Acta Materialia 286 (2025) 120735

Contents lists available at ScienceDirect

Acta Materialia

journal homepage: www.elsevier.com/locate/actamat

Full length article

Management of reference data in materials science and engineering exemplified for creep data of a single-crystalline Ni-based superalloy

L.A. Ávila Calderón [a,*], Y. Shakeel [b], A. Gedsun [c], M. Forti [d], S. Hunke [e], Y. Han [a], T. Hammerschmidt [d], R. Aversa [b], J. Olbricht [a], M. Chmielowski [f], R. Stotzka [b], E. Bitzek [f], T. Hickel [a], B. Skrotzki [a]

[a] Bundesanstalt für Materialforschung und -prüfung (BAM), Berlin, Germany
[b] Karlsruhe Institute of Technology, Kalrsruhe, Baden-Württemberg, Germany
[c] Albert-Ludwigs-Universität Freiburg, Freiburg im Breisgau, Baden-Württemberg, Germany
[d] Ruhr-Universität Bochum, Bochum, Nordrhein-Westfalen, Germany
[e] RWTH Aachen, Aachen, Nordrhein-Westfalen, Germany
[f] Friedrich-Alexander-Universität Erlangen-Nürnberg, Fürth, Bayern, Germany

ARTICLE INFO

Keywords:
Creep
Ni-Based Superalloy
CMSX-6
Reference Data
Research data management

ABSTRACT

The identification of process-structure-property relationships of materials inevitably requires the combination of research data from different measurements. Therefore, the concepts related to FAIR (findable, accessible, interoperable, reusable) data handling, increasingly reported in literature, are particularly important in the materials science and engineering domain. However, they have not yet been integrated into a single, overarching methodological framework, particularly for reference data. Here, we introduce such a framework. Our concept covers data generation, documentation, handling, storage, sharing, data search and discovery, retrieval, and usage. Furthermore, we prototypically implement it using a real dataset with creep data of a single-crystal CMSX-6 Ni-based superalloy. The presented implementation is traceable and permanently accessible through open repositories. The individual elements considered in the framework ensure the functionality and usability of the data and, thus, the adherence to the FAIR principles. In conjunction with this, we present a definition for reference data of materials. Our definition underlines particularly the importance of a comprehensive documentation, e.g., on material provenance, data processing procedures, and the software and hardware used, including software-specific input parameters, as these details enable data users or independent parties to assess the quality of the datasets and to reuse and reproduce the results. Reference data that is managed according to the proposed framework can be used to advance knowledge in the materials science and engineering domain, e.g., by identifying new process-structure-property relations.

1. Introduction

The field of materials science and engineering (MSE) is broad and interdisciplinary, and so are its research methods. As a result, the data generated by MSE is highly heterogeneous. The multitude of data formats and storage options, and the often-incomplete documentation of research results and related testing or simulation and data processing procedures complicates the reuse of data. Hence, there is an effort to generate data according to the FAIR (findable, accessible, interoperable, and reusable) principles [1]. These principles contribute to making the data citable, discoverable, and available for reuse. In addition, data structures need to be harmonized to enable future machine-actionability [1]. The FAIR principles are a prerequisite for handling the research data management (RDM) data lifecycle, which includes the collection, analysis, preservation, sharing and reuse of the data [2].

Within MSE research projects, newly generated research data are usually compared to existing literature data. However, such data are scarcely available in some cases, due to, e.g., the particular experimental conditions at which they are acquired or the rarity of the material under interest. Besides, available literature data may either be unreliable or incomplete or inaccessible. In this context, trustful research data of high quality which are well-described, well-documented, and located in data

* Corresponding author.
E-mail address: luis.avila@bam.de (L.A. Ávila Calderón).

https://doi.org/10.1016/j.actamat.2025.120735
Received 25 July 2024; Received in revised form 29 November 2024; Accepted 9 January 2025
Available online 10 January 2025

repositories or archives providing long-term accessibility [3], can play a key role in serving as a benchmark, e.g., to verify new results. In this article, we introduce the concept of reference data in the MSE domain to refer to research data (experimentally generated or the result of simulations) that meets a particularly high-quality standard.

While the term reference data is used in various contexts [4–9], formal definitions are rather rare. In the context of science and technology, however, there are publicly available formal definitions [10–12]. In that regard, Lee [9] presents a comprehensive summary of the global activities regarding the term Standard Reference Data (SRD) including the activities of the United States (US) and the Soviet Union in the 1960s, and the national SRD program of Korea. The author highlights the published definitions for SRD and the scope of the respective programs. The mentioned definitions are provided by the International Vocabulary of Metrology (VIM) [10], the SRD program of the National Institute of Standards and Technology (NIST) in the US [8] and the Russian national SRD service (RNSRDS) [11]. In the case of the latter two definition for standard reference data are institutionalized by public law and a council of ministers' decree, respectively. In the case of NIST, in addition to covering the type of information and characteristics or properties that SRD can represent and the possible usages, the definition highlights that composition and structure of the substance of system must be known and the data must be critically evaluated for reliability [12,13]. In addition to these activities, where formal definitions are available, the National Institute for Materials Science (NIMS) in Japan is also actively involved in the topic, though not strictly using the term reference data. NIMS runs a data platform called DICE to "promote materials science research and accelerate materials development" which provides different services including databases [14]. Their data catalogue includes databases and materials data sheets that are used by industry as a fundamental source of reliable data to be used as a reference for design and materials selection related tasks [15–18]. Similarly, the NIST and Korean SRD programs and the RNSRDS provide SRD through their own databases, data centers, and dataset publications [11,19,20].

In Germany, the decree of the Federal Ministry of Economics and Climate Protection on the Bundesanstalt für Materialforschung und -prüfung (BAM) states: "In addition to the reference materials and processes [...], BAM also develops and provides reference data" [21]. BAM has an internal audit process that the data must undergo before publication. The audit criteria include aspects such as: the necessity of the data, the documentation of data collection, data accuracy, measurement uncertainty, and data structure, storage and curation. Examples of such data sets are given in [22–25]. Since these datasets were created and published well before the start of the conceptual considerations that will be described here, they do not necessarily represent reference data of materials as defined later in this article. In particular, the generation process did adhere to all the steps of the framework proposed in section 3. However, it should be pointed out that the documentation of the metadata of these datasets is already very extensive and the measurements are of high precision. The here cited existing definitions are rather widely formulated. They cover physical, or chemical, or biological properties of substances or systems [12] and have an emphasis on the metrological aspects of data quality [9–11]. Although these approaches indeed do not exclude the MSE domain and in some cases MSE-relevant data have been shared as reference data, e.g., by NIMS, a framework including a definition specific for the MSE domain and its related challenges in terms of a comprehensive documentation for data reuse and knowledge advancement, is still missing. Our concept, which aims to close this conceptual gap, is explained in detail throughout this article.

Strict adherence to RDM policies is especially important in case of reference data to ensure their high quality and *FAIRness* (i.e., to what extent the FAIR principles are met). Therefore, it is reasonable to develop a conceptual methodological framework for the generation, distribution, and utilization of such data. This article presents how such a framework and its implementation can look like. Even though some approaches to data *FAIRification* (i.e., the process of aligning data to the FAIR principles) and the creation of robust data ecosystems exist [26–32], and some previous works in this area have shared practical examples in terms of research data generation and downstream usage [33–37], currently no comprehensive framework or RDM concept for reference data covering technical and semantical interoperability aspects has been systematically presented to the MSE community. The here presented framework and related methods, recommendations, and implementation can close this gap by providing guidelines on (i) the creation of reference datasets following relevant experimental standards and using well-characterized materials and calibrated equipment, (ii) their distribution, including documentation, creation of FAIR Digital Objects, semantic representation, and creation of data schemas, and (iii) best practices related to, among others, data handling and versioning. To develop and elucidate the proposed framework, a prototypical example from the field of mechanical materials testing is used. The method of interest is creep testing, and the material of interest is a Ni-based superalloy (CMSX-6). Creep of high temperature materials is a representative use case in materials science. The complete description of the data can be correlated, e.g., to microstructure data (or to other research data) and thus contribute to a gain in knowledge and scientific advancement.

Creep is a time-dependent plastic deformation under a constant load at elevated temperatures, which, for metals, often means temperatures above 0.5 T_m, where T_m is the absolute melting temperature [38]. A creep test is a mechanical test method to characterize the creep behavior, in which a laboratory specimen with a defined geometry is subjected to a constant load at elevated temperature [39–41]. Its resulting elongation is recorded over a long period (hundreds, thousands of hours, or more) until failure [42]. Ni-based superalloys, such as CMSX-6, are metallic alloys known for their exceptional combination of high strength, corrosion resistance, and microstructural stability at high service temperatures [43]. Due to these properties, they have been widely used for demanding high-temperature applications like gas turbines in aviation and power generation. They are well characterized by a wide range of methods, and detailed data are available [44–51]. These superalloys are typically composed of more than 10 chemical elements that open a correspondingly wide space of optimization regarding the chemical composition. Moreover, the manufacturing process offers further degrees of freedom for material design. The involved combination of casting and heat/pressure treatment at different temperatures strongly influence the microstructures, which as a result can lead to polycrystalline or single crystal alloys with wide variations of microstructure properties like the ratio of the volume fractions of the present phases [52].

Creep data of Ni-based superalloys are well suited to showcase the proposed framework for reference data due to their technical relevance and huge optimization potential. Furthermore, creep data are invaluable from a sustainability point of view since they emerge from energy-intense experiments with runtimes of weeks to years at high temperature. Despite their importance, the documentation of published creep data is often insufficient to enable a straightforward reuse of existing data. For example, important metadata on the material (e.g., actual analyzed chemical compositions), the specimen's history (e.g., manufacturing process, heat treatment), and the test procedure (e.g., measuring and test equipment used and its calibration[1] status) are not always available. As an example, the need for a comprehensive documentation and a standardized data reporting in the MSE domain has

[1] Calibration: comparison between an equipment and an established standard to ensure traceability. Specifically, calibration is defined as an "operation that [...] establishes a relation between the quantity values [...] provided by measurement standards and corresponding indications [...] for obtaining a measurement result from an indication" [53] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, OIML, International vocabulary of metrology – Basic and general concepts and associated terms (VIM). Joint Committee for Guides in Metrology, JCGM 200:2012. (3rd edition), 2012.

recently been highlighted for the case of mechanical fatigue data [54].

Overall, this article presents in section 2 a general definition and further considerations for reference data in the MSE domain and in sections 3 to 8 systematically outlines a framework for their generation, distribution, and utilization. More specifically, section 3 presents a general description of the proposed framework and the related challenges, and the subsequent sections 4 to 8 address the individual steps involved. Overall, our approach is developed in strong interdisciplinary interaction and includes guidelines regarding data handling, including the data schema concept, which is presented in section 6, and can substantially enhance the data provision process in MSE-related research activities due to the wide variety of the required content. To elucidate how research data, and ultimately reference data, in the MSE domain can be handled using the proposed framework, we prototypically implement the presented concepts for the specific materials science example of creep testing of a CMSX-6 Ni-based superalloy, which includes the publication of a real reference dataset. This contribution aims to represent a best practice example on how to handle reference data and lays the ground for further contributions and collaborations that merge data from different subdomains to help advance knowledge in the MSE domain.

2. General considerations and definition of *reference data of materials*

Reference *products* such as reference *materials*, reference *organisms*, and reference *methods* are used to verify the reliability and accuracy of measurements, tests, and analyses of materials in their technical applications [55]. They can serve as benchmark for quality assurance in science, technology, and industry [56] and are supplied by various providers (examples given in [57]). A reference *material* is, for example, a material or substance of sufficient homogeneity with one or more characteristic values that are so precisely defined that they can be used to calibrate measuring instruments, to assess measurement procedures, or to assign substance values [58]. Certified reference materials (CRMs) are high quality materials which represent certain characteristics with high accuracy. In addition, uncertainty contributions, e.g., resulting from material inhomogeneity, are assessed in extended analyses, and provided in a certificate with information on the measurement uncertainty and traceability[2] of the characteristic values to a unit [53]. A reference *method* is a thoroughly documented and demonstrably mastered test, measurement, or analysis method for (i) quality assessment of other methods for comparable tasks, or (ii) characterization of reference materials including reference objects, or (iii) determination of reference values. More details on the definition and application of reference methods and the closely related *reference procedures* are available in [59].

Currently, as the new paradigm of data driven research is becoming increasingly established, the concept of *reference data* should be introduced into the group of reference products to enable the validation of emerging data generation, processing and analysis methods, and tools. In this article, we distinguish between *reference* data and *research* data. Research data are defined as "data collected, created or examined [...] to be analyzed or considered as a basis for reasoning, discussion or calculation in a research context, with the purpose of generating, verifying and validating original scientific claims that support the answer to some specific research question" [60]. In general terms, research data neither necessarily strive to achieve the highest quality possible nor aim for a complete assessment in terms of reliability, comparability, and accuracy. Regarding reference data, we report here for the first time an initial proposal (version 1.0) for an elaborated definition in the MSE domain [61]. The definition was developed within the project NFDI-MatWerk [62] (National Research Data Infrastructure in the MSE domain).

