

RESEARCH ARTICLE

# Lessons learned from implementing FAIRification workflows in diabetes research in Germany

Esther Thea Inau<sup>1\*</sup>, Angela Dedié<sup>2</sup>, Ivona Anastasova<sup>2</sup>, Renate Schick<sup>2</sup>, Brigitte Fröhlich<sup>2</sup>, Michael Roden<sup>3,4,5</sup>, Andreas L. Birkenfeld<sup>6,7,8</sup>, Martin Hrabě de Angelis<sup>2,9,10</sup>, Martin Preusse<sup>2</sup>, Dagmar Waltemath<sup>1</sup>, Atinkut Alamirrew Zeleke<sup>1,11</sup>

**1** Medical Informatics Laboratory, University Medicine Greifswald, Greifswald, Germany, **2** German Center for Diabetes Research (DZD), München, Neuherberg, Germany, **3** German Center for Diabetes Research (DZD), Düsseldorf, Germany, **4** Department of Endocrinology and Diabetology, Medical Faculty and University Hospital Düsseldorf, Heinrich-Heine-University Düsseldorf, Düsseldorf, Germany, **5** Institute for Clinical Diabetology, German Diabetes Center, Leibniz Center for Diabetes Research at Heinrich-Heine-University Düsseldorf, Düsseldorf, Germany, **6** German Center for Diabetes Research (DZD), Tübingen, Germany, **7** Institute for Diabetes Research and Metabolic Diseases of the Helmholtz Zentrum München at the University of Tübingen (IDM), Tübingen, Germany, **8** Department of Diabetology, Endocrinology, and Nephrology, University Clinic Tübingen, Eberhard Karls University Tübingen, Tübingen, Germany, **9** Institute of Experimental Genetics and German Mouse Clinic, Helmholtz Munich, Neuherberg, Germany, **10** Chair of Experimental Genetics, TUM School of Life Sciences (SoLS), Technische Universität München, Freising, Germany, **11** Department of Research and Publication Support, University and City Library, University of Cologne, Cologne, Germany

\* [esther.inau@stud.uni-greifswald.de](mailto:esther.inau@stud.uni-greifswald.de)



 OPEN ACCESS

**Citation:** Inau ET, Dedié A, Anastasova I, Schick R, Fröhlich B, Roden M, et al. (2026) Lessons learned from implementing FAIRification workflows in diabetes research in Germany. *PLOS Digit Health* 5(1): e0001139. <https://doi.org/10.1371/journal.pdig.0001139>

**Editor:** Louis Agha-Mir-Salim, Charité Universitätsmedizin Berlin Campus Charite Mitte: Charite Universitätsmedizin Berlin, GERMANY

**Received:** February 19, 2025

**Accepted:** December 1, 2025

**Published:** January 13, 2026

**Copyright:** © 2026 Inau et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data availability statement:** An archived record of the previous version of the DZD

## Abstract

The FAIR principles guide data stewardship towards maximizing the value of scientific data while offering a high level of flexibility to accommodate differences in standards and scientific practices. Research communities have developed and implemented domain-specific workflows to make their data FAIR. This work compares the implementation of two externally developed structured FAIRification workflows—a generic workflow and a domain-specific workflow—using the example of meta-data captured in diabetes research in Germany and applying the FAIR data maturity model developed by the Research Data Alliance. Interestingly, the implementation of both workflows required similar resources and led us to achieve the same FAIRness rating. We therefore conclude that the adaptations made in the FAIRification workflow for health research data improve efficiency but do not necessarily lead to higher FAIRness scores when applied to core data sets. Based on the results of our workflow comparison, we identified a list of requirements that should be met for the FAIRification of a core data set regardless of the workflow employed. In the future, FAIR data strategies and infrastructure should be planned and implemented as early as possible in the FAIRification journey. It is anticipated that this comparative analysis will help establish standard operating procedures for the FAIRification of core data sets for health studies.

CDS prior to FAIRification is available in Zenodo at:

<https://doi.org/10.5281/zenodo.12526690>.

The FAIRified version of the DZD CDS has been deposited in the MDM Portal and are available at:

<https://medical-data-models.org/46011>.

The related metadata and SOPs have been deposited in Zenodo and are available at:

<https://zenodo.org/record/7360000>.

**Funding:** This work was partially funded by the NFDI4Health – Nationale Forschungsdateninfrastruktur für personenbezogene Gesundheitsdaten (DFG-funded project 442326535).

**Competing interests:** The authors have declared that no competing interests exist.

## Author summary

Implementation of the FAIR principles in research data management improves the yield of scientific data across diverse research contexts. This study compared two structured FAIRification workflows—one generic and one specific to health research—using diabetes research data from Germany. Both workflows required similar effort and resulted in the same FAIRness score. This suggests that while domain-specific adjustments can improve efficiency, they do not necessarily increase FAIRness for core data sets. The authors outline key requirements for FAIRifying core health data and emphasize that FAIR data planning should begin early. Their findings aim to support the development of standard procedures for FAIRifying core data sets in health research.

## Introduction

The FAIR principles guide data stewardship towards improved findability, accessibility, interoperability, and reusability (FAIRness) of research objects [1,2]. These principles systematically usher research data management (RDM) towards maximizing the value of scientific data and reproducibility of research findings [3]. FAIR RDM strengthens data sharing among heterogeneous data silos and enables collaborative research [4–8].

FAIRification is described as the process of making research objects FAIR [9]. It has been implemented in different sectors of biomedical research from computational biology where the results have helped to improve the quality of modelling pipelines [10] to clinical epidemiology where the results have enabled the combination of metadata on epidemiological studies in Germany [11].

FAIRification typically starts with a FAIR assessment in which a tool is used to determine how FAIR the research object is using a structured assessment [12,13]. The results of this assessment are used to inform measures that ought to be taken for the research object to become FAIRer. After these measures have been applied, a second assessment is conducted to determine their impact. A higher FAIR score is often the expected outcome of FAIRification [6,12,14]. The FAIRification journey, which is often embarked on to fulfil requirements set by funders or research institutions, has been described as intensive [5,6]. Therefore, work has been done to prescribe simpler steps to improve FAIRness [12,15]. In this context, FAIRification workflows promise to provide the needed guidance on how to implement the FAIR data principles in a practical and gradual manner [6,16,17].

The Research Data Alliance (RDA) has collected maturity indicators and determined evaluation levels that together serve as the assessment criteria for FAIRness [14]. This common set of assessment criteria has been used in FAIR assessments across different scientific communities and has also shown to be useful for other tasks such as the calibration of reporting guidelines for machine learning models in health research [18–22]. The RDA further incorporated these assessment criteria into the FAIR Data Maturity Model (FDMM), which enables comprehensive manual FAIR assessment and guides FAIRification [19,23]. The FDMM evaluates both data and

metadata compliance with each FAIR data principle through one or more indicators [19]. Each indicator is associated with (often domain-dependent) impact levels described as essential, important, or useful [14]. Communities have developed standardised FAIRification templates and workflows for their respective research domains [2,5,24–27]. For example, Jacobsen et al. developed a generic FAIRification workflow that has been instructive for the FAIRification of electronically captured data on vascular anomalies and COVID-19 [16,28–30]. To accommodate differences in contexts and scientific practices, Sinaci et al. proposed a FAIRification workflow which applies restrictions considerate of the technical, ethical and legal requirements specific to health research data [17]. This workflow has shown promise in improving health research management outcomes in terms of time and costs [4].

The German Center for Diabetes Research (known locally as Deutsches Zentrum für Diabetesforschung - DZD) is a federal government and state-funded national association that brings together experts to conduct translational research across the full spectrum of diabetes and metabolism [31]. In 2021 the DZD established a minimum data set known as the DZD CORE DATA SET (CDS), which is geared toward research projects in the fields of diabetes and metabolism research [32]. The DZD CDS enables the harmonization of data items, labels, definitions, and documentation across all the DZD clinical studies, which further enables data comparability between related studies. Insights gained from our earlier retrospective FAIRification efforts informed the decision to implement structured, established FAIRification workflows in the FAIRification of the DZD CDS [12,33]. The enhanced data sharing that has been facilitated by the FAIRification of the DZD CDS is expected to contribute to its increased uptake among relevant stakeholders within the DZD and the wider diabetes research community [12]. The objective of this work is to explore the retrospective implementation of a generic [16] and a domain-specific [17] FAIRification workflow using the DZD CDS as a case study. While previous work detailed the FAIRification of the DZD CDS [12], the novelty of the present work lies in its use of established FAIR infrastructure to uniquely compare the implementation of these two externally developed FAIRification workflows in the DZD CDS. This work further distills a set of minimum requirements for successful FAIRification, derived from the practical insights gained through this comparative implementation. Stakeholders considering the implementation of FAIRification workflows may use the experiences and results obtained from this work to inform the development of FAIRification standard operating procedures (SOPs) for their research domains.

