

PIA

A concept for a **P**ersonal **I**nformation **A**ssistant for
Data Analysis

Christopher Schnur, Tanja Dorst, Kapil Deshmukh, Sarah
Zimmer, Philipp Litzenburger, Lennard Margis, Tizian Schneider,
Rainer Müller and Andreas Schütze

10/26/2022

About



Originated in the EFRE projects :

- MessMo – Messtechnisch gestützte Montage
- iTecPro – Erforschung und Entwicklung von innovativen Prozessen und Technologien für die Produktion der Zukunft

Gefördert mit den Mitteln des Saarlandes und der Europäischen Union



Further developed within the project:

- NFDI4Ing – National Research Data Infrastructure for Engineering Sciences; Task Force ALEX



Involved Departments:

- Lab for Measurement Technology
Prof. Dr. rer. nat. Andreas Schütze
- Chair of Assembly Systems
Prof. Dr.-Ing. Rainer Müller



Motivation

- (Process) Knowledge in industry is widely distributed within the company
- Only specific people have expert knowledge (accessibility problems)
- Metadata are either not documented or not accessible
- Lack of resources and knowledge in the domain of machine learning

Objective:

PIA should allow authorized persons to efficiently perform a machine learning project, record high-quality data and carry out a first analysis

Research data compared to industrial data

Research data

- Rigorous design of experiments (DoE)
- Data as objective
- High quality of data
- Metadata well described



Industrial data

- Usually control and regulation data
- Data as by-product
- Poor quality of data
- Metadata incomplete

The industries' focus is on their product or service, whereas the objective of most experiments in research is to obtain high quality data

Main challenges in industrial data

1. Manufacturing domain specific

- Heterogeneous product families and operating equipment
- Concept drifts
- Imbalanced data distributions

3. Data specific challenges

- Complex data fusion from multimodal data sources
- Missing primary and foreign key relations
- Unsynchronized timestamps along manufacturing process
- Duplicate and invalid data samples
- Missing specifications and explanations of data
- Insufficient data acquisition

2. Software specific challenges

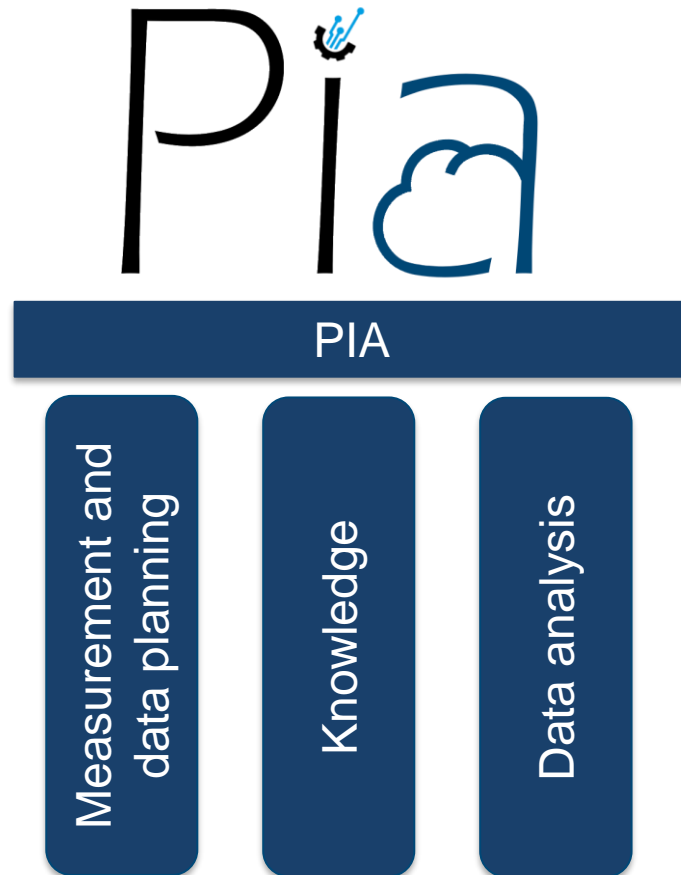
- Missing modern data access interfaces
- Proprietary software

