



Addressing limitations in the aggregation of archival fatigue data for an additively manufactured titanium alloy

I. J. Wietecha-Reiman^{1,2,*} , X. Zhao³, and T. A. Palmer^{1,4,*}

¹ Department of Material Science and Engineering, The Pennsylvania State University, University Park, PA 16802, USA

² National Institute of Standards and Technology, Gaithersburg, MD 20889, USA

³ legacy BlueHalo, AV Inc, Germantown, MD 20876, USA

⁴ Department of Engineering Science and Mechanics, The Pennsylvania State University, University Park, PA 16802, USA

Received: 8 May 2026

Accepted: 12 September 2026

This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply, 2026

ABSTRACT

A large body of archival fatigue testing data for additively manufactured Ti-6Al-4V is available but spread across many independent and unrelated studies. They can, however, be aggregated with other smaller datasets into a single large dataset that can then be used to identify sources of uncertainty and scatter in fatigue life. To identify larger trends, the underlying material properties, processing conditions, and testing procedures, which comprise the meta-data that accompany the stress amplitude and fatigue life data, must be captured. These meta-data are more general than the correlative information included in databases used to develop design allowable properties but are still valuable. However, the current practices for handling and processing such meta-data are plagued by recording and structural errors that limit their usefulness for traditional analysis and will inhibit the high-fidelity application of future machine learning (ML) and large language model (LLM) tools. Therefore, formalized procedures for compiling these datasets and maximizing the amount of usable data have been developed. Unreported meta-data values that meet certain criteria are replaced with reasonable estimates through a systematic imputation process to enhance the completeness of the accompanying meta-data. Building upon this existing dataset, a structured approach for introducing new datasets in a seamless fashion is also proposed to support a database that can then be used as a reliable source for evaluating the efficacy of emerging ML and LLM tools.

Handling Editor: Megumi Kawasaki.

Address correspondence to E-mail: ian.wietecha-reiman@nist.gov; tap103@psu.edu

<https://doi.org/10.1007/s10853-026-13789-3>

Published online: 16 October 2026

Introduction

Statistically derived design allowable fatigue properties [1] for common structural alloys are obtained using standardized mechanical testing procedures [2, 3] on materials processed under highly controlled conditions [4–6] and post-process heat treatments [7] to produce consistent microstructures and mechanical properties. The Metallic Material Properties Development and Standard (MMPDS) Handbook [8] and the National Institute for Materials Science Materials Database (MatNavi) [9] are two prominent curated databases used to define these material requirements for critical applications. A key component of these databases is the accompanying correlative information, which can be more generally described as meta-data, that capture critical information on the testing, specimen, material, and thermal and mechanical post-processing conditions. By taking this comprehensive and controlled approach, the design community maintains a high level of confidence in the fatigue properties of cast and wrought alloys.

Interest in using additive manufacturing (AM) processes to produce critical components across different industries is driving the development of similar databases that can be used to support the eventual implementation of these materials [10]. However, applying the same methodology to develop these databases for AM-fabricated materials will be much more difficult, given the complexity of the processing conditions in powder bed fusion (PBF) and directed energy deposition (DED) fusion-based AM processing routes that requires a high degree of quality control [11]. Both processes are characterized by complex solidification behaviors and thermal histories that promote high levels of alloying element segregation [12], inhomogeneous microstructural features [13–15], and process-induced defect structures [16, 17] that affect mechanical properties [18]. Any attempt at manipulating these microstructures and defect structures in a controlled manner, even with significant post-process heat treatments, is hindered by the lack of standardization of AM machine characteristics [19], processing parameters, and build path planning [20], as well as sensitivity to small changes in feedstock composition [21].

As a result, high levels of scatter emerge in the available mechanical properties, even for AM-fabricated specimens subjected to post-process machining and hot-isostatic pressing (HIP) steps that are designed to reduce the amount and severity of porosity and

brittle microstructural features that develop in the as-deposited condition [22–26]. This high level of scatter is presented in Figure 1a, in which stress-controlled fatigue testing data for AM PBF Ti-6Al-4V subjected to HIP post-processing are compiled and compared to the design allowable data of its wrought counterpart in the solutionized and aged and annealed conditions [8]. The difference in the scatter between the AM and wrought materials is obvious, with the scatter across selected stress amplitude levels for the AM materials pointed out in Figure 1b. Within each range of stress amplitudes, the standard deviation of fatigue life for the archival data spans two orders of magnitude, independent of the presence of multiple outliers.

Efforts are currently being undertaken to develop approaches that will capture the complexity inherent in the AM process and enable the development of high confidence design allowable values for AM materials [10, 27]. The modification of established techniques used for cast and wrought materials is one methodology currently being pursued. To address the unique characteristics of the AM process and the lack of standardization, the use of multiple AM machines and specimens fabricated using different processing parameters and heat treatments are being considered. With this approach, a larger number of material heats will be required, and the resulting datasets will need to be supplemented with microstructural characterization [27]. This multistep approach will demand a significant investment in cost, time, and effort to reach the same level of maturity as cast and wrought material datasets.

In the near-term, the absence of traditional highly pedigreed datasets limits the ability to identify trends in fatigue properties for AM materials. On the other hand, data available in the archival literature represent a potential source of information for meta-analysis that can provide insight into the sources of scatter. These archival fatigue testing datasets, however, are often plagued by small sample sizes within individual studies and inconsistent reporting of meta-data [28]. Despite these challenges, their utility in providing baseline data for traditional statistical analysis tools has been demonstrated, albeit on a limited scale [29]. By systematically aggregating both fatigue testing data and the corresponding meta-data for AM-fabricated 316L austenitic stainless steel, sources of scatter across multiple datasets were identified using a statistical model to estimate the mean fatigue life. Further analysis of the fatigue response using clustering

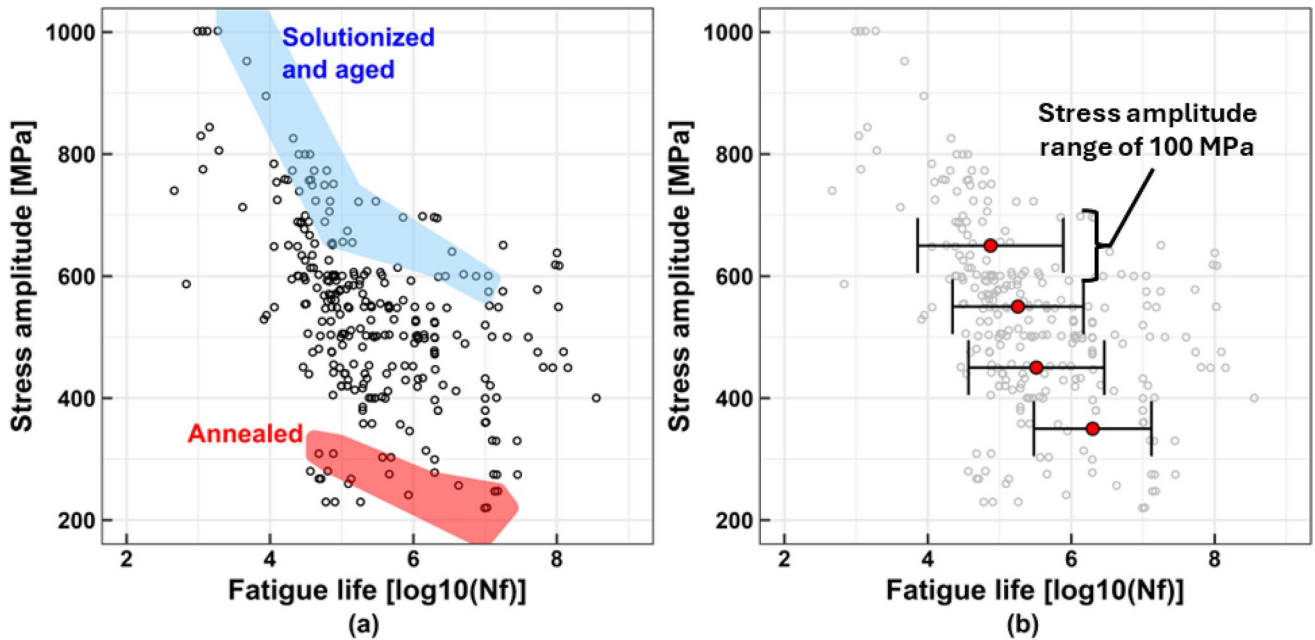


Figure 1 a&b: **a** Comparison of archival stress-controlled PBF Ti-6Al-4V fatigue data generated at a load ratio of -1 for machined specimens subjected to HIP post-processing to simi-

larly tested MMPDS datasets [8]. **b** Median values and standard deviations for ranges of stress amplitude levels.

algorithms allowed for the identification of the roles of microstructure and defects on fatigue life across a range of equivalent stress amplitudes. Despite the relatively small number of datasets evaluated, broad trends relating material performance to testing, specimen, and thermal and mechanical post-processing conditions could be identified.

As larger archival datasets become available [30], new challenges will be encountered, requiring the application of more robust strategies to capture and organize the testing conditions, processing parameters, and material properties that comprise the meta-data accompanying the fatigue testing data. The handling of fatigue testing data and meta-data [31] in AM processing has been previously addressed using various schema [32, 33] and interoperable data formats [34, 35]. These tools, however, have been primarily directed at capturing processing parameters and in situ monitoring data to support part qualification. While this approach can capture large amounts of process-specific data, it has not been uniformly applied to archival testing data, resulting in the inconsistent reporting of critical meta-data across individual studies [28, 36].

In the absence of the consistent reporting of meta-data for individual studies, the number of usable

datasets can be significantly impacted by listwise deletion, in which any datapoint with unreported meta-data is removed from analysis. A schematic diagram highlighting the mechanics of listwise deletion is shown in Figure 2a. In the case when all meta-data values are reported, no records, which are defined as a combination of a datapoint and its associated meta-data, are removed. Once unreported meta-data values are encountered, each affected record becomes incomplete and is removed from the analysis through listwise deletion. An example is provided in Figure 2b, which considers an arbitrary database consisting of 23 meta-data describing approximately 800 records. Of these 23 meta-data, only 5 can be combined to produce complete records. When any of the remaining 18 meta-data with incomplete reporting are taken into account, the number of complete records is reduced by varying degrees, depending on the sparsity of the meta-data being considered. As more meta-data are appended to the subset, the likelihood that a record contains all of those meta-data rapidly decreases in a logarithmic manner, yielding little usable information if all 23 available meta-data are considered.

