

# Data Science in der Werkstoffprüfung

## Data science in material testing

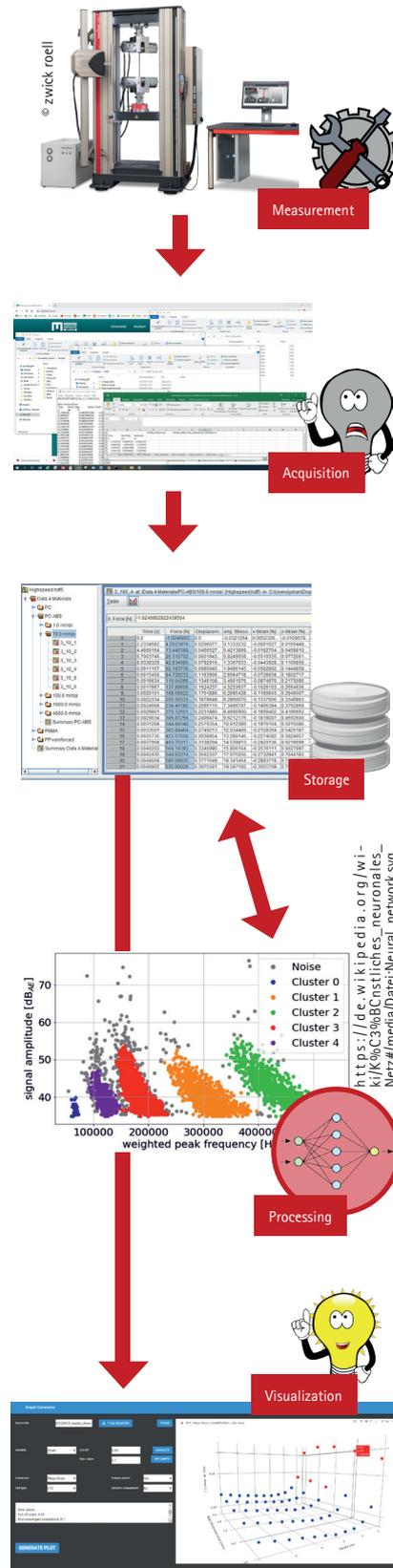
In der modernen Prüftechnik wird immer mehr Sensorik verwendet, um genügend Rohdaten über komplexe Abläufe zu sammeln. Trotz des Mehrwertes an Information sind die entstehenden Datenmengen oft unübersichtlich und schwer handzuhaben. Hier müssen Methoden der Data Science ansetzen, um den bestmöglichen Nutzen aus den generierten Daten zu ziehen. Das Ziel ist ein reibungsloser Ablauf von der Messung bis zur Darstellung und Diskussion der Ergebnisse.

Dabei ist der erste Schritt zum Erfolg oftmals die strukturierte Ablage von Informationen, die aus verschiedenen Quellen stammen (z. B. Maschinen-daten, optische Messtechnik, Simulationsergebnisse etc.). Das kann beispielsweise durch Datenbanken oder spezielle, hierarchische Datenformate geschehen, in denen verschiedenste tabellarische Daten mit Metadaten kombiniert werden können. Im selben Schritt können auch erste Auswertungen automatisiert ablaufen. Der Vorteil ist hierbei, dass diese Routinen schnell, fehlerfrei und ohne subjektive Einflüsse geschehen.

Auf Basis der archivierten Messergebnisse sind weiterführende Auswertungen möglich, die die Gesamtheit aller Daten benötigen. Beispiele dafür sind Clustering, multivariate Datenanalyse oder neuronale Netze. Durch diese Methoden werden Trends erkannt, die dem menschlichen Auge oft verborgen bleiben. Anwendungsgebiete sind unter anderem das Kategorisieren von Messpunkten sowie das Erkennen von Stör-signalen oder Ausreißern.

Schlussendlich können die Ergebnisse großer Messserien oder umfangreicher Simulationen durch interaktive Viewer besser visualisiert werden. Dadurch ist es möglich auch mehrdimensionale Daten schnell und gezielt zu analysieren.

In modern material testing, an increasing number of sensor systems is used in order to gather enough raw data on



complex processes. Despite the benefits of additional information, the accumulated amounts of data are often confusing and difficult to handle. Here, methods from data science are required to make the best use of the generated data. The goal is a smooth work flow from the first measurement to the visualization and discussion of the results.

Often times, the first step towards success is the structured storage of information that originated from different sources (e. g. machine data, optical sensors, simulation results etc.). One approach is to use data bases or certain hierarchical data formats to combine various spreadsheets with meta data. In the same step, basic evaluations can also be performed automatically. These routines have the advantage of working fast and are free of mistakes and subjective influences.

Based on the archived test results, it is possible to do more advanced computations, which rely on large, collective datasets. Examples are clustering, multivariate data analysis or artificial neural networks. These methods detect trends that cannot be recognized by the human eye. This enables the categorization of data points or the detection of noise and outliers.

Finally, the results of large measurement series or complicated simulations can be visualized better by using interactive viewers. As a result, even multidimensional data can be analyzed quickly and systematically. ■

### Ansprechpartner



**Dipl.-Ing. Dr. mont. Johannes Wiener**  
 johannes.wiener@unileoben.ac.at  
 +43 3842 402-2134