4. Analytical challenges

- Backtracking of analysis results to root causes

Source: *Wilhelm, Y. et al., 2020, doi: 10.1007/978-3-030-64846-6_4*

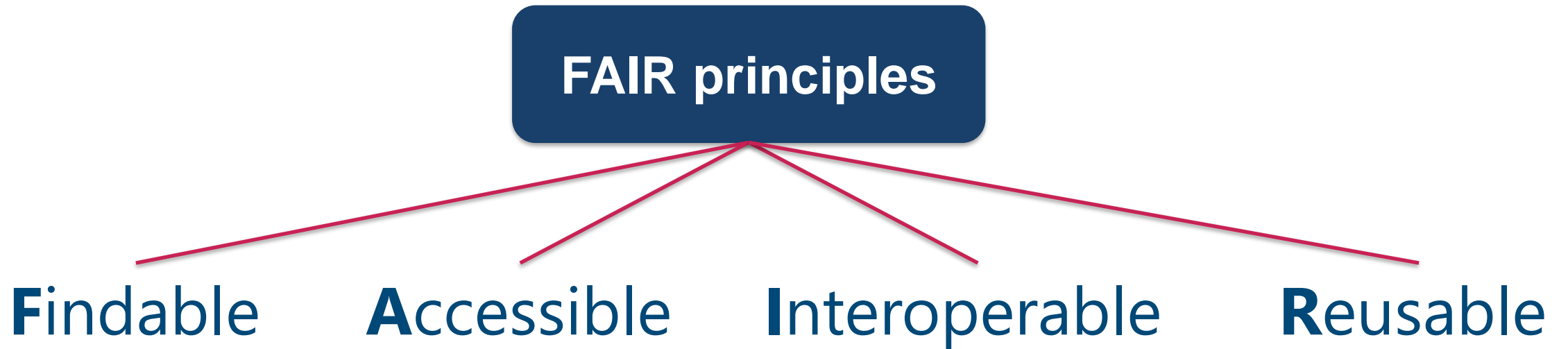
Schematic representation of PIA



Three pillars:

- Measurement and data planning checklist
- Accessibility of knowledge and data
- Low to no code data analysis

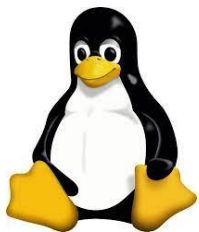
In accordance with the
FAIR principles



Source: *Wilkinson, M. et al., 2016 doi: doi.org/10.1038/sdata.2016.18*

Implementation of the PIA demonstrator

- VirtualBox as virtual machine
- Linux as operating system
- Angular as web application framework



Linux

+



VirtualBox

+



Angular



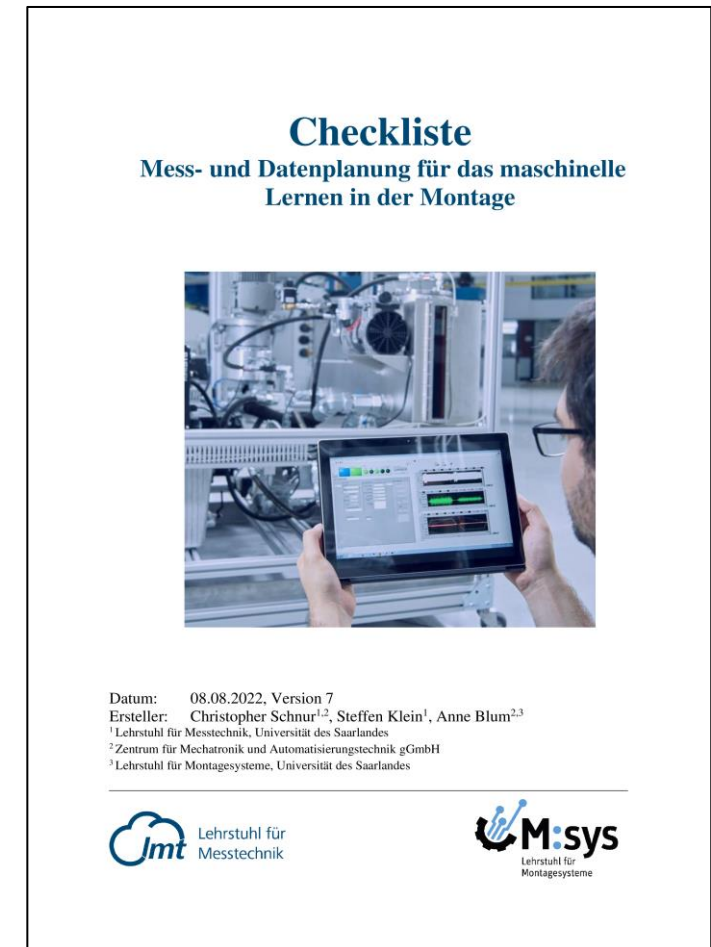
GUI of PIA in the browser

Pillar 1: Measurement and data planning

Measurement and data planning

Checklist for measurement and data planning for machine learning in assembly lines