Missing and unreported meta-data are particularly impactful on traditional statistical regression techniques [37]. Machine learning (ML) [36] and emerging

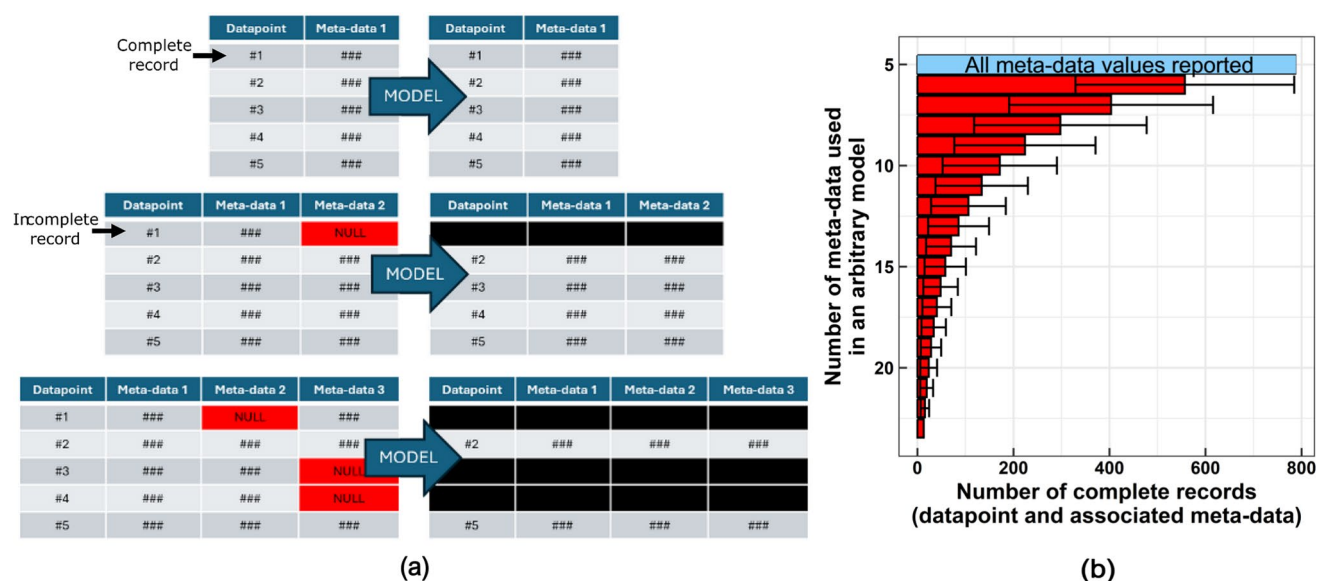


Figure 2 a&b: **a** Models developed using sparse datasets cannot make use of the full dataset due to listwise deletion of incomplete records. **b** The reduction in the number of retained records will drastically decrease as more meta-data with unreported values are added.

large language models (LLM) [38, 39] are being used to augment pre-existing workflows by retrieving and summarizing large datasets [40] and are also affected by many of these same issues. While these models can access scientific and engineering information [41], they lack the domain-specific knowledge needed to provide insight into the relationships between process-specific information and material properties and performance [42]. The presence of missing, unreported, or erroneous meta-data can not only affect the completeness of the available datasets but can also lead to model hallucinations, which are incorrect statements or values provided by LLM tools [38], affecting model reproducibility [39] and predictive capabilities. While tolerable for non-critical tasks, the presence of these errors will become a major impediment for the usage of archival data in the design and fielding of critical components.

A robust and comprehensive approach to aggregating, structuring, and handling large aggregated archival datasets with inconsistent reporting of meta-data will be required to expand the utility of these archival datasets for both conventional and emerging applications. By manually compiling and analyzing archival stress-controlled fatigue datasets for AM-fabricated Ti-6Al-4V for both accuracy and completeness, tools for mitigating the impact of unreported meta-data were evaluated and used to establish protocols for compiling and aggregating the archival data and meta-data. With these tools in place, high-quality and

low-sparsity archival databases can be produced and expanded as new datasets are generated, thus providing tools for the larger materials community that can be used as testbeds for emerging ML and LLM models.

Procedure

Creating a database from archival datasets begins with identifying and aggregating data and meta-data from archival sources and is followed by additional steps directed at detecting and correcting errors introduced during the transcription process. This sequenced approach, as portrayed in Figure 3, builds upon recent efforts that compiled archival fatigue data using manual [28, 29] and automated [30] approaches. The first step in the development of the aggregated dataset involved the identification of applicable sources. Several prominent citation indices, which include Web of Science^{1,2} and Google Scholar³, were used to manually

¹ Certain equipment, instruments, software, or materials are identified in this paper in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement of any product or service by NIST, nor is it intended to imply that the materials or equipment identified are necessarily the best available for the purpose.

² Clarivate PLC (Philadelphia, PA).

³ Alphabet Inc. (Mountain View, CA).

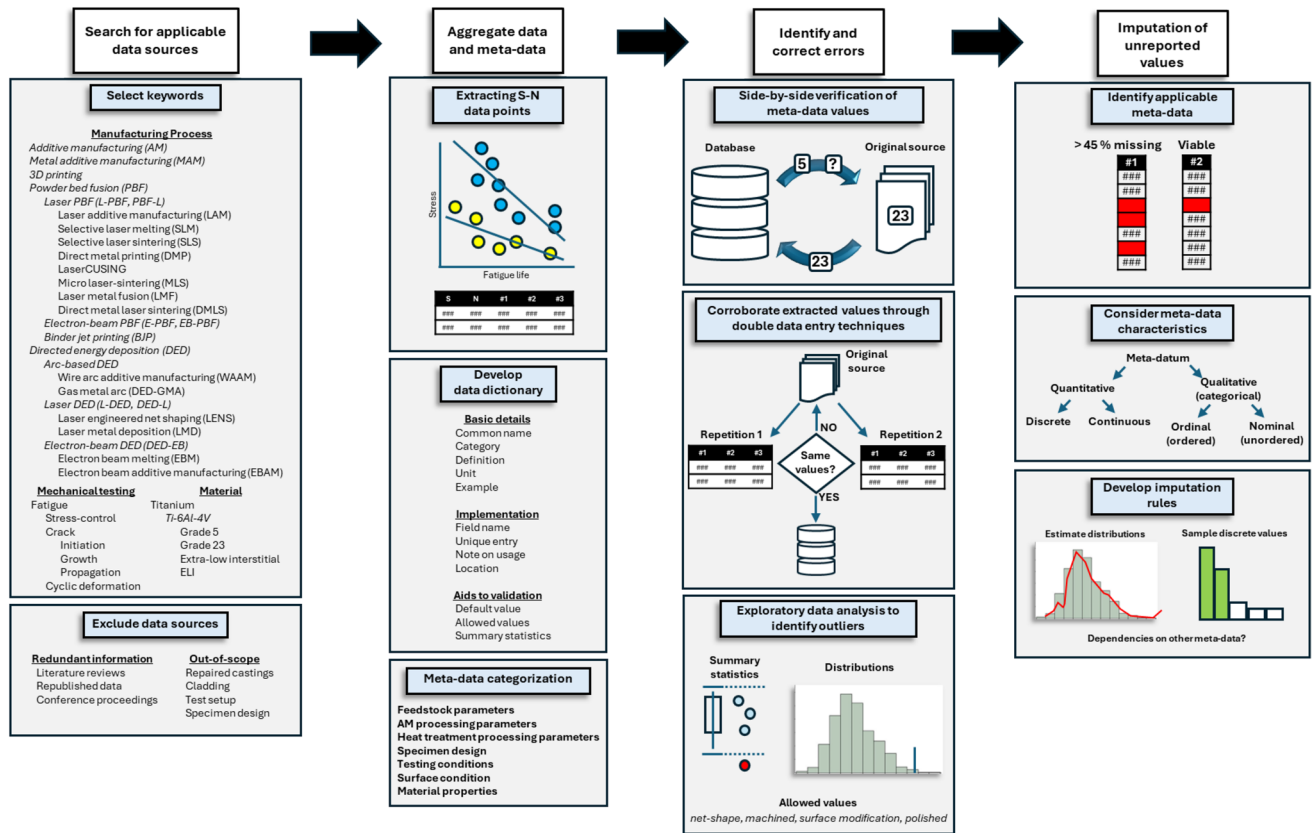


Figure 3 Representation of the major steps for identifying and aggregating data and meta-data from archival sources.

identify data sources in articles from scholarly journals, conference proceedings, and books. Common Boolean search operators with a combination of keywords were then used to search these indices for applicable data sources based on matches in the article title, author provided keywords, and abstract.

Defining appropriate keywords that allow these Boolean searches to capture the broadest number of data sources is hindered by the lack of standard terminology in the archival literature for the manufacturing process, mechanical testing technique, and type or grade of material. An iterative process was utilized to identify appropriate keywords across these three categories. This process begins with defining broad terms for each category based on commonly used verbiage, for example, “additive manufacturing” or “powder bed fusion”. New keywords were then introduced with a greater level of granularity to expand the number of sources captured. In this example, 15 terms and acronyms describing different PBF processes were enumerated. The final list of keywords is provided in Figure 3. After the data sources were

compiled, out-of-scope and duplicate records were identified and removed. Out-of-scope records, within the context of the fatigue of AM-fabricated Ti-6Al-4V, included materials not conforming to the stated alloy grade and fabrication route, experimental specimen geometries excluding lattice specimens, and testing under strain-controlled, elevated temperature, or aggressive environments. Additional filtering is at the discretion of the end user. Duplicate records predominantly appeared in literature reviews and other sources utilizing common datasets.

Stress–fatigue life (S-N) data, which were available in graphical and tabular forms, were then extracted from each publication. Graphical S-N data were manually digitized using WebPlotDigitizer⁴, with each S-N datapoint consisting of a fatigue life and an applied stress. While computer vision (CV) and natural language processing (NLP) tools have been used to extract these datapoints [30], errors arose in

⁴ <https://automeris.io/WebPlotDigitizer> (Pacifica, CA).

the resulting data [30]. A more time- and effort-intensive manual approach was selected to avoid these errors, particularly when changes in loading condition or in the representation of the applied stress were encountered. Differences in loading conditions were addressed qualitatively, with 67 % of fatigue testing performed under uniaxial loading conditions and the remaining tests performed under rotating–bending (17 %), vibratory (10 %), or other complex loading conditions (6 %).

Given the lack of common terminology and inconsistent reporting of testing, processing, and material conditions, extracting the associated meta-data from each source presented another significant challenge. To overcome these issues, a data dictionary, which provides rules for organizing, implementing, and validating meta-data values, was developed to identify and manage the meta-data in a consistent manner. When non-standard terms or jargon were used, they were replaced with standard terminology [43] or grouped into an applicable category using semantic normalization to capture the underlying meaning in a unified manner [44].

