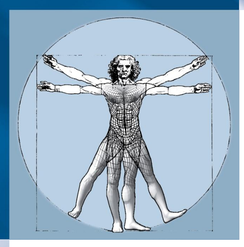

Measurement and Analysis of Data Quality for Simulations

Michael Reiter[1], Jun.- Prof. Dimka Karastoyanova[1], Prof. Frank Leymann[1], Dr. Hong-Linh Truong[2]

[1] IAAS, University of Stuttgart, Germany; [2] Infosys, Vienna University of Technology, Austria

Motivation

Application Independent Data Quality Determination

- Data quality determination within simulation workflows [1]:
  - multi-domain, multi-scale, multi-physic, multi-tool
- Support for all six main data quality dimensions [2]:
  - Accuracy, completeness, currency, timeliness, volatility, and consistency
- Separation of analysis and interpretation of data quality [3]
- Application independent data quality language [3]
- Application independent data quality framework [3]

Data Dependencies within Complex Simulation

- Example of FEM based Simulation

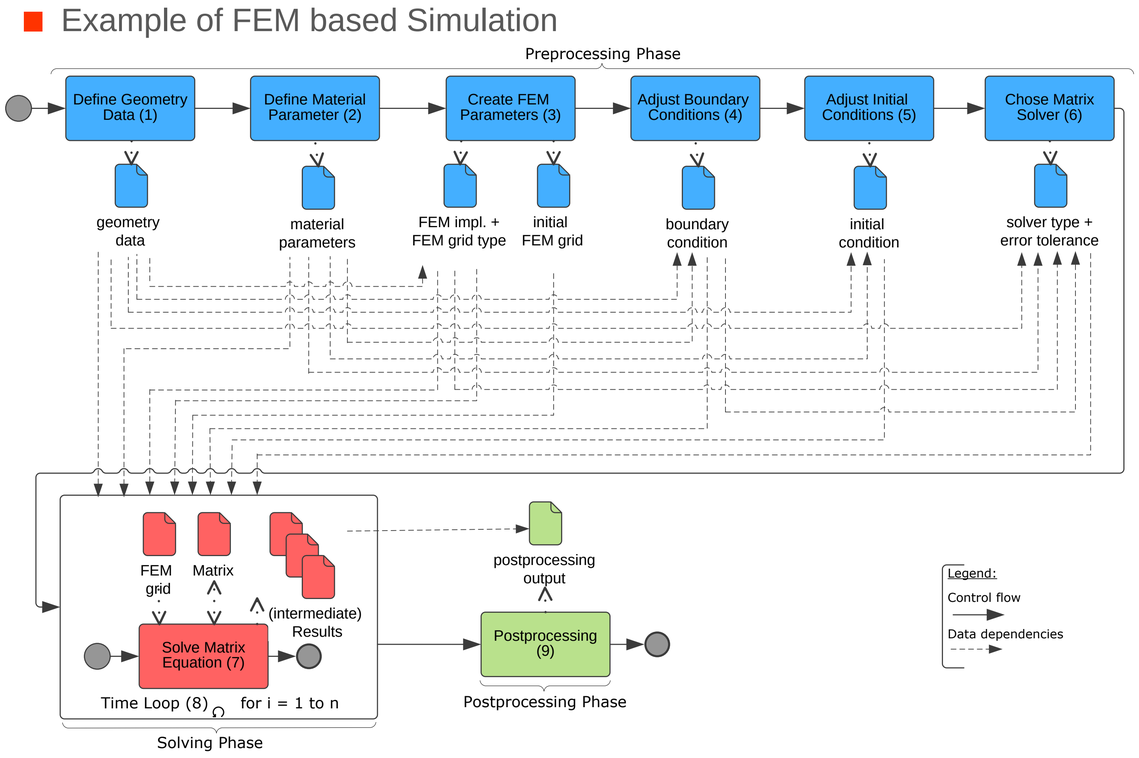

Figure 1: Typical activities of a finite element method (FEM) based simulation and their associated data as well as the dependencies between the data [4]

- Simulations consist of different phases:
  - preprocessing, solving, postprocessing
- The phases consist of different activities:
  - activities depend on the type of simulation
- Activities can have associated data:
  - all data have a quality, e.g. accuracy, completeness, currency, timeliness, volatility, and consistency
- Data flow between activities defines the data dependencies between activities
- Dependencies between the quality of used data must be derived

Process to Determine the Data Quality

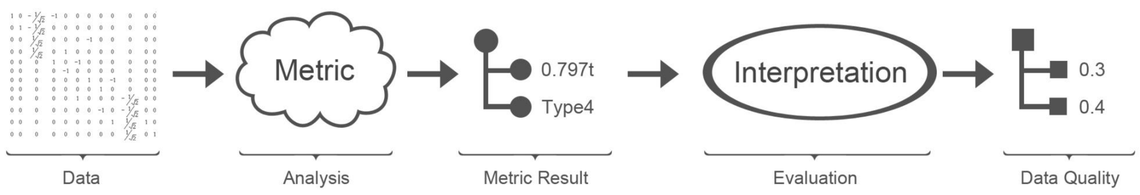

Figure 2: First, a metric determine (context free) characteristics of data. Secondly, the metric result can be interpreted explicitly in the context of the particular simulation [3]