The here presented definition "*reference data of materials*" [61] complements the existing definitions for (standard) reference data1 [10–12] as it is specifically targeted for the MSE domain, in contrast to the existing wider approaches which also cover, e.g., chemical and physical related data. Indeed, by limiting the definition of reference data to MSE, our approach intends to stimulate the sharing of research and reference data within this domain. Besides, most of the actual approaches have an emphasis on the metrological aspects of data quality [9–12]. They do not explicitly address the issue of comprehensive documentation regarding, e.g., material's history, test setup, and data processing that our approach covers as prerequisite for the reuse of reference data in digital environments.

In the context of this work, reference data are, in principle, considered as measured or simulated data that must meet a particularly high-quality standard, not only in measurement precision but also in terms of comprehensive documentation according to certain standards. Furthermore, our definition [61] covers, among others, the following requirements: (i) reference data must be generated, maintained, preserved, and verified according to certain standards, and describe one or multiple materials that have either undergone a well-documented manufacturing process or are a reference material whose composition and structure has been fully characterized, (ii) reference data of materials can also be derived data from reference data. Furthermore, (iii) the standards for documentation, generation, maintenance, preservation, validation, and measurement need to be defined for specific material characteristics (e.g., creep properties) in terms of domain standards, which are the result of a steered community process, and should integrate as much as possible existing standards. Finally, (iv) reference data shall be distributed with a license (e.g., CC BY 4.0 or GPL [63–65]) which clearly determines the options of their future use. In addition to those requirements, we believe that reference datasets should also be versionable to document and track changes or additions.

A *dataset* is defined as "collection of scientifically related (depending on the research context) Research Data" [60]. Thus, a *reference dataset* (RDS) represents a specific collection of reference data. Our definition implies that in addition to the data obtained from measurements or from simulations, they contain *all* the contextual data-related information that is required to be reusable, e.g., as benchmark. In the case of experimental data, this information includes material provenance, data generation and processing, and test history. For simulated data, these criteria include physical and numerical methods, software, hardware, and digital samples including software-specific input parameters to facilitate reproducibility.

As a complement to the proposed definition for reference data of materials [61], Fig. 1 summarizes mandatory inputs and possible usage scenarios. Evidently, the usage scenarios elucidate why reference data are needed in this domain. As shown in the figure, reference data of materials consist of (content-wise) material-specific, method-specific, data itself-specific, and data processing-specific information. Usages of reference data include (i) the validation of own measurement and simulation procedures, e.g., by carrying out tests or analyses on a material for which sufficient reference data are available so that the results can be compared, (ii) the verification of individual experimental and simulation results that require a cross-check with reference data, e.g., in cases where test repetitions are not feasible or an unusual deviation from expected values has occurred, (iii) their application as best practice example for the documentation of research data obtained in experimental or simulation campaigns, and (iv) their use as precise input for computational materials science. While we use the example of creep testing of a metallic Ni-based superalloy, this concept is fundamentally applicable to all type of materials and testing procedures within the MSE

[2] The metrological traceability of its measurement results is established by means of a documented unbroken chain of calibrations, each contributing to the measurement uncertainty, linking them to an appropriate reference [58] ISO/IEC 17025:2017 General requirements for the competence of testing and calibration laboratories, 2018, p. 65.

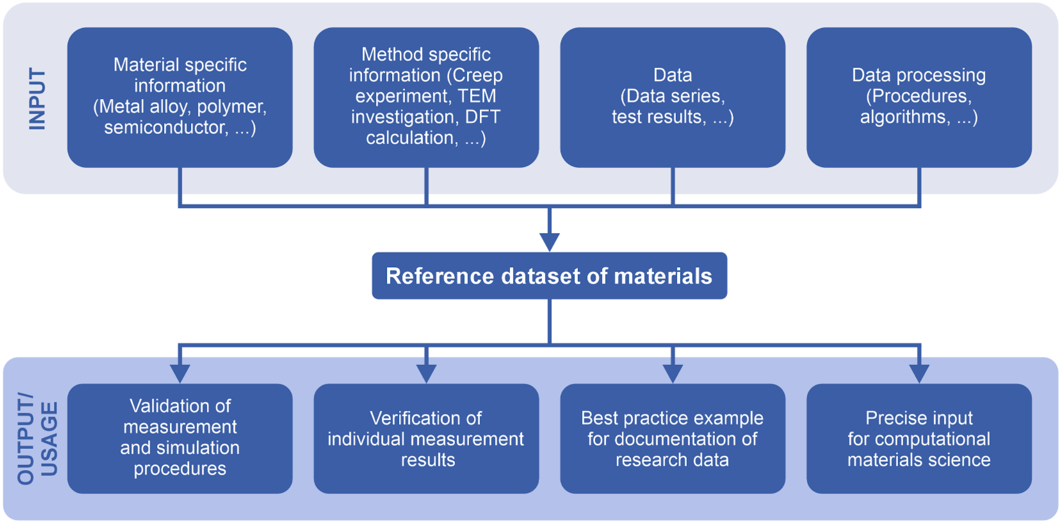

Fig. 1. Content and exemplary usage of reference data of materials. TEM: transmission electron microscopy, DFT: density functional theory.

domain.

In addition to the already mentioned aspects, reference data encompasses, in our view, two complementary dimensions: the content and the shape. The content dimension represents the scientific content from the MSE domain perspective. This dimension ensures the functionality and completeness of the dataset and provides the basis for answering specific research questions [66]. It requires an agreement to identify the criteria (in terms of content) for reference data in a community-driven process (e.g., (reference) material, equipment, procedures, results, required documentation). The shape dimension covers the RDM perspective further ensuring the alignment of the dataset and its contents (data, metadata) to FAIR principles [66]. This dimension necessarily includes digital formatting and storage to facilitate later discovery and repurposing.

The agreement of the community on criteria for reference data is a challenging task but of crucial importance for their acceptance and reusability. This agreement leads to a requirement profile (see section 6), which as a result represents the mandatory content that a dataset should comprise to be regarded as a reference dataset in the specific MSE domain, considering multiple intended usages and users. This requirement profile can serve as a guideline for data providers at creation time and helps distinguish reference datasets from research datasets in order to publish them with the appropriate classification. In fact, the requirement profile linked to a reference dataset implies high demands in terms of completeness and accuracy of the related documentation. Certainly, this level of quality cannot always be achieved by research data.

The differentiation between reference and research datasets and, for reference datasets, the introduction of a classification system regarding the level of fulfillment of the requirement profile will facilitate the process of discovering and selecting proper datasets for a given task. For instance, reference datasets generated using non-standardized test methods might only be useful for verifying measurement procedures (see Fig. 1) performed with a similar setup. In that sense, multiple quality classes will apply to reference datasets since these may differ, e.g., in the extent to which a testing standard was followed (see example above) or in the completeness of the given documentation. Such a classification system must also be agreed upon in the domain. In this regard, not every reference dataset needs to be derived from tests on a reference material or obtained according to an existing testing standard (e.g., the standard ISO 204 "Metallic materials – Uniaxial creep testing in tension – Method of test" [67] in the case of creep testing). This means that datasets created by different types of creep testing, e.g., the creep testing of miniaturized specimens [68], creep testing for a bending type of loading [69] or by small punch tests [70], or creep testing with accelerated testing strategies [71,72], could be classified as reference datasets, provided that they include all the required information agreed within the community. A subsequent analysis of available reference datasets would then allow to reveal the impact of the well-documented test details, e.g., the specimen size, on the reported test results.

Reference datasets could be classified of the highest quality class if all the measuring and test equipment or software and workflows used have been calibrated and/or validated and standardized specimen geometries have been used. Their use can be recommended for the calibration or the validation of measurement devices, procedures, or algorithms. If appropriate, the use of a CRM could also be required, which might lead to the creation of an additional class. Following up on that idea, a reference dataset of lower quality class could have less restrictive requirements for the description of measurement devices and procedures but could still be used for interpretation of individual measurement results or as precise input for machine learning-based data analytics and for computational materials science. In that case, however, all used data processing and analysis procedures must be extensively documented. Besides, note that comparing data between two laboratories still requires knowledge of the precision of the results. In general, reference datasets should adhere to the highest standards in terms of shape, e.g., using open machine-readable formats, to ensure an easy and smooth data reuse according to the FAIR principles.

To ensure long-term acceptance of this refence data concept by the MSE community, clearly defined decision rules and processes for the quality assessment and classification of datasets will be mandatory, as it has been implemented elsewhere with the focus on big data [9]. The details of the assessment process will depend on the agreed criteria, which are yet to be determined. It may be anticipated that the classification process will include automated procedures to evaluate, e.g., the completeness of the documentation. A fully automated quality assessment would, however, represent a major challenge. The process may involve an additional peer-review step to ensure the significance and applicability of the provided data, following the traditional approach in scientific publishing.

3. Framework and related challenges

While the previous section presents our definition of reference data and some general considerations, this section is devoted to our framework to generate, distribute, and utilize reference datasets of materials as well as the related implementation challenges. This framework, comprehensively encompassing the steps needed to fulfil all the

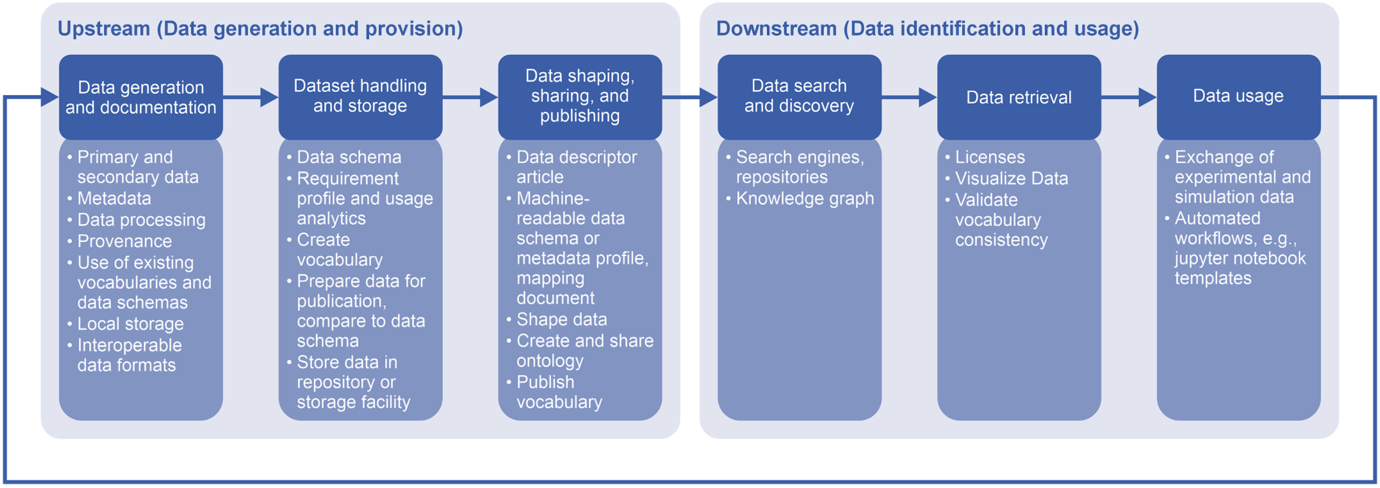

Fig. 2. Proposed framework for generation, distribution, and utilization of reference datasets of materials in the MSE domain.

requirements for reference data, is expected to have a strong impact on supporting research, development, applications and corresponding collaborations within the MSE community.

In Fig. 2, a simplified representation of the proposed framework is shown. It consists of six steps: (i) data generation and documentation, (ii) dataset handling and storage, (iii) data shaping, sharing, and publishing, (iv) data search and discovery, (v) data retrieval, and (vi) data usage. The first three steps, referred to in this article as the upstream part of the framework, are performed by the data provider, while the latter three steps, the downstream part, are performed by the data user. It is worth noting that the definition of reference data is a prerequisite for a successful upstream process.

Researchers in companies, research institutions, or universities can act as both data providers (upstream) and data users (downstream). In certain cases, a data user can act as data provider as well and generate new reference data, e.g., by using a reference dataset as precise input for computational materials science (see Fig. 1) for alloy optimization. This is indicated by an arrow that joins the last step with the first one in Fig. 2, suggesting a cyclic process. A third important player in the proposed framework is the provider of the digital infrastructure and services that facilitate both the upstream and the downstream processes and becomes especially relevant in the central four steps. In this general form, the proposed framework is independent of any specific technology, tool, service, or brand. Finally, prior to sharing the data, a quality assessment procedure to evaluate whether the dataset fulfills the criteria for reference data (of any class) in terms of content and shape (see section 2) should take place. Further details about such a process are outlined in section 9.

In the first step of the upstream scenario (see section 4), the data are generated and eventually processed, and all related data and metadata regarding material and specimen provenance, as well as measuring and test equipment including calibration, are documented following existing terminologies, where applicable. The comprehensive documentation of data will ensure their reusability, thereby substantially contributing to comply with the FAIR principles. Ultimately, the data are typically stored locally (on the data provider's personal hard drive or in the institution's storage system) in interoperable data formats.