Meta-data were then placed into the broad categories listed in Table 1, which include material feed-stock parameters, AM processing parameters, heat treatment processing parameters, specimen design, testing conditions, surface condition, and material properties. This initial structure provides the basis for identifying commonalities between meta-data. Once the meta-data were compiled, errors were identified using a side-by-side verification methodology [30]. Since this approach is susceptible to overlooking errors in the recorded meta-data values [45], a double data entry technique, in which the same information was extracted twice, was used to resolve any discrepancies. Double data entry is typically performed by multiple people or programs [46] but can also be performed by extracting the same data twice with sufficient time between processes to avoid memory of the first effort influencing the decisions made during the second effort. Any remaining errors in the recorded meta-data values were then identified using exploratory data analysis (EDA) techniques by generating summary statistics and graphical analyses of these datasets [47]. Violations of the rules or characteristics that describe a particular meta-datum, as defined in the data dictionary, were also identified and corrected.

After correcting these errors, unreported meta-data values were addressed. To maximize the number of

usable datasets, imputation, which is a process in which unreported values can be replaced with estimated values that provide a reasonable approximation of what unreported meta-data could have been, was used to increase the number of complete and usable records [48]. Imputed values fill gaps in the meta-data but also introduce noise into the dataset, thus decreasing the precision of any subsequent analysis. Although a correctly specified multiple imputation technique may be determined such that up to 90% of values may be missing [49], the risk of the values not being truly missing at random (MAR) or the use of an unoptimized multiple imputation algorithm to estimate bias [49, 50] was deemed to be greater than the potential benefit of recovering an infrequently or rarely reported meta-datum. Within the cited literature, model bias was maximized for a proportion of missing data greater than 60 % under an unoptimized algorithm, giving a potential upper limit under a conservative implementation. Imputation will tend to be less susceptible to bias with additional meta-data used in imputation algorithms, enabling value-added results up to the upper limit. To avoid excessive noise being introduced into any future statistical models, a 45% threshold value was selected as a compromise. This value may be altered and optimized at the discretion of the database engineer.

Once appropriate meta-data are identified, different imputation techniques can be used to obtain plausible estimates of missing values, depending on the meta-data characteristics. At the highest level, a distinction may be made between discrete meta-data with separate and countable values (*e.g.*, alloy grade or qualitative surface condition) and continuous meta-data (*e.g.*, ultimate tensile strength or heat treatment temperature). Discrete meta-data may be further split into ordinal, or ordered, and nominal, or un-ordered, meta-data. In the simplest case for either type, values could be taken from a relevant standard or through a consensus-driven methodology [51]. However, this strategy risks biasing the current distribution of meta-data values and should be used as a last recourse. More data-rich options involve describing a probability mass function (PMF) or a probability density function (PDF) that describes the distribution of reported values for discrete or continuous meta-data, respectively, may be defined. From the PMF or PDF, random values may be drawn, while calculation of a measure of central tendency such as the mean, median, or mode

Table 1 Number of recorded meta-data values before and after verification using double data entry, and how many complete S-N curves were provided by imputation.

Meta-data category	Meta-data	Number of S-N curves with meta-data recorded before re-verification [30]	Number of S-N curves with meta-data recorded after verification		Total number of S-N curves with meta-data after verification	Imputed?	Number of viable S-N curves provided by imputation
			Automated	Manual			
Feedstock parameters	Powder size	216 (61%)	212 (69 %)	325 (68 %)	537 (68%)	yes	249
	Grade	–	Impute as grade 5	Imputed as grade 5	786 (100%)	–	
	Al	–	108 (35%)	167 (35%)	275 (35%)	no	
	V	–	108 (35%)	167 (35%)	275 (35%)	no	
	Fe	–	98 (32%)	153 (32%)	251 (32%)	no	
	O	–	99 (32%)	201 (42%)	300 (38%)	no	
	N	–	78 (25%)	167 (35%)	245 (31%)	no	
	C	–	80 (26%)	134 (28%)	214 (27%)	no	
	H	–	87 (28%)	125 (26%)	212 (27%)	no	
	AM machine	326 (92%)	295 (96%)	402 (84%)	697 (89%)	no	
AM processing parameters	Heat source	354 (100%)	308 (100%)	478 (100%)	786 (100%)	no	129
	Process category	354 (100%)	308 (100%)	478 (100%)	786 (100%)	no	
	Energy density	–	34 (11%)	86 (18%)	120 (15%)	no	
	Layer thickness	273 (77%)	270 (88%)	387 (81%)	657 (84%)	yes	
	Power	142 (40%)	159 (52%)	268 (56%)	427 (54%)	yes	
	Hatch spacing	135 (38 %)	143 (46%)	210 (44%)	353 (45%)	yes	
	Scan speed	174 (49%)	165 (54%)	272 (57%)	437 (56%)	yes	
	Beam diameter	–	75 (24%)	72 (15%)	147 (19%)	no	
	Layer rotation	81 (23%)	91 (30%)	143 (30%)	234 (30%)	no	
	Contour pass	–	62 (20%)	62 (13%)	124 (16%)	no	
	Plate heat	81 (23%)	33 (11%)	86 (18%)	119 (15%)	no	
	Chamber temperature	81 (23%)	22 (7%)	57 (12%)	79 (10%)	no	
	Temperature	297 (84%)	Impute as no treatment	Impute as no treatment	786 (100%)	–	
Heat treatment processing parameters	Time	297 (84%)	Impute as no treatment	Impute as no treatment	786 (100%)	–	
	Pressure	297 (84%)	Impute as no treatment	Impute as no treatment	786 (100%)	–	
	Stress concentration factor or notch presence (Kt)	Multivalued field, no analysis	Impute as no notch (1)	Imputed as no notch (1)	786 (100%)	–	45
Specimen design	Thickness	287 (81%)	292 (95%)	449 (94%)	741 (94%)	yes	
	Cross section shape	340 (96%)	306 (99%)	459 (96%)	765 (97%)	yes	
	Orientation	319 (90%)	290 (94 %)	473 (99%)	763 (97%)	yes	

Table 1 continued

Meta-data category	Meta-data	Number of S-N curves with meta-data recorded before re-verification [30]	Number of S-N curves with meta-data recorded after verification		Total number of S-N curves with meta-data after verification	Imputed?	Number of viable S-N curves provided by imputation
			Automated	Manual			
Testing conditions	Loading method	354 (100%)	308 (100%)	473 (99 %)	781 (99%)	yes	5
	Load ratio	350 (99%)	304 (99%)	473 (99%)	777 (99%)	yes	9
	Frequency	287 (81%)	279 (91%)	440 (92%)	719 (91%)	yes	67
Surface condition	Qualitative surface	326 (92%)	308 (100%)	473 (99%)	781 (99%)	yes	5
	Ra (or equivalent)	–	167 (54%)	296 (62%)	463 (59%)	yes	323
	Rz (or equivalent)	–	110 (36%)	129 (27%)	239 (30%)	no	
Material properties	Elongation to failure (EL)	120 (34%)	148 (48%)	268 (56%)	416 (53%)	yes	370
	Hardness (HV)	–	63 (20%)	124 (26%)	187 (24%)	no	
	Ultimate tensile strength (UTS)	131 (37%)	176 (57%)	325 (68%)	501 (64%)	yes	285
	Yield strength (YS)	135 (38%)	168 (55%)	282 (59%)	450 (57%)	yes	336
	Alpha lath width	–	65 (21%)	124 (26%)	189 (24%)	no	
	Bulk porosity	–	98 (32%)	148 (31%)	246 (31%)	no	
	<i>Total number of S-N curves</i>	354	308	478	786		

Meta-data in bold are either reported for 100 % of S-N curves or are amenable to imputation

also risks biasing the distribution, but could be applicable for low variance meta-data.

When possible, tools that utilize more information, such as linear regression or Multiple Imputation by Chained Equations (MICE), should be used to provide more robust estimates of unreported values. As additional information is incorporated into the imputation technique, downstream predictive performance tends to improve [52], although multivariate approaches that simultaneously impute several meta-data may experience degraded performance [53] that must be considered by the database engineer. With the MICE technique, a series of multiple linear regression models that use other meta-data as inputs are created and are then averaged together to obtain robust estimates [54]. In other cases, a multistage approach, in which several imputation techniques are required, is used.

All imputation functions and associated statistical analyses were performed using the R statistical programming language⁵.

The distribution of imputed meta-data values was then compared to the original meta-data value distribution to verify that no gross distributional shifts occurred and that the estimates are plausible. A two-sample Kolmogorov–Smirnov (KS) test was used to compare the difference between the empirical cumulative distribution functions (CDF) generated by the reported and imputed values. A test statistic was then calculated using the following relationship [55, 56]:

⁵ The R Foundation for Statistical Computing (Vienna, Austria).

$$D_{n,m} = \sup |F_{1,n}(x) - F_{2,m}(x)| \quad (1)$$

where $\sup$ is the supremum function, F represents the value of the empirical CDF at meta-data value x , and n and m are sample sizes. These two distributions will differ when:

$$D_{n,m} > \sqrt{-\ln(\alpha/2) \left(\frac{1 + m/n}{2m} \right)} \quad (2)$$

where α is the statistical significance taken as 0.05, which corresponds to a 95 % confidence level. If Equation (2) is true, then it can be concluded that the incorporation of the imputed meta-data values significantly biases the distribution of the values. When it is not true, it is assumed that imputation did not meaningfully bias the meta-datum. For the relatively simple or well-established imputation techniques summarized in Appendix 1, verification is adequate. Imputation techniques that attempt to more accurately recover the unreported value, which are not among those addressed herein, must be validated by benchmarked against accepted techniques and will require additional verification steps to ensure the veracity of the results, such as a masking analysis where known values are discarded and the accuracy and precision of recovery are quantified.