- Two step process to determine the data quality
  - Step 1: determining characteristics of data, e.g. a metric calculates the Jacobian determinant of a matrix
  - Step 2: interpretation of the metric result to the context of the simulation, e.g. the significance of a Jacobian determinant to a specific simulation or solver
- Java Data Quality Framework supports both steps

WS-DataQuality

- XML based Language to describe Data Quality Requirements and Assurances

```
<DataQualityAssertion>
    <Statement Type="Requirement">
        <DataQualityId>
            org.example.metrics.metric1
        </DataQualityId>
        <DataQuality>
            <MinimumValue xsi:type="xs:double">
                0.5
            </MinimumValue>
        </DataQuality>
    </Statement>
</DataQualityAssertion>
```

```
<DataQualityAssertion>
    <Statement Type="Assurance">
        <DataQualityId>
            org.example.metrics.metric1
        </DataQualityId>
        <DataQuality>
            <Value xsi:type="xs:double">
                0.7
            </Value>
        </DataQuality>
    </Statement>
</DataQualityAssertion>
```

Figure 3: A compatible data quality requirement and data quality assurance assertion [3]

- WS-DataQuality based on the Web Services Policy (WS-Policy) Framework
- Java Data Quality Framework supports WS-DataQuality

Use Cases

- Simulation of structure changes within a human bone
  - FEM based simulation, using PANDAS
  - Metrics exist for preprocessing, solving, postprocessing phase
  - Metric implemented for Main Diagonal Condition of a matrix
  - Visualization of data quality

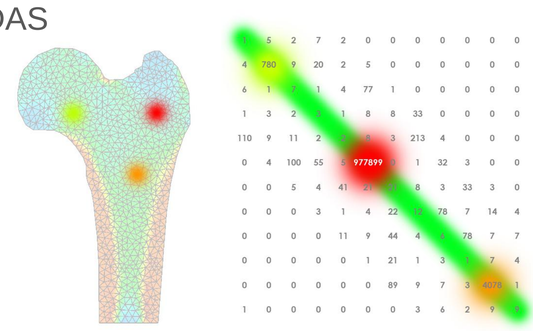

Figure 4: Visualization of data quality findings regarding the Main Diagonal Condition of a matrix at solving phase and the corresponding degrees of freedom (DoF) within the FEM Grid

- Simulation of driving dynamics
  - Comparing of simulation and experiment results using legacy tools
  - Metric exist for postprocessing phase
  - calculates the maximum and the average deviation between simulation and experiment

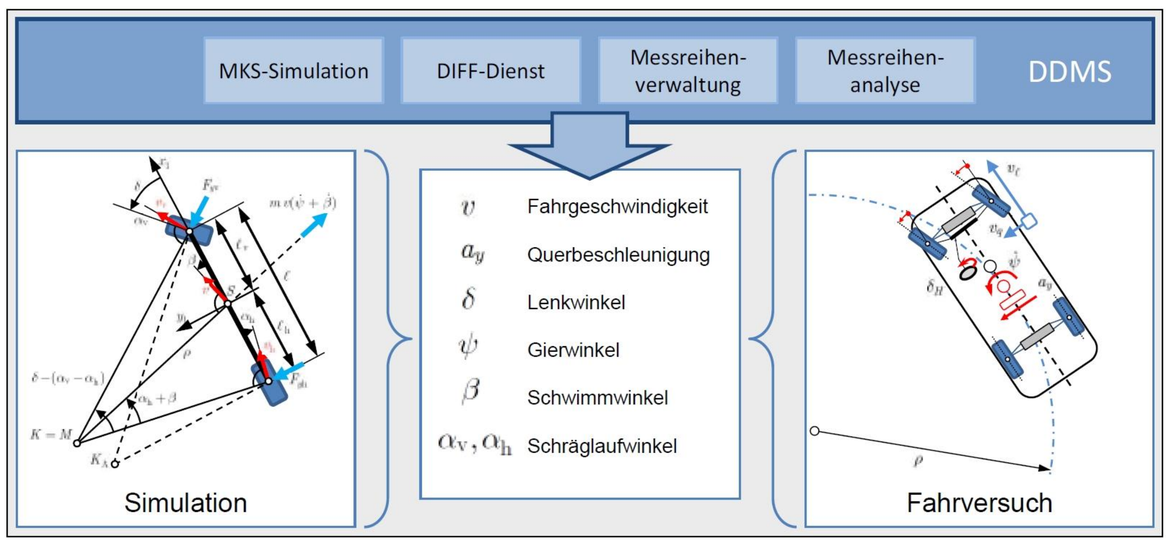

Figure 5: Degrees of freedom (DoF) within simulation and experiment to explore the driving dynamic of a car. We determine the quality of the DoF-data calculated by a simulation [3]

Literature

[1] K. Görlach, M. Sonntag, D. Karastoyanova, F. Leymann, M. Reiter: Conventional Workflow Technology for Scientific Simulation. 2011

[2] C. Batini, M. Scannapieco: Data Quality – Concepts, Methodologies and Techniques. 2006

[3] U. Breitenbücher: Datenqualität in Simulation-Workflows, Diploma Thesis, 2011

[4] P. Reimann, M. Reiter, H. Schwarz, D. Karastoyanova, F. Leymann: SIMPL – A Framework for Accessing External Data in Simulation Workflows, BTW, 2011