



Systematisches Datenmanagement als Grundlage für KI

Ralph Müller-Pfefferkorn, Lincoln Sherpa, Valentin Khaydarov, Leon Urbas (TU Dresden)

Gregor Tolksdorf, Michael Kawohl, Michael Wiedau (Evonik)

Udo Enste (Leikon)

Marco Gaertler (ABB)

Martin Krawczyk-Becker (Krohne) & David Wagner-Stürz (Samson Group)

22.05.2023

Gefördert durch:



Bundesministerium
für Wirtschaft
und Klimaschutz

aufgrund eines Beschlusses
des Deutschen Bundestages

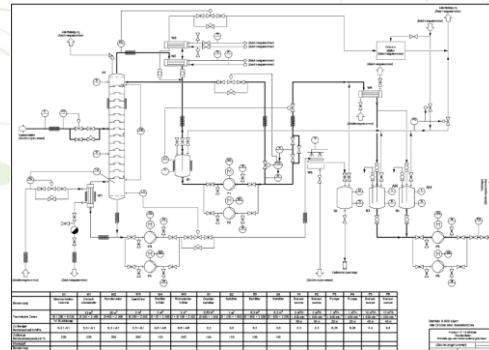
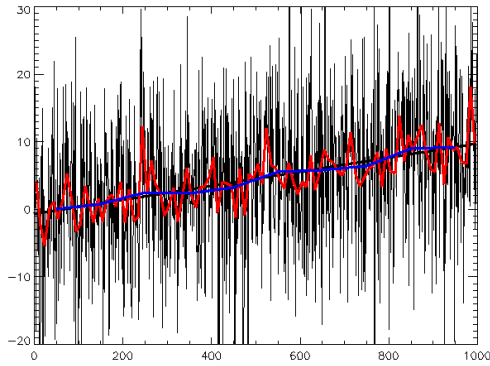


KE-3N  **KEEN**

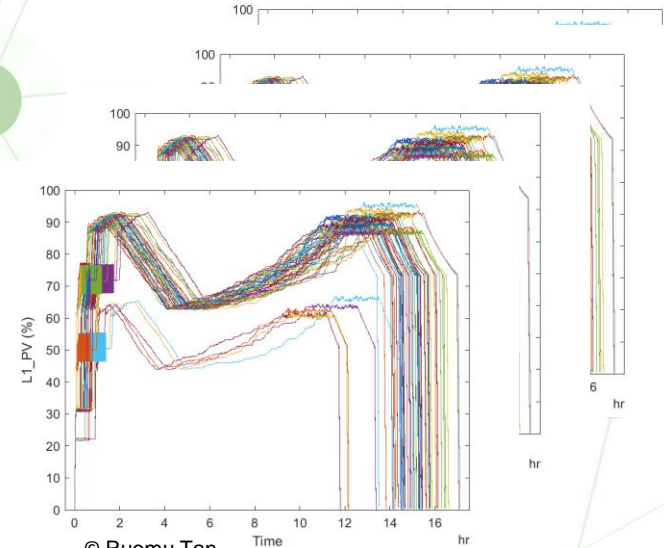
Datenvielfalt (für KI) in Prozessindustrie



©Valentin Khaydarov/TU Dresden



©ISO 10628



© Ruomu Tan,
<https://doi.org/10.57826/KEEN/ODU6MA>

FINDABLE
ACCESIBLE
INTER OPERABLE
REUSABLE

Wilkinson et al., 2016, "The FAIR Guiding Principles for Scientific Data Management and Stewardship"
Nature Scientific Data