- Guides users in a holistic approach through a machine learning project
- Covers the most important steps to achieve high quality data
- Also applicable in other domains



Cover of the checklist



Source: Schnur, C. et al., 2022, doi:10.5281/ZENODO.6943476


Measurement and data planning


Topics covered:

- Preparations and project planning
- Measurement and data planning
- Data acquisition
- Data validation and data cleansing
- Data evaluation and ML model building
- Project completion

Gefördert mit den Mitteln des Saarlandes
und der Europäischen Union

 SAARLAND 

 ZeMA
zentrum für Mechanik
und Automatisierungstechnik

 UNIVERSITÄT
DES
SAARLANDES

Inhalt

1. Vorwort	3
1.1. Ziele und Grenzen der Checkliste sowie Fokus der Betrachtung	3
1.2. Aufbau der Checkliste	3
2. Vorbereitung und Projektplanung	4
3. Mess- und Datenplanung	6
3.1. Prozesswissen aufbauen und nutzen	6
3.2. Normen und Standards nutzen	7
3.3. Aufbau der Daten	8
3.4. Datenablage	9
3.5. Manuell erfasste Daten einbinden	10
4. Datenaufnahme	11
4.1. Erste Testmessung und Überprüfung der Datenqualität	11
4.2. Langzeitdatenaufnahme	12
5. Datenprüfung und Datenbereinigung	13
6. Datenauswertung und Modellbildung	14
6.1. Datenverständnis	14
6.2. Auswahl von maschinellen Lernalgorithmen	15
6.3. Modellbildung	17
6.4. Modellanwendung	18
7. Projektabschluss	19
8. Fazit	19
9. Abkürzungsverzeichnis, Glossar und Literaturverzeichnis	20
9.1. Abkürzungsverzeichnis	20
9.2. Glossar	21
9.3. Literaturverzeichnis	22

Table of contents of the checklist

Source: Schnur, C. et al., 2022, doi:10.5281/ZENODO.6943476

Checklist as text document




Download here:



3. Mess- und Datenplanung


Nachdem die ersten Vorbereitungen abgeschlossen wurden, kann die eigentliche Mess- und Datenplanung beginnen. Zunächst ist es wichtig grundlegendes Prozesswissen zu erfassen bzw. aufzubauen und dabei bereits existierende Normen und Standards zu identifizieren. Ihre Daten werden in der Regel nicht nur innerhalb einer Abteilung genutzt, daher ist eine einheitliche und von allen Parteien verständliche Datenstruktur mit allen notwendigen Metadaten essenziell.

 **Hinweis** Die Mess- und Datenplanung ist, ein iterativer Prozess und darf (und sollte) angepasst werden

 **Tipp** Achten Sie zu **jeder Zeit** in Ihrem Projekt auf die **Wiederauffindbarkeit von Wissen/Informationen!** Nutzen Sie dazu in Ihrem Unternehmen vorhandene Strukturen und Vorgehensweisen zur Datenablage. Falls Sie eine eigene Struktur aufbauen müssen, orientieren Sie sich an den FAIR-Data Prinzipien unter <https://www.fairsharing.org/>. FAIR steht für **Findable** (Auffindbar), **Accessible** (zugänglich), **Interoperable** (interoperabel) und **Reuseable** (wiederverwendbar). Beachten Sie das „FAIRe“ Daten nicht gleichzusetzen sind mit „offenen“ Daten, da sich die Auffindbarkeit bzw. Zugänglichkeit auch auf Ihr Unternehmen bzw. einzelne Abteilungen beziehen kann.

