



Information reuse of nondestructive evaluation (NDE) data sets

Frank Leinenbach, Christopher Stumm, Fabian Krieg, and Aaron Schneider

Fraunhofer IZFP, Institute for Nondestructive Testing, Campus E3.1, Saarbruecken 66123, Germany

Correspondence: Frank Leinenbach (frank.leinenbach@izfp.fraunhofer.de)

Received: 22 September 2023 – Revised: 4 March 2024 – Accepted: 5 March 2024 – Published: 15 May 2024

Abstract. To achieve added value from data spaces and data sets in general, an essential condition is to ensure the high quality of the stored information and its continuous availability. Nondestructive evaluation (NDE) processes represent an information source with potential for reuse. These provide essential information for the evaluation and characterization of materials and components. This information, along with others such as process parameters, is a valuable resource for data-driven added value, e.g., for process optimization or as training data for artificial intelligence (AI) applications. However, this use requires the continuous availability of NDE data sets as well as their structuring and readability. This paper describes the steps necessary to realize an NDE data cycle from the generation of information to the reuse of data.

1 Motivation

Only a few aspects of digital transformation (Vrana and Singh, 2022) have been realized in nondestructive evaluation (NDE). For example, inspection results are often recorded manually on pre-printed documentation sheets, and the corresponding inspection results are stored in classic document folders or as PDFs on local hard drives. The captured raw data or processed data from the inspection system are often not archived. Regardless of the archiving method, however, these results are usually deleted again after a specified retention period. This circumstance contradicts current digitization strategies, in which the availability and networking of information represent an essential basis for optimizing process steps or predicting maintenance intervals, for example. At the same time, nondestructive evaluation methods are a valuable source of data, since NDE results always contain qualitative or quantitative information regarding the properties and defects of components (Böttger et al., 2020).

The reason for the differences in the degree of automation and the heterogeneity of the data structure is the diversity and complexity of the NDE methods. These include, among others, ultrasonic, eddy current, thermographic and X-ray testing methods but also penetrant and acoustic emission testing. In addition, there are combined testing methods such as the MAGNUS testing system (Szielasko et al., 2014), which

allows a combination of ultrasonic and micromagnetic testing. In the past, the diversity of test methods and their test data led to the fact that data formats and their structures were mostly defined by the device manufacturer, and thus, even in a group of test techniques, there are hardly any overlapping data formats. These boundary conditions complicate on the one hand the provision of information by the proprietary data format and on the other hand the interpretation and examination of the completeness of the data due to individual information models. In addition, the reading of the data sets is also tied to the manufacturer, so that, in the worst-case scenario, data sets will no longer be able to be used in the future due to missing software. Changing this circumstance is the goal of the ongoing digital transformation of NDE towards NDE 4.0 (Vrana and Singh, 2021) and includes, among other things, the structured digital archiving of NDE data sets with the long-term goal of establishing a NDE data space using current networking technologies.

In addition to the direct use of the information, e.g., in the context of KPI (key performance indicator) analyses, the reuse of the collected data sets offers significant long-term advantages over the current state of the art, from which AI (artificial intelligence) applications in particular benefit. For example, training an artificial neural network (ANN) requires a large number of suitable and diverse data sets, and when

there is a lack of real data sets, synthetic data sets are often generated and included, although real data sets are always preferred (Faltings et al., 2021). Furthermore, NDE data sets covering the entire history of a component could also be used to optimize recycling processes. In order to establish a suitable archive of relevant data records in the long term, it is necessary to build up a data ecosystem in which all aspects, from the generation of information to its reuse, are considered and described.

In this paper we propose a general workflow for storing NDE data in a sustainable manner. This allows the generation of additional value from the archived data by engineers and data scientists similar to cyber-physical loops (Vrana, 2021). Based on the insights gained, processes can be optimized, and even new business possibilities might arise for company management. Use cases where such an archiving strategy is beneficial are manifold. One example is the monitoring of components over their whole lifecycle, providing knowledge about aging and external influences. Such insights can ultimately lead to an improvement in the design or materials of components. Also, monitoring of flaws and their influence and growth over time can lead to a better understanding of their severity. Furthermore, this allows the automated comparison of multiple components, allowing improved monitoring of production, production quality and development over time.

The necessary prerequisites for a data cycle can be derived from the requirements of FAIR data sets (Wilkinson et al., 2016). The acronym FAIR stands for findable, accessible, interoperable and reusable. These are requirements for predominantly scientific data sets in order to make them accessible for usage scenarios that go beyond the current project. The criteria of the FAIR guiding principles are as follows.

To be findable,

- F1. (meta)data are assigned a globally unique and persistent identifier,
- F2. data are described with rich metadata (defined by R1 below),
- F3. metadata clearly and explicitly include the identifier of the data it describes, and
- F4. (meta)data are registered or indexed in a searchable resource.

To be accessible,

- A1. (meta)data are retrievable by their identifier using a standardized communications protocol,
 - A1.1 the protocol is open, free, and universally implementable,
 - A1.2 the protocol allows for an authentication and authorization procedure where necessary, and
- A2. metadata are accessible, even when the data are no longer available.

To be interoperable,

- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation,
- I2. (meta)data use vocabularies that follow FAIR principles, and
- I3. (meta)data include qualified references to other (meta)data.