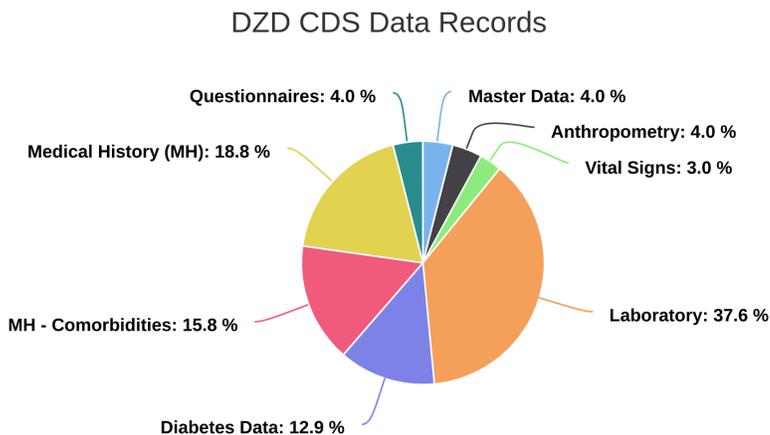
## Methods

### Data set

The DZD CDS is a mandatory component for the design of all upcoming DZD clinical studies and it has been implemented in various DZD studies since its conception [34–36]. It also includes the common CDS that was established among the German Centers for Health Research (*Deutsche Zentren der Gesundheitsforschung*) [37]. It contains 147 data items that have been selected by a group of medical experts and categorized into eight modules. It also contains optional modules relevant for special studies as shown in the following Fig 1.

### Analysis setup

A structured comparison was performed by retrospectively implementing the generic FAIRification workflow and the FAIRification workflow for health research in the context of the DZD CDS [16,17]. The decision to employ the FDMM as an evaluation instrument for this work was arrived at after a series of tools were tried, tested and eliminated for various reasons such as poor interpretation of the FAIR data principles, poor FAIRness assessment methods, and poor user friendliness [38]. We also considered the success reported by other researchers who used this tool for their own FAIR assessments [14,39–42]. The FDMM allowed us to discard the data indicators and instead focus on the metadata indicators (26 out of 41 indicators) as is necessary for the FAIR assessment of a CDS. Two data curators (ETI and AD) collaboratively assessed the DZD CDS FAIRness based on the publicly available guidance provided for this tool [19]. The curators assessed the DZD CDS FAIRness in three iterations for surety. Prior to this effort, the clinical relevance of data items and



**Fig 1. DZD CDS data records (base set).** Included modules: master data (contains patient information), anthropometry (contains patients' height, weight, waist & hip circumference, as well as the techniques used to obtain these measurements), vital signs (contains blood pressure and heart rate), laboratory (contains selected laboratory blood, plasma, serum and urine tests), diabetes data (contains date and type of diabetes diagnosis as well as treatment offered), medical history (contains a record of a patient's health background and the occurrence of disease events), medical history- comorbidities (indicates the presence of co-existing diseases with reference to the diagnosis) and questionnaires (contains the Baecke Index for sports, leisure, work and the total score) [12].

<https://doi.org/10.1371/journal.pdig.0001139.g001>

their alignment with study design requirements was validated by the stakeholder clinicians and study designers. In our assessment of the CDS we only used the FDMM metadata indicators shown in [S1 Table](#) (supporting information).

### Selection of FAIRification workflows

The results of our previous scoping review of FAIRification workflows [5] led us to select the workflows developed by Jacobsen et al. and Sinaci et al. for the FAIRification of the DZD CDS [16,17]. Both workflows have already been successfully implemented in various health research settings [4,28,29].

### Selection of a FAIR metadata repository

We included a structured selection process for a FAIR metadata repository in our FAIRification process prior to the implementation of the workflows. More specifically, we embarked on a search for a FAIR repository in which we would house the DZD CDS as part of the FAIRification. The Swiss National Science Foundation (SNSF) released a checklist for the selection of FAIR repositories [43].

Using the criteria from the SNSF checklist, we evaluated the Medical Data Models (MDM) Portal against the SNSF checklist for FAIR repositories [44,45]. [Table 1](#) shows the results of this evaluation.

Based on the results of this evaluation, we identified the MDM Portal as a FAIR repository employable in the FAIRification of the DZD CDS.

## Results

### Implementation of FAIRification workflows

**1. Generic FAIRification workflow** The generic FAIRification workflow developed by Jacobsen et al. consists of 7 steps which are categorised into the pre-FAIRification phase, the FAIRification phase and the post-FAIRification phase [16].

**Table 1. Adherence of the German Portal for Medical Data Models (MDM Portal) to the SNSF's Criteria for FAIR Repositories [43].**

Repository criteria	MDM portal
Is the repository non-commercial?	The MDM Portal serves as an infrastructure for academic (non-commercial) medical research.
Are datasets (or ideally single files in a dataset) given globally unique identifiers (IDs) and persistent IDs?	The MDM Portal offers digital object identifiers (DOIs) as globally unique and persistent IDs which can be assigned after users' request. The MDM Portal also provides guidelines for its citation in publications, presentations and registration of studies in study registries such as ClinicalTrials.gov [46].
Does the repository allow the upload of intrinsic and submitter-defined metadata?	Yes, basic intrinsic data has to be entered in specific mandatory fields. "Free text" fields are also available for the user to enter additional information.
Is it clear under which licence the data will be available, or can the user upload/choose a licence?	The user may upload the data models under one of four different Creative Commons Licenses including [47]: <ul style="list-style-type: none"> <li>• Creative Commons BY-NC 4.0</li> <li>• Creative Commons BY-NC-SA 4.0</li> <li>• Creative Commons BY-SA 4.0</li> <li>• Creative Commons BY 4.0</li> </ul> The terms and conditions of these licenses have been stipulated in the portal for the submitter's reference.
Are the citation information and metadata always (even in the case of datasets with restricted access) publicly accessible?	Yes
Does the repository provide a submission form requesting intrinsic metadata in a specific format (to ensure machine readability)?	Yes, basic intrinsic metadata has to be entered in a structured way (online form) in required fields. Checks are performed on some fields to ensure that the proper format is used (for example, email address and date).
Is there a long term preservation plan for the archived data?	Yes

<https://doi.org/10.1371/journal.pdig.0001139.t001>

**Identifying the FAIRification objectives:** We conducted a baseline FAIR assessment on the DZD CDS and then considered the results of this assessment (shown in Fig 2), along with the DZD CDS context, stakeholders' priorities and the desired scientific value of the data to formulate the objectives of the DZD CDS FAIRification as follows:

- Findability: To improve the searchability and findability of the DZD CDS items for users and across future DZD CDS versions
- Accessibility: To maintain the DZD CDS items in a manner that allows for them to be accessed under well-defined access conditions
- Interoperability: To annotate the DZD CDS items with biomedical ontologies, data standards, terminologies and a structured format so as to facilitate interoperability and automatic extraction of relevant data items across different future DZD CDS versions
- Reusability: To represent the data in a concise manner that allows for reuse of the data collected across all the different future DZD CDS versions

**Analysis of the data and metadata:** The data fields, types, and values were characterized. The data elements and data fields were extracted and the curated data was validated by the clinical data experts. We identified missing metadata with regard to temporal and spatial factors as well as contact persons, keywords and information on the target group.

**Defining the semantic data and metadata model:** We used the version of the DZD CDS provided by the MDM Portal which offers a bottom-up standardization process that facilitates the semantic enrichment of all the data items with codes from LOINC, SNOMED CT and UMLS [45,48,49]. All previous versions of the DZD CDS are listed under <https://medical-data-models.org/46011>.

**Making data and metadata linkable:** Contextual knowledge (persistent IDs and references to other data sets/publications) was added to the dataset in the form of meaningful links. The metadata and SOP were registered and hosted in Zenodo and the related Zenodo link was also added to the dataset [34,50].

**Hosting FAIR data:** The DZD CDS has been hosted on the MDM Portal for purposes of making it a community resource available for human and machine consumption. Hosting the DZD CDS on the MDM Portal allows downloading and exporting the file in most common technical formats. Hosting the CDS on the MDM Portal also required us to assign it a licence that stipulates its reuse. In all DZD multicentred clinical trials, the direct identifying data (IDAT) are handled spatially and organisationally separated from the medical data (MDAT) to comply with legal, organisational and technical requirements regarding data protection [51]. For this reason the IDAT is not part of the CDS. This, along with the fact that the DZD CDS does not contain personal data, influenced the decision to choose an open machine-readable licence (the Creative Commons BY-NC-SA 4.0) [52]. Hosting the DZD CDS on the MDM Portal also led to data versioning and indexing.

**Assessing FAIR data:** Finally, we determined if the defined objectives have been achieved and conducted a final FAIR assessment. The final FAIR assessment results of the DZD CDS are shown in Fig 2.

