Materials data management in research & development: Improving innovation and product liability

T. Ehrig*, L. Klepsch**

*WIAM GmbH, 01109 Dresden, Germany, W.-Reichard-Ring 4, toni.ehrig@wiam.de
**IMA Materialforschung und Anwendungstechnik GmbH, 01109 Dresden, Germany, linda.klepsch@ima-dresden.de

ABSTRACT

This paper introduces WIAM® ICE as a software solution to manage all data with the structural flexibility and content complexity required for the research and product development process. The target is, to have the development process data available fully audit proof for approval processes, quality auditing or liability trials. Even further WIAM® ICE is introduced to provide an infrastructure for data structuring, classification and to drive evaluation processes up to machine learning and reasoning.

Keywords: materials data management, liablity, approval, product development, machine learning

1 INTRODUCTION

Product development is a very challenging process, especially for companies with:

- Short innovation cycles, like e.g. automotive manufacturers and suppliers
- Strong safety and design constraints, like e.g. aerospace industry
- Heavy regulatory boundaries, like e.g. food, medical or chemical industry

Companies are forced by law to always keep track on each piece of information, knowledge and decision-making during product development. Even further, companies must proof that all such used information pieces are corresponding to the state of art. Also potential safety implications need to be considered - of cause documented - during development as far as possible. Part of this challenge are the following requirements:

- Safely store documents, protocols and decision papers related to product or project
- Keep history of available data related to the product or project, which can be:
    - Standards and delivery conditions providing target values used for product design
    - Regulations influencing processes for design, testing, validation and production
    - Data sheets for materials, supplied parts and machinery used for testing and production
    - Study results for finding the optimal combination of material, geometry and manufacturability
    - Analyses to minimize risks of market, availability, error rate, change of legal guidelines
    - Test result data providing real values for material specimen, prototypes, parts, first articles, production samples and defects
- Somehow collect the technical communication related to the development process

Especially the last bullet is typically hard to catch, but mostly provides important insight on decisive reasons. Also from the remaining bullets only a minority can be covered by structured software solutions to store and retrieve information. There are software solutions existing for financial processes (Enterprise resource planning ERP), for product design drawings and build-up (Product data management PDM). Some of them or further software systems care for material standards, but the whole rest of data typically ends up being huge collection of PDFs or office documents. That is not a small portion of the product development information, it is the overwhelming majority of data.

More than this, just storing this information does not automatically mean having it available if needed. Availability requires the data to be accessible in the context someone is looking for it and approaching his research.

2 SOLUTION APPROACHES

A typical first approach is to put information into a file-server folder-structure. The chance to find information in such a location is limited to one dimension. Whatever is chosen for the top-folder gives you the chance to find data fast. Typically this is one of:

- Date
- Type of data
- Product-Designation
- Customer
- Project-Identification
- Standard or Regulation reference

Most data is located somewhere deep in the folder structure. The top-folder order is easy to see through. The deeper you dive into the structure the harder it is to find items of the same kind to collect, to compare or to build a proof or decision on. If you ever have been responsible for an auditing process, you know that auditors don't like to see you having no fast or no appropriate answer.

If the budget and amount of data is big enough, the valid second approach is to install a document management system. It does indexing of documents and allows you to find documents by contained words or - if available - by a given keyword cloud. The chance to find data fast by a combination of search terms grows. The limit of this approach is reached, when searched terms appear in many different contexts within the index. This may be e.g. dates, numbers for products and parts or names of companies doing a lot of varying business. Not to forget - you never can be sure to find all relevant data, after at least some of the documents may lack a part of your search terms. So this is also no safe approach to be audit proof in case of approval processes, quality auditing or liability trials. So what is needed to have fast and reliable answers in such situations. Which challenges are needed to be overcome and how time-consuming is it to do so?

3 OVERCOME CHALLENGES

This section lists the challenges of the research and development data management and introduces WIAM® ICE as engineering data management software to solve each of these challenges.

Challenge 1: Documents are just a bag of words and values. Document management does not know anything about relations and categories.