To be reusable,

- R1. meta(data) are richly described with a plurality of accurate and relevant attributes,
 - R1.1. (meta)data are released with a clear and accessible data usage license,
 - R1.2. (meta)data are associated with detailed provenance, and
 - R1.3. (meta)data meet domain-relevant community standards.

Even if these requirements are aimed in particular at research data, these principles are also transferable to other sectors and should be applied as far as possible in terms of interoperability and connectivity.

2 Technical background of data management

Technologies for managing, archiving, and distributing data are required to achieve the above goals. In the following, we therefore provide an overview of relevant technologies for the implementation of a data ecosystem. In glycoinformatics, which is a field of informatics that deals with the acquisition, analysis, and interpretation of information about glycans, data set interoperability was realized by structuring data sets and using a SPARQL database (Egorova and Toukach, 2018). A similar benefit of databases was described in Moser et al. (2023). By linking structured data in the field of medical technology, relevant information can be exchanged between different actors in a secure and timely way, thus avoiding delays in medical treatment. Another example of the benefits of structured data for healthcare is discussed in Lohler et al. (2014). Here, an ABHAB database (Abbreviated Profile of Hearing Aid Benefit) enables the assessment of individual patient data against the background of a larger data set. While previous approaches have focused on highly structured data sets, the interest in managing unstructured data sets is increasing in the context of current Industry 4.0 activities. To address this, Klettke et al. (2014) discuss fundamental challenges in using NoSQL database systems in an agile application development. The preservation of schema flexibility is emphasized as an important prerequisite. Another application example for NoSQL databases can be found in Bach et al. (2016). Here, the use of the database

with web applications was investigated, and the performance in the case of unpredictable load peaks was highlighted. In addition, Vrana et al. (2018) described in their paper about smart data analysis of ultrasonic inspections for probabilistic fracture mechanics the successful usage of an SQL database which led to a more realistic lifetime estimation due to this enhanced database approach.

Data spaces represent another important aspect of data interoperability. A general overview of a European data infrastructure is presented in Otto and Burmann (2021). The authors discuss the definition and characteristics of data spaces, the current state of the art, as well as possible future research fields. For example, Strieth-Kalthoff et al. (2022) use the lack of networking and data sets to discuss the limitations of chemical prediction models. In Noll et al. (2020), the benefits of data spaces for the circular economy are presented using the example of cellphones and server boards. In this context, information from the classical production chain is linked to the inverse production chain in the data network. Security aspects of data spaces are addressed in Eckert (2017). The security aspects of the Industrial Data Space (IDS) are presented, which realizes the workflow of data through external rights management.

Data workflow is also an important aspect of a data ecosystem, describing the hierarchy of organization of individual process steps, whereby such workflows are very individual, depending on the industry. The following examples show how such a workflow can succeed, e.g., in Zander et al. (2015) with a data workflow for the acquisition of diffraction values of randomly oriented crystals. Similarly, Yuan and Fischer (2021) show a workflow to predict effective diffusion coefficients of radionuclides in clay-rich formations. In Kobos et al. (2021), an optimized data workflow is used to provide critical information in a timely manner and thus optimize the quality of research results. In the maritime research environment, an example of structuring and managing heterogeneous environmental data sets can be found (Bruder et al., 2017). This work also highlights the need for research data management. For the NDE, classical workflows but no data workflows are currently described by various standards, although initial ideas have already been presented (Valeske et al., 2020, 2022).

The need for long-term archiving of data sets in the NDE is also demonstrated by the example of the Inspection and Revision Management System (IRMS) database, which was used, for example, for the archiving of ultrasonic inspection results of railroad components (Völz et al., 2014).

3 Data acquisition and generation

The very first step of every data cycle is the gathering of data. Regardless of the source of the data, be it from ultrasonic measurements with thousands of data points per specimen or from humidity measuring with one value per hour, the data

are acquired under specific conditions. Every measurement takes place at a concrete location at a specific point in time with certain device settings and different environmental parameters. Most of these conditions can and should be defined beforehand and should be held constant during the experiments, but some are intentional variations by design such as increasing age of the specimens by an artificial aging process. In order to gain real added value from these data, all these conditions should be attached to the data set as metadata. The same procedure is also applicable in a real-life scenario outside a laboratory where a component is inspected over time continually. A comparative analysis at a later date can only be performed properly if the metadata describe the conditions during the experiment properly and as completely as possible.

A scientific study always starts with the design of the experimental phase. In this phase the measurement methodologies and their different measurement effects, whether intentional or unintentional, such as the influence of the environmental temperature fluctuation, should be known and explicitly described. In addition, the specimens' types of material, dimensions, reference status and variation of one or multiple properties follow and can also influence the choice of a specific measurement technique or procedure. The variation of material properties should then be chosen considering how many steps in the variation are statistically necessary and realizable, which unintentional influences should be considered in the production of these specimens and how these can be eliminated or compensated for, such as in randomization. Additional further considerations can ensure the homogeneity of the samples or the reproducibility of the entire experiment.

The next phase is the measurement process itself. The first or all specimens are available, and their current status like dimensions or composition should be documented. Also, photographs with rulers can be useful documentation which can be attached to the data set, but the metadata should also contain these dimension values in a machine-readable manner. If the measurement procedure changes some material properties (maybe only temporarily like in ultrasonic testing with the specimen placed in a water bath), this should be digitally documented as well. Any further environmental conditions, assuming they are known, should also be part of the metadata. In addition, every measurement is performed at a specific position and arrangement between the measurement probe and specimen and is a mandatory part of the final data set. A rule of thumb is to document everything imaginable to enrich the data set for later analysis.