3.1. Prozesswissen aufbauen und nutzen

Unterstützung von Experten angefordert
Die jeweiligen Experten kennen i.d.R. die Produkte, Prozesse und Anlagen am besten. Bitten Sie diese um Rat bei der Bearbeitung der folgenden Punkte

 **Hinweis** Oftmals lässt sich (spezifisches) Prozesswissen nur bei einzelnen Personen im Unternehmen finden und wird dadurch schwer zugänglich, was die Datenauswertung/-analyse enorm behindern kann.

Bereits existierende Merkmale identifiziert
Welche Merkmale werden bereits in Ihrer Anlage bzw. Ihrem Prozess aufgezeichnet und für wen (welche Abteilung) sind diese relevant? Z.B.: Welche Merkmale sind für die Montage-, Qualitäts-

 Best practice checkbox

 Obligatory checkbox

Source: Schnur, C. et al., 2022, doi:10.5281/ZENODO.6943476

Implementation of the checklist in PIA

Completed task/chapter

Info-box

The screenshot shows a checklist interface with a left sidebar and a main content area. The sidebar contains a list of tasks with checkboxes. The main content area contains two informational boxes: 'INFO' and 'TIPP'. Red arrows point from external labels to specific elements: 'Completed task/chapter' points to the 'Vorbereitung' task, 'Info-box' points to the 'INFO' box, and 'Tipps/hint boxes' points to the 'TIPP' box.

Completed task/chapter

- Vorbereitung +
- Mess und Datenplanung +
 - Prozesswissen aufbauen und nutzen +
 - Unterstützung von Experten angefordert
 - Kritische/Interessante (Montage-) Prozesse i
 - Bereits existierende Merkmale identifiziert
 - Qualitätsanforderungen recherchiert
 - Relevante Prozesse/Anlagen ausgewählt
 - Ursache-Wirkungs-Diagramm erstellt
 - Einfluss von Störgrößen bestimmt
 - Einfluss von Störgrößen minimiert
 - Anbringung zusätzlicher Sensorik

Info-box

INFO

Wenn Sie bspw. aufgrund Ihres Ursache-Wirkungsdiagramm oder einer Empfehlung Ihrer Prozessexperten interessante Messgrößen oder Kennwerte identifizieren konnten, erwägen Sie die Anbringung zusätzlicher Sensorik. Zusätzliche Sensoren müssen nicht zwangsweise physisch sein, es können auch virtuelle Sensoren genutzt werden, die ihre Zielgröße aus physikalischen Messgrößen ableiten.

TIPP

Wenn Sie sich nicht sicher sind, welche Sensoren Sie benötigen, probieren Sie zunächst „genauere“ Sensoren aus. Viele Sensorik Hersteller unterstützen Sie auch bei der Auswahl von Sensoren und stellen Ihnen diese für Testmessungen zur Verfügung. Alternativ bieten manche Unternehmen auch Sensorik zum Mieten an.

Tipps/hint boxes

Screenshot of the implementation of the checklist in PIA

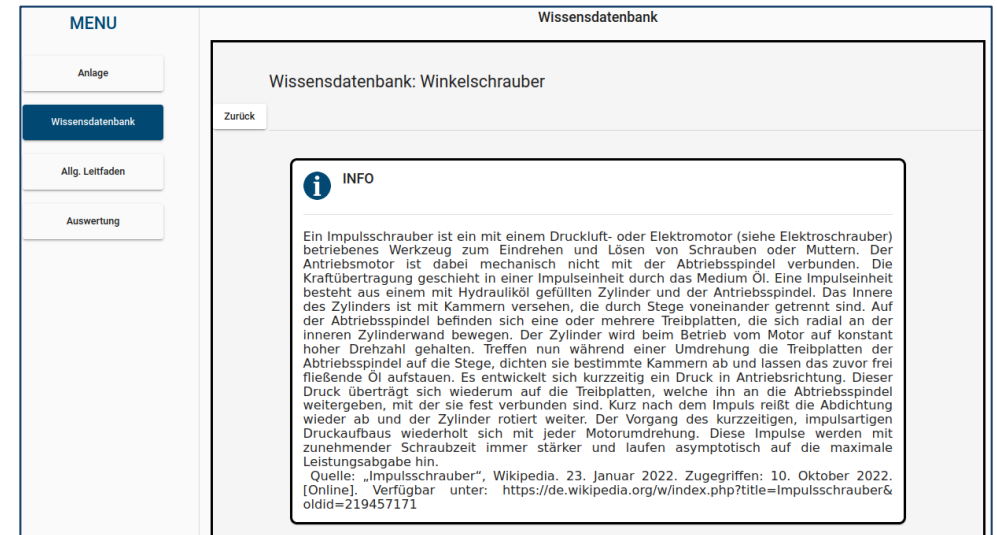
Pillar 2: Accessibility of knowledge and data

