-dortmund/technical-support/tudodata/>
- RfII – Rat für Informationsinfrastrukturen. (2019). Herausforderung Datenqualität – Empfehlungen zur Zukunftsfähigkeit von Forschung im digitalen Wandel.
<https://rfii.de/download/herausforderung-datenqualitaet-november-2019/>
- Rothfritz, L., Matthias, L., Pampel, H., & Wrzesinski, M. (2025). Current challenges and future directions for institutional repositories: A systematic literature review. An annual review of information science and technology (ARIST) paper. *Journal of the Association for Information Science and Technology*, 1–22.
<https://doi.org/10.1002/asi.70016>
- Rules of Good Scientific Practice at TU Dortmund University. (2019, December 12). Retrieved April 16, 2025, from
https://www.tu-dortmund.de/storages/tu_website/Referat_1/Dokumente___Ordnungen/Rules_of_Good_Scientific_Practice_EN.pdf
- Schumann, L. E., & Zey, B. (2024). How do I publish my data in RESOLVdata?
<https://doi.org/10.17877/RESOLV-2024-M18ZX3SP>
- Umbrich, J., Neumaier, S., & Polleres, A. (2015). Quality Assessment and Evolution of Open Data Portals. *2015 3rd International Conference on Future Internet of Things and Cloud*, 404–411. <https://doi.org/10.1109/FiCloud.2015.82>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). The FAIR guiding principles for

scientific data management and stewardship. *Scientific Data*, 3(1), 160018.
<https://doi.org/10.1038/sdata.2016.18>

Zey, B. (2024). Wie publiziere ich meine Daten in TUDOdata? / How do I publish my data in TUDOdata? <https://doi.org/10.17877/TUDODATA-2024-LYPSTGCJ>

Zey, B., Kibies, P. J., Schumann, L. E., Stapels, J., Kleina, W., & Kletke, O. (2025). TUDOdata: Bridging Institutional Standards and International Visibility in Research Data Management. <https://doi.org/10.5281/ZENODO.16735838>