In the second step, the data are methodologically organized and further stored in preparation for publication, as further detailed in section 5. This data handling goes along with the generation or the (re)use of data schemas (a structured representation of the community-agreed content and criteria for reference data) and a related requirement profile (which defines what descriptive information is mandatory). These concepts are addressed in detail in section 6. Evidently, the data documentation and handling process implies reaching a community agreement on content and vocabulary. Reaching this agreement represents the biggest implementation challenge. Nevertheless, beyond being merely a challenge, community action should be recognized as both essential and, in the long term, highly rewarding. In [73], it is even described as the driving force behind a revolution in materials science. Finally, the datasets are stored in a data repository or any other storage system (see section 5). At this point the data might be publicly available, e.g., if stored in an open repository.

In the third step, the data acquire a digital shape (RDM dimension, see section 2) and are (further) shared and published. Sharing and publishing can take place via open repositories and a data descriptor article. Examples of such articles can be found in [74–76]. To enable data to be handled during the downstream in a more automated manner, two complementary paths are considered: the implementation of FAIR Digital Objects (FDOs) to shape the data and the implementation of ontologies and semantic structuring. For both paths, a data schema serves as a basis for the structuring and shaping of the data (see section 6). These processes represent an implementation challenge from the perspective of an MSE researcher as they require a specialized RDM-skillset. Here, the support of RDM experts or infrastructure providers is highly beneficial. FDOs and ontologies play a crucial role in enhancing the discovery, retrieval, and exchange of data (and thus to successfully implement the FAIR principles) and are addressed in more detail in section 7. Although steps two and three are, for better representation, handled separately in Fig. 2, in practice some of the tasks, such as the creation of the machine-readable data schema and the ontology, might start in either of the two steps.

An essential and challenging task during all the upstream steps is the use of controlled vocabularies in the data documentation, in the data schema, and in the ontology, as this enhances the data interoperability and in turn facilitates the subsequent integration into the related digital infrastructure services. A controlled vocabulary is a list of terms or words, together with their definitions, which are organized using semantic relations (e.g., hierarchical, similarity or associative relations) in a standardized representation and in agreement with the domain experts [77]. The controlled vocabulary can be further extended to an ontology by implementing additional relations and logical constraints. An ontology is a formal and explicit description of concepts within a domain, where the information is organized in triples of an object (class), a predicate (relations, attribute, or properties), and a subject (class or instance of object), and logical rules and axioms are used [78]. Both the ontology and the controlled vocabulary should be finished and shared in the third step.

In the downstream scenario (last three steps in Fig. 2), the data user may search, discover, retrieve, and use the data. The data search and discovery might occur through search engines, data repositories [79], or repositories of repositories [80,81]. In perspective, a knowledge graph (KG) could also be used for data search and discovery. A knowledge graph is a structured representation of knowledge that is used to provide a comprehensive and interconnected view of a specific domain [82], with the instantiated ontology as its backbone.

The data retrieval and usage include the visualization of data, a manual or automated validation of the vocabulary consistency, and the data exchange and analysis. The data exchange and analysis could be performed through automated workflows. The FDO-scope [83] is a powerful tool that, in perspective, will enhance the data retrieval process by including the visualization of the single FDOs and of the related data (see section 7). By the data retrieval, the license will delimitate to what extent the discovered dataset can be reused. The reference data usage can be automated, e.g., through Jupyter Notebooks.

4. Data generation and documentation

Research data that are established by physical experiments may be categorized, from the MSE perspective, as primary, secondary, and metadata, depending on their origin, their processing state, and the type of included information. As demonstrated in section 6 through the data schema developed for our considered example of creep data, this categorization facilitates ensuring complete documentation of an experiment by systematically organizing the often-extensive available information.

Primary data refers to data that is directly acquired by sensors or measuring instruments during or after a test including, e.g., time, force, length, or temperature. This acquisition is carried out once or periodically within a certain interval, defined by (run-) time or other test parameter's change. The data may be acquired using a data acquisition unit and related data acquisition software. An automated integration of the acquired data in Electronic Laboratory Notebooks (ELNs) or other databases is possible. Although primary data are often considered to represent the unbiased "raw" data of an experiment, it should be noted that the contained information may already be influenced by, e.g., device internal processing or the selection of an insufficient acquisition rate. A strict definition of the term primary data does therefore not exist; more general, existing definitions of primary data would even allow normalization and/or harmonization steps [84].

The category *secondary data* refers to test results that cannot be directly measured. Instead, they are determined or converted by processing primary data and/or metadata by means of equations or algorithms. For example, the applied mechanical stress cannot be directly measured in mechanical testing. It needs to be determined by relating the actual force (which is often continuously measured) to the original cross-sectional area of the tested specimen. Primary and secondary data can represent single (characteristic) values (e.g., tensile strength) or data series.

The third data category, *metadata,* refers to all attributes and additional information characterizing the provenance and history of the investigated material and specimen as well as detailed information on the applied measuring and test equipment, and the data provenance.

Data provenance is connected to the fact that the generated data may be processed before publication, e.g., to identify characteristic values from an extended data series. As a general definition, data processing involves the application of any procedure (equations, algorithms, methods, unit conversions) used to transform data from its original into a new state. Our definition aligns with existing definitions, such as the one found in [60]: "set of actions [...] performed on research data to prepare it for one or more further processes". It is important to differentiate online processing from later, post-test data analyses. As an example, force transducers in mechanical (including creep) testing will typically provide a voltage output. The voltage signal is, however, not interpretable by third parties, even though it would indeed represent the "raw" data. Typically, based on a pre-recorded calibration relationship, the measured voltage is immediately converted to a force signal by the data acquisition unit during the running experiment. Additional analysis procedures can be performed later, e.g., to convert the recorded force values to mechanical stress based on the cross-sectional area of the test piece. It is important to identify and document all involved processing steps to allow data users to correctly interpret and/or re-analyze the published data, and to identify possible sources of error. Documenting the performed data processing procedures is required for reference data and is also part of the metadata documentation (see Fig. 2). The extension of these concepts to material simulations is also straightforward: technological properties are usually quantities derived from numerical or visual analysis performed after simulation runtime and after utilizing complex algorithms, and the reproducibility of the results depends on the documentation of such procedures.

Overall, a comprehensive metadata description includes: (i) a full description of the involved measuring equipment, including calibration information, (ii) information about the material type, its manufacturing history, and results of additional characterizations (mechanical, microstructural), and (iii) a comprehensive documentation of the applied test and data processing procedures. Comprehensive metadata records are mandatory for any later evaluation of the quality and/or reliability of the measurements. Most importantly, they allow one to search, identify, and retrieve best-matching datasets from large-scale databases for specific analyses.

Metadata, primary and secondary data shall be reported using, as much as possible, already existing agreed controlled vocabularies, e.g., from testing standards, and further normative documents. If measurement uncertainties are reported for test results, the calculation procedure must be similarly documented if not performed according to an existent standard.

The data generation and documentation processes go along with the (typically local) first storage steps, e.g., on the computer that controls the involved instrument, on the personal hard disk of the data provider, or on central data server facilities of the institute or the university. At this point, the data provider should already store and document the results in a well-organized manner. Open file formats, that do not require proprietary software, should be preferred . In this way, further structuring towards achieving a machine-readable interoperable representation can take place (see section 5). A description of how the data are organized (in a readme file, for instance), as well as the use of existing vocabularies whenever feasible, is important to ensure the reusability of the data in follow-up projects. Available data schemas may serve as a guideline to identify all relevant data, to organize the available pieces of information, and to agree on a common terminology from the very beginning. Data schemas require to be community-agreed to ensure their completeness, correctness, and the future reusability of the respective datasets by third-party users. The definition and role of a data schema within the proposed framework are introduced in section 6.

In our example of creep data of the single crystal Ni-based superalloy CMSX-6, the tests were performed in an accredited test laboratory using calibrated measuring equipment, and according to the standard ISO 204. A lever-arm creep testing machine was used, and the tests were performed at 980 °C under constant force and initial stresses between 140 MPa and 230 MPa. The as-tested specimens were single crystal investment casts from a vacuum induction refined ingot provided by PCC Airfoils, Inc., which were subsequently heat treated (annealed and aged). During the data generation and documentation process, one of the main challenges was the collection of the material related metadata, as the material was delivered in the 1990s, and the information was mostly available in non-machine-readable formats. At some points new data had to be generated, e.g., micrographs of the microstructure. A close collaboration with the testing engineer was the key factor to effectively gather all the relevant information during the collection of metadata on data processing, as well as on measuring and test equipment. Test results were determined according to the specifications in ISO 204. To document and structure all the generated and documented data, we used our own developed data schema as a template [85]. Details about this data handling and the data schema are given in sections 5 and 6, respectively. The dataset is available at [86].

5. Dataset handling and storage

Once the data has been generated and properly documented in a well-organized manner using interoperable file formats (see section 4), they must be further handled for publication. Besides, the storage of the data following the 3-2-1 backup rule [87–90] can be advantageous at this point. The data handling includes: (i) the creation of a case-specific data schema and a related requirements profile, (ii) the creation and usage of a case-specific vocabulary, and (iii) the final comparison and matching of the to-be-published results to both.

A data schema is a structured representation of the community-agreed content and criteria for reference data. Furthermore, it represents the basis for the further storing, shaping, and sharing of the reference datasets. A data schema includes a so-called requirements profile (see section 6), which represents the mandatory content of a reference dataset in the specific MSE domain and is directly linked to the intended usage. Therefore, a community-steered usage analytics process is required. Within this process, domain experts discuss and define the possible usages of the planned reference dataset and the related required data. To enable the data to be automatically processed during the downstream, adhering to the FAIR principles, and to promote interoperability, the experimental or simulated data must be structured and mapped to the case-specific data schema, which in turn must be available in a machine-readable format. Given its major significance, the data schema is treated in further detail in the context of the selected example in section 6.

Parallel to the creation of the case-specific data schema, the agreement on concepts and the creation process of the controlled vocabulary take place. The vocabulary must be defined to have consistency in the intended representations, which in our case are a JSON schema and a semantic representation (such as an ontology; see section 7). In other words, the agreement on the controlled vocabulary is a prerequisite for successfully creating the JSON schema and the related ontology. The vocabulary creation process relies on existing terminologies from, e.g., testing standards, further normative documents, and related literature or taxonomies. If this is not sufficient, the domain experts must work on a community-specific agreement. This could be done, for instance, in dedicated platforms or working groups. An example for this in the MSE domain is the working group for "harmonized terminologies and schemas for FAIR data in materials science and related domains" of the research data alliance (RDA) [91]. In its final form, the controlled vocabulary provides definitions and hierarchical as well as equivalence relationships between the terms. A vocabulary can be created using, e.g., the Editor for Vocabularies to Know Semantics (EVOKS) [92] and can be published, so that the single terms are accessible via a Uniform Resource Identifier (URI). In our case, the controlled vocabulary is represented in the Simple Knowledge Organization System (SKOS) format, providing a flexible lightweight semantic structuring [93]. The publication of the agreed terminology, which we propose to perform in the third step of Fig. 2 along with the publication of the ontology, enables gaining the attention of the broader domain-specific community, which might promote its involvement in the development process.

The last step of the data handling comprises the preparation of the reference data for publication and includes the final comparison and matching of the to-be-published results to both the case-specific data schema and related requirements profile and vocabulary. In our prototypical example of creep data for the CMSX-6 Ni-based superalloy, we do this by using the developed data schema [85] (see section 6) as a template for data documentation, structuring, and provision of all the relevant information and test results. Additionally, we provide LIS-type files, which are simple text files, containing a header with selected mandatory metadata and test results and columns with the required data series. Furthermore, we provide data about the measurement uncertainty and selected PDF files regarding the as-tested material and the test piece [86]. In this last step, it is important to double-check that community-common units are used and that the decimal point setting is based, if possible, on standard specifications and the resolution of the used measuring devices. Generally, required information can be embedded as images in image file formats (e.g., JPEG) or as text in PDFs (e.g., in technical drawings). However, all the required information (as defined in the requirement profile, see section 6), even when available in, e.g., a PDF, must be included as alphanumeric type of data, for instance in the header of a CSV file, an attached README file or in dedicated keywords in a JSON file. In our case [86], this is covered in the provided files, including the LIS-type files. In addition, at this point, if already existing, Digital Object Identifiers (DOIs) or FDO Persistent Identifiers (PIDs) can be provided (for details, see section 7).