S2 Table (supporting information) illustrates the implementation of the generic FAIRification workflow in the DZD CDS and the decisions made in adapting it to this context.

## FAIR assessments

We assessed the FAIRness of the DZD CDS before (left diagram) and after (right diagram) application of the FAIRification workflows using the FDMM. The results are shown in the following Fig 2.

We observe that all FAIR indicators recorded significant improvement after the implementation of the FAIRification workflows. The implementation of both workflows led us to achieve the same FAIRness score. The DZD CDS contains neither patient identification data nor sensitive data. For this reason, the RDA-A1-01M and RDA-F3-01M indicators (defined in S1 Table, supporting information) were not applicable [14].

**2. Domain-specific FAIRification workflow** The FAIRification workflow specific to health research developed by Sinaci et al. consists of 10 steps [17].

**Analysing the data and metadata:** The implementation of this workflow began with raw (meta)data analysis similar to the one performed in the implementation of the generic FAIRification workflow.

**Curating and validating the data:** This step was performed as already indicated in the implementation of the generic FAIRification workflow.

**De-identifying and pseudonymising data:** We skipped the pseudonymization and de-identification step of this workflow because the DZD CDS does not contain any sensitive patient data.

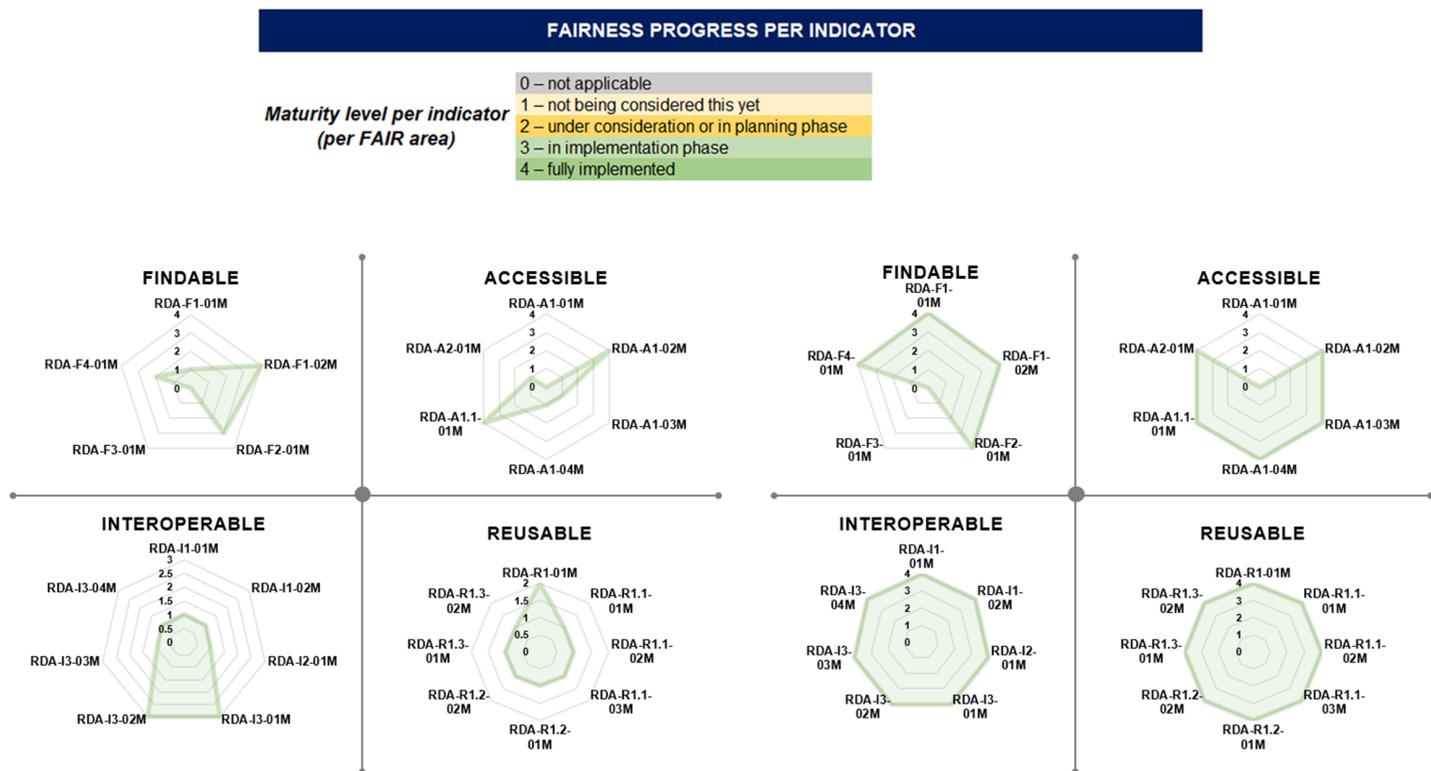
**Semantic modelling:** This step was performed as already indicated in the step that calls for "defining of the semantic data and metadata model" in the generic FAIRification workflow.

**Making data and metadata linkable:** The data was then enriched by adding contextual knowledge in the form of meaningful links as performed in the implementation of the generic workflow.

**Attributing a licence, data versioning and indexing:** These steps were already performed in the implementation of the generic workflow when the DZD CDS was registered in the MDM Portal. The same was done for the metadata and SOP in Zenodo where they were registered as shown in the implementation of the generic workflow [50].

**Aggregating the metadata:** The metadata was then aggregated, leading to the provision of the DZD CDS readme file and provenance information that contains the data origin, citations for reused data, description of the data collection, data processing history and version history.

**Publishing:** The DZD CDS was then published in the MDM Portal for consumption by the audience.



**Fig 2.** The diagram on the left illustrates the results of the baseline DZD CDS FAIR assessment while the diagram on the right illustrates the results of the final DZD CDS FAIR assessment. The indicator maturity levels are defined in the legend above the radar charts. This legend is used to indicate the FAIRness progress per indicator [12].

<https://doi.org/10.1371/journal.pdig.0001139.g002>

**S3 Table** (supporting information) illustrates the implementation of the health research FAIRification workflow in the DZD CDS and the decisions made in adapting it to this context.

### Similarities in workflows

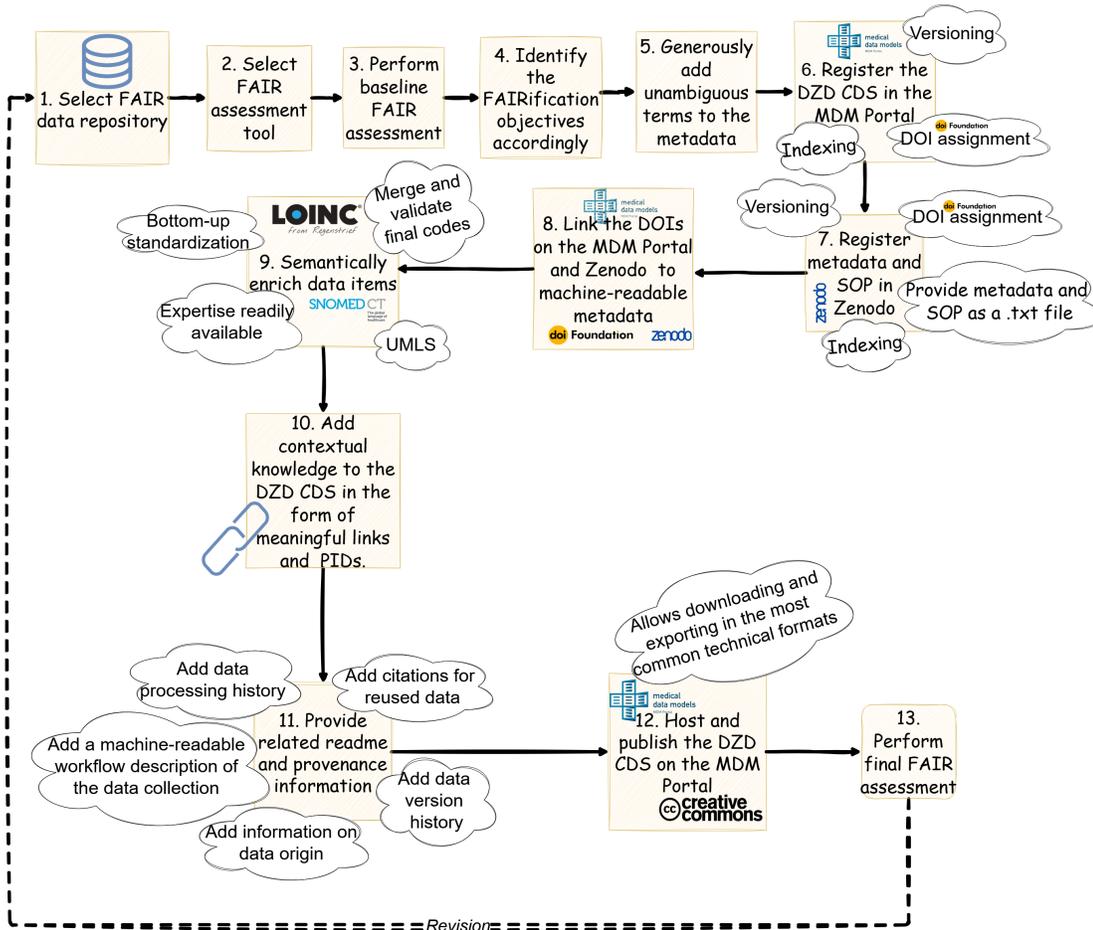
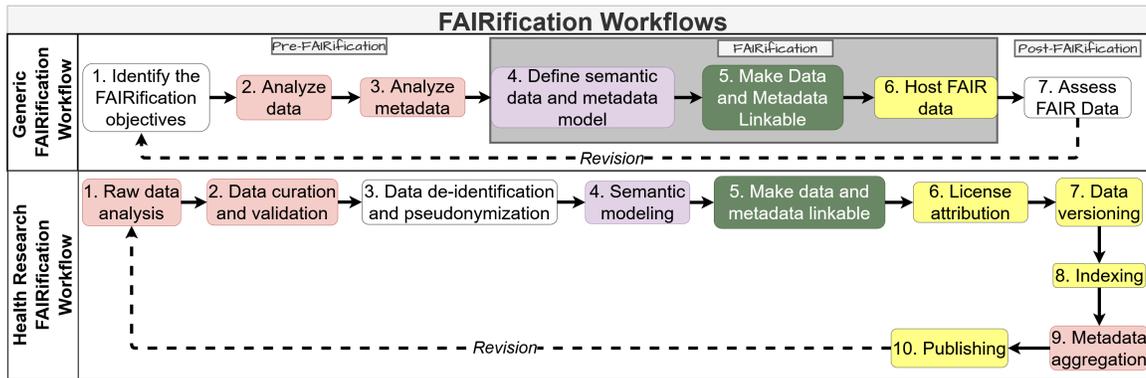
The following **Fig 3** illustrates the steps taken in both FAIRification workflows and the principles achieved by implementing each step. The steps with similar colours are representative of the steps we found to be similar in our implementation. The bottom sketch shows our implementation of these workflows in the FAIRification of the DZD CDS.