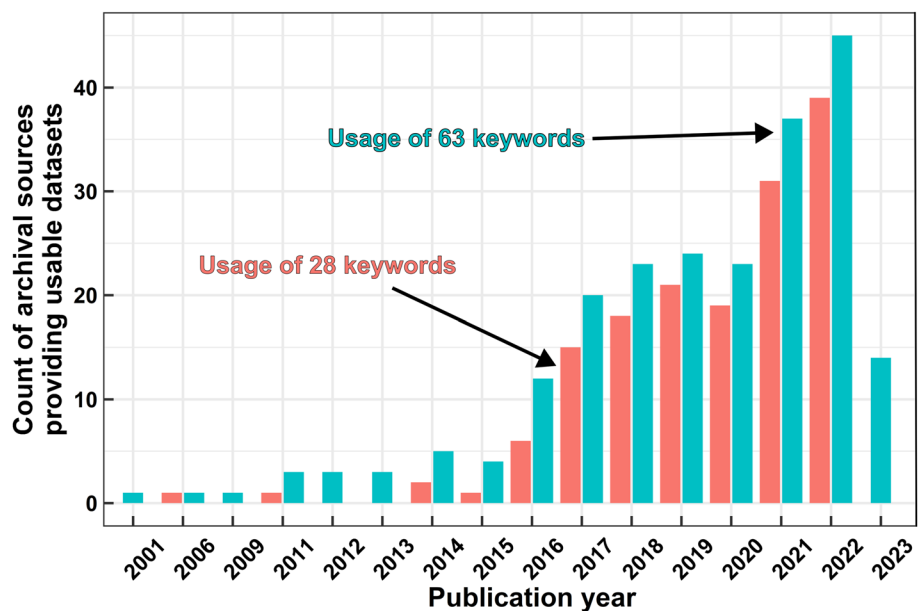
Results and discussion

Identification and aggregation of archival testing data

The feasibility of aggregating archival data sources has been demonstrated for several materials systems using a combination of manual and automated techniques [28–30]. While manual aggregation techniques have proven useful for smaller datasets, the use of automated techniques that utilize application programming interfaces (API) to automate source identification and downloading of the content is much more attractive for material systems with larger available datasets. One material candidate with a large archival fatigue dataset is AM Ti-6Al-4V. A summary of the number of data sources produced on an annual basis and captured with an automated approach [30] is highlighted in Figure 4. This automated approach employed a combination of 28 keywords to identify different AM fabrication techniques and the terms “fatigue” and “Ti-6Al-4V” to capture the testing condition and material of interest, yielding a total of 154 sources and 5580 datapoints across 354 S-N curves published between 2006 and 2022.

Once relevant data sources were identified, unique S-N curves were extracted from each source, and a verification procedure was undertaken to remove duplicate S-N curves from the aggregated dataset. A summary of the number of S-N curves obtained from

Figure 4 A histogram comparing the number of archival sources containing usable Ti-6Al-4V stress-controlled fatigue datasets identified in a previous database [30] and the current database.



an earlier automated database [30] and the number of non-duplicated S-N curves remaining after verification is provided in Table 1. Through this verification process, a total of 46 duplicate S-N curves were identified and removed, reducing the size of the data contributed by the earlier database by 13 % as the original 354 S-N curves were reduced to a total of 308 unique S-N curves, reducing the contribution to 5027 datapoints. Similar verification processes were performed on the larger manual aggregation dataset to identify missing values and correct typographical errors.

While a large amount of data was identified using this automated approach, it was difficult to determine how comprehensive it was or whether the broadest possible collection of available data sources was identified. To evaluate the ability of these automated techniques to capture as much of the available data as possible, a manual aggregation effort directed at expanding the number of relevant data sources was performed. Since the keywords used in the earlier automated database were rather arbitrarily selected and limited by the absence of a common terminology, a hierarchical approach focused on identifying relevant keywords that provide a broader coverage of terminology was used in this manual aggregation effort. Keywords with general meanings were used first, followed by increasingly specific ones, resulting in the development of a much-expanded keyword list with 50 manufacturing, seven mechanical testing, and six material keywords. The entire list of keywords used in this approach is provided in Figure 3. This broader approach to keyword selection allowed an additional 41 sources to be identified that were published between 2001 and 2022, as highlighted in Figure 4, providing an additional 2085 datapoints and increasing the size of the database by approximately 29 %.

After the duplicate S-N curves were removed, the datapoints within the S-N curves were identified, allowing data extraction and identification errors to be eliminated. A typical extraction error involved the inclusion of unverifiable datapoints from a S-N plot with overlapping symbols. In one instance, several hundred unverifiable datapoints were captured [57] and subsequently removed. Identification errors included cases where runout specimens, in which the test was terminated at a predetermined stopping point, were improperly labeled as fractured specimens. Other cases involved multiaxial tests that were recorded as uniaxial ones, and logarithmic stress

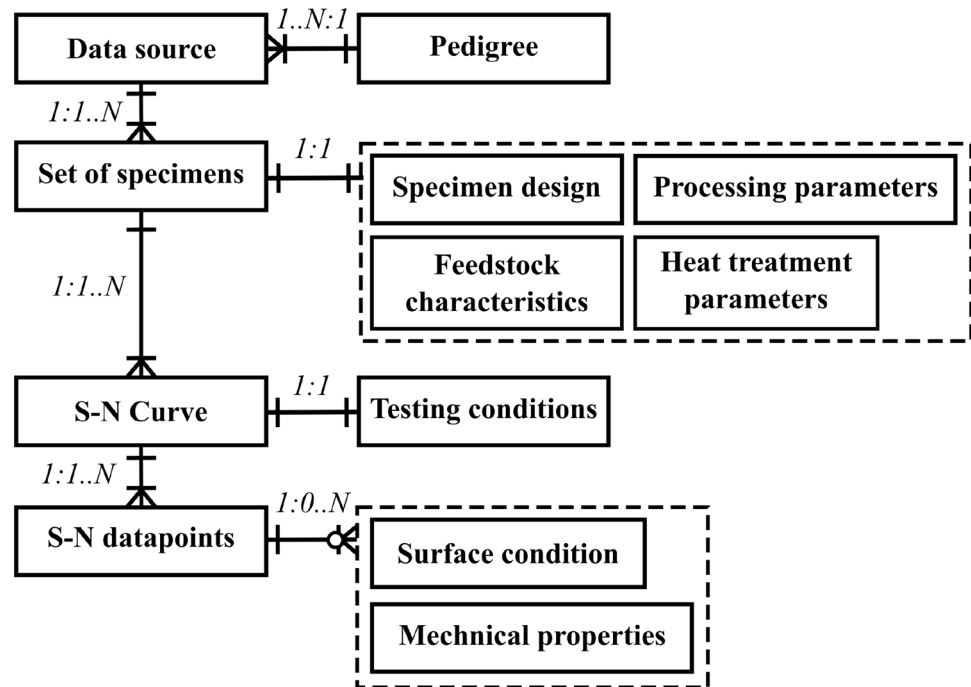
amplitude or maximum stress axes were improperly scaled on the reported S-N plots. Combining the unique results of the automated and manual aggregation efforts allows the comprehensiveness of the automated approach to be leveraged with the expanded scope and capability of the manual approach to survey non-standard publications. A total of 7112 unique datapoints published between 2001 and 2022 were obtained using this combined approach.

Compilation and verification of accompanying meta-data

After aggregating and verifying the published S-N data, the selection and compilation of associated and appropriate testing, processing, and material property meta-data must be undertaken. It is not a straightforward exercise, however, and there are major challenges that will be encountered when attempting to produce usable databases. One such challenge originates from the lack of standardized terminology for the meta-data reported in archival sources, leading to differences in the terminology used to describe the meta-data and creating significant issues in how the meta-data are categorized. To address these issues and improve the usability of the database, a data dictionary was developed. The data dictionary includes basic details, notes on implementation, and aids to validation, which are summarized in Figure 3 and provided in the supplementary material. When taken together, these details provide a structure to allow each meta-datum field to be understandable and navigable within the overall database structure.

Building off an earlier data aggregation and analysis effort [29], seven meta-data categories, as summarized in Table 1, were selected, and a total of 41 unique meta-data were identified and collected. Even though the meta-data categories are somewhat arbitrary, they compensate for inconsistent reporting across the different individual studies and are intended to capture as much of the reported meta-data as possible within the bounds of the data dictionary. These meta-data were placed in a flat-file structure and then separated into distinct tables in a stepwise normalization process to eliminate redundant information and interdependencies [58], yielding the beginnings of a relational database. In a relational database, unique identifiers, or keys, connect information in different tables in a parent-child hierarchy, which is depicted in Figure 5. For fatigue datasets, this enforced hierarchy relates

Figure 5 Conceptual model diagramming the basic structure of a potential relational database for archival fatigue datasets, as well as the inherent hierarchy in fatigue datasets.



sets of specimens produced by a given experimental campaign to individual S-N curves and the datapoints contained within them. The connections between the tables also provide information on the cardinality between the databases using Crow's foot notation [59]. For example, a single data source may contribute many S-N curves in a mandatory one-to-many (1:1..N) relationship. Additionally, to improve the manual navigability of the relational database in an intuitive manner, some of these tables were de-normalized, or combined [59]. Within a database management system, de-normalization introduces some amount of redundant information but reduces the complexity of how various data and meta-data that would otherwise be stored in different tables are joined, which can speed up queries and simplify reporting [60, 61].

Transferring meta-data from the previous automated aggregation efforts [30] was hindered by the lack of predictable structure or contents [62] that inhibited the integration of the meta-data and compromised the ability to query that database for specific information. Several examples of these inappropriate meta-data structures are summarized in Table 2, including the presence of ordered and unordered lists of single or multiple meta-data within a single field and the use of multipurpose fields whose meaning changed depending on the value of another field. When these structures are present, not even a rigorous data

dictionary can unambiguously contextualize these meta-data. A full revision of the affected portion of the database was performed by placing the meta-data into separate fields and eliminating interdependencies between remaining fields. When the database structure was developed for the manual aggregation effort, these complex features were deliberately avoided to maximize transferability of the data and meta-data to other applications or database structures (Table 3).

After the structuring of the meta-data was completed, a systematic verification process was used to identify and mitigate any errors that occurred during transcription. An independent re-verification using side-by-side, double data entry, and EDA verification techniques revealed that errors remained uncorrected in the original database. Such errors included, but were not limited to, typographical errors and incorrectly transcribed information. Depending on the meta-data, between 1 and 5% of records were affected by these errors. Table 1 provides a summary of the reported meta-data both before and after the independent re-verification step. The number of recorded meta-data values increased by 7%, on average, with re-verification. For other meta-data, such as those for mechanical properties that are reported in either graphical or non-graphical forms, re-verification increased the number of transcribed values by up to 20%.