Verbesserte Datensammlung

Integration und semantische Aufbereitung von Daten

Plattform zum Teilen und Nachnutzen der Daten



Plattform zum Teilen und Nachnutzen der Daten

22.05.2023

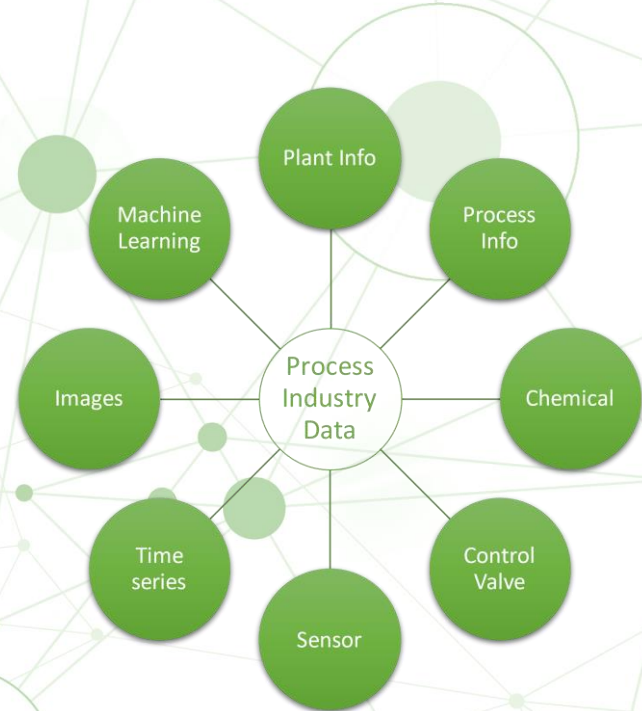
Unbeschränkt freigegeben für die Verwendung durch die KEEN-Partner

4



Daten dokumentieren - ProMetaS

- Daten müssen dokumentiert werden – zum Verständnis und zur (Nach)Nutzung
- Gemeinsame (Ingenieure + Informatiker) Entwicklung eines modularen Metadaten-schemas zur Beschreibung von Prozessindustriedaten:
ProMetaS - Process Engineering/Industry Metadata Schema
- Nutzt existierende Standards wie DEXPI oder ISA-95 – nur fehlende Beschreibungen ergänzt



Metadatenkategorien von ProMetaS

Zentrale Datenplattform zum Teilen

- Daten teilen (z.B. in einer Gruppe), archivieren oder publizieren (für alle)
- Basiert auf Open Source Software dataverse (Repositorium)
- Flexible Metadaten: ProMetaS integriert
- Suche nach Daten in Metadaten
- Zugang über Webbrowser
- Programmier-Schnittstelle (REST) zum direkten Ablegen, Suchen oder Lesen von Daten

