Using Automated Data Quality Check Tools

Josef Kriegmair, MTU Aero Engines, Munich

1 Introduction

The presentation is based on my experience with CAD/CAM in manufacturing. The examples are also from manufacturing.

2 What is data quality?

Data quality is from my point of view perceived customer satisfaction, which means that requirements of the business and the processes are met. The customers are the users of the data.

Data quality rest on two pillars:

- Content
- Data structure

Important quality criteria for the content are accuracy, completeness, consistency, and trustworthiness.

Important quality criteria for the data structure are suitability for the process steps, integrity, reliability, and performance.

3 What is meant by using automated data quality check tools?

The term ‘using’ means combining tools and processes to create solutions. Automation means several checks are performed after a start triggered by a user, a workflow, or a routine that processes the data.

Using automated quality check tools is combining tools and processes to create solutions.

4 Combining tools and processes to create solutions

4.1 Benefits

By combining tools and processes to create solutions, processes are organized in such a way that the full potential of the technology / tool is realized. The benefit is in this case much higher than in the case of installing a tool and pressing the start button. If the tool is misaligned with the quality requirements of the process step, the benefit will be even lower than in the case of installing a tool and pressing the start button, since the tool refuses data with sufficient quality.

4.2 Need for data quality check tools

CAD models for instance have tolerances due to rounding and truncation caused by a limited amount of digits in computer systems for the representation of geometry. Appropriate data quality along the value chain is required if well running processes are to be achieved. Data quality tools help in guaranteeing data / geometric integrity as well as conformance to standards. Both items lead to a quality virtual product / CAD model / data set.

4.3 Approach for combining tools and processes to create solutions

One method of combing tools and processes consists of the following steps:

1. Plan – design of a solution in order to improve data quality
2. Implement – designed solution is tested with use cases and rolled out
3. Use – customers / users apply the solution to real life tasks
4. Analyze – analysis of results delivered by the solution in order to improve it
5. Act – decision process for improving the existing solution

A data creation process should be accompanied by a quality process for assessing the quality of the created data.

The focus of this paper is on the phase ‘Plan’ since I firmly believe that quality is designed in and not checked in.

4.3.1 Planning phase

The central task is to establish a robust and reliable solution which delivers a set of quality items, minimizes the checking effort through good methods, provides suitable usability, and has good performance. Quality is designed in, not checked in. The central task consists of five major blocks:

- Analyze needs along the value chain
- Analyze existing solutions for solving potential
- Specify solution
- Define use cases
- Manage user acceptance

There are two main processes within the company:

- Flow of the virtual product with the major steps: research and design, process engineering, tool making, factory layout planning
- Flow of the physical product starting with the raw material and ending with the final product

In both flows the tasks are to process the part in order to achieve the specified conditions, as well as to check whether these conditions have been met. Both flows may encompass several companies.

An elaborate quality assurance system with standard units for characterizing and documenting the quality exists for physical products. For instance dimensions and roughness are measured, and cracks, nicks, and dents are identified. Process quality is supervised by key items like Cp index, reliability and repeatability. A full set of quality items is available characterizing the quality of a physical part.

Key items for virtual product quality such as geometric integrity and model structure integrity are necessary. Concepts can often be transferred from physical to virtual products. For example, define a set of named quality items for each operation process, guaranteeing a quality virtual product as long as these key items are within the specification.

Virtual products are often stored in and managed with a PLM system. CA-systems are used to generate virtual products. There are two different types of data:

- Metadata – document number, project, document type, etc.
- Primary data – e. g. CAD data

At MTU Aero Engines, a lot of metadata have to be selected from lists whenever a new virtual product is created within our PLM system. The document number is checked against allowed character sets – e. g. some characters are refused, lower case letters are converted to capital letters.

The data structure of CAD files and some content of the primary data can be checked with some special check routines.

A check routine keeps metadata stored in the PLM consistent with transferred metadata to the CAD file. Thereby, the correct metadata is displayed in the frames. A modification flag displays the status.

User acceptance is contingent on four items:

1. Solve a problem – users embrace a solution since checking data against standards and guaranteeing data integrity are tedious tasks
2. Involve key users to promote the solution – these users promote the solution, because they regard the solution as their own solution
3. Provide good performance – if a solution is more an obstacle than a help, it will be bypassed by users
4. Deliver appropriate usability – users find the solution logical and intuitive

4.3.2 Implementation phase

Implementation is divided into three stages:

- Realization
- Testing
- Roll out

The specification (output of the phase 'Plan') is implemented in a development environment during the stage realization.

Use cases, which have also been defined during the phase 'Plan', are the basis for testing the implemented solution. This work is often done by key users and second level support staff since they know the problems and requirements very well.

Before and after roll out of the solution from the test environment to the production environment, user training (if necessary) occurs. The project plan contains a contingency plan for remigration to the former production environment in case the tool causes problems in the production environment despite the intensive testing.

4.3.3 Use phase

Solutions are started manually, by workflows for instance during the release process of virtual products, and by modeling routines. Automated data quality check tools are part of a modeling routine, or are user functions within the data creation system, or are separate programs applied to data.

If the output of the tools is correct or a warning, the data will be accepted. If the output is an error, the virtual product / data set will be modified in order to produce appropriate data.

4.3.4 Analysis phase

The driving force for the phase 'Analyze' is transparency before improvement. Therefore, if a problem has been localized, a root cause analysis will follow. This provides the potential for improvement – methodologically and monetarily.

4.3.5 Action phase

The potential benefit – methodological and monetarily – provides the basis for the change request. Based on this document, there will be a decision whether the potential will be realized. If the decision is go ahead, a new phase 'Plan' will start.

5 Summary

The intention of this paper is to provide a brief overview of the field of running automated data quality check tools and to explain the necessity of such tools based on a few examples from jet engine manufacturing as well as on my experience with CAD/CAM/PLM in this area.

Physical components assembled – e. g. into a jet engine – have undergone stringent quality checks and approvals. Virtual components that build the DMU and provide the base for manufacturing and testing of the physical components require appropriate quality checks and approvals. If the data quality of virtual products is insufficient, downstream process will run into very serious problems.

Running automated data checking tools is much more than installing a tool and pressing the start button. Checking data quality has to address metadata and primary data, has to analyze the correctness of the data structure, has to run tools, has to analyze processes in order to find out the appropriate data quality, has to ensure user acceptance so that the solution is used in a proper manner, and needs a framework for transforming an idea into a robust solution.

A full set of quality data similar to the one for a physical product will be required if fully digital value chains are to run well.

The key message is: quality is designed in, not checked in!