Accessibility of knowledge

- General knowledge database “wiki” with relevant information
- Expandable with specific knowledge



General overview about assembly operations

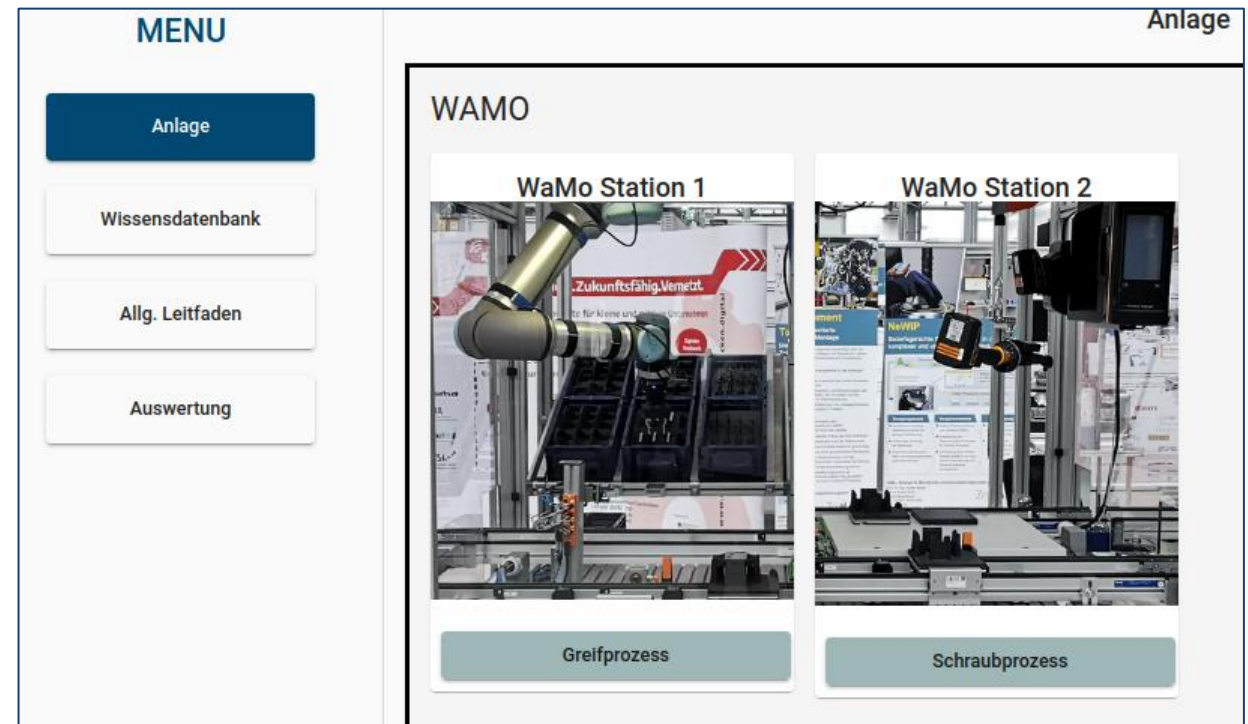


Specific entry about a tool

PIA groups information into lines and processes

Information included:

- Data
- Sensors used
- Data sheets
- Shift books
- Technical drawings
- Etc.



Example processes demonstrated in PIA

Accessibility of data and metadata

PIA provides additional documents to support the user:

e.g., a list of sensors used with relevant metadata:

- Sensor type
- Manufacturer
- Position
- IP-Address
- Sampling rate
- Channels etc.