Once the data schema is created using the controlled vocabulary, and the data are generated and handled (i.e., further prepared for publication), the data can be stored in (open) repositories that allow version tracking (like Zenodo [79]) or in storage facilities. In our case the data are published in Zenodo [86]. The usage of storage facilities has advantages in terms of storage capacity, preservation, or data management compliance with the FAIR principles, depending on the used facility. An example of such facilities is Coscine [94], which provides the possibility to share and manage the data according to the FAIR principles and is suitable for experimental as well as for simulation data. Additionally, this service allows the creation of so-called metadata profiles for new applications. For simulation data, similar approaches have been adopted in the last years. For instance, the NOMAD repository [95] provides parsers for a variety of simulation software that collect all the mandatory metadata including input files, software version, physical methods, and workflow steps with semantic connection between individual single point calculations, as well as interfaces in ELN format for some experimental techniques on the same principles. The Materials Project [96], on the other hand, provides a database for DFT calculations which can derive in reference data in the terms explained here. These calculations span over more than 150,000 materials and include results for many properties available from DFT like band gaps, enthalpies of formation, and elastic constants.

6. Data schema

Identifying the general criteria for reference data (see section 2) and later the contents of reference datasets for specific methods is a critical task (see sections 2 and 3). A data schema is a powerful element to convey this community-agreed information in one structured representation and is therefore of utmost importance within the reference dataset generation process. It represents the basis for the further storing, shaping, and sharing of the reference dataset. In the following paragraphs, we elaborate on relevant details of the suggested data schema approach.

In general, a data schema primarily consists of a list of any possibly relevant concepts for a specific method covering primary, secondary, and metadata as defined in section 4, thereby modeling the specific domain knowledge. These entries are grouped according to a basic categorization and a related hierarchy of concepts. The resulting overarching structure mimics the way a domain expert would structure the data of the characterization (or simulation) method of interest. Furthermore, for reference data, we propose that the data schema comes along with a requirement profile, which defines the minimum required information and the level of testing rigor that is mandatory to qualify a dataset as a reference dataset in the specific MSE domain. Fig. 3 exemplifies the basic structure of the data schema generated for the considered example of creep testing on single-crystal Ni-based superalloys. For the considered example, the "NFDI-MatWerk/IUC02 Data schema for creep data of Ni-based superalloys including a comprehensive documentation of test results and metadata", which was used as a template for data documentation, structuring, and provision of the CMSX-6 data in [86], is available in Zenodo [85] and in a git repository [97] both in CSV and JSON format. Further schemas for materials science, though not developed under the current project and under the considerations presented here, can be found in, e.g., [98].

To avoid any misinterpretation of the data schema, the individual relevant pieces of information (not shown in Fig. 3) within each category (I to IV in Fig. 3) should be delivered with the respective units, if applicable, and exemplary answers. In principle, one or multiple requirement profile(s) can be added to the schema to indicate the minimum required information for certain purposes. For reference data, we propose that the requirement profile differentiates only between "mandatory" and "optional" fields, avoiding a "recommended" type of grading.

The presented data schema for the considered example [85] was developed based on the international ISO 204 standard for uniaxial creep testing of metallic materials [67]. The minimum required information and level of testing rigor that needs to be demonstrated for the considered example of creep data of Ni-based superalloys is defined by the mandatory fields of the related requirement profile. Special care was taken to ensure that the vocabulary, symbols, units, and recommended information and results to be reported, which are defined within ISO 204, are similarly used in the developed data schema. However, to meet the community-agreed requirements for the considered example, the data schema goes indeed beyond the recommendations of the standard ISO 204, which is rather focused on the details of the test method (the description of the test equipment and test results). In that sense, our data schema additionally includes information regarding the description of the material's manufacturing history and the microstructure. This is especially important for the considered example of CMSX-6 single-crystal superalloy, as no reference material was used. Our thorough characterisation and documentation of manufacturing history and microstructure includes references to literature [99–101] and covers, e.g., the chemical composition and the single crystal orientation (deviation of [100] axis) and related measurement procedures [99], and the description of the manufacturing process, and heat treatment procedures, including protocols, when available. However, it does not cover in detail the description of, e.g., micrographs or chemical composition measurement procedures, as this is beyond the scope of the developed data schema. Instead, this information is considered metadata itself for the CMSX-6 creep data and thus included as individual entries of the data schema. Nevertheless, special care was taken to provide metadata for these and, in general, for all metadata entries. In the future, these fields could include links to other data schemas (e.g., on scanning electron microscopy data) or related data publications. The longer extension of the data schema compared to the ISO 204 standard is reflected in the number of entries: it consists of more than 200 individual pieces of information that may be documented for a creep experiment, while in ISO 204 less than 15 items are listed as mandatory elements of a test report. The usage of the developed data schema as a template in the publication of the CMSX-6 reference dataset [86] can be considered as a best practice example for generation and documentation (cf. input part of Fig. 1).

As depicted in Fig. 3, the concepts of metadata, primary data, and secondary data introduced in section 4 were applied as the first level of categorization. The three subsequent levels of the data structure displayed in Fig. 3 classify the available information in further detail. The metadata part exhibits the most complex hierarchical structure, covering a wide range of information from material history to details of the test setup and data processing. While parts of this metadata do not seem to be closely related to the creep experiment, they are still of central importance for the future reuse (represented by "R" in FAIR) of the datasets. For instance, any later study of the impact of impurities on the creep behavior of the considered alloy CMSX-6 would require not only the chemical analysis of the tested batch but also information on the heat treatment or the distribution of the impurities in the microstructure.

Also, with respect to the future use of the datasets as reference datasets, special emphasis was placed on the detailed documentation of all aspects that could possibly impact the accuracy of the experimental results, as also at least partly considered in the already existing approaches for SRD mentioned in section 1. This includes a full description of the whole test setup, including manufacturer information and model numbers in case of commercial products, since these allow to obtain further technical details from independent sources, e.g., measurement ranges or resolution. Furthermore, a statement about the metrological traceability, and information about the calibration status of the test rig, including alignment checks and all sensors, is described in detail to allow more elaborated cross-checks, e.g., on the match between observed loads and the calibration range of the load sensor or the fulfillment of the calibration intervals of a thermal sensor as defined in the ISO standard. Beyond the existing approaches, full disclosure of the test piece geometry and the location of the test piece within the as-tested material from the original material, to be provided preferably as technical drawings, is included to assess the suitability of the applied sensors and/or support the interpretation of obtained local deviations in material properties.

A similar attempt is made to document all data acquisition, processing, and analysis steps. Compared to the physical details of the test setup and the test procedure, such aspects are typically covered to a lesser extent in the testing standards. Therefore, in this context, a generalized format is not required, and the information is currently requested via free-text boxes. It is acknowledged that this represents a limitation for any future automated analyses of the datasets, which would presumably be restricted to assessing their completeness until more advanced techniques, including, e.g. natural language processing, become available. Similarly, the variety of data types and input formats, ranging from single numerical results like a rupture time to technical drawings or free-text, represents a challenge for the transformation of the datasets into digital objects, see section 7. However, the data schema presented in [85,97] represents a first approach for collecting all the information that allows an independent, third-party evaluation of the quality of reference datasets. Depending on the usage of the data, this quality assessment may be focused on different aspects and metrics, e.g., the completeness of documentation or the level of measurement accuracy, as outlined in section 9.

It should be pointed out here that *research* datasets will typically not contain all the information expected in the data schema. As already discussed in section 2, such datasets will possibly fail to meet the requirement profile defined for reference data but may still serve as valuable contributions for less demanding research tasks. Overall, one intention behind the suggested data structure (Fig. 3) is clearly to encourage all domain experts to collect and provide as much information as possible and to consider extending the documentation of their experiments in future test campaigns.

The chosen hierarchical categorization offers the advantage of making incrementally more specific information easy to identify by locating it at multiple description levels. This modular structure, with nested groups, serves as the backbone of a comprehensive documentation and the foundation for further storing, shaping, and sharing of reference datasets. The flexibility that a modular structure ensures allows for the creation of additional data schemas for other MSE experimental and simulation domains, requiring, e.g., only adjustments to the method-specific parts. Thus, the provided data schema is useful not only for reference data but also as a template for organizing *research* data in general in daily laboratory practices.

When creating a data schema for the reference dataset generation process, different scenarios are possible. In case the reference dataset has already been published for the respective method (i), the data provider can use the existing schema, without modifications. In the opposite case (ii), when a reference dataset is generated for the first time for a given method, a data schema must be generated, but existing data

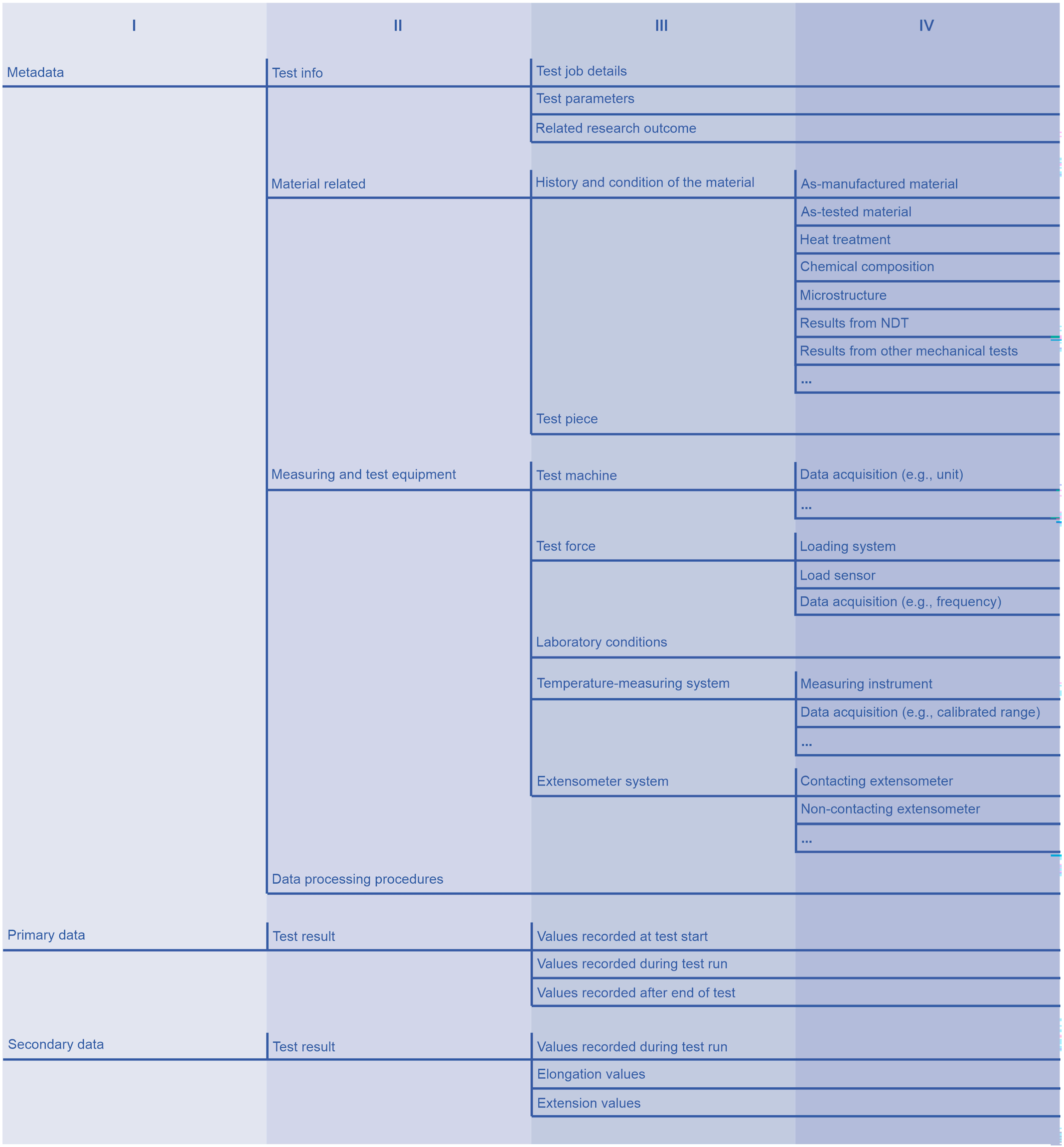

Fig. 3. Data schema structure generated for the chosen example (creep data of Ni-based superalloys).

schemas, such as the one presented here, can be used as reference. The intermediate case (iii) implies the partial modification of existent data schemas to create a new version. The latter scenario is the most expected one in the future, as the process of generating reference datasets evolves, and the proposed framework becomes established. In cases (ii) and (iii), for reference data, a community agreement is needed for tailoring the data structure, content, and requirement profile.

The community agreement on content, structure, and requirement profile, as mentioned in section 3, is a challenging and time-consuming task. New categories and related individual entries can, in principle, always be included if a corresponding agreement is reached and the data schema is correspondingly updated. Also, new releases of the respective testing standards may include variations of the experimental methods or modifications of the vocabulary which require an update of the data schema. Therefore, the data schema must be understood as a product of a dynamic development process and a service from and for the community. Each new release of the data schema requires a community process where domain experts (data providers and data users) work together. This challenging generation process makes the adoption of a versioning system of the data schema essential. Reference datasets that were generated based on a specific data schema version can, however, continue being used provided that such a version is clearly documented.