With the raw measurement data recorded, the next step is data reduction or compression. Data reduction can be achieved in several ways. For example, if there are any areas which do not show the specimen or any reference area, they could be removed from the data set. This can be considered for all dimensions the measurement technique provides. In the end, any suitable compression functionality can be ap-

plied to the data set (either the raw data set or the whole data set, including the metadata), depending on which application or archive system is used.

4 Data structuring and data archiving

In addition to the criteria already mentioned for the acquisition of data, the structuring and archiving of data sets represents a significant challenge, which directly addresses the requirements mentioned in the motivation. The structuring of a data set is defined by its information model. It describes which information is recorded in the data record, at which position (tag) the data are to be found, and which data type is intended. The importance of a clear description of a tag is often overlooked. If none is available, it is the task of the user to decide which information is to be provided. For example, consider a tag named “Component”. Without further information, this could be a technical measuring element of a system. It could also be a component to be tested. While for this example it can often be inferred from the context of the information model, other tags without a description are less obvious to classify. Thus, the explicit description of an information model is an essential prerequisite for the correctness, searchability and completeness of the data records.

The information model is usually coupled with the data format. Proprietary data formats usually have information models with a unique structure. This is necessary because these data formats are often used in closed-data ecosystems of a manufacturer and are therefore optimized for the present specific requirements. This ensures the completeness of a data set and optimized computational performance for a specific use case but at the same time limits the availability of the data to participants in the ecosystem. For use in other data ecosystems, manufacturers usually offer the export of their data sets in generic open-data formats such as CSV (column-separated value), JSON (Javascript Object Notation), XML (Extensible Markup Language) or HDF5 (Hierarchical Data Format 5).

Many open-data formats are text-based to allow easier access. Text-encoded formats especially introduce several limitations. For example, tag descriptions are often coupled with values and described as text. The description of the tags is mostly limited to the tag name, which means that without documentation the definition of the values is missing. The file size is several times larger than the original because both values and tags are represented as text. Furthermore, processing the information as text requires more time and computing resources than data in other encodings due to additional processing steps.

An alternative to generic open-data formats is structured open-data formats, which both have a predefined structure and at the same time enable interoperability within open-data ecosystems. These properties make structured open formats an ideal system for archiving information, as they fulfill all

criteria for the reuse of data sets. In contrast to proprietary formats, they can be supported by arbitrary parties, and accessibility is ensured. In contrast to generic data formats, the predefined structure allows interpretability and reuse. One structured open-data format that is of particular importance for the NDE and that fulfills the criteria mentioned above is DICONDE (Brierley et al., 2023), which is explained in more detail below.

In DICONDE, NDE data sets of various test procedures can be recorded, transmitted and archived in a structured manner (Vrana, 2021). The first version of the standard was published in 2004 and is based on the medical standard DICOM (Digital Imaging and Communications in Medicine) from 1985, which is used today in almost every imaging or processing system in medical technology and which utilizes the PACS (Picture Archiving and Communication System). In the hierarchy of a DICONDE data set, the component to be examined represents the highest-order element. Each component can then be assigned studies (measurement campaigns), which in turn consist of several series (e.g., individual measurements). This hierarchy is also found in the underlying DICOM format, where, for example, instead of the component, the patient is the highest-order element. Equivalent to DICOM, DICONDE is thus designed to track the history of a component over its entire lifecycle. Figure 1 schematically shows the basic elements of the DICONDE information model.

Based on this structure, general and method-specific standards have been published by Subcommittee E07.11 of ASTM International. General standards include instructions as well as explanations of implementation and interoperability, while the method-specific standards extend the general descriptions with additionally required fields. Table 1 lists general standards, and Table 2 lists method-specific standards.

For all the methods mentioned in Table 2, except for eddy current testing (ASTM E2934), a counterpart exists from medical technology. In addition to the meta-information in the methods mentioned, data are saved either as an image, a PDF or a raw data format. This step is independent of the process, which essentially differs in the meta-information used. The scope of the standards makes it clear that a basis for archiving NDE records already exists. However, this basis does not currently cover all known test methods, but since it is an open standard, additional use-case-specific information can be added as private tags. Examples of this are current developments by the ASTM regarding an extension of the standards for eddy current (E2934) and ultrasound (E2663) in order to map waveforms. This allows, for example, phased-array ultrasound data to be mapped in the DICONDE standard. Another feature is the combination of DICONDE with a PACS server known from medicine. Like DICOM, the DICONDE standard provides for a combination with the PACS and thus also for revision-safe, structured and central storage of NDE data sets.