Notably, information about licensing options is more distinct in the health research workflow than in the generic FAIRification workflow. Although the generic FAIRification workflow does not explicitly indicate licensing as a FAIRification step, we still licensed the CDS in our implementation of this workflow because hosting it on the MDM Portal required us to do so. Hosting the DZD CDS on the MDM Portal also simultaneously led to its versioning and indexing.

### Minimum requirements for successful FAIRification of core data sets in health research

This work has shown that regardless of the selected FAIRification workflow, there are some bare minimums that should be present for successful metadata FAIRification of core data sets in health research. They include:

1. A thorough understanding of the processes by which the metadata is defined, adapted, and expanded: The data analysis step helps to understand the project capabilities and resources. This step is also useful in developing an



**Fig 3. Comparison of generic and domain-specific workflows:** The upper lane shows the generic workflow (Jacobsen et al.) and the lower lane shows the steps specific to the health research workflow (Sinaci et al.) [16,17]. The sketch at the bottom shows our implementation of these workflows in the FAIRification of the DZD CDS.

<https://doi.org/10.1371/journal.pdig.0001139.g003>

understanding the processes by which the dataset is defined, adapted, and expanded. It also helps to identify the data characteristics that should be improved based on the defined FAIRification goal. The capabilities that a FAIR data management environment should exhibit to enable and support the realization of a FAIR dataset can also be

determined, including data access, data hosting, ontology services and data sharing. Finally, the findings of the raw data analysis play a critical role in determining the FAIR assessment tool and requirements needed to achieve the desired FAIRification outcome.

2. Collaborative development of standards: A cultural shift is required to enable the implementation and co-development of standards for metadata, centrally managed and organized [23,53]. This cultural change needs to be broadly accepted in order to implement continuous FAIR RDM throughout the data lifecycle [27,53].
3. Collaborative data stewardship: A shift away from the culture of individualised data ownership towards one of collaborative data stewardship is necessary for efficient data sharing that facilitates impactful research [54,55]. Consultations and training for data publication pipelines should be offered to ensure sustainably operated FAIR infrastructures.
4. An investment of resources: A significant investment of time, money and domain knowledge is necessary to implement the described FAIRification steps [23,56]. No significant difference in personnel requirement was observed in the implementation of both workflows. Customised incentives are important to encourage stakeholders to engage in data sharing beyond their moral obligations, especially if tangible motivators for this remain limited [55].

## Discussion

The comprehensive approach including initial FAIR assessments, FAIRification activities, post-FAIRification adjustments, and being intentional about implementing FAIR-enabling infrastructure can significantly improve the FAIRness of health data sets, as demonstrated in this work using the example of the DZD CDS. The application and comparison of two common FAIRification workflows underscores both the adaptability and domain-specific nuances of such processes. The domain-agnosticism of the generic workflow provides a flexible framework applicable to diverse datasets. Conversely, the health-specific workflow addresses the vital additions that are necessary to safeguard the integrity of health research data.

The inclusion of metadata aggregation and systematic data versioning in the health-specific workflow emphasizes the importance of robust data management practices to address the complexities of health research data. Domain-specific requirements shape FAIRification, balancing technical interoperability with adherence to contextual and legal constraints. Thereby, tailored health FAIRification workflows enhance the applicability of the FAIR principles and showcase how FAIRification efforts can be extended in other contexts, ensuring flexibility and precision.

The implementation of structured FAIRification workflows significantly improved the FAIRness of the DZD CDS. Previous FAIRification efforts, guided by experience rather than structured workflows, required extended timelines due to multiple consultations and iterative revisions [12,57]. Contrarily, the two workflows applied in this work were straightforward to implement, with the FDMM proving instrumental. The FDMM facilitated both binary and scaled approaches to FAIR assessment, enabling objective evaluations with minimal reliance on personal judgement while allowing for progress measurement towards FAIRer scores. Baseline and final FAIR assessments, though not explicitly included in the FAIRification workflow for health research, were invaluable in evaluating improvements and demonstrating the impact of FAIRification efforts [13,42]. Interestingly, the implementation of both workflows led us to achieve the same FAIRness rating and required similar resources for implementation. However, adapting these workflows to specific objectives and contexts may improve cost and time efficiency, particularly when applied to real-world medical data.

De-identification and pseudonymization are measures implemented to preserve the data privacy rights of the subjects and is performed based on the purpose for which the dataset has been developed [58,59]. Unlike the generic FAIRification workflow, the FAIRification workflow for health research includes pseudonymization and de-identification as a distinct FAIRification step. This indicates the workflow's consideration of the heightened sensitivity of health research data, especially in the wake of the implementation of the General Data Protection Regulation in the European Union [60, 61]. Skipping this step in the implementation of the FAIRification workflow for health research did not adversely affect

the final results of the final FAIR assessment. This may indicate that de-identification and pseudonymization of health research data does not directly contribute to data FAIRness. Pseudonymization and de-identification processes have been described as slow and cumbersome and would likely increase the effort required in the FAIRification of DZD CDS-related clinical research datasets that contain patient information [58,62–65]. Other components unique to the health-specific workflow—such as metadata aggregation and structured licensing—contributed to improved procedural efficiency but had no discernible impact on the FDMM score. We therefore deduce that the adaptations made in the FAIRification workflow for health research data are useful to improve efficiency but do not necessarily lead to higher FAIRness scores when applied to core data sets. We were able to incorporate all the other metadata indicators without modifying them.

### Goal-oriented and flexible implementation

While the health research FAIRification workflow does not explicitly include goal setting, defining the objectives has shown to be a critical component of FAIRification [6,59]. In this context, collaboratively setting objectives helped to ensure that the objectives are inclusive of the perspectives of the various pertinent stakeholders and served as a means to justify the expenditure of resources in the FAIRification exercise. The business interest that is expected to evolve as the number of exports from the MDM Portal increases also served as a factor that encouraged collaboration among the stakeholders. There are also previous success stories that served as reference points [66–72]. The objectives also served as the basis for feedback on the effectiveness of individual FAIRification tasks. This feedback can be used to determine the overall success of the process and set a clear endpoint for this FAIRification iteration. Jacobsen et al. indicated that the generic FAIRification workflow steps need not follow a strict sequence [73]. This flexibility allowed us to start the DZD CDS FAIRification by selecting FAIR infrastructure, evaluating the baseline FAIRness and identifying the FAIRification objectives accordingly (step 1). We then proceeded to enrich the metadata as informed by our metadata analysis (steps 2 and 3) and then registered the DZD CDS in the MDM portal (step 6). The related metadata and SOP were registered in Zenodo. After this, we semantically annotated the data items since the MDM Portal provided codes and we were able to discern these codes (Step 4). The data stewards ensured the choice of terminologies aligned with international standards in real-world clinical practice. Accordingly, LOINC and SNOMED CT concepts were included in addition to UMLS codes. The semantic enrichment of DZD CDS raw data items also required retrospective reconciliation of terminology preferences between the data curators at the DZD and the German Centers for Health Research. As a result, all numeric parameters were annotated using LOINC codes. Several collaborative sessions regarding the choice of standard concepts between data curators at the DZD but also with the German Centers for Health Research ensured semantic accuracy while preserving domain usability. However, establishing a common understanding of the parameters presented a substantial clerical burden [49,74]. Once this process was complete, subsequent workflow implementation (shown in Fig 3) required comparatively less time and effort.