Finally, for reference data as defined in this article, the data schema should be aligned to a domain-level or application-level ontology and the related controlled vocabulary, as this helps enhance machine-assisted data search and discovery (see section 3). From the downstream usage perspective, the same data schema may be mapped to different ontologies. However, special care should be taken to keep the vocabulary of both variants aligned to avoid any misinterpretation of information items in the automated data discovery and analysis.

7. Data shaping

The processes that lead to creating a reference dataset in fulfillment of all required content and shape quality standards (see section 2) are illustrated in Fig. 4. The proposed data journey contains the steps of the data workflow that are needed to create a reference dataset from the RDM perspective. These steps are valid for both experimental and simulated data and represent a technical workflow not only to generate but also to search and discover reference datasets. For the considered example of creep data of the CMSX-6 superalloy, the implementation of the concept presented here is a dynamic, ongoing development and is continuously documented in a git repository available at [97]. For research data, some previous practical, standard-compliant, efforts partly addressing the concepts presented here can be found at [33,37, 102].

The data journey starts with the data provider, who generates, documents, and organizes the reference data, as described in section 4. This is shown by the two white boxes at the top of Fig. 4. At this point, the data are stored in open, non-proprietary, and machine-readable formats, such as CSV or LIS, and uses vocabulary accepted within the domain experts. Besides, the data and metadata are stored in an accessible location, preferably in a research data repository with long-term archiving (see section 5). This storing is also a prerequisite for the subsequent creation of FDOs [103].

Afterwards, the data journey continues with two possible, complementary paths, represented by blue and green boxes in Fig. 4. The blue boxes represent the data shaping workflow (third step in Fig. 2) to create FDOs for a reference dataset, while the green boxes illustrate the ontology development and the implementation into a KG for data search, discovery, and retrieval (fourth and fifth steps in Fig. 2), respectively. FDOs add a layer of interoperability and provide a way to technology-independent search and access to data [104]. A FDO is a high-level common representation of a data object. Its unique structure allows machine-interpretability enabling technical interoperability as depicted in the left side of Fig. 4, whereas semantic interoperability is achieved by common ontologies as shown on the right side of Fig. 4. If data providers adopt the proposed framework, FDOs interoperability can be possible. This means that if FDOs that are created in line with the corresponding community accepted data schema, vocabulary, and ontology, only a technical step would be required to enable their interoperability or mutual interpretation. For instance, mappings from locally implemented data formats and data structure to the community accepted ontology would have to be developed for a seamless communication between the parts. However, this would not be directly related to FDO technology but rather to ontology implementation. A KG enhances the search and discovery of data, and the alignment to related ontologies ensures semantic interoperability, thus opening possibilities for knowledge creation and discovery. In our vision, the search and discovery of the data, which could happen through search engines, repositories, or the NFDI-MatWerk KG [105], will include a search and visualization of FDOs and will be supported by the application- or domain-level ontologies.

Following the blue boxes in Fig. 4, the process starts with converting

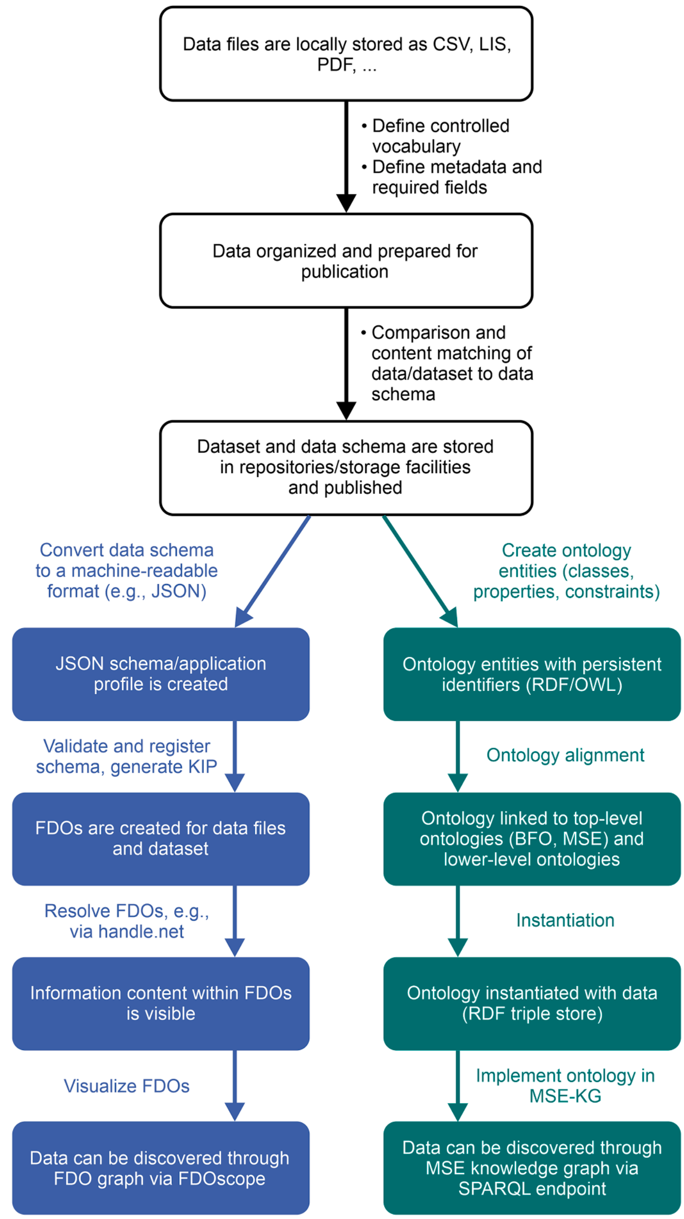

Fig. 4. Schematic view of the data journey, outlining the conversion of an original data file to its structured digital representation. See text for details and abbreviations.

the data schema into a machine-readable and actionable format, in our case JSON[3]. The "JSON schema", for our implementation example available under [107], consists of the community-agreed properties and hierarchy. In addition, those information parameters that are considered mandatory for the dataset (i.e., requirement profile) are defined. Overall, the JSON schema provides a structure to organize the information parameters from the experimental data files, retains the information regarding mandatory fields, and defines the information type.

From the downstream perspective, to automate the process of creating JSON files with content, a mapping document (see step 3 in

[3] JSON stands for *JavaScript Object Notation* and is a lightweight, text-based, language-independent data interchange format; JSON objects are formatted using a relatively simple key/value pair structure that is easy for humans and machines to read and write [106] M.C. Munro, Using JSON, in: M.C. Munro (Ed.), Learn FileMaker Pro 19: The Comprehensive Guide to Building Custom Databases, Apress, Berkeley, CA, 2021, pp. 319-330.

Fig. 2) should accompany the schema. The mapping document provides a path for entering the information content directly into the JSON schema and generating new JSON files with the content. This mapping process and related documents can be performed in several ways, e.g., through a python [108] script. Since it is primarily relevant for the downstream process, this mapping process is not explicitly shown in Fig. 4, which rather focuses on the upstream perspective. Furthermore, the exact implementation depends on the specific research needs of the data user.

Following the data journey, the machine-readable version of the data schema (e.g., in JSON) is then validated and registered. The validation should be performed in at least two perspectives before it can be registered (i.e., published) in a repository (for instance, MetaStore [109]): (i) a rather technical validation, where online JSON validators, such as [110] can be used for checking the syntax, and (ii) a validation of the content, where domain experts can be involved. After validation of the JSON schema, the FDOs can be created.

Technically, FDOs are a sequence of bits, identified by (i) a globally unique Persistent Identifier (PID) and (ii) an unambiguously defined information record that points to the data and metadata required to comply with the FAIR principles. PIDs (i) are alphanumeric sequences typically accompanied by a PID information record, which is a key-value map for holding the PID-related information.

The PID of an FDO allows the identification of the digital content (data and their metadata) in a globally unique and persistent manner. Each FDO resolves to a machine-readable PID information record containing the kernel metadata (i.e., metadata for machines to make decisions). A Kernel Information Profile (KIP) defines the kernel metadata entries and their types. The KIP acts as a predefined template that is made globally available using the Data Type Registry [111] and is the key element to make digital content machine actionable [81]. The PID information record holds machine-readable information content, defined by the KIP, including the keys that are supported, their cardinality, and value ranges for each key [112]. To generate a KIP, all mandatory attributes are defined with their correct data type, and the PID information record is registered via a PID service using a PID-Maker [113].

To create FDOs, a PID information record is created using the KIP [114]. For each FDO, a Kernel Information Record is filled and stored with a PID in the PID Handle Record, which can be represented and downloaded as JSON using a REST API [115]. The created FDOs can be resolved via a PID service, for example [82]. The Handle system is a federated system used for generating PIDs. The GWDG (*Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen*) [116] hosts handle servers that are connected to this federated system. The GWDG servers (and their prefixes) are part of the federated handle system and are resolvable using any handle resolver (similar to DOIs). The created and resolved FDOs can be visualized using the FDO-Scope [117] or in combination with a graphical user interface (GUI) like Coscine [94].

All the steps covered by the blue path in the data journey (in Fig. 4) can also be performed by RDM platforms such as Coscine (cf. section 5), which, to adhere to the FAIR principles, offers upstream data management across multiple levels.

Parallel to the representation of the FDO creation, the ontology development process to ensure semantical interoperability is represented in the data journey (green boxes in Fig. 4). Thereby, the scope (top,-application- or domain-level) and requirements of the ontology must be defined. An overview of ontology-related concepts is presented in [32]. This is done by defining the intended use-cases, the intended end users, the competency questions, and the language, e.g., Resource Description Framework (RDF) [118] and the Web Ontology Language (OWL) [119]. The controlled vocabulary is created in consideration of the data schema and the existing taxonomies or ontologies in the domain, which represent the same or similar concepts. Definitions for the terms are included to avoid any misinterpretation or to identify similar concepts. The controlled vocabulary is then further developed towards an ontology, which is represented in OWL, which is more expressive than SKOS [93] and RDF [118].

An important aspect of ontology development is to link the ontology concepts to other domain, application, or top-level ontologies. A standalone ontology can work for a use-specific application, such as retrieving knowledge from one specific document. To ensure the interoperability of the ontology for reference datasets (exemplarily described on a creep dataset) the developed ontology will be linked, inter alia, to the top-level Basic Formal Ontology (BFO) [120], which represents general concepts, the MSE-domain related ontologies developed within the NFDI-MatWerk consortium, such as the MatWerk Ontology (MWO) [121], the Platform Material Digital (PMD) Core Ontology (PMDco) [122,123] and to the application-level ontologies for tensile test (TTO) [124,125] and for heat treatment (HTO) [126], which were developed within the PMD [127]. For the specific example of creep testing a respective ontology involving the general concepts of these higher-level ontologies is under development.

Finally, the classes can be instantiated or filled with actual data and, in perspective, could be implemented into the NFDI MatWerk KG [105]. The KG will be able to interpret the data with the help of the concepts defined in the ontology and ultimately provide a SPARQL endpoint allowing access to data or metadata (e.g., on the calibration class) via specific queries.

8. Data search, discovery, retrieval, and usage

Research and development-related activities start with a research question as motivation. Usually, the first step towards answering this question is taken by performing queries in search engines or dedicated databases. The final step is usually the sharing of the results. The steps in between include the data search and discovery, retrieval and access, usage, and citation, which can lead to generation of new data, in which case, the newly generated reference dataset would ideally be published in a dedicated database together with all necessary metadata and in the form of an FDO. We globally refer to this process as a user journey.

Within the user journey, we distinguish three roles: the data user, the data provider, and the infrastructure provider. As mentioned in section 3, researchers in companies, research institutions, or universities can act as data providers, and a data provider can be also a data user. An infrastructure provider is an organization or entity that offers the necessary technological resources and services to support data storage, processing, sharing, and preservation. The role of the data infrastructure provider in the user journey is to provide technical solutions tailored to the needs of both data user and data provider allowing them to discover, access, cite, share, and publish data. In addition, infrastructure providers are responsible for funding and maintaining the infrastructures that support sustainable RDM and serve as technology stewards, raising awareness about the importance of managing and sharing research data [128,129]. Providing a solution for every single step of the user journey might imply requirements for the infrastructure provider. For instance, they should implement notifications of missing required metadata and data quality quantifications. Publishing data usually requires the selection of a licensing framework, and the data provider should also get support from the infrastructure provider about the implications of the selected option.

For the considered example of creep data of the CMSX-6 Ni-based superalloy, Fig. 5 shows two variants of a user journey, each of them driven by an exemplary use case. The use cases have different research goals, which imply the need for specific queries and data infrastructure requirements to address the research tasks (see first step box). The implementation of the here presented use cases is a dynamical, ongoing development and is continuously being documented in a git repository available at [97]. For research data, some downstream usage practical examples using ontological concepts can be found at [33–36].