Table 2 Summary of complex structures containing meta-data that were identified in an earlier automated database [30]

Type of structure	Field name	Example	Number of cells with values	Number of cells with issues
Multiple meta-data; ordered values	Specimen description	Combines cross section shape with use of drawing, standards, notch presence, and additional details	340	273 (80.3%)
	AM environment	Combines shielding gas with measurements of residual O ₂ levels	89	2 (2.2%)
	Processing sequence and parameters	Combines heat treatment parameters with qualitative surface condition, the presence of internal structures, and processing route	353	295 (83.6%)
Multiple meta-data; unordered values	Types of AM	Combines feedstock form, type of energy source, deposition, and post-process wrought manufacture	353	3 (0.8%)
	AM machine	Combines manufacturer and model	327	327 (100%)
	Type of fatigue test	Combines nominal stress state with test standard and frequency	354	44 (12.4%)
Single meta-datum; unordered values	Source power		142	4 (2.8%)
	Scan speed		175	4 (2.3%)
	Hatch spacing		133	2 (1.5%)
	Preheat		82	4 (4.9%)
	Size of feedstock	Combines average size, ranges, span, and other metrics	217	164 (75.6%)
	Frequency		287	13 (4.5%)
	Cross section size		285	90 (31.6%)
Multipurpose	Source power	Accounts for fill, contour, pre-scan, and unknown components		
	Scan speed	Accounts for fill, contour, pre-scan, and unknown components		
	Scan pattern	Conveys pattern, delays, or contour details		
	Voltage	Changes based on AM process category		
	Current	Changes based on AM process category		
	Size of feedstock	Changes based on AM process category		

A large amount of unreported meta-data was identified across the different categories highlighted in Table 1, with only seven of the 41 meta-data displaying a 100% reporting rate. The frequency of reporting for each of the 41 meta-data is captured in Fig. 6 through the use of a correlation matrix for different pairs of meta-data organized with respect to the different categories described in Table 1. It is evident in this figure that meta-data within the same category are co-reported at much higher rates than those across different categories. While constructing a record using meta-data from a single category might create a large number of complete records, doing so would severely

restrict any subsequent analysis. If all viable meta-data from different categories were considered for such an analysis, every record would contain unreported information, resulting in their elimination through listwise deletion and leaving no usable records.

The impact of imputation on the size of the usable dataset can therefore be significant, even when only considering a small subset of meta-data. For example, if data are taken only from tests performed at a load ratio of 0.1, a total of 4360 datapoints across 368 S-N curves are initially available for a hypothetical analysis. If only the ultimate tensile strength, elongation to failure, power, and surface roughness are considered

Table 3 Summary of high-level requirements for AM Ti-6Al-4V meta-data to promote uniformity and backwards compatibility

Meta-data category	Meta-data	Type	Allowed values	Unit	Note
Feedstock parameters	Powder size	Continuous	Precision to 0.1	µm	Dn(50) or Dv(50) will be specified.
	Grade	Nominal	5; 23		5 denotes standard Ti-6Al-4V, 23 denotes extra low interstitial Ti-6Al-4V.
AM processing parameters	Heat source	Nominal	laser; electron beam; arc		
	Process category	Nominal	powder bed; deposition		
	Layer thickness	Continuous	Integer values	µm	
	Power	Continuous	Precision to 0.1	W	
	Hatch spacing	Continuous	*Precision to 0.1	µm	
	Scan speed	Continuous	Precision to 0.1	mm/second	
Heat treatment processing parameters	Temperature	Continuous	25 for no heat treatment Integer values	° C	
	Time	Continuous	0 for no heat treatment Integer values	minutes	
	Pressure	Continuous	0.101 for no heat treatment or heat treatment without applied pressure Integer values	MPa	
Specimen design	Kt (notch presence)	Continuous	1 for no notch Precision to 0.01		
	Thickness	Continuous	Precision to 0.01 Smaller dimension for a rectangular cross section	mm	For specimens with a scaffold structure, provide the nominal thickness of the specimen's outer dimensions.
	Cross section shape	Nominal	round; rectangle; internal structure; scaffold; other		
	Orientation	Continuous (Discrete)	Integer values (90, 45, 0)	°	Orientation of loading axis relative to build plate.
Testing conditions	Loading method	Nominal	axial; rotating; bending; vibration; in-phase; compression; out of phase		
	Load ratio	Continuous (Discrete)	− 1 to 0.99 (− 1 and 0.1)		
	Frequency	Continuous	Above 10, integer values Below 10, precision to 0.1	Hz	Ultrasonic frequencies will be listed with vibration loading method.
Surface condition	Qualitative surface	Ordinal	net-shape; machined; polished; modification		
	Ra (or equivalent)	Continuous	Precision to 0.1	µm	Lineal or areal technique will be noted. Detail on instrumentation as database expands.
	Rz (or equivalent)	Continuous	Precision to 0.1	µm	
Material Properties	EL	Continuous	Precision to 0.1	% elongation	Detail on instrumentation as database expands.
	UTS	Continuous	Precision to 0.1	MPa	
	YS	Continuous	Precision to 0.1	MPa	

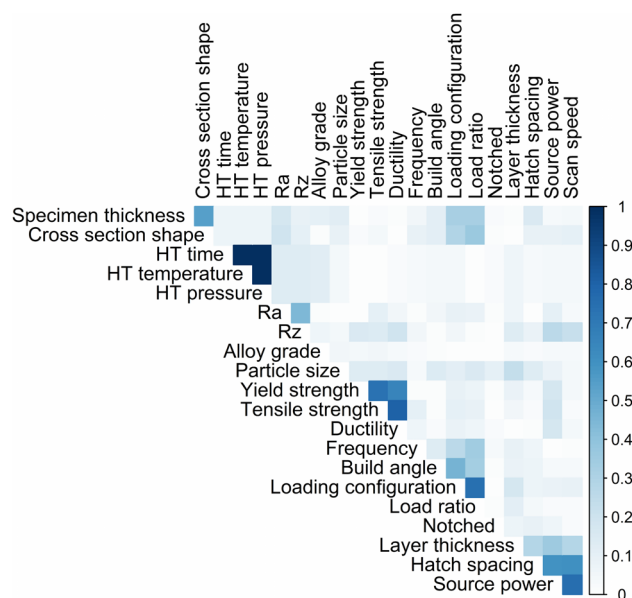


Figure 6 Correlation between the rates of reported meta-data values. The acronyms HT, Ra, and Rz represent heat treatment, arithmetic mean surface roughness, and maximum asperity height, respectively.

as meta-data of interest, the number of usable datapoints without imputation falls to 511 across 59 S-N curves, as shown in Fig. 7a. With the application of imputation, all 4360 datapoints become available, as highlighted in Figure 7b. Increasing the number

of available datasets improves the sensitivity of the analysis and allows subtle effects to be detected [63, 64] and increases the variety and combinations of conditions that may be evaluated [65].

Given the positive impact of imputation on the size of the available dataset with only five meta-data, this approach can be expanded to the overall aggregated dataset and the much larger body of meta-data. Considering the 34 meta-data with incomplete reporting rates, only 16 displayed rates higher than 45%, which is the threshold for imputation currently applied. When combined with the seven meta-data that displayed 100% reporting rates, a total of 23 viable meta-data were identified, with at least one meta-datum from each category included in the final tabulation. Of these available meta-data, the heat treatment processing parameters, specimen design, and testing conditions were the most widely reported. With the application of imputation across all 23 viable meta-data, each of which did not significantly affect the distribution of available values based on the KS test, the rules of which are provided in Appendix 1, the usable dataset grew to the maximum number of available datapoints.

When these datasets are used for model development, the end user must decide whether individual S-N datapoints are treated independently or in a groupwise manner. Usage of the latter technique would limit the variety of meta-data that the model

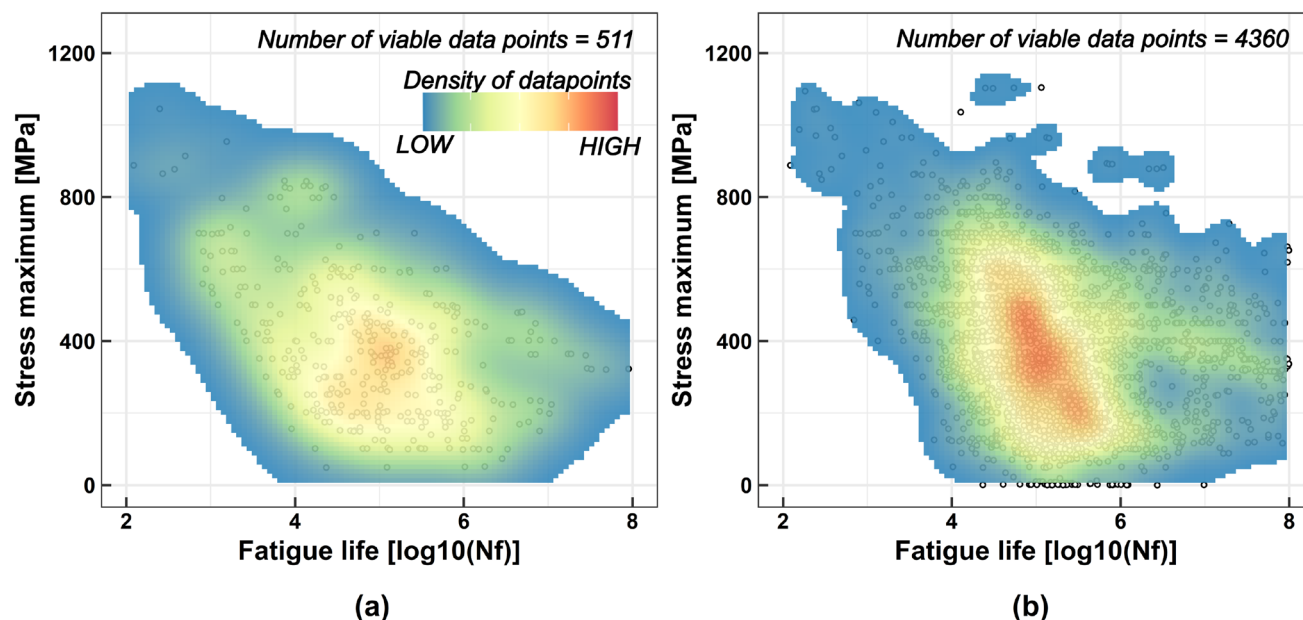


Figure 7 a&b: Plots showing the density of datapoints representing complete records with four meta-data with **a** no imputation and **b** imputation applied.

is exposed to, yielding more conservative training and testing performance metrics, while usage of the former technique assumes a higher degree of variability between individual AM-fabricated specimens than those specimens extracted from a bar or sheet. Dataset partitioning could be performed based on S-N point membership to an individual S-N curve or to an experimental campaign, as illustrated in Figure 5.