### FAIR infrastructure

Registering the DZD CDS in the MDM Portal alongside metadata and SOPs in Zenodo enabled simultaneous completion of multiple workflow steps as shown in Fig 3. This approach streamlined the process and demonstrated the importance of selecting FAIR infrastructure prior to the onset of a FAIRification journey. Jacobsen et al. have highlighted that the generic FAIRification workflow specifically targets the implementation of principles F1, F2, R1, R1.1, R1.2 and R1.3 [1,73]. In our context these principles target the corresponding metadata sub principles as defined by the RDA (see S1 Table, supporting information). Depositing the metadata and SOPs in Zenodo also led to the fulfillment of the corresponding F3 and F4 metadata principles (defined in S1 Table, supporting information). Registering the DZD CDS in the MDM Portal led to the fulfillment of the “accessibility” metadata sub-principles. The DZD CDS can now be downloaded and exported in most common technical formats and provides UMLS codes for semantic enrichment, which further enables the implementation of an ontology matching service for querying FAIR data [12,75]. Qualified references are available in the form of links

to the comprehensive metadata and SOP in Zenodo, which further fulfils the “interoperability” sub-principles. An interesting finding in the implementation of the generic workflow is that key FAIRification measures were already implemented while we were still in the pre-FAIRification phase. These include addition of rich explicit metadata, licensing, versioning, DOI assignment, as well as registration and indexing of the metadata and SOPs. This raises the question, “when does FAIRification actually begin?”

### **Expected impact of a FAIRer DZD CDS**

FAIRification has increasingly become an important part of the DZD RDM priorities and this has necessitated FAIR data sharing of the DZD CDS. The FAIRer DZD CDS has been made available to the community on the MDM portal and supports data sharing among DZD sites [76]. FAIRification of the DZD CDS aligns with the DZD’s commitment to comprehensive data stewardship [12], enhanced data sharing and data analysis across DZD sites [12,57]. A machine-actionable framework to describe and structure the CDS has been established and the DZD CDS is better interoperable [49,77]. The expected return on the investment of the efforts made to provide machine-actionable, heterogenous data is that there are now wider possibilities for data integration and larger-scale analyses. Expanding access via alternative authentication and authorisation procedures could further enhance utility for non-MDM portal users.

The transferable lessons derived from this work include the importance of dedicated multidisciplinary collaboration that fosters mutual understanding, the clear definition of stakeholder roles, rigorous version control and documentation throughout the FAIRification process, the early and deliberate selection of appropriate FAIR infrastructure, as well as the anticipation of resource demands and justification of the effort by articulating the expected return on investment. However, this work only addresses a single CDS. We did not have any particular domain-specific requirements that influenced the FAIRification process therefore we cannot comment on any domain-specific challenges. We expect, however, that such challenges may occur in the context of real-world data related to the DZD CDS.

### **Is the FAIRification journey of medical research data worthwhile?**

FAIRification of medical research data is essential for evidence-informed decisions and has proven beneficial for datasets such as the DZD CDS, which has garnered significant interest (725 views and 379 downloads as of November 2024) [5,78,79]. Enhanced reusability increases the value of the CDS, provides a basis for profitable reuse in other contexts, and may eliminate the need for a new data collection process. The improved reusability of the data set over an extended period of time may further result in the development of new therapeutic regimens by the secondary data user and increased business value for the DZD.

Interoperability, supported by standardized technical formats and metadata, supports integration with heterogeneous datasets, broadening research opportunities, and simplifying cross-evaluation [80]. Since all current DZD clinical studies launched in 2021 and later use the same CDS, the data pool is enlarged and the possibility of cross-evaluation is simplified [12]. Novel research based on integrated and analyzed heterogenous data is anticipated once the FAIRified DZD CDS is connected to routine data from data integration centers [81,82]. In addition, the semantic enrichment of the DZD CDS supports seamless interoperability with patient-generated health data and routine electronic health records. This enhanced interoperability is anticipated to support more inclusive and diverse study designs, and to accelerate the generation of evidence for improved therapeutic strategies in diabetes care. Taken together, these efforts improve discoverability and readiness for artificial intelligence, contributing to greater visibility and impact of the DZD CDS.

The key resources required for the implementation of both FAIRification workflows included funding, incentives and expertise for the metadata curation and enrichment. The DZD CDS FAIRification required numerous consultations and iterative meetings between data owners and FAIR experts to make key decisions, which contributed to a significant time investment that spanned several months. The quantified investment amount that should be made in running FAIRification

cycles in a manner that keeps it beneficial to stakeholders remains open to discussion. We recommend that this discussion should include the stakeholders of other clinical core data sets such as the one developed by the German Medical Informatics Initiative and the French CDS for geriatric oncology studies [33,83,84].

## Conclusion

In this work, we compared the implementation of two FAIRification workflows for the core data set applied to German diabetes studies. We also identified minimums that will help to reduce efforts and costs for FAIRification, when applied. We recommend that more FAIRification workflows that take into account the nature of domain-specific data should be developed for other scientific domains that have embraced FAIRification as a necessary journey. Although this work focused on the DZD CDS, we hypothesize that the identified lessons and minimum requirements are applicable to the FAIRification of core datasets of other clinical domains.

Retrospective data FAIRification is a cumbersome task regardless of the FAIRification workflow implemented. For this reason, we resonate with current recommendations that encourage scientists and data owners to design their scientific projects in a way that takes into account the FAIRification of prospective data right from the infancy stage and continuously improves FAIRness throughout the lifecycle of the project [23,85]. One aspect of these preparatory steps is the thoughtful collection of appropriate FAIRification infrastructure as early as possible [38].

However, this is a multi-stakeholder engagement and different stakeholders may have different preferences with regard to the FAIRification process. For example, the funders may deem it more cost-effective to FAIRify data all in one cycle as opposed to gradually and iteratively while other participant stakeholders may have different preferences on the order in which the FAIRification steps should be implemented. The implementation of both FAIRification workflows in the DZD CDS would not have been possible without the tremendous involvement of the data owners, data stewards and the pertinent stakeholders. We therefore recognise the importance of harmonizing the stakeholders' perspectives and expectations.

Templates for FAIR data management plans (DMPs) continue to be developed as funders and policy makers continue to require DMPs to be prospectively FAIR inclusive [5,26,43,86]. It remains to be seen how these prospective FAIR DMPs can be integrated into retrospective FAIR workflows. It may also be necessary to develop FAIRification workflows for prospective implementation specific to the health research data domain. It also remains to be explored what the resultant differences would be to the FAIRification workflows if the respective steps were implemented prospectively.

Quite a lot of work has already been done to automate the semantic enrichment of health data [22,87–90]. It would be interesting to further research how automated semantic enrichment can be incorporated into FAIR workflows, how many more of the steps in the FAIR workflows can be automated, and what would be the consequent changes, if any, to the workflows once the steps are automated. In 2019 the European Commission estimated the cost of not having FAIR research data in the European research economy at € 10.2 billion [91]. Another interesting area of research would be to determine how much time and resources have been saved by FAIRifying the DZD CDS. It remains to be seen which steps will be iterated or eliminated in subsequent FAIRification cycles as the priorities of this FAIRification journey evolve and new insights are obtained.

## Supporting information

**S1 Table. FDMM metadata indicators.** S1 Table illustrates the RDA FDMM metadata indicators that we employed in the FAIRness assessment of the DZD CDS [14].

(TEX)

**S2 Table. Implementation of the generic FAIRification workflow.** S2 Table illustrates the implementation of the generic FAIRification workflow in the DZD CDS and the decisions made to adapt it to this context [16].

(TEX)

**S3 Table. Implementation of the health research FAIRification workflow.** S3 Table illustrates the implementation of the health research FAIRification workflow in the DZD CDS and the decisions made in adapting it to this context [17].

(TEX)

## Acknowledgments

This work was partially funded by the NFDI4Health – *Nationale Forschungsdateninfrastruktur für personenbezogene Gesundheitsdaten* (DFG-funded project 442326535) (AAZ).