- Structured information storage needs tree-structured items in multiple trees. Each item must be connectable to any number of tables holding parameters, categories, documents and links to other items in the same or different tree. WIAM® ICE data management is build upon this principle. For the theory behind refer to [1].
- Whichever pedigree information is stored - it must be possible to find this information in the managed data.

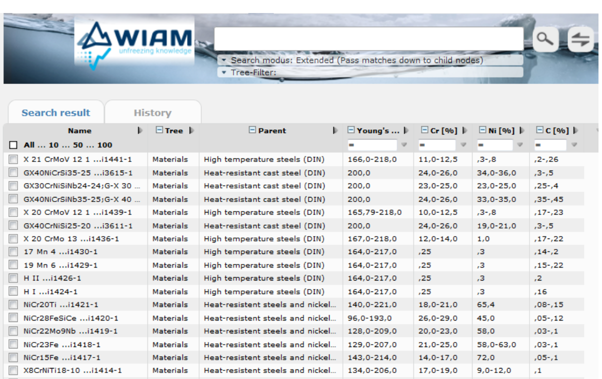


Figure 1: Example for search on parameter combination

- Merely this structuring of information in trees, with interconnected items carrying the tabular master data and having all related documents attached, in many cases warrants a good part of the required and missing transparency already.
- **Efforts:** None at all, if you have put your information into WIAM® ICE

Challenge 2: Especially Excel [2] documents typically contain not just information about one related test/part/ product - they often contain lists of tests/parts/products with information related to many of them in different rows.

- WIAM® ICE provides the “Simple Excel Import”, where individual lines can become individual structured items and further values of these lines can be put to any of the item's tables.
- This is also a good way to put master data of products, parts, materials, tests from ERP/PDM systems to WIAM® ICE, where then all the objects can be interconnected and documents can be attached to.
- **Effort:** Adding 1-3 columns to a typical Excel-tab to designate a material/part/product leading the data following in each row - in many cases it is there already.

Challenge 3: Not only data points inside a document relate to each other. Also data from different documents is related to the same test/part/product potentially.

- The “Standard Excel Import” finds relations between multiple items holding manifold parameters in different Excel-tabs. The relations are detected automatically inside the same or even between different documents.
- This generates complex structure dependencies like:

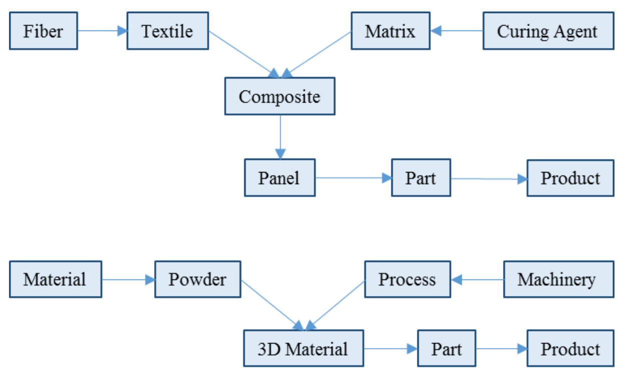


Figure 2: Complex structure examples in development

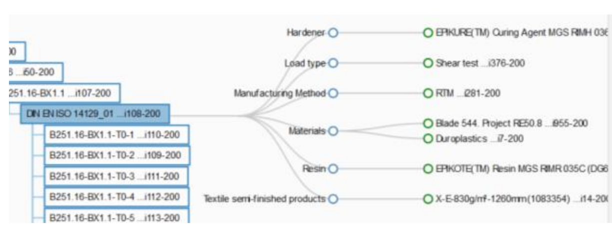


Figure 3: Complex structures in WIAM® ICE

TechConnect Briefs 2018, TechConnect.org, ISBN 978-0-9988782-5-6

- Even for multiple test results a single template can take information about the result data, machine, specimen, material, components, etc. This way a single import can read all of this in one step and a single export can generate a dataset covering all the results, information and details you want to pass to a customer or autitor.
- **Effort:** Make standard Excel templates for your typical process items - in most cases the one-time time-exposure is much lower than the time-saving in the middle-term. Not to forget about the quality improvement and the reduction of risk to loose data.