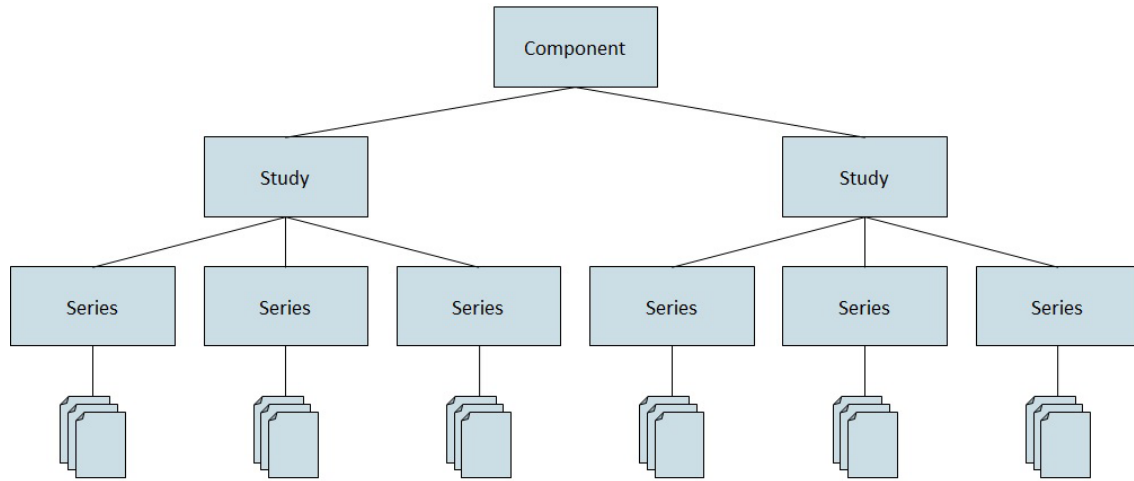


Figure 1. Representation of the basic elements of the information model of DICONDE starting from the component up to the contents of a series.

Table 1. General digital imaging and communication in nondestructive evaluation (DICONDE) standards.

Document	Description
ASTM E3169	Standard guide for digital imaging and communication in nondestructive evaluation
ASTM E3147	Standard practice for evaluating DICONDE interoperability of nondestructive testing and inspection systems
ASTM E2339	Standard practice for digital imaging and communication in nondestructive evaluation
ASTM E1316	Standard terminology for nondestructive examinations
ASTM E3267	Standard guide for building information models and archiving for digital imaging and communication in nondestructive evaluation

In addition to DICONDE, there are other approaches to storing NDE data sets in a harmonized way. The ZEUS data format was developed in 1997 and aimed to link different hardware and software systems (Völz et al., 2007). NDTML (XML for NDT) was another attempt in 2001 to establish a common data format and data structure based on XML (Gilbert and Onoe, 2001). VMAP (Michels et al., 2022) and the NDE data format (Peloquin, 2024) are HDF5-based approaches for securing test data. While the ZEUS or NDTML approaches have not been able to establish themselves, the approaches based on HDF5 are also relevant data formats for NDE. Also, non-specific NDE data formats such as AQDEF (Dietrich, 2019) are possible solutions for structuring NDE data sets. The advantage of DICONDE lies in its standardization for NDE, while the other formats mentioned are currently only used by small groups or, as in the case of AQDEF, have not been developed specifically for NDE applications.

5 Database handling and data extraction

The archived data are usually organized into some kind of database. Given the use case of larger amounts of NDE data

to be archived and reused, both the database management system and the user interfaces must meet certain criteria. The database must be accessible to all stakeholders to perform data archiving or data analysis tasks. At the same time, however, access must be restricted so that, for example, only authorized persons within the company have access rights. In addition, the availability of the database itself must be ensured, while the contents must also be protected against unwanted changes, including those from unsanctioned external access or destructive forces. Finally, the database must be searchable so that data sets can be identified for reuse in an efficient manner.

In the case of DICONDE, the operations for accessing the database, which are usually based on the PACS, are precisely defined as part of the standard. These are so-called services. Each DICONDE system implements a store operation for records that send data, including all metadata, to a PACS. Similar operations are defined for searching and retrieving data, thus realizing the already mentioned requirements regarding accessibility and searchability. Furthermore, data cannot be edited by design, which is mandatory for keeping archived data valid.

Table 2. Method-specific DICONDE standards.

Document	Description
ASTM E2738	Standard practice for DICONDE for computed radiography (CR) test methods
ASTM E2699	Standard practice for DICONDE for digital radiography (DR) test methods
ASTM E2767	Standard practice for DICONDE for X-ray computed tomography (CT) test methods
ASTM E2663	Standard practice for DICONDE for ultrasonic test methods
ASTM E2934	Standard practice for DICONDE for eddy current (EC) test methods

Regarding the safety aspects of the system, however, the following issues must be considered. As the DICOM network protocol was originally designed in the 1990s, some security risks exist with the default configurations (Eichelberg et al., 2020). By default, network traffic is not encrypted, though DICOM also defines TLS-encrypted messages. Furthermore, user access control on the PACS server was defined in 2004 but is not enforced. Also, the access control mechanism is not fixed – different mechanisms can be implemented (NEMA, 2023). For this reason, the use of encrypted communication with the PACS, as well as the establishment of a user authentication system, is mandatory to fulfill the integrity of the data. Also, the physical access to the server has to be limited. Furthermore, there are limitations regarding queries on the database, which are imposed in the DICOM standard (NEMA, 2013 – Sect. 6.7). When searching for specific studies or series, the tags that are searchable are limited. Since this tag was originally defined for medical use, the DICONDE equivalents of the tags do not cover all query use cases in the NDE. For example, the name of a component (in the DICOM of the patient) is queryable. In contrast, in the NDE more common attributes of where an inspection took place and who conducted it are not searchable in the standard. Consequently, DICONDE has shortcomings in terms of data set searchability and accessibility.

