



## AiF proposal outline

### "Knowledge based pre-processing of process data"

Due to the progress in information technology that has been made in the last years the amount of measurement data which are available from running production processes has increased remarkably. The analysis of these measurement data offers optimization potential with respect to cost and quality and the possibility to detect faults in a production process in case of problems. In that field especially small and medium industry (SMI) companies are operating which offer special software tools, project management and consulting for process optimization. Despite the high availability of process measurement data the economic potential of process data analyses is only rarely exploited. The reason for that phenomenon lies in the enormous initial effort, often up to 80% of the whole data analysis, for data pre-processing.

The submitted research project aims at a dramatic reduction of this initial effort by developing a knowledge-based methodology for the efficient pre-processing of process data:

- Consistency of different steps for pre-processing when using various tools (subsequently) will be assured in order to prevent errors due to data conversion that is needed between these tools
- Pre-processing steps will be made comprehensible so that the use of the pre-processed data can be assessed even at a later date
- Reuse of algorithms and procedures for pre-processing will be supported
- Possible errors in the data (missing values etc.) or in pre-processing procedures will be prevented or at least be detected at an earlier stage

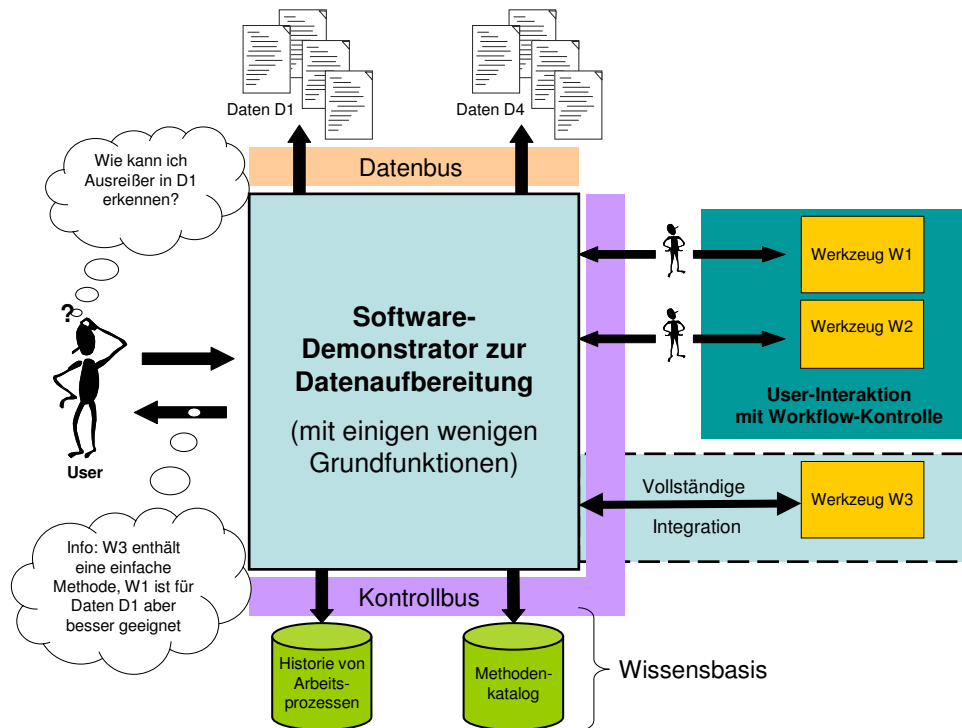
In order to achieve these goals the project focuses on three key result components:

- a) A catalogue of methods that contains knowledge about best practice procedures in data pre-processing
- b) An architecture that makes use of this knowledge and supports the consistent interactive use of different external tools (that are not required to be fully integrated, see figure below) for pre-processing
- c) A software demonstrator that implements this architecture

Ad a) The catalogue of methods will be a kind of manual that contains a list of known steps in the field of data pre-processing, e.g. "variance estimation" and "denoising". The manual will be augmented by important properties, such as input and output parameters, e.g. "variance" as an input for "denoising", and tools that support the user in performing the corresponding pre-processing step. Established procedures will be given in the manual, too, in order to document best practice procedures and make them available to the users. We have suggested to publish the catalogue of methods for pre-processing e.g. via the internet as a Wiki (public domain) in order to reach a broad range of users who would be able to extend this knowledge base in that way. Note that the catalogue is completely independent from the proposed software in this research project.

Ad b) The architecture (see figure below) will support the comprehensive data pre-processing using different (external) tools interactively based on the knowledge in the catalogue of methods. Key components are a data bus for the consistent and transparent access of data independent from their storage location and format, and a control component that documents work steps performed on the data, to make these steps comprehensive and re-usable and that suggests possible next steps using the knowledge base (catalogue + historical work processes).

Ad c) Feasibility will be demonstrated by the proposed software demonstrator for data pre-processing that will implement the architecture. Practical applicability of the software demonstrator will be tested during the project and with increasing complexity together with the corresponding partners from industry. The software could be made available to the public under a free license. The concrete license issues would have to be discussed in the board that accompanies the project and with consultation of the AixCAPE members.



Economic benefit would be obtained especially in process systems engineering where the analyzed measurement quantities influence product quality or production costs directly. All industry in the field of process data analysis but especially SMI which could focus on their key competences and do not have specialized departments with the corresponding human resources would gain benefit from the results of the submitted research project.

The requested staff assignment for the project is 2 person-years (one scientific employee) and additionally a student worker (~20 hours per week).