Even with this large expansion of the usable dataset with the implementation of imputation for selected meta-data, more meta-data can be made available with improved and more uniform reporting procedures. For example, among the feedstock parameters, powder size and grade can be reliably obtained, but there has been little consistency in the archival literature in how powder size is reported, with different percentiles and particle size distribution (PSD) spans and even characterization methods producing inconsistent results. In the short term, the median particle diameter needs to be captured, but as the field matures and adherence to reporting standards [66] in research improves, the scope of the database will need to accommodate more comprehensive measurements. When considering mid- and long-term needs, improvements in the consistency of reporting of AM processing parameters, which should include the type of heat source and process category, layer thickness, power, hatch spacing, scan speed, and material properties, such as surface roughness and tensile properties, will be needed.

Integration of new data into a consolidated database

While the aggregation of these datasets from current archival sources is a valuable contribution, they represent just a snapshot in time. With new datasets being continuously produced, the amount of new data that are unavailable to the existing database can become rather significant over time. As highlighted in Fig. 4, even a single year after the publication of the original archival database [30], an additional 13 data sources became available bringing the total of additional datasets to 54, representing an increase of 540 datapoints (8%) and 51 S-N curves (7%) to the existing data set. The aggregated database now consists of 7652 datapoints with article publication dates up to 2023. As more data become available, the dataset can be expanded.

To overcome the limitations of data aggregation efforts performed over limited time frames, a procedure is needed for integrating new datasets into a central repository [67] that can be maintained and updated. Adding new S-N data to the dataset can be rather straightforward, but identifying and reporting meta-data is a more difficult endeavor that will require the development of standardized procedures that can be adapted to emerging datasets. As a starting point, the 23 meta-data identified in the current aggregation effort can be used to promote the standardization of meta-data reporting in new datasets. When new sources are added, each datapoint should be accompanied by these 23 meta-data, thus minimizing the need for imputation and allowing a more highly detailed database to be developed. This level of reporting would need to be enforced, in part, by the relevant scholarly journals. Such an action is not unprecedented. Indeed, reporting requirements for systematic reviews and meta-analyses have become standard practice in many research fields [68]. Of course, the protocols and critical meta-data enumerated herein serves as a preliminary outline for such requirements, and further refinement is required.

Building upon the existing aggregated dataset, new datasets can be identified through either automated means using citation index APIs [30] or as direct submissions to the database using a template, as outlined in Fig. 8, following the example of the NIST Materials Data Curation System. The submitted meta-data are then verified to determine if they conform to the requirements outlined in the data dictionary and are re-formatted as needed. As illustrated herein, verification is by necessity a multistage process. Future iterations of an archival database should leverage CV and NLP techniques such that they replicate double data entry validation. A human operator would then perform an additional extraction of data and meta-data, following the protocols outlined in Fig. 3. These three sets of datapoints and meta-data are then compared algorithmically with cases exceeding a predefined threshold investigated by the human operator. By applying the process outlined in Fig. 8, a higher level of data and meta-data quality than what is currently available in the archival literature may be maintained, improving reproducibility and supporting the development of higher-fidelity ML and LLM tools.

As new datasets are added, the distribution of values for each meta-datum will change, especially for those in which imputation was used to expand

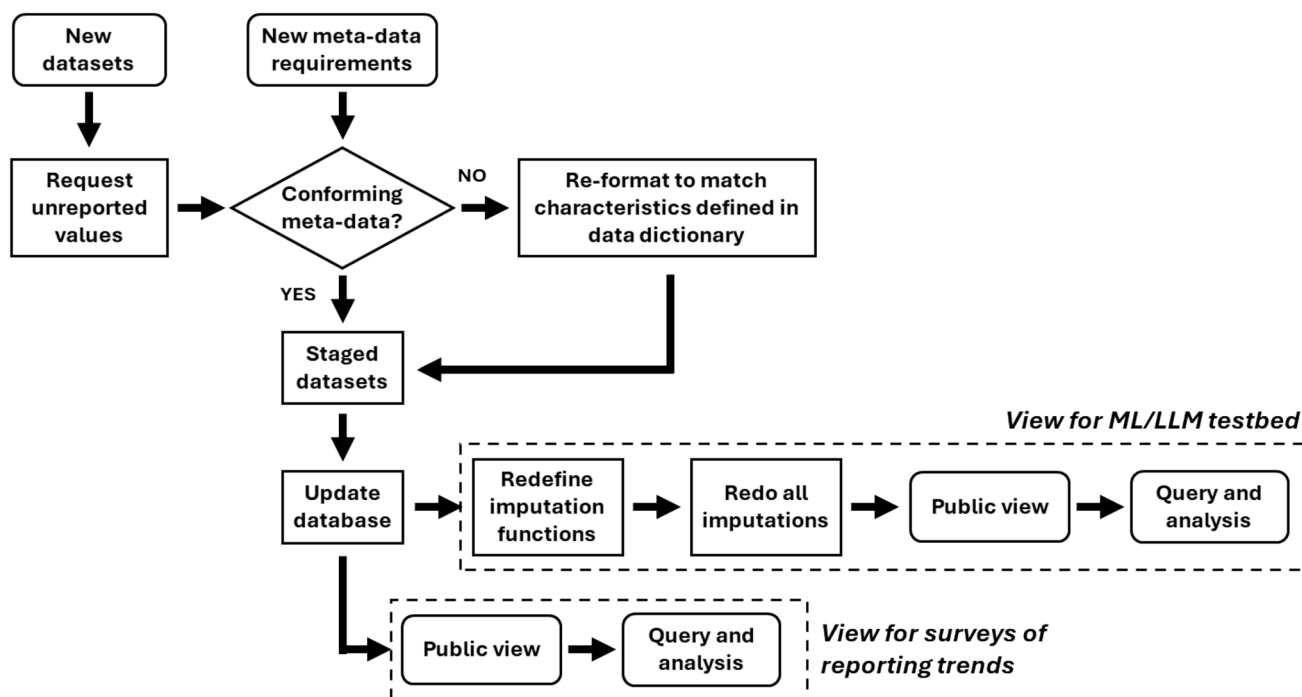


Figure 8 Schematic for a data processing pipeline to ingest new datasets and meta-data.

the original archival database. With the addition of new meta-data values, the imputation functions will change, requiring a recalculation of the overall distribution so that a more accurate representation of the meta-data values can be obtained. An estimate of the time complexity of the operation and a recommendation of the frequency of recalculation is provided in the supplementary material. Recalculation of the imputed data distributions will also prevent ML and LLM tools from incorrectly recognizing non-existent patterns [69] and maintaining the randomness of the operation.

It can also become possible to expand the number of meta-data that can be reported and used to identify new trends, particularly for meta-data that did not meet the minimum requirements for imputation during the initial aggregation efforts. For example, oxygen and nitrogen contents, which can have significant impacts on material properties in Ti-6Al-4V, are reported in only 38 and 31 % of the datasets, respectively, falling just below the minimum amount for imputation and precluding their inclusion in the original meta-data compilation. If the inclusion of new datasets increases these values to 45 %, these meta-data may then be imputed, increasing the number of meta-data that need to be reported in archival sources, but also providing another valuable parameter for ML and LLM tools to utilize.

Summary and conclusions

The large amount of archival fatigue testing data available for AM-fabricated Ti-6Al-4V represents a valuable resource for evaluating emerging statistical, ML, and LLM tools and supplementing ongoing efforts to develop pedigreed experimental datasets. While statistical analysis of a smaller dataset comprising archival S-N data for AM-fabricated 316L austenitic stainless steel has been successfully demonstrated, the aggregation of larger datasets presents several significant challenges prior to analysis. A formalized procedure for aggregating, extracting, and structuring a large dataset has been developed. While the procedure produces a reliable database, issues in reporting key meta-data require the imputation of selected meta-data values to produce complete records. Building upon this existing dataset, a structured approach for incorporating new datasets in a seamless fashion is proposed to support a continuously expanding database. Additional conclusions are provided below in more detail:

- Given the inconsistency in applying terminology for AM processes and materials, a formalized procedure to identify applicable data sources requires careful consideration of keywords to produce a comprehensive survey of available data. By

Table 4 Summary of suggested meta-data imputation rules for the current aggregated archival AM Ti-6Al-4V fatigue dataset

Meta-data	Meta-data type	Imputation rule type	Imputed value
Powder size	Continuous	Hierarchical random	$\begin{cases} 35 + 2N(0, 1) & \text{if L-PFB} \\ 75 + 10N(0, 1) & \text{if E-PBF} \\ 100 + 5N(0, 1) & \text{if L-DED} \\ 70 + 3N(0, 1) & \text{if E-DED} \end{cases}$
Grade	Discrete	Mode	Grade 5
Layer thickness	Continuous	Hierarchical random	$\begin{cases} S\{(30 + 2N(0, 1), 60 + 2N(0, 1)) : p_i = 0.46, 0.54, i = 1, 2\} & \text{if L-PFB} \\ \frac{x_{n/2} + x_{1+n/2}}{2} + 10N(0, 1) & \text{if E-PBF} \\ \frac{x_{n/2} + x_{1+n/2}}{2} + 150N(0, 1) & \text{if L-DED} \\ S\{(1000, 1200) : p_i = 0.66, 0.33, i = 1, 2\} & \text{if Arc-DED} \\ S\{(70, 1500) : p_i = 0.83, 0.16, i = 1, 2\} & \text{if E-DED} \end{cases}$
Power	Continuous	Hierarchical random	$\begin{cases} 300 + 50N(0, 1) & \text{if L-PFB} \\ 1000 + 250N(0, 1) & \text{if E-PBF} \\ S\{(300 + 50N(0, 1), 5000 + 750N(0, 1)) : p_i = 0.66, 0.33, i = 1, 2\} & \text{if L-DED} \\ S\{(2100, 2200) : p_i = 0.33, 0.66, i = 1, 2\} & \text{if Arc-DED} \\ 120 + 5N(0, 1) & \text{if E-DED} \end{cases}$
Hatch spacing	Continuous	Hierarchical random	$\begin{cases} 125 + 25N(0, 1) & \text{if L-PFB} \\ S\{(100, 180, 200) : p_i = 0.47, 0.09, 0.47, i = 1, 2, 3\} & \text{if E-PBF} \\ S\{(50 + 100N(0, 1), 2500 + 500N(0, 1)) : p_i = 0.66, 0.33, i = 1, 2\} & \text{if L-DED} \\ 5500 + 100N(0, 1) & \text{if E-DED} \end{cases}$
Scan speed	Continuous	Hierarchical random	$\begin{cases} 1250 + 100N(0, 1) & \text{if L-PFB} \\ 3000 + 500N(0, 1) & \text{if E-PBF} \\ S\{(10 + 2N(0, 1), 1000 + 100N(0, 1)) : p_i = 0.75, 0.25, i = 1, 2\} & \text{if L-DED} \\ S\{(2.5, 5.0, 5.5) : p_i = 0.09, 0.55, 0.36, i = 1, 2, 3\} & \text{if Arc-DED} \\ S\{(5, 125, 240) : p_i = 0.16, 0.16, 0.66, i = 1, 2, 3\} & \text{if E-DED} \end{cases}$
Temperature	Continuous	Mode	Room temperature
Time	Continuous	Mode	0 minutes
Pressure	Continuous	Mode	1 atm
Notch presence	Continuous	Mode	1 (no notch)
Smallest dimension	Continuous	Random	$S\{f(x) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-x/\theta}; k = 3.10, \theta = 0.610\}$
Cross section shape	Discrete	Random	$S\{(\text{circular, rectangular, hollow, scaffold}): p_i = 0.69, 0.22, 0.06, 0.03, i = 1, 2, 3, 4\}$
Orientation	Discrete	Random	$S\{(90, 45, 0): p_i = 0.75, 0.06, 0.19, i = 1, 2, 3\}$
Loading method	Discrete	Random	$S\{(\text{axial, bending, rotating-bending, vibration, other}): p_i = 0.68, 0.07, 0.10, 0.10, 0.05, i = 1, 2, 3, 4, 5\}$
Load ratio	Discrete	Random	$S\{(-1, 0.1): p_i = 0.53, 0.47, i = 1, 2\}$
Frequency	Continuous	Random	$S\{f(x) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-x/\theta}; k = 0.89, \theta = 0.019\}$
Qualitative surface	Discrete	Random	$S\{(\text{net-shape, machined, polished, surface modified}): p_i = 0.23, 0.45, 0.20, 0.12, i = 1, 2, 3, 4\}$