Due to the lack of a complete query for all attributes of the data sets, the PACS is an excellent data archive, although it is only suitable to a limited extent for the reuse of data sets outside a defined data workflow. To overcome this circumstance, it is possible to combine the PACS with other database technologies, whose functionality, especially regarding the realization of complex search queries, represents a supplement to this system. SPARQL or NoSQL, for example, could be used here. In this case, the complementary database could mirror the meta-information of the PACS, with the actual measured values remaining in the archive. As a result, search queries in the complementary database can output unique identifiers (UIDs) of data sets, which can then be obtained directly from the PACS. This also enables connectivity to data ecosystems outside the local network. In addition, NoSQL in particular allows the management of various data structures, which also

enables the integration and linking of non-NDE data structures.

6 Data reuse and potential impacts

The last step in the cycle shown represents its transition point. An important difference here compared to classic cycles such as material cycles is that the reuse of data does not change or destroy it. This circumstance leads to the fact that in data cycles new information is constantly won and thus the data basis for the utilization grows. Data sets from similar sources are usually used in reuse, and therefore new data sets of the same type are usually generated. For example, ultrasound data sets are likely to be used predominantly for training an evaluation algorithm for ultrasound. Conversely, only new ultrasound data sets are likely to be created as part of the application. Exceptions to this could be the use simply as comparative data sets or for validation for novel procedures without an existing database. This limit of a data ecosystem must be considered and furthermore shows the relevance of synthetically generated data sets. Regardless of this limitation, the continuous increase in the data treasure holds advantages. In the following, short-term, medium-term, and long-term added values are mentioned. There is no claim to completeness in this naming. Also, the definition of the mentioned categories is subjective, since both the time limits and the required number of data sets are individual.

A first short-term advantage is the consistent documentation of all the data sets. The unambiguous structure of the data sets and their assignment allow data sets to be viewed in relation to one another and, in the case of measurement data, can reflect the history of a component. At the same time, the comparability of data sets is increased by the information models. For example, results from a thermographic inspection can be compared with the results of an ultrasonic inspection of the same component, since the essential information is organized in the same way.

In the medium term, the benefit of a data cycle will lead to sufficient databases being obtained, which are needed for the calibration of systems or the training of AI procedures. In many cases, this task is addressed as an individual solution approach, and measurement campaigns are conducted

to generate training data. By using existing data sets from past tests, the generation of additional training data can be reduced although probably not completely avoided. It must also be considered that the data sets used must be relevant data sets for the application.

The exact implications for long-term use over a period of years are difficult to assess, and therefore only conjecture is possible. However, it can be assumed that the data treasure will reach a point where many modalities and variants of data sets are available, thus minimizing the effort required for additional acquisition of data sets. Whereas previously a wide variance of data sets had to be reacquired, in the long run only certain constraints will have to be considered. At the same time, the question of the relevance of certain data sets will arise due to the storage requirements. The problem is already known from the development away from “big data” to “relevant data” and will presumably also be important for data cycles and data spaces in the future.

7 Implementation of data cycles

The implementation of data loops for NDE data sets is currently being realized at Fraunhofer IZFP to achieve long-term added value from the collected information. In this context, DICONDE was chosen as the structured open-data format for the acquisition of all the measurement results due to its focus on NDE processes and the standards defined by ASTM. In the area of data acquisition, this means that already existing data acquisition programs were extended by functions for creating DICONDE-conform files and interfaces for communication with the PACS. For this purpose, free libraries for the creation of DICOM files as well as for the communication with residual interfaces were used and implemented for the backup of ultrasound data sets. Since a complete implementation of such software modules in all the NDE procedures at Fraunhofer IZFP is a time-consuming process, a simple manual tool was created for the transition, which allows employees to upload their data sets directly, including the essential relevant information regardless of the method. For this purpose, the following tags of ASTM guideline 2339 are recorded in the first step and linked to the data sets in order to be able to assign them to specific components as well as series and projects.

The last entry (0050,0020) represents a self-defined tag, which gives the user the option of specifying additional information about the inspection system used.

To perform complex search queries, a Mongo DB NoSQL server is set up in parallel, and the elements listed in Table 3 are linked by means of a BSON file. This allows users to search specifically for relevant records for their application scenarios and extract them from the PACS server based on the UID of each record. At present, this access is still rudimentary but will be optimized by means of suitable tools, so that a direct download of the data sets will be possible. The

Table 3. List of relevant DICONDE tags for the first implementation at Fraunhofer IZFP.

Attribute category	DICONDE field name	DICONDE tag
Component	Component ID number	0010,0020
Component	Component name	0010,0010
Component	Component manufacturer	0014,0028
Material	Material name	0010,2160
Material	Material notes	0014,0046
Study	Accession number	0008,0050
Study	Component owner name	0008,0090
Study	Study description	0008,1030
Study	Study date	0008,0020
Study	Study time	0008,0030
Study	Performed location	0040,0243
Study	Examination notes	0032,4000
Series	Inspector name	0008,1050
Series	Modality	0008,0060
Series	Series description	0008,103E
Series	Series number	0020,0011
Series	Series date	0008,0021
Equipment	Software version	0018,1020
Equipment	Manufacturer	0008,0070
Equipment	Company name	0008,0080
Equipment	Company address	0008,0081
Equipment	Manufacturers model name	0008,1090
Equipment	Device serial number	0018,1000
Equipment	Device description	0050,0020

described implementation thus makes it possible to introduce already existing inspection data sets into the new circuit, to extract specific data sets for new application scenarios, and to return new results to the circuit in the same way.