Übersicht der Sensorik

Ersteller: _____
Projekt: _____
Datum: _____

UNIVERSITÄT DES SAARLANDES
Gefördert mit den Mitteln des Saarlandes und der Europäischen Union

		Sensor 1	Sensor 2	Sensor 3	Sensor 4	Sensor 5	Sensor 6
Allgemein	Sensorbezeichnung						
	IP-Adresse						
	Hersteller						
	Messgröße						
	Messposition						
	Erwartete Effekte						
	Anzahl Sensoren						
Sensordaten	Sensortyp						
	Messprinzip						
	Abtastrate						
	Bandbreite						
	Auflösung						
	Kanäle						
	Ausgangssignal						

Ersteller _____ Verantwortlicher _____

Overview of sensors used with their metadata

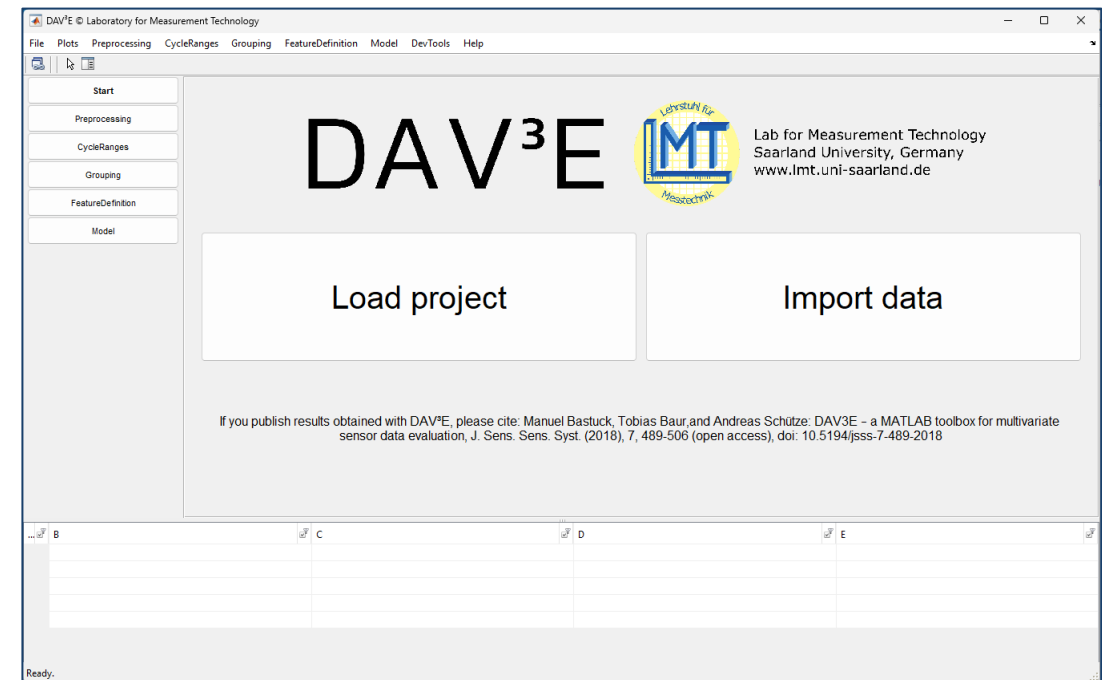
Pillar 3: Data analysis

Data Analysis with DAV³E

DAV³E – Data Analysis and Verification/Visualization/Validation Environment

DAV³E is a MATLAB toolbox for:

- Feature extraction from time series, especially cyclic sensor signals
- Sensor fusion
- Data preprocessing
- ML model building
- Evaluation & Validation



GUI of DAV³E

Currently stand-alone software – link to PIA pending

Source: Bastuck, M. et al., 2018, doi: 10.5194/jsss-7-489-2018

DAV³E – Automated Machine Learning Toolbox

Feature Extraction

Adaptive Linear Approximation
(local details)

Principal Component Analysis
(global features)

Best Fourier Coefficients
(frequency domain)

Best Daubechies Wavelets
(time-frequency domain)

Statistical Moments
(statistics)

Feature Selection

Correlation
(univariate)

**Recursive Feature Elimination
Support Vector Machine**
(multivariate)

ReliefF
(non-linear)

Correlation
(univariate)

Classification

**Linear Discriminant Analysis
+ Mahalanobis distance**
(pattern recognition)

15 combinations of feature
extractors and selectors:
best combination is chosen
automatically based on
minimum cross-validation error

Download here:



Source: Schneider, T. et al., 2018, doi: 0.1109/I2MTC.2018.8409763

Demonstration of PIA



**UNIVERSITÄT
DES
SAARLANDES**

Thank you for your attention!

c.schnur@imt.uni-saarland.de