In the first use case, a data user wants to search for reference data to assess the current state of a newly generated dataset after shaping and structuring it. The dataset is probably stored locally in the data user laboratory using in-house conventions and equipment. The data user starts the user journey by querying databases for datasets derived from the same experiments on similar or same material. As a result of the query, the data user gets a FDO from which the relevant data and metadata can be retrieved. To benefit from the advantages of a complete comparison, the locally stored dataset description needs to be aligned to the same ontology and/or data schema as the content of the FDO. An efficient way to achieve this task is to apply mappings between them. If not already existing, python scripts or Jupyter Notebooks can be implemented and made available for future use. Finally, a direct comparison between the datasets can be performed. From this comparison, data users can either draw new conclusions or decide to modify and reorganize their data to explore additional scientific questions and conduct further analysis and processing.

In a second use case, the data user is interested in providing evidence supporting, e.g., a working thesis. For this, the data user queries the databases specialized in creep experiments for all the available data on superalloys. This time, due to the nature of the performed query, the result consists of a list of FDOs. Data and metadata can be extracted in batches from such FDOs, and the user is able to perform correlation analysis between different quantities, e.g., primary, and secondary data. In this case, the task will be facilitated if the contents of the FDOs were already aligned with a unique ontology and data schema and using the same vocabulary. Otherwise, the data user will need considerable effort to find mappings between possible different FDOs. In this case, new primary data are not being generated, but new insights are brought into the domain after the complete analysis.

The highlighted use case scenarios show how reference data that is managed according to the proposed framework can be used to advance knowledge in the MSE domain, e.g., by developing new structure-property relations like relations between segregation and yield strength [130], or consistent predictions of creep lifetime [131]. Furthermore, it is clear in both use cases that the completeness of the retrieved FDOs is essential for a seamless workflow. The FDOs should then comply with the full description of the experiment of interest. This

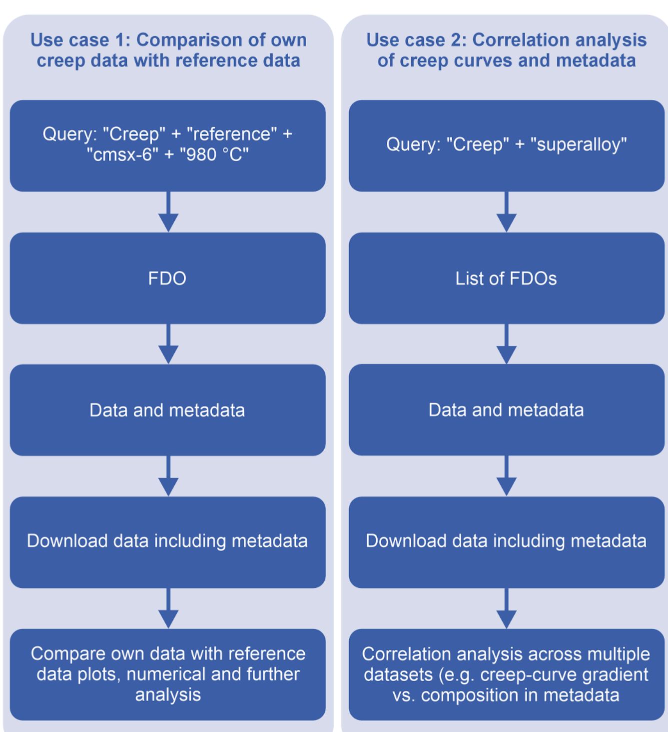

Fig. 5. Possible user journeys for the considered example of creep properties of Ni-based superalloys.

description is given by the data schema and the related controlled vocabulary. In this section, we presented two examples about FDOs, but it is worth noticing that the semantics-based KG approach would represent an alternative path.

9. Summary and outlook

In this article the concept of *reference data of materials* was explored in detail as part of the National Research Data Initiative for MSE, NFDI-MatWerk, in Germany [132]. A definition is proposed, and a framework for the management of such data is presented and prototypically implemented for an important and specific materials science example. The presented implementation is traceable and permanently accessible through open repositories. It includes the publication of a related dataset [86] and lays the ground for future collaboration and expansion of the concept. MSE reference data that is handled according to the presented framework can be used to reliably develop new structure-property relations like relations between segregation and yield strength, or consistent predictions of creep lifetime, which will be of relevance for determining component-relevant properties, e.g., the lifetime of turbine blades.

Reference data in the MSE community are research data that are characterized by high quality in terms of the precision of the experimentally determined or simulated data combined with comprehensive documentation of, e.g., material provenance, data generation and processing, and test history. Our framework concept addresses the generation of reference data, an agreement on the necessary content, and its digital representation with FAIR digital objects. The outlined framework also includes integration into an easily accessible digital infrastructure for annotation, discovery, and dissemination of datasets using a data schema, ontologies, and a controlled vocabulary.

The data schema, a crucial element of the proposed framework, is presented in this article for the example of creep data of Ni-based superalloys [85] and used as a template for the best practice example of reference data of a CMSX-6 Ni-based superalloy [86]. It is a structured representation of all potentially relevant information for a specific experimental or simulation method in the materials science and engineering domain and forms the basis for further storage, shaping, and sharing of the reference dataset. With the help of an application ontology, the reference dataset can be formally described adopting the digital representation and reuse of shared concepts from domain experts. The ontology is aligned with the data schema and with existing ontologies. The interplay between the digital representation with FAIR digital objects, the agreement on necessary metadata and related data schema, and an underlying ontology ensures functionality, usability, and technical and semantical interoperability. These data and results can be easily and efficiently accessed, shared, taken up, and reused, fulfilling therefore the FAIR principles.

Further work on the presented reference data concept is necessary to refine its details and establish the required procedures for its continued development. This includes the definition of the minimum requirements for reference data regarding the most important test and analysis methods in the field of materials science and engineering, e.g., electron microscopy or tensile testing, for which efforts towards *FAIRification* have been recently shared [37,98,133,134] and which, to fit into the proposed framework, could also be expressed in respective data schemas as outlined for the example of creep testing in this article. It is assumed that the parallel development of the digital infrastructure will allow these requirement profiles to be provided as input masks for the various Electronic Laboratory Notebooks and Laboratory Information Management Systems used in the materials science and engineering domain. Leveraging the presented framework to, e.g., microstructural or simulation data, will increase the impact of the presented concept and help to advance knowledge in the MSE domain.

A crucial element for the long-term acceptance of the presented reference data concept is the introduction of suitable metrics and methods for the quality assessment and classification of reference datasets. These must at least involve the following criteria: (i) status and transparent description of the involved equipment, including statements on the sensor calibration and traceability (metrological requirements), (ii) suitability and concise documentation of all applied procedures during the experiment or simulation and the subsequent data processing, if applicable including a statement on which agreed procedures, like ISO or ASTM standards, etc., have been followed (procedural requirements), (iii) complete description of the material's provenance, including suppliers and processing history or current material condition (material requirements), and (iv) comprehensive documentation of all further relevant details of the investigation and the subsequent shaping of the dataset (documentation requirements). Similar requirements apply to simulation data.

As outlined in section 2, not all criteria must be fulfilled by a reference dataset, depending on the intended usage. The classification system proposed there should take this aspect into account. To cover all possible scenarios, a deep involvement of the materials science and engineering community in the development of such a classification system is required. In addition to the definition of classes, suitable classification procedures need to be identified. To date, it is expected that a combination of automated procedures (checking the completeness of the datasets and their integrity, e.g., calibration ranges which match the measurand) and a peer-review step will be most beneficial to ensure the accuracy and significance of published reference datasets. The NFDI-MatWerk consortium [62] provides a suitable interaction platform for discussing such challenging tasks and for developing guidelines for the content and shape of reference data, and for infrastructure components to share them.

CRediT authorship contribution statement

L.A. Ávila Calderón: Writing – original draft, Writing – review & editing, Visualization, Methodology, Investigation, Conceptualization. **Y. Shakeel:** Writing – review & editing, Writing – original draft, Visualization, Investigation, Conceptualization. **A. Gedsun:** Writing – review & editing, Writing – original draft, Visualization, Investigation, Conceptualization. **M. Forti:** Writing – review & editing, Writing – original draft, Visualization, Investigation, Conceptualization. **S. Hunke:** Writing – review & editing, Writing – original draft. **Y. Han:** Writing – review & editing, Writing – original draft, Visualization. **T. Hammerschmidt:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Methodology, Investigation, Conceptualization. **R. Aversa:** Writing – review & editing, Writing – original draft, Supervision, Conceptualization. **J. Olbricht:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Methodology, Investigation, Conceptualization. **M. Chmielowski:** Project administration. **R. Stotzka:** Supervision, Conceptualization. **E. Bitzek:** Supervision, Conceptualization. **T. Hickel:** Writing – review & editing, Supervision, Conceptualization. **B. Skrotzki:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Investigation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was carried out in the framework of the consortium NFDI-MatWerk, funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under the National Research Data Infrastructure – NFDI 38/1 – project number 460247524. Mariano Forti and Thomas Hammerschmidt acknowledge financial support by the DFG through project C1 of the collaborative research centre SFB/TR 103 (project number 190389738). Rossella Aversa acknowledges financial support by the EU's H2020 framework program for research and innovation under grant agreement n. 101007417, NFFA-Europe Pilot Project. All authors acknowledge Klaus Neuking and Gunther Eggeler from the Institute for Materials, Ruhr-Universität Bochum, Germany, for fruitful discussions on the detailed description of creep test setups and creep data of single-crystal superalloys. Finally, we would like to thank Silke Richter from Bundesanstalt für Materialforschung und -prüfung (BAM) for the fruitful discussion on the definition of reference data of materials.

References

[1] M.D. Wilkinson, M. Dumontier, I.J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L.B. da Silva Santos, P.E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C.T. Evelo, R. Finkers, A. Gonzalez-Beltran, A.J.G. Gray, P. Groth, C. Goble, J.S. Grethe, J. Heringa, P.A.C. 't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S.J. Lusher, M. E. Martone, A. Mons, A.L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M.A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, B. Mons, The FAIR Guiding Principles for scientific data management and stewardship, Sci. Data 3 (1) (2016) 160018.

[2] S. Nakhaie, A.E. Mansour, K. Helbig, M. Bierwirth, C. Draxl, M. Aeschlimann, FAIRmat Guide to Legal Aspects in Research Data Management, 2024. https://doi.org/10.5281/zenodo.11083303.

[3] A.E. Mansour, L. Rotheray, K. Helbig, S. Botti, H.B. Weber, M. Aeschlimann, C. Draxl, FAIRmat, Guide to Writing a Research Data Management Plan, 2023. https://doi.org/10.5281/zenodo.7936477.

[4] EU Vocabularies. https://op.europa.eu/en/web/eu-vocabularies/semantic-knowledge-base, (accessed 2024-07-24.).

[5] G. Firican, Reference data needs a standard definition – here is one. https://www.lightsondata.com/reference-data-needs-standard-definiti, (accessed 2024-07-24.).

[6] S. Baumann, What the heck is reference data? https://jobs.natwestgroup.com/posts/what-the-heck-is-reference-data, (accessed 2024-07-24.).

[7] K.D. Foote, Master Data vs. Reference Data. https://www.dataversity.net/master-data-vs-reference-data/, 2024 (accessed 2024-07-24.).

[8] NIST Standard Reference Data. https://www.nist.gov/srd, (accessed 2024-07-24.).

[9] D. Lee, Big data quality assurance through data traceability: a case study of the national standard reference data program of Korea, IEEe Access. 7 (2019) 36294–36299.

[10] IEC BIPM, ILAC IFCC, IUPAC ISO, OIML IUPAP, International vocabulary of metrology – Basic and general concepts and associated terms (VIM), in: Joint Committee for Guides in Metrology, 3rd edition, 200, JCGM, 2012, p. 2012.

[11] K. A. Yu., K. AD, Status and results of russian national standard reference data service (50 years since foundation). Factory laboratory. Diagnost. Mater.. (2018) 84(2):73-76.

[12] SRD Definition Per Public Laws. https://www.nist.gov/srd/srd-definition, 2021 (accessed 2024-10-31.).

[13] 15 U.S. Code § 290b - Collection, compilation, critical evaluation, publication and dissemination of standard reference data. https://www.law.cornell.edu/uscode/text/15/290b, (accessed 2024-11-28.).

[14] About DICE. https://dice.nims.go.jp/about.html, 2019 (accessed 2024-10-31.).

[15] New NIMS structural materials data sheets released. 2021 (accessed 2024-10-31.).

[16] K. Sawada, K. Kimura, F. Abe, Y. Taniuchi, K. Sekido, T. Nojima, T. Ohba, H. Kushima, H. Miyazaki, H. Hongo, T. Watanabe, Catalog of NIMS creep data sheets, Sci. Technol. Adv. Mater. 20 (1) (2019) 1131–1149.

[17] Y. Furuya, H. Nishikawa, H. Hirukawa, N. Nagashima, E. Takeuchi, Catalogue of NIMS fatigue data sheets, Sci. Technol. Adv. Mater. 20 (1) (2019) 1055–1072.

[18] M. Yamazaki, Y. Xu, M. Murata, H. Tanaka, K. Kamihira, K. Kimura, NIMS structural materials databases and cross search engine, MatNavi, Finland, 2007.

[19] National standard reference data series. 2018 (accessed 2024-10-31.).