Another example of such a data cycle described above, which was not based on DICONDE but on JSON-based attached metadata, was a European-funded research project with 13 partners across Europe with the intention of finding a measurement method or a combination of multiple methods which is able to determine the current age of a cable (TeaM Cables H2020, 2023). On average, about 25 000 cables with a total length of 1500 km are installed in a nuclear power plant. In part, these cables are exposed to rough environmental conditions such as high temperatures and radioactive radiation. This can result in aging of the cable insulation with embrittlement and thus cracks and short circuits as possible consequences. All thinkable destructive and nondestructive methods starting from the microscale up to the macroscale were used to measure almost 1200 different specimens. This total number of specimens comes from the fact that all the specimens varied with seven different material compositions (e.g., with or without specific antioxidants or flame retardants), four different geometrical forms (tape, sheet, coaxial cable, twisted-pair cable), seven different ag-

ing types (thermal- and/or radiation-aged), and six successive aging steps.

The data set of each measurement contains the above-mentioned specimen variation, its dimensions, its status after the aging process, and all the properties and settings of the specific measurement method along with the measurement data themselves. A closer look at the experimental design reveals that this entire data set containing all measurements and metadata per sample does not describe one single experiment but many experiments together, so it is mandatory to consider this fact in the design of the algorithms for later analysis. The analysis in this project had the main goal of finding a correlation between the measurements of a particular measurement method of a specimen and its age, which was defined as its measured elongation at the break value.

Considering all the above components of the data cycle, the project consortium was able to find several appropriate methods to measure the age of a cable and decide whether the cable should be replaced or kept in service (Hettal et al., 2021; Suraci et al., 2022).

8 Summary and outlook

Within the scope of this paper, essential steps for the establishment of a data cycle were discussed, boundary conditions were defined, and possible technological solutions for the achievement of the objective were presented. Aspects of data acquisition, generation, structuring, archiving, database handling, data extraction, and reuse were discussed and potential approaches identified. Special focus was on the data format DICONDE, whose structuring and archiving techniques primarily address test data, as well as on the combination of the PACS with other database technologies such as SPARQL and NoSQL. In addition, two examples were presented in which this schema was implemented using DICONDE and JSON.

While initial steps have been presented in these examples, further steps are required to unlock the potential of a data loop within a data ecosystem. On the one hand, archiving all relevant information of diverse NDE methods in the DICONDE format requires the development of a DICONDE-conforming description of methods for which no standard exists yet, such as the 3MA testing method (Wolter et al., 2019). At the same time, raising the awareness of users of the data cycle is an essential prerequisite for its success. Relevant data can only exist in the loop if the participants also enter it. Furthermore, for the long-term success and benefit of an NDE data loop, concepts must be implemented that extend beyond the NDE community and connect the data loop to other data spaces, thus making the NDE's own know-how available to other parties while at the same time opening their know-how to the NDE's own applications. For this purpose, the linking of the metadata with the reference to the data sets using the DICONDE UID and suitable database technologies such as SPARQL or NoSQL was presented as a solution. By achiev-

ing this stated goal, quality information from NDE technology can be made available in existing data spaces and thus have the potential to generate new added value and take on the role of a valuable data provider in data ecosystems.

Code availability. As no algorithms for analyzing or evaluating data sets are addressed in this paper, unfortunately no sample code is available. The basis of the data cycle was achieved using the following technologies: a commercial PACS server as a data archive (<https://www.dimate.de/en/what-is-pacs.html>) (Dimate GmbH, 2024) and <https://www.mongodb.com/> (NoSQL server, 2024).

Data availability. No data sets were generated as part of this paper that are necessary for the plausibility of the results mentioned. Relevant information regarding the data structure can be found in the standards mentioned in Tables 1 and 2. If required, sample data sets can be provided by the corresponding author on request.

Author contributions. FL, CS, FK, and AS designed the concept and wrote the manuscript draft. In addition, FL, CS, FK, and AS reviewed and edited the manuscript. The aforementioned implementations were realized by FL, CS, and FK.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors.

Special issue statement. This article is part of the special issue "Sensors and Measurement Science International SMSI 2023". It is a result of the 2023 Sensor and Measurement Science International (SMSI) Conference, Nuremberg, Germany, 8–11 May 2023.

Financial support. The project TeaM Cables has received funding from the Euratom research and training program 2014–2018 under grant agreement no. 755183.

Review statement. This paper was edited by Sebastian Wood and reviewed by three anonymous referees.