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Author contributions

**Conceptualization:** Esther Thea Inau, Angela Dedie, Ivona Anastasova, Brigitte Fröhlich, Dagmar Waltemath, Atinkut Alamirrew Zeleke.

**Data curation:** Esther Thea Inau.

**Formal analysis:** Esther Thea Inau.

**Investigation:** Esther Thea Inau, Angela Dedie, Ivona Anastasova, Atinkut Alamirrew Zeleke.

**Methodology:** Esther Thea Inau, Ivona Anastasova, Atinkut Alamirrew Zeleke.

**Project administration:** Renate Schick, Brigitte Fröhlich, Martin Preusse, Dagmar Waltemath.

**Resources:** Michael Roden, Andreas L. Birkenfeld, Martin Hrabe de Angelis.

**Supervision:** Dagmar Waltemath.

**Writing – original draft:** Esther Thea Inau, Dagmar Waltemath, Atinkut Alamirrew Zeleke.

**Writing – review & editing:** Esther Thea Inau, Ivona Anastasova, Dagmar Waltemath, Atinkut Alamirrew Zeleke.

## References

1. Wilkinson MD, Dumontier M, Aalbersberg IJJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data*. 2016;3:160018. <https://doi.org/10.1038/sdata.2016.18> PMID: 26978244
2. Wilkinson SR, Eisenhauer G, Kapadia AJ, Knight K, Logan J, Widener P, et al. F\*\*\* workflows: when parts of FAIR are missing. In: 2022 IEEE 18th International Conference on e-Science (e-Science). 2022. p. 507–12. <https://doi.org/10.1109/escience55777.2022.00090>
3. Alharbi E, Gadiya Y, Henderson D, Zaliani A, Delfin-Rossaro A, Cambon-Thomsen A, et al. Selection of data sets for FAIRification in drug discovery and development: which, why, and how?. *Drug Discov Today*. 2022;27(8):2080–5. <https://doi.org/10.1016/j.drudis.2022.05.010> PMID: 35595012
4. Martínez-García A, Alvarez-Romero C, Román-Villarán E, Bernabeu-Wittel M, Luis Parra-Calderón C. FAIR principles to improve the impact on health research management outcomes. *Heliyon*. 2023;9(5):e15733. <https://doi.org/10.1016/j.heliyon.2023.e15733> PMID: 37205991
5. Inau ET, Sack J, Waltemath D, Zeleke AA. Initiatives, concepts, and implementation practices of the findable, accessible, interoperable, and reusable data principles in health data stewardship: scoping review. *J Med Internet Res*. 2023;25:e45013. <https://doi.org/10.2196/45013> PMID: 37639292
6. Welter D, Juty N, Rocca-Serra P, Xu F, Henderson D, Gu W, et al. FAIR in action - a flexible framework to guide FAIRification. *Sci Data*. 2023;10(1):291. <https://doi.org/10.1038/s41597-023-02167-2> PMID: 37208349

7. Batista DO, Leiva Mederos AA, Peralta González MJ. FAIRification: a necessary practice for research data management. *Advanced Notes in Information Science. Pro-Metrics*. 2024. <https://doi.org/10.47909/978-9916-9974-5-1.92>
8. Waithira N, Mukaka M, Kestelyn E, Choththanawathit K, Thi Phuong DN, Thanh HN, et al. Data sharing and reuse in clinical research: are we there yet? A cross-sectional study on progress, challenges and opportunities in LMICs. *PLOS Glob Public Health*. 2024;4(11):e0003392. <https://doi.org/10.1371/journal.pgph.0003392> PMID: 39565766
9. Gehrman J, Herczog E, Decker S, Beyan O. What prevents us from reusing medical real-world data in research. *Sci Data*. 2023;10(1):459. <https://doi.org/10.1038/s41597-023-02361-2> PMID: 37443164
10. Niarakis A, Waltemath D, Glazier J, Schreiber F, Keating SM, Nickerson D, et al. Addressing barriers in comprehensiveness, accessibility, reusability, interoperability and reproducibility of computational models in systems biology. *Brief Bioinform*. 2022;23(4):bbac212. <https://doi.org/10.1093/bib/bbac212> PMID: 35671510
11. Pigeot I, Ahrens W, Darms J, Fluck J, Golebiewski M, Hahn HK, et al. Making epidemiological and clinical studies FAIR using the example of COVID-19. *Datenbank Spektrum*. 2024;24(2):117–28. <https://doi.org/10.1007/s13222-024-00477-2>
12. Inau ET, Dedié A, Anastasova I, Schick R, Zdravomyslov Y, Fröhlich B, et al. The journey to a FAIR CORE DATA SET for diabetes research in Germany. *Sci Data*. 2024;11(1):1159. <https://doi.org/10.1038/s41597-024-03882-0> PMID: 39433782
13. Waltemath D, Beyan O, Cramer K, Dedié A, Gierend K, Gröber P, et al. FAIRe Gesundheitsdaten im nationalen und internationalen Datenraum. *Bundesgesundheitsbl*. 2024;67(6):710–20. <https://doi.org/10.1007/s00103-024-03884-8>
14. RDA FAIR Data Maturity Model Working Group B. FAIR Data Maturity Model: specification and guidelines. Research Data Alliance; 2020. <https://doi.org/10.15497/rda00050>
15. Inau ET, Zeleke A, Waltemath D. Harvesting the low hanging fruits from the FAIRtree. *Stud Health Technol Inform*. 2023;302:390–1. <https://doi.org/10.3233/SHTI230155> PMID: 37203700
16. Jacobsen A, Kaliyaperumal R, da Silva Santos LOB, Mons B, Schultes E, Roos M, et al. A generic workflow for the data FAIRification process. *Data Intelligence*. 2020;2(1–2):56–65. [https://doi.org/10.1162/dint\\_a\\_00028](https://doi.org/10.1162/dint_a_00028)
17. Sinaci AA, Núñez-Benjumea FJ, Gencturk M, Jauer M-L, Deserno T, Chronaki C, et al. From raw data to FAIR data: the FAIRification workflow for health research. *Methods Inf Med*. 2020;59(S 01):e21–32. <https://doi.org/10.1055/s-0040-1713684> PMID: 32620019
18. Shiferaw KB, Balaur I, Welter D, Waltemath D, Zeleke AA. CALIFRAME: a proposed method of calibrating reporting guidelines with FAIR principles to foster reproducibility of AI research in medicine. *JAMIA Open*. 2024;7(4):ooae105. <https://doi.org/10.1093/jamiaopen/ooae105> PMID: 39430802
19. Bahim C, Casorrán-Amilburu C, Dekkers M, Herczog E, Loozen N, Repanas K, et al. The FAIR data maturity model: an approach to harmonise FAIR assessments. *Data Science Journal*. 2020;19:41. <https://doi.org/10.5334/dsj-2020-041>
20. Islam S, Hardisty A, Addink W, Weiland C, Glöckler F. Incorporating RDA outputs in the design of a european research infrastructure for natural science collections. *Data Science Journal*. 2020;19:1–14. <https://doi.org/10.5334/dsj-2020-050>
21. Shiferaw KB, Zeleke A, Waltemath D. Assessing the FAIRness of deep learning models in cardiovascular disease using computed tomography images: data and code perspective. *Caring is Sharing—Exploiting the Value in Data for Health and Innovation*. IOS Press; 2023. p. 63–7.
22. Zeleke AA, Inau E, Gebhardt T, Clemens V, Golebiewski M, Waltemath D. Implementation of a FAIR metrics concept for health data. Zenodo; 2025. <https://doi.org/10.5281/zenodo.15654481>
23. Waltemath D, Inau E, Michaelis L, Satagopam V, Balaur I. Experiences from FAIRifying community data and FAIR infrastructure in biomedical research domains. *Proc Conf Res Data Infrastr*. 2023;1. <https://doi.org/10.52825/cordi.v1i.415>
24. Chen X, Jagerhorn M. Implementing FAIR Workflows along the research lifecycle. *Procedia Computer Science*. 2022;211:83–92. <https://doi.org/10.1016/j.procs.2022.10.179>
25. de Visser C, Johansson LF, Kulkarni P, Mei H, Neerincx P, Joeri van der Velde K, et al. Ten quick tips for building FAIR workflows. *PLoS Comput Biol*. 2023;19(9):e1011369. <https://doi.org/10.1371/journal.pcbi.1011369> PMID: 37768885
26. World Health Organization. Sharing and reuse of health-related data for research purposes: WHO policy and implementation guidance. WHO; 2022. <https://www.who.int/publications/i/item/9789240044968>
27. Bloemers M, Montesanti A. The FAIR funding model: providing a framework for research funders to drive the transition toward fair data management and stewardship practices. *Data Intelligence*. 2020;2(1–2):171–80. [https://doi.org/10.1162/dint\\_a\\_00039](https://doi.org/10.1162/dint_a_00039)
28. Kersloot MG, Jacobsen A, Groenen KHJ, Dos Santos Vieira B, Kaliyaperumal R, Abu-Hanna A, et al. De-novo FAIRification via an electronic data capture system by automated transformation of filled electronic case report forms into machine-readable data. *J Biomed Inform*. 2021;122:103897. <https://doi.org/10.1016/j.jbi.2021.103897> PMID: 34454078
29. Van Reisen M, Oladipo FO, Mpezamihigo M, Plug R, Basajja M, Aktau A. Incomplete COVID-19 data: the curation of medical health data by the Virus Outbreak Data Network-Africa. 2022.
30. Queralt-Rosinach N, Kaliyaperumal R, Bernabé CH, Long Q, Joosten SA, van der Wijk HJ, et al. Applying the FAIR principles to data in a hospital: challenges and opportunities in a pandemic. *J Biomed Semantics*. 2022;13(1):12. <https://doi.org/10.1186/s13326-022-00263-7> PMID: 35468846
31. Das Deutsche Zentrum für Diabetesforschung. DZD Website. 2021. <https://www.dzd-ev.de/en/the-dzd/index.html>
32. Deutsches Zentrum für Diabetesforschung e V D. DZD Core Data Set. 2022. <https://medical-data-models.org/46011>
33. Michaelis L, Inau ET, Muzoora MR, Wodke JA, Ganslandt T, Thun S, et al. Applying the RDA data maturity model on the core dataset of the German medical informatics initiative. In: *The 15th International Conference on Semantic Web Applications and Tools for Health Care and Life Sciences*. vol. 1613; 2024. p. 0073.