[20] National reference standard. https://www.srd.re.kr/css/index.do, 2022 (accessed 2024-10-31.).

[21] F.M.o. Justice, Decree on the Federal Institute for Materials Research and Testing (Erlass über die Bundesanstalt für Materialforschung und -prüfung), Bundesanzeiger Verlag, Berlin, 2022.

[22] B. Skrotzki, S. Schriever, BAM reference data: results of ASTM E139 -11 creep tests on a reference material of Nimonic 75 nickel-base alloy. Zenodo, v1.1, 2023, https://doi.org/10.5281/zenodo.7764161.

[23] B. Rehmer, F. Bayram, L.A. Ávila Calderón, G. Mohr, B. Skrotzki, BAM reference data: Temperature-dependent Young's and shear modulus data for additively and conventionally manufactured variants of austenitic stainless steel AISI 316L. Zenodo, v1.0, 2023, https://doi.org/10.5281/zenodo.7813836.

[24] B. Rehmer, F. Bayram, L.A. Ávila Calderón, G. Mohr, B. Skrotzki, BAM reference data: Temperature-dependent Young's and shear modulus data for additively and conventionally manufactured variants of Ni-based alloy Inconel IN718. Zenodo, v1.0, 2023, https://doi.org/10.5281/zenodo.7813825.

[25] B. Rehmer, F. Bayram, L.A. Ávila Calderón, G. Mohr, B. Skrotzki, BAM reference data: Temperature-dependent Young's and shear modulus data for additively and conventionally manufactured variants of Ti-6Al-4V. Zenodo, v1.0, 2023, https://doi.org/10.5281/zenodo.7813733.

[26] T. Rodani, E. Osmenaj, A. Cazzaniga, M. Panighel, A. Cristina, S. Cozzini, Towards the FAIRification of scanning tunneling microscopy images, Data Intell. 5 (1) (2023) 27–42.
[27] N. Blumenröhr, N. MacKinnon, R. Aversa, FAIR Data Management workflow for MRI data, in: Workshop on Metadata and Research (objects) Management for Linked Open Science, 2023.
[28] M. Tanifuji, A. Matsuda, H. Yoshikawa, Materials data platform - a FAIR system for data-driven materials science, in: 2019 8th International Congress on Advanced Applied Informatics (IIAI-AAI), 2019, pp. 1021–1022.
[29] M.K. Jamshid, Alexander; Ribic, Brandon; Seifi, Mohsen strategic guide for additive manufacturing data management and schema., 2020.
[30] H.B. Nasrabadi, T. Hanke, M. Weber, M. Eisenbart, F. Bauer, R. Meissner, G. Dziwis, L. Tikana, Y. Chen, B. Skrotzki, Toward a digital materials mechanical testing lab, Comput. Ind. 153 (2023) 104016.
[31] A. Valdestilhas, B. Bayerlein, B.Moreno Torres, G.A.J. Zia, T. Muth, The intersection between semantic web and materials science, Adv. Intell. Syst. 5 (8) (2023) 2300051.
[32] B. Bayerlein, T. Hanke, T. Muth, J. Riedel, M. Schilling, C. Schweizer, B. Skrotzki, A. Todor, B.Moreno Torres, J.F. Unger, C. Völker, J. Olbricht, A perspective on digital knowledge representation in materials science and engineering, Adv. Eng. Mater. 24 (6) (2022) 2101176.
[33] J.F. Unger, A. Robens-Radermacher, E. Tamsen, Data provenance - from experimental data to trustworthy simulation models and standards, in: Materials Today: Proceedings, 2023.
[34] J. Maas, M. Leemhuis, J. Mertens, H. Schmidtke, R. Courant, M. Dahlmann, S. Stark, A. Böhm, K. Pagel, M. Hinze, D. Pinkal, M. Wegener, M.F.-X. Wagner, T. Sattel, H. Neubert, Ö. Özçep, A novel digitalization approach for smart materials – ontology-based access to data and models, Adv. Eng. Mater. 2302208.
[35] B. Mieller, S.B. Hassine, J. Töpfer, C. Priese, A. Bochmann, B. Capraro, S. Stark, U. Partsch, C. Fresemann, Ontology-based data acquisition, refinement, and utilization in the development of a multilayer ferrite inductor, Adv. Eng. Mater. 2401042.
[36] B. Bayerlein, M. Schilling, P. von Hartrott, J. Waitelonis, Semantic integration of diverse data in materials science: Assessing Orowan strengthening, Sci. Data 11 (1) (2024) 434.
[37] M. Schilling, B. Bayerlein, P. von Hartrott, J. Waitelonis, H. Birkholz, P. Dolabella Portella, B. Skrotzki, FAIR and structured data: a domain ontology aligned with standard-compliant tensile testing, Adv. Eng. Mater. (2024) 2400138.
[38] M.E. Kassner, Chapter 1 - Fundamentals of creep in materials, in: M.E. Kassner (Ed.), Fundamentals of Creep in Metals and Alloys, Third Edition, Butterworth-Heinemann, Boston, 2015, pp. 1–6.
[39] W. Blum, P. Eisenlohr, F. Breutinger, Understanding creep—a review, Metallurg. Mater. Trans. A 33 (2) (2002) 291–303.
[40] M.E. Kassner, Fundamentals of Creep in Metals and Alloys, 3 ed., Butterworth-Heinemann is an imprint of Elsevier, 2015.
[41] F.R.N. Nabarro, F. de Villiers, Physics Of Creep And Creep-Resistant Alloys, CRC Press, London, 1995.
[42] B. Skrotzki, J. Olbricht, H.-J. Kühn, High temperature mechanical testing of metals, in: S. Schmauder, C.-S. Chen, K.K. Chawla, N. Chawla, W. Chen, Y. Kagawa (Eds.), Handbook of Mechanics of Materials, Springer Singapore, Singapore, 2018, pp. 1–38.
[43] R.C. Reed, The Superalloys: Fundamentals and Applications, Cambridge University Press, Cambridge, 2006.
[44] A.B. Parsa, D. Bürger, T.M. Pollock, G. Eggeler, Misfit and the mechanism of high temperature and low stress creep of Ni-base single crystal superalloys, Acta Materialia 264 (2024) 119576.
[45] L.H. Rettberg, T.M. Pollock, Localized recrystallization during creep in nickel-based superalloys GTD444 and René N5, Acta Materialia 73 (2014) 287–297.
[46] S.G.R. Brown, R.W. Evans, B. Wilshire, Creep strain and creep life prediction for the cast nickel-based superalloy IN-100, Mater. Sci. Eng. 84 (1986) 147–156.
[47] E. Fleischmann, C. Konrad, J. Preußner, R. Völkl, E. Affeldt, U. Glatzel, Influence of solid solution hardening on creep properties of single-crystal nickel-based superalloys, Metallurg. Mater. Trans. A 46 (3) (2015) 1125–1130.
[48] M. Segersäll, J.J. Moverare, D. Leidermark, K. Simonsson, Creep and stress relaxation anisotropy of a single-crystal superalloy, Metallurg. Mater. Trans. A 45 (5) (2014) 2532–2544.
[49] A. Pineau, S.D. Antolovich, High temperature fatigue of nickel-base superalloys – A review with special emphasis on deformation modes and oxidation, Eng. Failure Anal. 16 (8) (2009) 2668–2697.
[50] A.I. Epishin, B. Fedelich, B. Viguier, S. Schriever, I.L. Svetlov, N.V. Petrushin, R. Saillard, A. Proietti, D. Poquillon, A. Chyrkin, Creep of single-crystals of nickel-base γ-alloy at temperatures between 1150°C and 1288°C, Mater. Sci. Eng.: A 825 (2021) 141880.
[51] A. Epishin, B. Fedelich, M. Finn, G. Künecke, B. Rehmer, G. Nolze, C. Leistner, N. Petrushin, I. Svetlov, Investigation of elastic properties of the single-crystal nickel-base superalloy cmsx-4 in the temperature interval between room temperature and 1300°C, Crystals 11 (2) (2021) 152.
[52] M. Durand-Charre, The Microstructure of Superalloys, Routledge, London, 1997.
[53] IEC BIPM, ILAC IFCC, IUPAC ISO, OIML IUPAP, International vocabulary of metrology – Basic and general concepts and associated terms (VIM), in: Joint Committee for Guides in Metrology, JCGM, 3rd edition, 200, 2012, p. 2012.
[54] Z. Xu, Z. Zhang, The need for standardizing fatigue data reporting, Nat. Mater. 23 (7) (2024) 866–868.
[55] H. Czichos, T. Saito, L. Smith, Springer Handbook of Materials Measurement Methods, Springer Berlin, Heidelberg, 2006.
[56] B. Skrotzki, H. Czichos, Materialprüfung, in: M. Hennecke, B. Skrotzki (Eds.), HÜTTE – Das Ingenieurwissen, Springer Berlin Heidelberg, Berlin, Heidelberg, 2020, pp. 1–27.
[57] BAM, Welcome to the BAM Webshop! https://webshop.bam.de/webshop_en/, 2024 (accessed 2024-07-24.).
[58] ISO/IEC 17025:2017 General requirements for the competence of testing and calibration laboratories, 2018, p. 65.
[59] M.H. Ramsey, S.L.R. Ellison, H. Czichos, W. Hässelbarth, H. Ischi, W. Wegscheider, B. Brookman, A. Zschunke, H. Frenz, M. Golze, M. Hedrich, A. Schmidt, T. Steiger, Quality in measurement and testing, in: H. Czichos, T. Saito, L. Smith (Eds.), Springer Handbook of Metrology and Testing, Springer Berlin Heidelberg, Berlin, Heidelberg, 2011, pp. 39–141.
[60] R. Aversa, A. Boubnov, D. De Angelis, C. Eschke, S. Irvine, R.E. Joseph, M. Kabbe, N. MacKinnon, M. Irene, M. Panighel, R. Thelen, D. Valentinis, The MDMC-NEP glossary of terms, 2024. https://doi.org/10.5281/zenodo.10663833.
[61] T. Hickel, S. Richter, E. Bitzek, L.A. Ávila Calderón, A. Gedsun, M. Forti, T. Hammerschmidt, J. Olbricht, B. Skrotzki, NFDI-MatWerk/IUC02 definition for reference data of materials, 2024. https://doi.org/10.5281/zenodo.11667674.
[62] NFDI-MatWerk, On the route towards a national research data infrastructure for materials science & engineering. https://nfdi-matwerk.de/, (accessed 2024-07-24.).
[63] T. Margoni, P. Tsiavos, Toolkit for researchers on legal issues, 2018. https://doi.org/10.5281/zenodo.2574619.
[64] GNU general public license. https://www.gnu.org/licenses/gpl-3.0.html, 2007 (accessed 2024-07-24.).
[65] CC BY 4.0 Attribution 4.0 International. https://creativecommons.org/licenses/by/4.0/, (accessed 2024-07-24.).
[66] A. Ariza, E. Asef, J. Jacob, A. Mühlichen, K. Peters-von Gehlen, H. Schranzhofer, U. Trautwein-Bruns, Data reuse in practice, 2023. https://doi.org/10.5281/zenodo.7568266.
[67] ISO 204, Metallic materials - Uniaxial creep testing in tension - Method of test (ISO 204:2023), ISO, 2023.
[68] J. Olbricht, M. Bismarck, B. Skrotzki, Characterization of the creep properties of heat resistant 9–12% chromium steels by miniature specimen testing, Mater. Sci. Eng.: A 585 (2013) 335–342.
[69] S.P. Singh, V. Jayaram, D. Srinivasan, P. Kumar, High throughput determination of creep behavior of additively manufactured AlSi10Mg alloy, Addit. Manufact. 58 (2022) 103049.
[70] H. Nobakhti, N. Soltani, Evaluating small punch test as accelerated creep test using larson–miller parameter, Exp. Tech. 40 (2) (2016) 645–650.
[71] J.A. Cano, M.S. Haque, M.A. Hossain, C.M. Stewart, Accelerated qualification of creep-resistant materials using a datum temperature method (DTM) to calibration, Int. J. Press. Vessels Piping 199 (2022) 104746.
[72] C.M. Stewart, M.A. Hossain, J. Pellicotte, R. Mach, D. Alexander, S.F. Siddiqui, Accelerated creep testing of inconel 718 using the stepped isostress method (SSM), Mater. Perform. Character. 11 (2) (2022) 281–300.
[73] L.C. Brinson, L.M. Bartolo, B. Blaiszik, D. Elbert, I. Foster, A. Strachan, P. W. Voorhees, Community action on FAIR data will fuel a revolution in materials research, MRS Bull. 49 (1) (2024) 12–16.
[74] B. Rehmer, F. Bayram, L.A. Ávila Calderón, G. Mohr, B. Skrotzki, Elastic modulus data for additively and conventionally manufactured variants of Ti-6Al-4V, IN718 and AISI 316 L, Sci. Data 10 (1) (2023) 474.
[75] N. Sonntag, M. Jürgens, P. Uhlemann, B. Skrotzki, J. Olbricht, Experimental data from service-like creep-fatigue experiments on grade P92 steel, Data Brief. 49 (2023) 109333.
[76] C. Rockenhäuser, P. von Hartrott, B. Skrotzki, Brinell-Hardness data (HBW 2.5/62.5) of aluminum alloy EN AW-2618A after different aging times and temperatures, Data Brief. 46 (2023) 108830.
[77] U.o. Pittburgh, Taxonomies and Controlled Vocabularies. https://pitt.libguides.com/metadatadiscovery/controlledvocabularies, (accessed 2024-07-24.).
[78] N. Guarino, D. Oberle, S. Staab, What Is an ontology? in: S. Staab, R. Studer (Eds.), Handbook on Ontologies Springer Berlin Heidelberg, Berlin, Heidelberg, 2009, pp. 1–17.
[79] E.O.F.N. Research, Zenodo. https://www.zenodo.org/, (accessed 2024-07-23.).
[80] P. Ost, Y. Shakeel, P. Tögel, Data collections explorer – an easy-to-use tool for sharing and discovering research data, (2023).
[81] P. Ost, Data collections explorer, (2022).
[82] L. Ehrlinger, W. Wöß, Towards a definition of knowledge graphs, in: 12th International Conference on Semantic Systems - SEMANTiCS2016, 2016.
[83] T. Jejkal, FAIR-DOscope - Explore the facets of FAIR digital objects, repository KITopen, 2022. https://www.doi.org/10.5445/IR/1000151433.
[84] CWA 17815 - Materials characterisation - Terminology, metadata and classification, 2021.
[85] L.A. Ávila Calderón, Y. Shakeel, S. Schriever, A. Gedsun, M. Forti, H. Jordan, Y. Han, O. Baer, R. Aversa, J. Olbricht, T. Hammerschmidt, T. Hickel, B. Skrotzki, NFDI-MatWerk/IUC02 Data schema for creep data of Ni-based superalloys including a comprehensive documentation of test results and metadata, Zenodo, 2024, https://doi.org/10.5281/zenodo.11668375.
[86] [dataset] L.A. Ávila Calderón, S. Schriever, Y. Han, J. Olbricht, P.D. Portella, B. Skrotzki, BAM reference data: creep of single-crystal ni-based superalloy CMSX-6. Zenodo, v1.0, 2024, https://doi.org/10.5281/zenodo.13937987.
[87] H. Bruderer, A. Vakulov, The dangers of digitization, and the importance of data backup, Commun. ACM 67 (5) (2024) 23–25.
[88] J.M. Perkel, 11 ways to avert a data-storage disaster. https://www.nature.com/articles/d41586-019-01040-w, 2019 (accessed 2024-07-24.).
[89] P. Ruggiero, M.A. Heckathorn, Data Backup Options, 2012.