Challenge 4: Data from development processes needs much more flexibility in terms of structure changes and relations than production and quality checking data.

- Each Excel import or data management interaction through the WIAM® ICE user interface can extend the data structure and provide further options for parameters, relations and dependencies.
- Independent from the extension in terms of structure or data, the integrity and content of previous revisions is safely preserved.
- **Effort:** None at all, this comes out of the box - to be fair, to operate a software system is never free of efforts.

Challenge 5: Technical communication is typically widespread over e-mails, social networks, messengers, talks on phone/corridor/lobby. The project related information should be collected and stored.

- This is naturally a hard one, but the main challenge of all of the named communication pathes is: It is running between two or a group of persons.
- It makes sense to add a communicatin path that enables communication with - or better related to - an object within the data management system, may it be a material, a product or a project.
- This does not guarantee, but enables - if enforced - a nearly complete gathering of implications thought through or of reasons a decision is based on. These are always related to the project or product.
- **Effort:** A mix of force and kind requests, plus a lot of patience to make your people use it.

After this point it is possible - without major changes to processes and work-flows - to have a full and well accessible overview on all data and information related to the development process. This is sufficient to proof compliance to and supervision of nearly all existing legal or supplier obligations. This alone is worth considering an engineering data management solution as WIAM® ICE.

4 EVALUATION, MACHINE LEARNING

There is more. If you successfully started your research and development data management, you will see lots of additional effects of having data at hand where ever needed. Having all this data in one place provides the option to overcome further challenges of data evaluation and processing:

Challenge 6: Small internal evaluation tasks are easy to be made within Excel. Just as easy such evaluations are lost on file modification or by accident.

- The internal formula engine of WIAM® ICE with java scripting allows to do evaluation live or by safely stored and - if required - revisioned formula code.

Challenge 7: Data from typically different sources is needed for evaluation processes within development, e.g. management reporting, technical reporting, test evaluation, calculation interfaces.

WIAM® ICE provides an HTTP/XML API to trigger search processes with full functionality as through the user interface and to access or even modify data using Excel or XML formats. This enables:

- Simple derived web-services to access information in a specialized or simplified way
- “Jupyter notebook” [3] reports through Python [4] scripting
- Excel templates or forms through VBA scripting
- Interfaces into programming languages like C/C++/C#/ Python/Java/... to generate interfaces to/from calculation applications (even specific and tailored solutions), test machinery, etc.
- Interfaces to programming stacks, e.g. for machine learning

Challenge 8: Though mentioned in the last bullet, this one is a separate challenge itself. Machine learning provides a huge potential of industrial applications but currently lacks an infrastructure for integration. Automated approval, failure detection or parameter estimation is possible, but requires well-structured and -categorized learning sets.

- Looking at the answer to the first challenges - especially **Challenge 3** - WIAM® ICE provides structure trees with structured items = hierarchic categories | item connected tables to relate to other items = relation categories | tables to hold textual, numeric values and even documents = valued categories, dependencies for learning and target parameter sets
- Through the API learning sets can be extracted in a standardized way to be passed to a learning stack, like e.g. TensorFlow [5], to generate a suitable model
- By calling the WIAM® ICE formula interface, the model evaluation can be run on given datasets of test results, materials, production specimens, ...

The challenges and chances of machine learning as described in **Challenge 8** are a key focus of industry and governmental interest. An integration example of machine learning and WIAM® ICE will be provided in this paper.

The example shows material’s young’s modulus estimation based on material’s chemical composition through standard API procedures.