References

- Bach, C., Kundisch, D., Neumann, J., Schlangenotto, D., and Whittaker, M.: Dokumentenorientierte NoSQL-Datenbanken in skalierbaren Webanwendungen, *HMD Praxis Der Wirtschaftsinformatik*, 53, 486–498, <https://doi.org/10.1365/s40702-016-0229-6>, 2016.
- Böttger, D., Stampfer, B., Gauder, D., Straß, B., Häfner, B., Lanza, G., Schulze, V., and Wolter, B.: Concept for soft sensor structure for turning processes of AISI4140: DFG priority program 2086, project: In-process soft sensor for surface-conditioning during longitudinal turning of AISI4140, *tm – Technisches Messen*, 87, 745–756, <https://doi.org/10.1515/teme-2020-0054>, 2020.
- Brierley, N., Casperson, R., Engert, D., Heilmann, S., Herold, F., Hofmann, D., Kuchler, H., Leinenbach, F., Lorenz, S., Martin, J., Rehbein, J., Sprau, B., Suppes, A., Vrana, J., and Wild, E.: Specification ZfP 4.0 – 01E: DICONDE in Industrial Inspection, *DGZfP e.V, Berlin*, ISBN: 978-3-947971-32-9, 2023.
- Bruder, I., Klettke, M., Möller, M., Meyer, F., Heuer, A., Jürgensmann, S., and Feistel, S.: Daten wie Sand am Meer – Datenerhebung, -strukturierung, -management und Data Provenance für die Ostseeforschung, *Datenbank Spektrum*, 17, 183–196, <https://doi.org/10.1007/s13222-017-0259-4>, 2017.
- Dietrich, E.: Geeignete Messprozesse – Valide Informationen, *tm – Technisches Messen*, 86, 528–539, <https://doi.org/10.1515/teme-2019-0104>, 2019.
- Dimate GmbH, <https://www.dimate.de/en/what-is-pacs.html>, last access: 21 April 2024.
- Eckert, C.: Cybersicherheit beyond 2020, *Informatik Spektrum*, 40, 141–146, <https://doi.org/10.1007/s00287-017-1025-6>, 2017.
- Egorova, K. and Toukach, P.: Glycoinformatics: Bridging Isolated Islands in the Sea of Data, *Angew. Chem. Int. Edit.*, 57, 14986–14990, <https://doi.org/10.1002/anie.201803576>, 2018.
- Eichelberg, M., Kleber, K., and Kämmerer, M.: Cybersecurity Challenges for PACS and Medical Imaging, *Acad. Radiol.*, 27, 1126–1139, <https://doi.org/10.1016/j.acra.2020.03.026>, 2020.
- Faltings, U., Bettinger, T., Barth, S., and Schäfer, M.: Impact on Inference Model Performance for ML Tasks Using Real-Life Training Data and Synthetic Training Data from GANs, *Information*, 13, 9, <https://doi.org/10.3390/info13010009>, 2021.
- Gilbert, D. and Onoe, M.: NDTML – XML for NDT, *Asia Pacific Conference on NDT – 2001 – Brisbane (Australia)*, *e-Journal of Nondestructive Testing*, Vol. 6, <https://www.ndt.net/?id=1260> (last access: 21 April 2024), 2001.
- Hettal, S., Suraci, S., Roland, S., Fabiani, D., and Colin, X.: Towards a Kinetic Modeling of the Changes in the Electrical Properties of Cable Insulation During Radio-Thermal Ageing in Nuclear Power Plants, *Application to Silane-Crosslinked Polyethylene*, *Polymers*, 13, 4427, <https://doi.org/10.3390/polym13244427>, 2021.
- Klettke, M., Scherzinger, S., and Störl, U.: Datenbanken ohne Schema? *Datenbank-Spektrum*, 14, 119–129, <https://doi.org/10.1007/s13222-014-0156-z>, 2014.
- Kobos, L., Ferreira, C., Sobreira, T., Rajwa, B., and Shannahan, J.: A novel experimental workflow to determine the impact of storage parameters on the mass spectrometric profiling and assessment of representative phosphatidylethanolamine lipids in mouse tissues, *Anal. Bioanal. Chem.*, 413, 1837–1849, <https://doi.org/10.1007/s00216-020-03151-0>, 2021.
- Lohler, J., Akcicek, B., Kappe, T., Schlattmann, P., Wollenberg, B., and Schonweiler, R.: Development and use of an APHAB database/Entwicklung und Anwendung einer APHAB-Datenbank, *HNO*, 62, 735, <https://doi.org/10.1007/s00106-014-2915-4>, 2014.
- Michels, P., Bruch, O., Evers-Dietze, B., Grotenburg, D., Ramakers-van Dorp, E., and Altenbach, H.: Shrinkage simulation of blow molded parts using viscoelastic material models, *Materialwiss. Werkst.*, 53, 449–446, <https://doi.org/10.1002/mawe.202100350>, 2022.
- Moser, K., Mikolajczyk, R., Bauer, A., Tiller, D., Christoph, J., Purschke, O., Lückmann, S. L., and Frese, T.: BeoNet-Halle – Aufbau einer multifunktionalen Datenbank zur automatisierten Extraktion von Versorgungsdaten aus Haus- und Facharztpraxen, *Bundesgesundheitsbla.*, 66, 569–577, <https://doi.org/10.1007/s00103-023-03691-7>, 2023.
- NEMA: DICOM PS3.18 2013 – Web Services, *DICOM Standards Committee*, 2013.
- NEMA: DICOM PS3.15 2023d – Security and System Management Profiles, *DICOM Standards Committee*, 2023.
- Noll, R., Bergmann, K., Fricke-Begemann, C., and Schreckenberger, F.: Inverse Produktion für nachhaltige Wertstoffkreisläufe – Aktuelle Entwicklungen zur automatisierten Demontage und Entstickung von Elektronikplatinen/Inverse Production for Sustainable Recycling Routes – New Developments for Automated Disassembly of End-of-Life Electronics, *Chem.-Ing.-Tech.*, 92, 360–367, <https://doi.org/10.1002/cite.201900123>, 2020.
- NoSQL server: Mongo DB, <https://www.mongodb.com/>, last access: 21 April 2024.
- Otto, B. and Burmann, A.: Europäische Dateninfrastrukturen, *Informatik Spektrum*, 44, 283–291, <https://doi.org/10.1007/s00287-021-01386-4>, 2021.
- Peloquin, E.: The NDE 4.0 Journey: How Adopting a Universal Open File Format Empowers the Whole Industry, *e-Journal of Nondestructive Testing*, Vol. 29, <https://www.ndt.net/?id=29033> (last access: 21 April 2024), 2024.
- Strieth-Kalthoff, F., Sandfort, F., Kühnemund, M., Schäfer, F., Kuchen, H., and Glorius, F.: Machine Learning for Chemical Reactivity: The Importance of Failed Experiments, *Angew. Chem. Int. Edit.*, 61, E202204647, <https://doi.org/10.1002/anie.202204647>, 2022.
- Suraci, S., Li, C., and Fabiani, D.: Dielectric Spectroscopy as a Condition Monitoring Technique for Low-Voltage Cables: Onsite Aging Assessment and Sensitivity Analyses, *Energies (Basel)*, 15, 1509, <https://doi.org/10.3390/en15041509>, 2022.
- Szielasko, K., Tschuncky, R., Altpeter, I., Dobmann, G., and Boller, C.: Real-Time Monitoring of Crack Growth Behaviour During Impact and Compact Tension Test with Non-Destructive Testing, *Stud. Appl. Electromag.*, 38, 238–246, <https://doi.org/10.3233/978-1-61499-354-4-238>, 2014.
- Team Cables H2020: Euratom research project, <https://www.team-cables.eu/>, last access: 13 September 2023.
- Valeske, B., Osman, A., Römer, F., and Tschuncky, R.: Next Generation NDE Sensor Systems as IIoT Elements of Industry 4.0, *Res. Nondestruct. Eval.*, 31, 340–369, <https://doi.org/10.1080/09349847.2020.1841862>, 2020.
- Valeske, B., Tschuncky, R., Leinenbach, F., Osman, A., Wei, Z., Römer, F., Koster, D., Becker, K., and Schwender, T.: Cognitive sensor systems for NDE 4.0: Technology, AI embedding, valida-