[90] C. Pauly, Research Data Management practices for every day, Zenodo, 2024, https://doi.org/10.5281/zenodo.12703358.

[91] Working Group Harmonised terminologies and schemas for FAIR data in materials science and related domains WG. https://www.rd-alliance.org/groups/harmonised-terminologies-and-schemas-fair-data-materials-science-and-related-domains-wg/activity/, 2015 (accessed 2021-11-11.).

[92] G. Abdildina, F. Ernst, R. Aversa, P.-J. Ost, Using EVOKS to build controlled vocabularies, Repository KITopen, 2023, https://doi.org/10.5445/IR/1000164983.

[93] W.W.W.C. (W3C), SKOS simple knowledge organization system reference. http://www.w3.org/TR/2009/REC-skos-reference-20090818/, 2009 (accessed 2024-07-24.).

[94] M. Politze, F. Claus, B.D. Brenger, M.A. Yazdi, B.P.A. Heinrichs, A. Schwarz, How to manage it resources in research projects? towards a collaborative scientific integration environment, Eur. J. Higher Educ. IT 1 (2020) 5.

[95] M. Scheidgen, L. Himanen, A.N. Ladines, D. Sikter, M. Nakhaee, Á. Fekete, T. Chang, A. Golparvar, J.A. Márquez, S. Brockhauser, S. Brückner, L. M. Ghiringhelli, F. Dietrich, D. Lehmberg, T. Denell, A. Albino, H. Näsström, S. Shabih, F. Dobener, M. Kühbach, R. Mozumder, J.F. Rudzinski, N. Daelman, J. M. Pizarro, M. Kuban, C. Salazar, P. Ondračka, H.-J. Bungartz, C. Draxl, NOMAD: A distributed web-based platform for managing materials science research data, J. Open Source Softw. 8 (90) (2023).

[96] A. Jain, S.P. Ong, G. Hautier, W. Chen, W.D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, K.A. Persson, Commentary: the materials project: a materials genome approach to accelerating materials innovation, APL Mater. 1 (1) (2013).

[97] IUC02 Git Repository. https://git.rwth-aachen.de/nfdi-matwerk/iuc02, (accessed 2024-11-12.).

[98] K.D. Manager, Metadata Schemas for Materials Science. https://github.com/kit-data-manager/Metadata-Schemas-for-Materials-Science, 2024 (accessed 2024-11-11.).

[99] A.T. Jones, C. Baxter, The rolls royce 'SCORPIO' system, Measur. Sci. Technol. 6 (1) (1995) 131.

[100] M. Gell, D.N. Duhl, A.F. Giamei, The development of single crystal superalloy turbine blades, 1980.

[101] M.J. Goulette, P.D. Spilling, R.P. Arthey, Cost effective single crystals, in: 5th Int. Symp. on Superalloys, Champion, PA, 1984.

[102] M. Schilling, S. Bruns, B. Bayerlein, J. Kryeziu, J. Schaarschmidt, J. Waitelonis, P. Dolabella Portella, K. Durst, Seamless science: lifting experimental mechanical testing lab data to an interoperable semantic representation, Adv. Eng. Mater. 2401527.

[103] Y. Shakeel, G. Abdildina, R. Aversa, L.A. Ávila Calderón, N. Blumenröhr, S. Diebels, M. Engstler, J. Fell, F. Fritzen, V. Hartmann, H.-G. Herrmann, T. Jejkal, R. Joseph, A. Kirar, A. Laadhar, F. Mücklich, J. Olbricht, P. Ost, C. Pauly, A. Pfeil, M. Roland, B. Skrotzki, M. Soysal, R. Stotzka, E.G. Vitali, Creating Exemplary RDM Reference Datasets: Technical Process Overview, Repository KITopen, 2023, https://doi.org/10.5445/IR/1000159368.

[104] N. Blumenröhr, R. Aversa, From implementation to application: FAIR digital objects for training data composition, Res. Ideas Outcomes 9 (2023).

[105] NFDI-MatWerk, MSE Knowledge Graph v1.0. https://msekg-nfdi-matwerk-02.web.vulcanus.otc.coscine.dev/matwerk/, (accessed 2024-07-24.).

[106] M.C. Munro, Using JSON, in: M.C. Munro (Ed.), Learn FileMaker Pro 19: The Comprehensive Guide to Building Custom Databases, Apress, Berkeley, CA, 2021, pp. 319–330.

[107] Build more. Break less. Empower ohters. https://json-schema.org/, (accessed 2024-07-24.).

[108] M. Pilgrim, Dive Into Python 3, SoHo Books, 2010.

[109] H.M. Collaboration, MetaStore. https://helmholtz-metadaten.de/en/tools/metastore, 2024 (accessed 2024-11-11.).

[110] J. Newton-King, JSON Schema Validator. https://www.jsonschemavalidator.net/, (accessed 2024-07-24.).

[111] Data Type Registry. https://typeregistry.org/#urls/objects/design?payload=About, 2021 (accessed 2024-11-11.).

[112] H.M. Collaboration, Realization of FAIR Digital Objects. https://helmholtz-metadaten.de/en/fair-data-commons/realizing-fair-digital-objects, (accessed 2024-07-24.).

[113] C. Curdt, G. Günther, T. Jejkal, C. Koch, F. Krebs, A. Pfeil, A. Pirogov, J. Schweikert, P. Videgain Barranco, M. Weinelt, HMC Cross-cutting topic working group "from development to deployment", helmholtz metadata collaboration, helmholtz kernel information profile, in: S. Lorenz, A. Finke, C. Langenbach, K. Maier-Hein, S. Sandfeld, R. Stotzka (Eds.), HMC Paper, Kiel, Germany, 2022.

[114] T. Jejkal, A. Pfeil, J. Schweikert, A. Pirogov, P.V. Barranco, F. Krebs, C. Koch, G. Günther, C. Curdt, M. Weinelt, A common PID kernel information profile for the german helmholtz association of research centres, 2022. https://doi.org/10.5445/IR/1000148263.

[115] REST API. http://hdl.handle.net/api/handles/21.T11998/0000-001A-3905-F, (accessed 2024-11-11.).

[116] GWDG - Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen. https://gwdg.de/, (accessed 2024-11-11.).

[117] H.M. Collaboration, FAIR-DOscope. https://helmholtz-metadaten.de/de/werkzeuge/fair-doscope, (accessed 24.07.2024.2024).

[118] S.W. Standards, Resource description framework (RDF). https://www.w3.org/2001/sw/wiki/RDF, (accessed 2024-07-24.).

[119] S.W. Standards, Web ontology language (OWL). https://www.w3.org/2001/sw/wiki/OWL, (accessed 2024-07-24.).

[120] BFO Basic Formal Ontology. https://basic-formal-ontology.org/, (accessed 2024-07-24.).

[121] A.A. Guzmán, A.Z. Ihsan, S. Fathalla, The MatWerk ontology (mwo). http://purls.helmholtz-metadaten.de/mwo/, (accessed 2024-07-24.).

[122] The PMD Core Ontology. https://matportal.org/ontologies/PMDCO, 2023 (accessed 2024-07-24.).

[123] B. Bayerlein, M. Schilling, H. Birkholz, M. Jung, J. Waitelonis, L. Mädler, H. Sack, PMD core ontology: achieving semantic interoperability in materials science, Mater. Design 237 (2024) 112603.

[124] Tensile test ontology. https://matportal.org/ontologies/TTO, 2024 (accessed 2024-07-24.).

[125] M. Schilling, B. Bayerlein, P. von Hartrott, J. Waitelonis, H. Birkholz, P. D. Portella, B. Skrotzki, FAIR and structured data: a domain ontology aligned with standard-compliant tensile testing, Adv. Eng. Mater. (2024).

[126] Heat Treatment Ontology HTO. https://github.com/materialdigital/application-ontologies/tree/main/heat_treatment_ontology_HTO, (accessed 2024-07-24.).

[127] M. Digital, The Material Digitalization Platform. https://material-digital.de/, (accessed 2024-07-24.).

[128] M. Bijsterbosch, D. Duca, M. Katerbow, I. Kupiainen, I. Dillo, P. Doorn, H. Enke, J. M.d. Lucas, Funding research data management and related infrastructures: Knowledge Exchange and Science Europe briefing paper, 2016.

[129] K. Posavec, D. Celjak, L.J. Musap, Role of a croatian national repository infrastructure in promotion and support of research data management, Data Sci. J., keyword = en_US (2020).

[130] A. Bezold, J. Vollhüter, N. Karpstein, M. Lenz, A.P.A. Subramanyam, C.H. Zenk, T. Hammerschmidt, E. Spiecker, M. Göken, S. Neumeier, Segregation-induced strength anomalies in complex single-crystalline superalloys, Commun. Mater. 5 (1) (2024) 8.

[131] A. Zare, R.K. Hosseini, A breakthrough in creep lifetime prediction: Leveraging machine learning and service data, Scr. Mater. 245 (2024) 116037.

[132] NFDI-MatWerk, IUC02 Framework for curation and distribution of reference datasets. https://nfdi-matwerk.de/infrastructure-use-cases/iuc02-framework-for-curation-and-distribution-of-reference-datasets, (accessed 2024-07-24.2024).

[133] R. Schröder, V. Hofmann, M. Kühbach, C. Pauly, R. Aversa, S. Brockhauser, A. Azócar Guzmán, R.E. Joseph, Ö. Özkan, P.J. Konijnenberg, M. Wollgarten, O. Mannix-Brendike, S. Sandfeld, The EM Glossary: a community effort towards harmonized terminology in electron microscopy (v1.0.0), 2024. https://doi.org/10.5281/zenodo.11071439.

[134] R. Joseph, A. Chauhan, C. Eschke, A.Z. Ihsan, M. Jalali, U. Jäntsch, N. Jung, C. N. Shyam Kumar, C. Kübel, C. Lucas, M. Mail, A. Mazilkin, C. Neidiger, M. Panighel, S. Sandfeld, R. Stotzka, R. Thelen, R. Aversa, Metadata schema to support FAIR data in scanning electron microscopy, in: A. Pozanenko (Ed.), Supplementary Proceedings of the XXIII International Conference on Data Analytics and Management in Data Intensive Domains (DAMDID/RCDL 2021): Moscow, Russia, October 26-29, 2021, Moskau, Russland, CEUR-WS.org, 2021, pp. 265–277.