Training the learning model using WIAM® ICE data:

- API selection of all materials providing chemical composition and physical parameters like young's modulus from the WIAM® METALLINFO dataset
- API extraction of chemical composition and physical parameters - this returns thousands of material entries within an Excel container file
- Cleaning and "hole-filling" of the dataset in a Python routine
- Generation of a learning set with 80% of the data for training and the remaining 20% for test and optimization
- Set up a fully connected, deep convolutional regression network in Python/TensorFlow to learn the estimation from chemical composition to young's modulus
- Run the training and the optimization process
- Save the model and in addition the linear regression data for confidence estimation

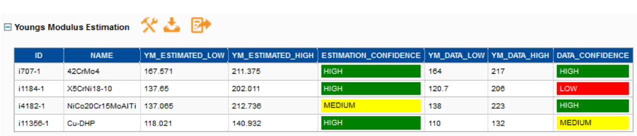
Figure 4: Set up and integration schema [4][5]

Integration of the learning model into WIAM® ICE:

- Set up a Python script to load the saved model into TensorFlow and to run the estimation from input chemical composition to estimated young's modulus
- Configure the Python script as external extension to the formula engine integrated into WIAM® ICE
- Set up a formula to extract chemical composition from the currently selected material or test
- Store the formula into the WIAM® ICE formula storage

Running the estimation in WIAM® ICE:

- Find the material/test you want to base the estimation on
- Run the formula evaluation
- Find an estimation of the young's modulus and use it for one of the following:
    - Having an estimate of the range to expect with a given confidence. So you can get information where no data would be available otherwise.
    - Having a validation of the measured value against the estimation. So you can verify data and process quality.



Youngs Modulus Estimation

ID	NAME	YM_ESTIMATED_LOW	YM_ESTIMATED_HIGH	ESTIMATION_CONFIDENCE	YM_DATA_LOW	YM_DATA_HIGH	DATA_CONFIDENCE
i707-1	42CrMo4	167.571	211.375	HIGH	164	217	HIGH
i1184-1	X5CrNi18-10	137.65	202.011	HIGH	120.7	206	LOW
i4182-1	NiCo20Cr15MoAlTi	137.065	212.736	MEDIUM	138	223	HIGH
i11356-1	Cu-DHP	118.021	140.932	HIGH	110	132	MEDIUM

Figure 5: Estimated and validated by machine learning

This is just an example. In the same way the set up and running of learning/estimation processes works with WIAM® ICE for all technical aspects in the product development, e.g.:

- Estimate technical parameters for materials with measured chemistry and known manufacturing process, e.g. strength, elasticity, lifetime, …
- Detect material type, material components, manufacturing processes from specimen testing
- Check material, part, product test result validity
- Detect test failures, e.g. slip of clamping, non-aligned specimen, geometrical inconsistencies
- Detect hidden production failures by quality testing results
- Estimate production quality by quality control and process monitoring values
- Estimate parameters of additive manufactured parts by powder composition and manufacturing parameters
- Validate quality of additive manufactured parts by manufacturing parameters monitored
- Estimate property influence for additive manufactured parts by manufacturing parameters detected drifting or out of range
- … any many, many more …

5 CONCLUSION

Current research and development processes often lack the needed transparency and availability of data. Where financial processes are typically backed by a fine software infrastructure, technical data from the development process often is hard to find, to prepare and to work on. WIAM® ICE provides the needed functions and work-flows to bring transparency into the product development process up to a level of being audit proof in case of approval processes, quality auditing or liability trials. Even further WIAM® ICE can provide the infrastructure for modern machine learning integrations between the well-controlled data, the innovative learning tool-kits and the challenging product development processes in each company.

REFERENCES

[1] Toni Ehrig, "Entwurf eines allgemeinen Konzepts zur uniformen Repräsentation und Verarbeitung von Daten in Wissensbanken", Shaker Verlag, ISBN 978-3-8322-6696-7

[2] “Excel” is a product and trademark of the Microsoft corporation, www.microsoft.com

[3] “Jupyter” notebook is an opensource project, www.jupyter.org

[4] “Python” is a programming language and logo of the Python Software Foundation, www.python.org

[5] “TensorFlow”, the TensorFlow logo and any related marks are trademarks of Google Inc., www.tensorflow.org