- tion and qualification, *tm – Technisches Messen*, 89, 253–277, <https://doi.org/10.1515/teme-2021-0131>, 2022.
- Vrana, J.: The Core of the Fourth Revolutions: Industrial Internet of Things, Digital Twin, and Cyber-Physical Loops, *J. Nondestruct. Eval.*, 40, 46, <https://doi.org/10.1007/s10921-021-00777-7>, 2021.
- Vrana, J. and Singh, R.: NDE 4.0 – A Design Thinking Perspective, *J. Nondestruct. Eval.*, 40, 8, <https://doi.org/10.1007/s10921-020-00735-9>, 2021.
- Vrana, J. and Singh, R.: Digitization, Digitalization, and Digital Transformation, in: *Handbook of Nondestructive Evaluation 4.0*, edited by: Meyendorf, N., Ida, N., Singh, R., Vrana, J., Springer, Cham, https://doi.org/10.1007/978-3-030-73206-6_39, 2022.
- Vrana, J., Kadau, K., and Amann, C.: Datenanalyse der Ergebnisse von Ultraschallprüfungen für die probabilistische Bruchmechanik, *VGB PowerTech* 07/2018, 38–42, 2018.
- Völz, U., Schenk, G., Montag, H., and Spruch, W.: Anwendung des ZEUS-Urdatenformats in der Praxis, *German Society of NDT – 2007 – Fürth*, <https://www.ndt.net/?id=5100> (last access: 21 April 2024), 2007.
- Völz, U., Heilmann, P., and Scholz, H.: New Generation of Test Benches for Ultrasonic Testing of Solid Axles, in: *11th European Conference on Non-Destructive Testing (ECNDT 2014)*, Prague, 6–11 October 2014, *e-Journal of Nondestructive Testing*, Vol. 19, <https://www.ndt.net/?id=16828> (last access: 21 April 2024), 2014.
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J. W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J., Groth, P., Goble, C., Grethe, J. S., Heringa, J., 't Hoen, P. A. C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., Mons, A., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S. A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., and Mons, B.: The FAIR Guiding Principles for scientific data management and stewardship, *Scientific data*, 3, 1–9, <https://doi.org/10.1038/sdata.2016.18>, 2016.
- Wolter, B., Gabi, Y., and Conrad, C.: Nondestructive Testing with 3MA – An Overview of Principles and Applications, *Appl. Sci.-Basel*, 9, 1068, <https://doi.org/10.3390/app9061068>, 2019.
- Yuan, T. and Fischer, C.: Effective Diffusivity Prediction of Radionuclides in Clay Formations Using an Integrated Upscaling Workflow, *Transport Porous Med.*, 138, 245–264, <https://doi.org/10.1007/s11242-021-01596-0>, 2021.
- Zander, U., Bourenkov, G., Popov, A., De Sanctis, D., Svensson, O., McCarthy, A., Round, E., Gordeliy, V., Mueller-Dieckmann, C., and Leonard, G.: *MeshAndCollect*: an automated multi-crystal data-collection workflow for synchrotron macromolecular crystallography beamlines, *Acta Crystallogr. D*, 71, 2328–2343, <https://doi.org/10.1107/S1399004715017927>, 2015.