Table 4 continued

Meta-data	Meta-data type	Imputation rule type	Imputed value
Ra	Continuous	Multistage	$S\{(1.7 + 0.15Rz) + 2N(0,1)\} \rightarrow \text{MICE}$
EL	Continuous	MICE	MICE
UTS	Continuous	Multistage	$S\left\{\left(\frac{H_V}{2.9A}\right)\left(\frac{n}{0.217}\right)^n\right\}; A = 1.12, n = 0.07 \rightarrow S\{(61 + YS) + 10N(0,1)\} \rightarrow \text{MICE}$
YS	Continuous	Multistage	$S\left\{\left(\frac{H_V}{3A}\right)0.1^n\right\}; A = 1.12, n = 0.07 \rightarrow S\{(32 + 0.8UTS) + 10N(0,1)\} \rightarrow \text{MICE}$

*Sampling spaces are denoted by $S\{\}$ for a vector of i probabilities p_i . $N(0,1)$ indicates the standard normal distribution.

expanding the number of manufacturing process related keywords from 28 to 50, and using seven testing and six material related keywords, the number of data sources increased by 27%, resulting in the development of a database with over 7000 data-points.

- While automated search tools for identifying data sources are powerful, a non-trivial number of errors were identified when the results were compared with a manual aggregation technique. The application of a three-stage verification process, including side-by-side, double data entry, and exploratory data analysis techniques, reduced the number of duplicated datasets by 13% and other errors in transcribed data by (1–5)%. Application of the additional verification steps also increased the number of recorded meta-data values by 7%, on average.
- The accompanying meta-data for the S-N curves compiled for AM Ti-6Al-4V are not well-defined or standardized and are characterized by significant differences in reporting rates. A data dictionary was developed to provide a standard set of terminology that could be used to improve the meta-data structure that could be more easily integrated into a usable database.
- Based on previous data aggregation work for AM 316L stainless steel, 41 meta-data were identified, but with inconsistent reporting, only seven meta-data display 100% reporting. By applying imputation, the number of viable meta-data increased to 23, thus significantly increasing the usable datasets and limiting the role of listwise deletion.
- The database structure developed for the current AM Ti-6Al-4V fatigue testing dataset will need to be expanded as new datasets and their accompanying meta-data become available. A structured

approach for introducing new datasets in a seamless fashion and expanding the available meta-data is proposed. This flexible and expanding database can be used as a reliable source for evaluating the efficacy of emerging ML and LLM tools with a comprehensive high-fidelity dataset.

Acknowledgements

This work was financially supported by the Navy Small Business Innovation Research/Small Business Technology Transfer (SBIR/STTR) program under contract N68335-20-C-0146. Support from the American Welding Society Graduate Research Fellowship is greatly appreciated. These opinions, recommendations, findings, and conclusions do not necessarily reflect the views or policies of NIST or the United States Government.

Funding

Small Business Innovative Research and Small Business Technology Transfer, N68335-20-C-0146, Xiaoliang Zhao, American Welding Society

Data availability

Contents of the database are provided as several de-normalized .CSV files that may be re-structured to be compliant with various database management systems or other methods of viewing and interacting with the information. A data dictionary and an alphabetized list of original sources are also provided.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10853-026-13789-3>.

Appendix 1

See Table 4

References

- [1] Furuya Y, Nishikawa H, Hirukawa H, Nagashima N, Takeuchi E (2019) Catalogue of NIMS fatigue data sheets. *Sci Technol Adv Mater* 20:1055–1072. <https://doi.org/10.1080/14686996.2019.1680574>
- [2] Chopra OK, Shack WJ (2009) A review of the effects of coolant environments on the fatigue life of LWR structural materials. *J Press Vessel Technol Trans ASME*. 1–21 <https://doi.org/10.1115/1.3027496>
- [3] Mandell JF, Samborsky DD (1998) DOE/MSU composite material fatigue database: Test Methods, Materials, and Analysis. 1–202
- [4] Chopra SO, W (2003) Review of the margins for ASME code fatigue design curve – effects of surface roughness and material variability. *Tech Rep NUREG/ CR6815, ANL-02/39*. U.S. Nuclear Regulatory Commission. 1–62
- [5] Wormsen A, Avicé M, Fjeldstad A, Reinås L, Macdonald KA, Muff AD (2015) Base material fatigue data for low alloy forged steels used in the subsea industry. Part 1: in air S-N data. *Int J Fatigue* 80:477–495. <https://doi.org/10.1016/j.ijfatigue.2015.07.007>
- [6] DeLa'O JD, Gundlach RB, Tartaglia JM (2003) Development of a Cast Iron Fatigue Properties Database for Use in Modern Design Methods, 1–48, <https://www.papers3://publication/uuid/EE49D0EB-58C5-49CE-A352-823AE A1CE2B9>
- [7] Beddoes J, Bibby MJ (2003) Principles of metal manufacturing processes. Elsevier Butterworth-Heinemann, Boston, MA
- [8] Battelle Memorial Institute (2019) MMPDS-14 : metallic materials properties development and standardization (MMPDS), 14th ed
- [9] Yamazaki M, Hosoya J (2012) Metallic Materials Database: Kinzoku, NIMS NOW International 10
- [10] America Makes (2022) Joint Metal Additive Database Definition (JMADD) Project Summary
- [11] Hu Z, Huang C, Xie L, Hua L, Yuan Y, Zhang LC (2025) Machine learning assisted quality control in metal additive manufacturing: a review. *Adv Powder Mater.* 100342 <https://doi.org/10.1016/j.apmate.2025.100342>
- [12] Iams AD, Keist JS, Giannuzzi LA, Palmer TA (2021) The evolution of oxygen-based inclusions in an additively manufactured super-duplex stainless steel. *Metall Mater Trans A* 52:3401–3412. <https://doi.org/10.1007/s11661-021-06311-8>
- [13] Lodhi MJ, Iams AD, Sikora E, Palmer TA (2022) Microstructural features contributing to macroscopic corrosion: the role of oxide inclusions on the corrosion properties of additively manufactured 316L stainless steel. *Corros Sci* 203:110354. <https://doi.org/10.1016/j.corsci.2022.110354>
- [14] Zuback JS, Iams AD, Zhang F, Giannuzzi LA, Palmer TA (2022) Stable nitride precipitation in additively manufactured nickel superalloys. *J Alloys Compd* 910:164918. <https://doi.org/10.1016/j.jallcom.2022.164918>
- [15] Wietecha-Reiman JJ, Iams AD, Sabol SM, Palmer TA (2025) Contributions of sub-surface intergranular phases on fatigue crack initiation in additively manufactured austenitic stainless steel. *J Mater Sci.* <https://doi.org/10.1007/s10853-025-11259-w>
- [16] Brennan MC, Keist JS, Palmer TA (2021) Defects in metal additive manufacturing processes. *J Mater Eng Perform* 30:4808–4818. <https://doi.org/10.1007/s11665-021-05919-6>
- [17] Wietecha-Reiman JJ, Segall A, Zhao X, Palmer TA (2023) Combining fractal and topological analyses to quantify fracture surfaces in additively manufactured Ti-6Al-4V. *Int J Fatigue* 166:107232. <https://doi.org/10.1016/j.ijfatigue.2022.107232>
- [18] Zhang M, Zhou X, Li R, Wang M, Yuan T, Xie S (2022) Study of microstructure and compressive strength of a novel Ti-0.75Al-9.5V alloys manufactured by laser deposition melting. *Mater Charact.* 112189 <https://doi.org/10.1016/j.matchar.2022.112189>
- [19] Reijonen J, Björkstrand R, Riipinen T, Que Z, Metsä-Kortelainen S, Salmi M (2021) Cross-testing laser powder bed fusion production machines and powders: variability in mechanical properties of heat-treated 316L stainless steel. *Mater Des* 204:109684. <https://doi.org/10.1016/j.matdes.2021.109684>
- [20] Chern AH, Nandwana P, McDaniels R, Dehoff RR, Liaw PK, Tryon R, Duty CE (2020) Build orientation, surface roughness, and scan path influence on the microstructure, mechanical properties, and flexural fatigue behavior of Ti-6Al-4V fabricated by electron beam melting. *Mater Sci Eng A* 772:138740. <https://doi.org/10.1016/j.msea.2019.138740>
- [21] Shaffer DJ, Wilson-Heid AE, Keist JS, Beese AM, Palmer TA (2021) Impact of retained austenite on the aging