34. German Center for Diabetes Research D. DZD Core Data Set - Metadata and SOP. 2022. <https://zenodo.org/records/12652100>
35. University Hospital Tuebingen. IFIS Website. 2020. <https://clinicaltrials.gov/ct2/show/NCT04607096>
36. Jumpertz von Schwartzberg R, Stefan N, Wagner R, Guthoff M, Sandforth A, Icks A, et al. SGLT2 inhibition in addition to lifestyle intervention and risk for complications in subtypes of patients with prediabetes - a randomized, placebo controlled, multi-center trial (LIFETIME) – rationale, methodology and design. Cold Spring Harbor Laboratory; 2023. <https://doi.org/10.1101/2023.11.18.23298622>
37. D Z G. DZG core data set. Medical Data Models. 2023. <https://medical-data-models.org/45851#model-model>
38. Inau E, Dedie A, Zeleke A, Waltemath D. A discussion on available software tools for the FAIR assessment of the CORE DATA SET for diabetes research in Germany. *Stud Health Technol Inform*. 2025;327:1328–32. <https://doi.org/10.3233/SHTI250618> PMID: 40380721
39. Jacob D, David R, Aubin S, Gibon Y. Making experimental data tables in the life sciences more FAIR: a pragmatic approach. *Gigascience*. 2020;9(12):giaa144. <https://doi.org/10.1093/gigascience/giaa144> PMID: 33319910
40. Amdouni E, Bouazzouni S, Jonquet C. O'FAIRe makes you an offer: metadata-based automatic FAIRness assessment for ontologies and semantic resources. *IJMISO*. 2022;16(1):16–46. <https://doi.org/10.1504/ijmso.2022.131133>
41. Bach JS, Limani F, Zhang Y, Latif A, Mathiak B, Mutschke P. FAIR assessment practices: experiences from KonsortSWD and BERD@ NFDI. In: *Proceedings of the Conference on Research Data Infrastructure*. 2023. p. 433–9.
42. Balaur I, Welter D, Rougny A, Inau ET, Mazein A, Ghosh S, et al. MINERVA FAIR assessment fosters open science & scientific crowd-sourcing in systems biomedicine. Cold Spring Harbor Laboratory; 2024. <https://doi.org/10.1101/2024.08.28.610042>
43. Swiss National Science Foundation. Data Management Plan (DMP) - Guidelines for Researchers. 2017. <https://www.snf.ch/en/FAiWVH4WVpKvohw9/topic/research-policies>
44. Dugas M, Neuhaus P, Meidt A, Doods J, Storck M, Bruland P, et al. Portal of medical data models: information infrastructure for medical research and healthcare. *Database (Oxford)*. 2016;2016:bav121. <https://doi.org/10.1093/database/bav121> PMID: 26868052
45. Riepenhausen S, Blumenstock M, Niklas C, Hegselmann S, Neuhaus P, Meidt A, et al. Europe's largest research infrastructure for curated medical data models with semantic annotations. *Methods Inf Med*. 2024;63(1–02):52–61. <https://doi.org/10.1055/s-0044-1786839> PMID: 38740374
46. Zarin DA, Fain KM, Dobbins HD, Tse T, Williams RJ. 10-Year update on study results submitted to ClinicalTrials.gov. *N Engl J Med*. 2019;381(20):1966–74. <https://doi.org/10.1056/NEJMSr1907644> PMID: 31722160
47. Kim M. The creative commons and copyright protection in the digital era: uses of creative commons licenses. *Journal of Computer-Mediated Communication*. 2007;13(1):187–209. <https://doi.org/10.1111/j.1083-6101.2007.00392.x>
48. Hegselmann S, Storck M, Gessner S, Neuhaus P, Varghese J, Bruland P, et al. Pragmatic MDR: a metadata repository with bottom-up standardization of medical metadata through reuse. *BMC Med Inform Decis Mak*. 2021;21(1):160. <https://doi.org/10.1186/s12911-021-01524-8> PMID: 34001121
49. Bodenreider O, Cornet R, Vreeman DJ. Recent developments in clinical terminologies - SNOMED CT, LOINC, and RxNorm. *Yearb Med Inform*. 2018;27(1):129–39. <https://doi.org/10.1055/s-0038-1667077> PMID: 30157516
50. van de Sandt S, Nielsen LH, Ioannidis A, Muench A, Henneken E, Accomazzi A. Practice meets principle: tracking software and data citations to Zenodo DOIs. *arXiv preprint 2019*. <https://doi.org/10.48550/arXiv.1911.00295>
51. Bahls T, Pung J, Heinemann S, Hauswaldt J, Demmer I, Blumentritt A, et al. Designing and piloting a generic research architecture and workflows to unlock German primary care data for secondary use. *J Transl Med*. 2020;18(1):394. <https://doi.org/10.1186/s12967-020-02547-x> PMID: 33076938
52. Morrison H, Desautels L. Open access, copyright and licensing: basics for open access publishers. *J Orthop Case Rep*. 2016;6(1):1–2. <https://doi.org/10.13107/jocr.2250-0685.360> PMID: 27299111
53. Mayer G, Müller W, Schork K, Uszkoreit J, Weidemann A, Wittig U, et al. Implementing FAIR data management within the German Network for Bioinformatics Infrastructure (de.NBI) exemplified by selected use cases. *Brief Bioinform*. 2021;22(5):bbab010. <https://doi.org/10.1093/bib/bbab010> PMID: 33589928
54. Sadeh Y, Denejkina A, Karyotaki E, Lenferink LIM, Kassam-Adams N. Opportunities for improving data sharing and FAIR data practices to advance global mental health. *Glob Ment Health (Camb)*. 2023;10:e14. <https://doi.org/10.1017/gmh.2023.7> PMID: 37860102
55. Modjarrad K, Moorthy VS, Millett P, Gsell P-S, Roth C, Kieny M-P. Developing global norms for sharing data and results during public health emergencies. *PLoS Med*. 2016;13(1):e1001935. <https://doi.org/10.1371/journal.pmed.1001935> PMID: 26731342
56. Oladipo F, Folorunso S, Ogundepo E, Osigwe O, Akindede A. Curriculum development for FAIR data stewardship. *Data Intelligence*. 2022;4(4):991–1012. [https://doi.org/10.1162/dint\\_a\\_00183](https://doi.org/10.1162/dint_a_00183)
57. Inau ET, Dedie A, Anastasova I, Birkenfeld A, Frohlich B, Hrabe de Angelis M, et al. First steps towards the FAIRification of the DZD CORE DATASET at the German Center for Diabetes Research. In: *14th International Conference on Semantic Web Applications and Tools for Health Care and Life Sciences (SWAT4HCLS 2023)*; 2023. p. 139–40. <http://ceur-ws.org/Vol-3415/#paper-28>
58. Mulder T, Tudorica M. Privacy policies, cross-border health data and the GDPR. *Information & Communications Technology Law*. 2019;28(3):261–74. <https://doi.org/10.1080/13600834.2019.1644068>
59. Rocca-Serra P, Sansone SA, Gu W, Welter D, Abbassi Daloui T, Portell-Silva L. D2.1 FAIR Cookbook. Zenodo; 2022.
60. Mostert M, Bredenoord AL, Biesart MCIH, van Delden JJM. Big Data in medical research and EU data protection law: challenges to the consent or anonymise approach. *Eur J Hum Genet*. 2016;24(7):956–60. <https://doi.org/10.1038/ejhg.2015.239> PMID: 26554881

