DOI 10.5162/SMSI2023/A2.3

Information recycling of NDE data sets

Frank Leinenbach[1], *Bernd Sprau*[1], *Christopher Stumm*[1]
[1] *Fraunhofer IZFP, Institute for Nondestructive Testing, Campus E3.1, 66123 Saarbruecken, Germany*
frank.leinenbach@izfp.fraunhofer.de

Summary:
NDE data provides essential information for the evaluation and characterization of materials and components. These could be used along the entire value chain, for example, for process optimization or as training data for AI applications. However, this requires not only the unique identification of the component or material, but also the continuous readability of the NDE data sets. In this presentation, technical challenges, and solutions for the development of a data cycle are presented.

Keywords: NDE 4.0, Data Cycle, DICONDE, SPARQL, NDE Data Space

Motivation

Today, only a few aspects of current Industry 4.0 goals are reflected in the performance of non-destructive evaluation (NDE). Files are often saved on hard disks and inspection reports are usually stored as PDFs and then deleted again after a specified safekeeping period.

But besides this, NDE methods can be valuable data suppliers for the optimization of e.g., manufacturing processes [1]. However, this is currently affected by the fact that NDE data sets usually do not guarantee a uniform structure, data format or the completeness of all relevant information. These changes are part of the ongoing digital transformation of the NDE towards NDE 4.0 [2] and include, among other things, the structured digital archiving of NDE data sets with the long-term goal of establishing an NDE data space using current networking technologies.

In addition to the direct use of the information, e.g., as part of KPI analyses, the recycling of the collected data sets offers significant advantages over the current state of the art in the long term, from which AI applications in particular benefit. For the training of an ANN, for example, a multitude of suitable and diverse data sets are needed and, in case of a shortage of real data sets, often synthetic data sets are generated and included, whereas real data sets are always preferred [3]. Also, NDE data sets of the whole history of a component could be used for optimization of recycling processes.

In the following, an approach is presented that allows the recycling of NDE data sets based on a data cycle.

The Data Cycle

The sub-aspects for realizing a data cycle can be divided into five essential steps. These are technically independent in their implementation but build on each other technologically. In the following, the sub-steps and their requirements are discussed in general terms, followed by the specific processes used to achieve the results. The sub-aspects are data generation, data archiving, database, data extraction, and the reuse of the data sets.

The data cycle always begins with the generation of data sets. At best, these are carried out during the product life cycle of the product as part of in-service inspection or in case of malfunctions. This already represents a first critical point, because whether data sets are suitable for reuse depends on several factors. The most obvious characteristic is thereby the correctness and valency of the measurement [4]. Since this is influenced by essentially the correct execution of the inspection, this point is unfortunately difficult to verify afterwards. However, it is equally important for reuse that all relevant meta data is recorded as completely as possible. In addition to the information for interpretation of the measurement data, this should also contain as much information as possible about the circumstances under which the results were generated. This includes, among other things, the measurement points, information about the inspected component and, if applicable, other environmental conditions. At the same time, it is difficult to estimate at the time of creation whether an information is relevant for later use, which makes it difficult to freely structure the measurement data and meta data.

The knowledge of the essential factors of suitable data generation has a significant impact on the second step of the data cycle, the archiving of the data sets and thus also the data format. There is a conflict between the proprietary data formats usually specified by the manufacturers of inspection systems and the manufacturer-independent generic data formats. Whereas manufacturer-dependent data formats already combine an extensive structure of raw and meta data in their data sets, the further processing of the data sets is usually only possible with company-specific software. Generic data formats such as XML or CSV, on the other hand, ensure readability by third parties, but the structures are user-specific and thus offer scope for misinterpretation. An alternative is provided by structured open data formats such as DICONDE (Digital Imaging and Communication for Nondestructive Evaluation) [5] and AQDEF (Advanced Quality Data Exchange Format) [6], which offer a specific data structure and open access. These can also be extended by further information fields, if necessary, whereby this again brings potential for misinterpretation of information with itself. The DICONDE data format is particularly suitable for NDE applications since it specifically addresses inspection tasks and is already standardized for common NDE methods.

The specific structuring of the data sets is important for the implementation of a database. This allows data sets to be separated based on their meta data and found by means of a suitable query. For this purpose, SQL (Structured Query Language) or SPARQL (SPARQL Protocol And RDF Query Language) are suitable database languages. These technologies thereby allow complex search queries, whereby relevant data sets can be identified based on the meta values. In a simple file system, on the other hand, it is usually only possible to search and select by file name, format, and creation date. The contents of the files can usually only be viewed in the case of text files, for example.

By structuring the data sets combined with the database, concrete queries can now be processed by the database, which then outputs a series of matching data sets for data extraction. Depending on the technology, the storage location or a UID (Unique Identifier) is used to extract the data from an archive or folder. What remains is the individual use of the data set for a new application. At this point the data cycle closes as soon as the new data sets, which are obtained in this new application, are archived again according to the described pattern, and transferred to the database.

Implementation and outlook

During implementation, ultrasound data was addressed as the first application. These were stored in DICONDE format and archived on a DIMATE DICONDE server. Since the search function of the DICONDE server only allows simple queries, e.g., several criteria cannot be specified as conditions at the same time, an Apache Jena Fuseki SPARQL server was set up in parallel to the DICONDE server, which allows structured queries of the archived data sets based on the DICONDE ontology. The result provides their UID, which allow extraction from the DICONDE server and can be used afterwards for further applications such as extending training data with real data sets for AI algorithms.

The described procedure will be transferred to other NDE methods in the next development iterations. In addition to established NDE methods such as thermography and eddy current, more complex methods such as 3MA will also be addressed. This work represents an essential basis for the implementation of an NDE data space, in which the mentioned setup allows a cross-company data access. However, this also requires a focus on the usage and access rights.

References

[1] D. Böttger, B. Stampfer, D. Gauder et al., Concept for soft sensor structure for turning processes of AISI4140, *tm - Technisches Messen* 87(12), 745-756 (2020); doi: 10.1515/teme-2020-0054

[2] J. Vrana, R. Singh, NDE 4.0—A Design Thinking Perspective, *Journal of Nondestructive Evaluation* 40 (8), (2021); doi: 10.1007/s10921-020-00735-9

[3] U. Faltings, T. Bettinger, S. Barth, M. Schäfer, Impact on Inference Model Performance for ML Tasks Using Real-Life Training Data and Synthetic Training Data from GANs, *Information 2022* 13 (9), (2022); Doi: 10.3390/info13010009

[4] D. Petri, P. Carbone, L. Mari, Quality of Measurement Information in Decision-Making, *IEEE Transactions on Instrumentation and Measurement* 70, 1-16 (2021); doi: 10.1109/TIM.2020.3047954

[5] B. Valeske, R. Tschuncky, F. Leinenbach, et al., Cognitive sensor systems for NDE 4.0: Technology, AI embedding, validation and qualification, *tm - Technisches Messen* 89 (4), 253-277 (2022); doi: 10.1515/teme-2021-0131

[6] D, Edgar. Geeignete Messprozesse – Valide Informationen, *tm - Technisches Messen* 86 (9), 528-539 (2019); doi: 10.1515/teme-2019-0104