- response of additively manufactured 17–4 PH grade stainless steel. *Mater Sci Eng A* 817:141363. <https://doi.org/10.1016/j.msea.2021.141363>
- [22] Masuo H, Tanaka Y, Morokoshi S, Yagura H, Uchida T, Yamamoto Y, Murakami Y (2017) Effects of defects, surface roughness and HIP on fatigue strength of Ti-6Al-4V manufactured by additive manufacturing. *Procedia Struct Integr* 7:19–26. <https://doi.org/10.1016/j.prostr.2017.11.055>
- [23] Lewandowski JJ, Seifi M (2016) Metal additive manufacturing: a review of mechanical properties. *Annu Rev Mater Res* 46:151–186. <https://doi.org/10.1146/annurev-matsci-070115-032024>
- [24] Nayir S, Keist JS, Palmer TA (2021) Role of changes in heat input on additively manufactured Ti-6Al-4V fabricated by directed energy deposition. *Mater Sci Eng A* 822:141541. <https://doi.org/10.1016/j.msea.2021.141541>
- [25] Keist JS, Nayir S, Palmer TA (2020) Impact of hot isostatic pressing on the mechanical and microstructural properties of additively manufactured Ti-6Al-4V fabricated using directed energy deposition. *Mater Sci Eng A* 787:139454. <https://doi.org/10.1016/j.msea.2020.139454>
- [26] Wolff S, Lee T, Faierson E, Ehmann K, Cao J (2016) Anisotropic properties of directed energy deposition (DED)-processed Ti-6Al-4V. *J Manuf Process* 24:397–405. <https://doi.org/10.1016/j.jmapro.2016.06.020>
- [27] Battelle Memorial Institute. (2024). *Metallic Materials Properties Development and Standardization (MMPDS), Volume II*. Federal Aviation Administration, Washington, D.C., Office of Aviation.
- [28] Rogers J, Qian M, Elambasseril J, Burvill C, Brice C, Wallbrink C, Brandt M, Leary M (2023) Fatigue test data applicability for additive manufacture: a method for quantifying the uncertainty of AM fatigue data. *Mater Des* 231:111978. <https://doi.org/10.1016/j.matdes.2023.111978>
- [29] Wietecha-Reiman IJ, Sabol SM, Palmer TA (2024) Developing statistical tools to analyze historical stress-controlled fatigue data in additively manufactured austenitic stainless steel. *Mater Sci Eng A*. 146580 <https://doi.org/10.1016/j.msea.2024.146580>
- [30] Zhang Z, Xu Z (2023) Fatigue database of additively manufactured alloys. *Sci Data* 10:1–15. <https://doi.org/10.1038/s41597-023-02150-x>
- [31] Rumble, J., "Standards for Materials Databases: ASTM Committee E49," *Computerization and Networking of Materials Databases: Second Volume*, ASTM STP 1106, 1991, p. 73-83.
- [32] Lu Y, Choi S, Witherell P (2015) Towards An Integrated Data Schema Design for Additive Manufacturing Conceptual Modeling, in: IDETC/CIE 2015, Boston, MA
- [33] Roberts C, Wells D, Sen S, Miller T, Serwecki R, Cordner S, Lambert D (2019) Overview of NASA AM Database Efforts Specialty Workshop Series: Additive Manufacturing Data Management & Schema Supporting organizations Sponsors, Tysons, VA
- [34] America Makes, Python-based data management tool parses deliverables, prepares XML files, and saves information to CDCS instance (2021)
- [35] Ashino T (2010) Materials ontology: an infrastructure for exchanging materials information and knowledge. *Data Sci J* 9:54–61. <https://doi.org/10.2481/dsj.008-041>
- [36] G. King, J. Honaker, A. Joseph, K. Scheve (1998) Listwise deletion is evil : what to do about missing data in political science, *Br J Polit Sci* 30. 1-33
- [37] Pearl J (2009) Causal inference in statistics: an overview. *Stat Surv* 3:96–146. <https://doi.org/10.1214/09-SS057>
- [38] Huang L, Yu W, Ma W, Zhong W, Feng Z, Wang H, Chen Q, Peng W, Feng X, Qin B, Liu T (2025) A survey on hallucination in large language models: principles, taxonomy, challenges, and open questions. *ACM Trans Inf Syst*. 1-58 <https://doi.org/10.1145/3703155>
- [39] Singh C, Inala JP, Galley M, Caruana R, Gao J (2024) Rethinking Interpretability in the Era of Large Language Models, 1-12 <http://arxiv.org/abs/2402.01761>.
- [40] Wolf, T, Debut, Lysandre (2020) Transformers: State-of-the-Art Natural Language Processing, in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*. <https://github.com/huggingface/transformers/blob/master/>
- [41] Petroni F, Rocktäschel T, Lewis P, Bakhtin A, Wu Y, Miller AH, Riedel S (2019) Language Models as Knowledge Bases?, in: *EMNLP 2019*, 1-11 <http://arxiv.org/abs/1909.01066>
- [42] Kuniyoshi F, Ozawa J, Miwa M (2021) Analyzing Research Trends in Inorganic Materials Literature Using NLP, in: *ECML PKDD*, 319-334 <http://arxiv.org/abs/2106.14157>
- [43] ASTM/ISO 52900 (2021) Additive Manufacturing - General Principles - Terminology, ASTM International 2021 1–14. <https://www.iso.org/obp/ui/#iso:std:69669:en%0Ahttps://www.iso.org/standard/69669.html%0Ahttps://www.astm.org/Standards/ISOASTM52900.htm>
- [44] Pathak J, Bailey KR, Beebe CE, Bethard S, Carrell DS, Chen PJ, Dligach D, Endle CM, Hart LA, Haug PJ, Huff SM, Kaggal VC, Li D, Liu H, Marchant K, Masanz J, Miller T, Oniki TA, Palmer M, Peterson KJ, Rea S, Savova GK, Stancel CR, Sohn S, Solbrig HR, Suesse DB, Tao C, Taylor DP, Westberg L, Wu S, Zhuo N, Chute CG (2013) Normalization and standardization of electronic health records for high-throughput phenotyping: the sharpn

- consortium. *J Am Med Inform Assoc.* e341–e348 <https://doi.org/10.1136/amiajnl-2013-001939>
- [45] Barchard KA, Pace LA (2011) Preventing human error: the impact of data entry methods on data accuracy and statistical results. *Comput Human Behav* 27:1834–1839. <https://doi.org/10.1016/j.chb.2011.04.004>
- [46] Day S, Fayers P, Harvey D (1998) Double data entry: what value, what price? *Control Clin Trials* 19:15–24
- [47] Tukey J (1977) *Exploratory data analysis*. Addison-Wesley Pub. Co., Reading, MA
- [48] Jäger S, Allhorn A, Bießmann F (2021) A benchmark for data imputation methods. *Front Big Data* 4:1–16. <https://doi.org/10.3389/fdata.2021.693674>
- [49] Madley-Dowd P, Hughes R, Tilling K, Heron J (2019) The proportion of missing data should not be used to guide decisions on multiple imputation. *J Clin Epidemiol* 110:63–73. <https://doi.org/10.1016/j.jclinepi.2019.02.016>
- [50] Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, Wood AM, Carpenter JR (2009) Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ* 338:b2393. 1–11 <https://doi.org/10.1136/bmj.b2393>
- [51] Keist JS, Palmer TA (2017) Development of strength-hardness relationships in additively manufactured titanium alloys. *Mater Sci Eng A* 693:214–224. <https://doi.org/10.1016/j.msea.2017.03.102>
- [52] Jäger S, Allhorn A, Bießmann F (2021) A benchmark for data imputation methods. *Front Big Data*. 1–16 <https://doi.org/10.3389/fdata.2021.693674>
- [53] Hadeed SJ, O'Rourke MK, Burgess JL, Harris RB, Canales RA (2020) Imputation methods for addressing missing data in short-term monitoring of air pollutants. *Sci Total Environ.* 139140 <https://doi.org/10.1016/j.scitotenv.2020.139140>
- [54] van Buuren S, Groothuis-Oudshoorn K (2011) mice: Multi-variate Imputation by Chained Equations in R. *J Stat Softw.* 1–67 <https://doi.org/10.18637/jss.v045.i03>
- [55] Marsaglia G, Tsang WW, Wang J (2003) Evaluating Kolmogorov's Distribution. *J Stat Softw.* 1–4 <https://doi.org/10.18637/jss.v008.i18>
- [56] Conover WJ (1971) *Practical nonparametric statistics*. John Wiley and Sons Ltd, New York, NY
- [57] Ghods S, Schur R, Montelione A, Schleusener R, Arola DD, Ramulu M (2022) Importance of build design parameters to the fatigue strength of Ti6Al4V in electron beam melting additive manufacturing. *Materials* (Basel). 1–21 <https://doi.org/10.3390/ma15165617>
- [58] Hernandez MJ (2020) *Database design for mere mortals*, 4th ed. Addison-Wesley Professional
- [59] Date CJ (2019) *Database design and relational theory: normal forms and all that jazz*. Apress Media LLC. <https://doi.org/10.1007/978-1-4842-5540-7>
- [60] Toukach P (2025) Carbohydrate structure database: current state and recent developments. *Anal Bioanal Chem* 417:1025–1034. <https://doi.org/10.1007/s00216-024-05383-w>
- [61] Taipalus T (2025) On the effects of logical database design on database size, query complexity, query performance, and energy consumption. 1–21 <http://arxiv.org/abs/2501.07449>
- [62] A. Gupta, S. Tyagi, N. Panwar, S. Sachdeva, U. Saxena (2017) NoSQL databases: critical analysis and comparison, 2017 Int Conf Comput Commun Technol Smart Nation, IC3TSN 2017 2017–Octob. 293–299. <https://doi.org/10.1109/IC3TSN.2017.8284494>.
- [63] Moher D (1994) Statistical power, sample size, and their reporting in randomized controlled trials. *JAMA* 272:122–124 <https://doi.org/10.1001/jama.1994.03520020048013>
- [64] Cohen J (1988) *Statistical power analysis for the behavioral sciences*, 2nd ed. Lawrence Erlbaum Associates
- [65] Egger M, Smith GD, Sterne JAC (2001) Uses and abuses of meta-analysis. *Clinical Medicine* 1. 478–484
- [66] F3184 - Specification for Additive Manufacturing Stainless Steel Alloy (UNS S31603) with Powder Bed Fusion (2016) <https://doi.org/10.1520/F3184-16>
- [67] Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M, Welch V (eds) (2024) *Cochrane handbook for systematic reviews of interventions*. John Wiley & Sons
- [68] Page MJ, Moher D, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, Chou R, Glanville J, Grimshaw JM, Hróbjartsson A, Lalu MM, Li T, Loder EW, Mayo-Wilson E, McDonald S, McGuinness LA, Stewart LA, Thomas J, Tricco AC, Welch VA, Whiting P, McKenzie JE (2021) PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. *The BMJ* 372. 1–36 <https://doi.org/10.1136/bmj.n160>
- [69] Lones MA (2024) Avoiding common machine learning pitfalls. *Patterns* (NY) 5:101046. <https://doi.org/10.1016/j.patter.2024.101046>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.