61. Inau ET, Nalugala R, Nandwa WM, Obwanda F, Wachira A, Cartaxo A. FAIR equivalency, regulatory framework and adoption potential of FAIR guidelines in health in Kenya. *Data Intelligence*. 2022;4(4):852–66. [https://doi.org/10.1162/dint\\_a\\_00175](https://doi.org/10.1162/dint_a_00175)
62. Rothstein MA. Is deidentification sufficient to protect health privacy in research?. *Am J Bioeth*. 2010;10(9):3–11. <https://doi.org/10.1080/15265161.2010.494215> PMID: 20818545
63. El Emam K, Arbuckle L, Koru G, Eze B, Gaudette L, Neri E, et al. De-identification methods for open health data: the case of the Heritage Health Prize claims dataset. *J Med Internet Res*. 2012;14(1):e33. <https://doi.org/10.2196/jmir.2001> PMID: 22370452
64. Mawji A, Longstaff H, Trawin J, Dunsmuir D, Komugisha C, Novakowski SK, et al. A proposed de-identification framework for a cohort of children presenting at a health facility in Uganda. *PLOS Digit Health*. 2022;1(8):e0000027. <https://doi.org/10.1371/journal.pdig.0000027> PMID: 36812586
65. Joo M-H, Kwon H-Y. Comparison of personal information de-identification policies and laws within the EU, the US, Japan, and South Korea. *Government Information Quarterly*. 2023;40(2):101805. <https://doi.org/10.1016/j.giq.2023.101805>
66. Beitia AO, Kuperman G, Delman BN, Shapiro JS. Assessing the performance of LOINC® and RadLex for coverage of CT scans across three sites in a health information exchange. *AMIA Annu Symp Proc*. 2013;2013:94–102. PMID: 24551324
67. Lee S, Han J, Park RW, Kim GJ, Rim JH, Cho J, et al. Development of a controlled vocabulary-based adverse drug reaction signal dictionary for multicenter electronic health record-based pharmacovigilance. *Drug Saf*. 2019;42(5):657–70. <https://doi.org/10.1007/s40264-018-0767-7> PMID: 30649749
68. Dugas M, Thun S, Frankewitsch T, Heitmann KU. LOINC codes for hospital information systems documents: a case study. *J Am Med Inform Assoc*. 2009;16(3):400–3. <https://doi.org/10.1197/jamia.M2882> PMID: 19261942
69. Sockolow P, Chou EY, Park S. Addressing the gap in data communication from home health care to primary care during care transitions: completeness of an interoperability data standard. *Healthcare (Basel)*. 2022;10(7):1295. <https://doi.org/10.3390/healthcare10071295> PMID: 35885822
70. Khan AN, Russell D, Moore C, Rosario AC Jr, Griffith SP, Bertolli J. The map to LOINC project. *AMIA Annu Symp Proc*. 2003;2003:890. PMID: 14728395
71. Lin MC, Vreeman DJ, McDonald CJ, Huff SM. A characterization of local LOINC mapping for laboratory tests in three large institutions. *Methods Inf Med*. 2011;50(2):105–14. <https://doi.org/10.3414/ME09-01-0072> PMID: 20725694
72. Inau ET, Radke D, Westphal S, Schafer C, Zeleke AZ, Nauck M. Semantic enrichment of the laboratory data dictionary of the study of health in Pomerania (SHIP-START-4) with LOINC; detailed mapping results. 2024.
73. Jacobsen A, de Miranda Azevedo R, Juty N, Batista D, Coles S, Cornet R. FAIR principles: interpretations and implementation considerations. 2020.
74. Inau E, Radke D, Westphal S, Zeleke AA, Waltemath D. Comparing voluntary LOINC mappings for the SHIP-4 medical laboratory data dictionary before and after domain expert review. In: 67 Jahrestagung der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie e V (GMDS), 13 Jahreskongress der Technologie- und Methodenplattform für die vernetzte medizinische Forschung e V (TMF). 2022. <https://doi.org/10.3205/22gmds058>
75. Bhatia K, Tanch J, Chen ES, Sarkar IN. Applying FAIR principles to improve data searchability of emergency department datasets: a case study for HCUP-SEDD. *Methods Inf Med*. 2020;59(1):48–56. <https://doi.org/10.1055/s-0040-1712510> PMID: 32535879
76. Bernabé C, Sales TP, Schultes E, van Ulzen N, Jacobsen A, da Silva Santos LOB. A goal-oriented method for FAIRification planning. In: CEUR Workshop Proceedings. 2023.
77. Musen MA, O'Connor MJ, Schultes E, Martínez-Romero M, Hardi J, Graybeal J. Modeling community standards for metadata as templates makes data FAIR. *Sci Data*. 2022;9(1):696. <https://doi.org/10.1038/s41597-022-01815-3> PMID: 36371407
78. Stellmach C, Muzoora MR. How to assess fairness of your data—a summary of testing two fair validators. In: MEDINFO 2023—The Future Is Accessible. 2024. p. 154–8.
79. Ouwerkerk J, Rasche H, Spalding JD, Hiltemann S, Stubbs AP. FAIR data retrieval for sensitive clinical research data in Galaxy. *Gigascience*. 2024;13:giad099. <https://doi.org/10.1093/gigascience/giad099> PMID: 38280189
80. Inau ET, Radke D, Bird L, Westphal S, Ittermann T, Schäfer C, et al. Semantic enrichment of Pomeranian health study data using LOINC and WHO-FIC terminology mapping principles. *JAMIA Open*. 2025;8(2):ooaf010. <https://doi.org/10.1093/jamiaopen/ooaf010> PMID: 40059976
81. Grundner J. Cross-Hospital Infrastructure for Research, Statistical Analysis and the Creation and Deployment of Statistical Models Based on Standardized Data of the German Data Integration Centers. 2022. <https://open.fau.de/handle/openfau/20771>
82. Alharbi E, Skeva R, Juty N, Jay C, Goble C. Exploring the current practices, costs and benefits of FAIR implementation in pharmaceutical research and development: a qualitative interview study. *Data Intelligence*. 2021;3(4):507–27. [https://doi.org/10.1162/dint\\_a\\_00109](https://doi.org/10.1162/dint_a_00109)
83. Semler SC, Wissing F, Heyder R. German medical informatics initiative. *Methods Inf Med*. 2018;57(S 01):e50–6. <https://doi.org/10.3414/ME18-03-0003> PMID: 30016818
84. Paillaud E, Soubeyran P, Caillet P, Cudennec T, Brain E, Terret C, et al. Multidisciplinary development of the Geriatric Core Dataset for clinical research in older patients with cancer: a French initiative with international survey. *Eur J Cancer*. 2018;103:61–8. <https://doi.org/10.1016/j.ejca.2018.07.137> PMID: 30212804
85. Rocca-Serra P, Gu W, Ioannidis V, Abbassi-Daloui T, Capella-Gutierrez S, Chandramouliswaran I, et al. The FAIR Cookbook - the essential resource for and by FAIR doers. *Sci Data*. 2023;10(1):292. <https://doi.org/10.1038/s41597-023-02166-3> PMID: 37208467

86. Grootveld M, Leenarts E, Jones S, Hermans E, Fankhauser E. Openaire and fair data expert group survey about horizon 2020 template for data management plans. 2018.
87. Dong H, Falis M, Whiteley W, Alex B, Matterson J, Ji S, et al. Automated clinical coding: what, why, and where we are?. *NPJ Digit Med.* 2022;5(1):159. <https://doi.org/10.1038/s41746-022-00705-7> PMID: 36273236
88. Venkatesh KP, Raza MM, Kvedar JC. Automating the overburdened clinical coding system: challenges and next steps. *NPJ Digit Med.* 2023;6(1):16. <https://doi.org/10.1038/s41746-023-00768-0> PMID: 36737496
89. Chen Y, Chen H, Lu X, Duan H, He S, An J. Automatic ICD-10 coding: deep semantic matching based on analogical reasoning. *Heliyon.* 2023;9(4):e15570. <https://doi.org/10.1016/j.heliyon.2023.e15570> PMID: 37151662
90. Newman-Griffis D, Fosler-Lussier E. Automated coding of under-studied medical concept domains: linking physical activity reports to the international classification of functioning, disability, and health. *Front Digit Health.* 2021;3:620828. <https://doi.org/10.3389/fgth.2021.620828> PMID: 33791684
91. European Commission, Directorate-General R and I. Cost-benefit analysis for FAIR research data – Cost of not having FAIR research data. Publications Office; 2018.
92. German Center for Diabetes Research. Withdrawn: DZD core data set - first version published at DZD website for internal use (obsoleted by DOI 10.21961/mdm:45923). 2024.