Data Citation with Persistent Identifier

Data Files

Metadata

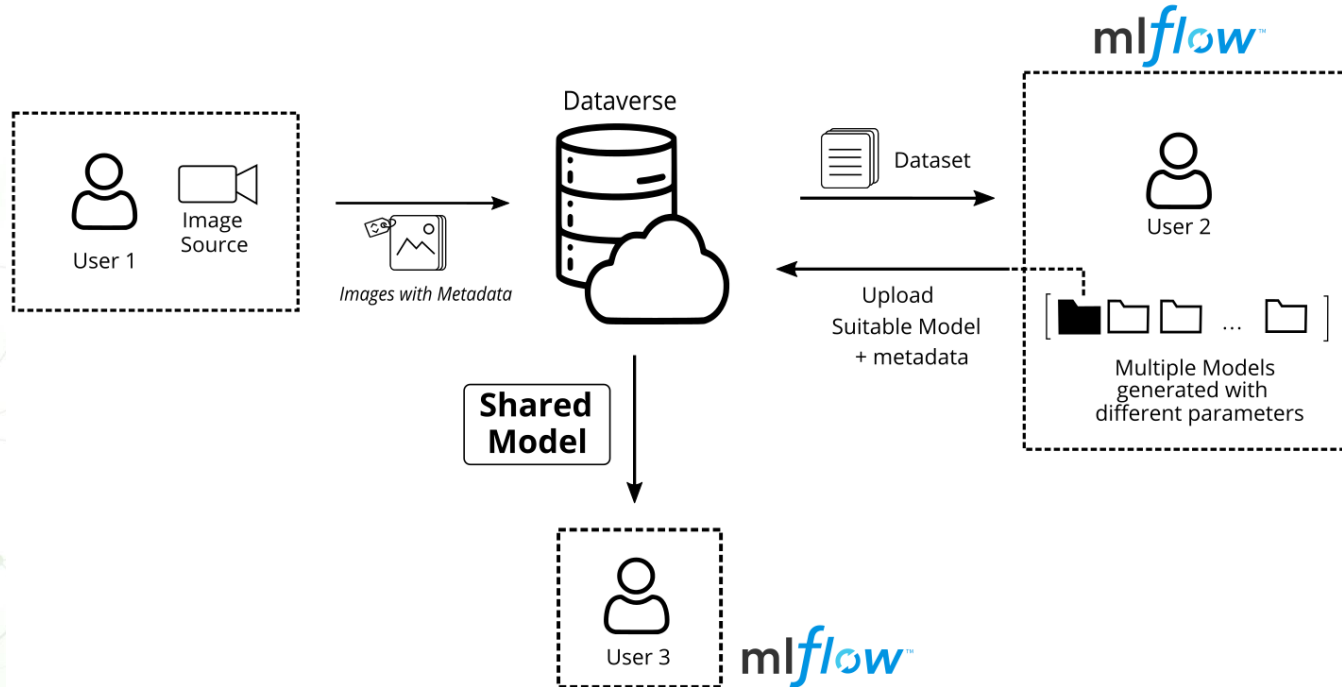
Data Licenses, User Agreements, Restrictions

Versions

The screenshot shows a Dataverse dataset page for 'DN15_Extraction_column_Images' from TU Dortmund. The page includes a header with the TU Dortmund logo and navigation links. The dataset title is 'DN15_Extraction_column_Images'. Below the title, there is a 'Cite Dataset' button and a 'Learn about Data Citation Standards' link. The description is 'Image Data set for binary classification of a DN15 extraction column (2022-05-04)'. The subject is 'Chemistry; Engineering'. The license is 'CC-BY-4.0'. There are tabs for 'Files', 'Metadata', 'Terms', and 'Versions'. The 'Files' tab is active, showing a list of files with columns for file name, type, size, publication date, and download count. Two files are visible: 'fluten000001.png' and 'fluten000002.png', both PNG images of 2.2 MB, published on June 30, 2022, with 1 and 0 downloads respectively. Annotations with orange arrows point to various elements: 'Data Citation with Persistent Identifier' points to the 'Cite Dataset' button; 'Data Files' points to the file list; 'Metadata' points to the 'Metadata' tab; 'Data Licenses, User Agreements, Restrictions' points to the 'License' section; and 'Versions' points to the 'Versions' tab.

keen.zih.tu-dresden.de

Beispiel: Automatisierte ML-Analyse von Bilddaten



Verbesserte Datensammlung

22.05.2023

Unbeschränkt freigegeben für die Verwendung durch die KEEN-Partner

8



Datenbasis für ML im Feld verbessern



Kosteneffiziente
Verbreiterung der
Datenbasis für
maschinelles Lernen!

Smart Data Hub – Kabellose Datensammlung
aus benachbarter Sensoren

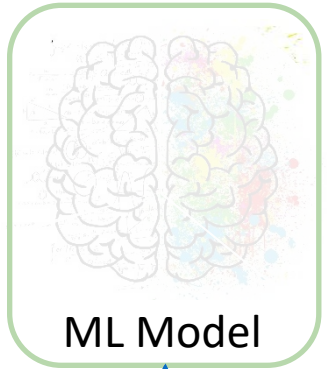


Anreichern mit Metadaten über 2. Kanal
(insbesondere Qualitätsmetriken)

SPS

SCADA

lediglich
primärer Messwert



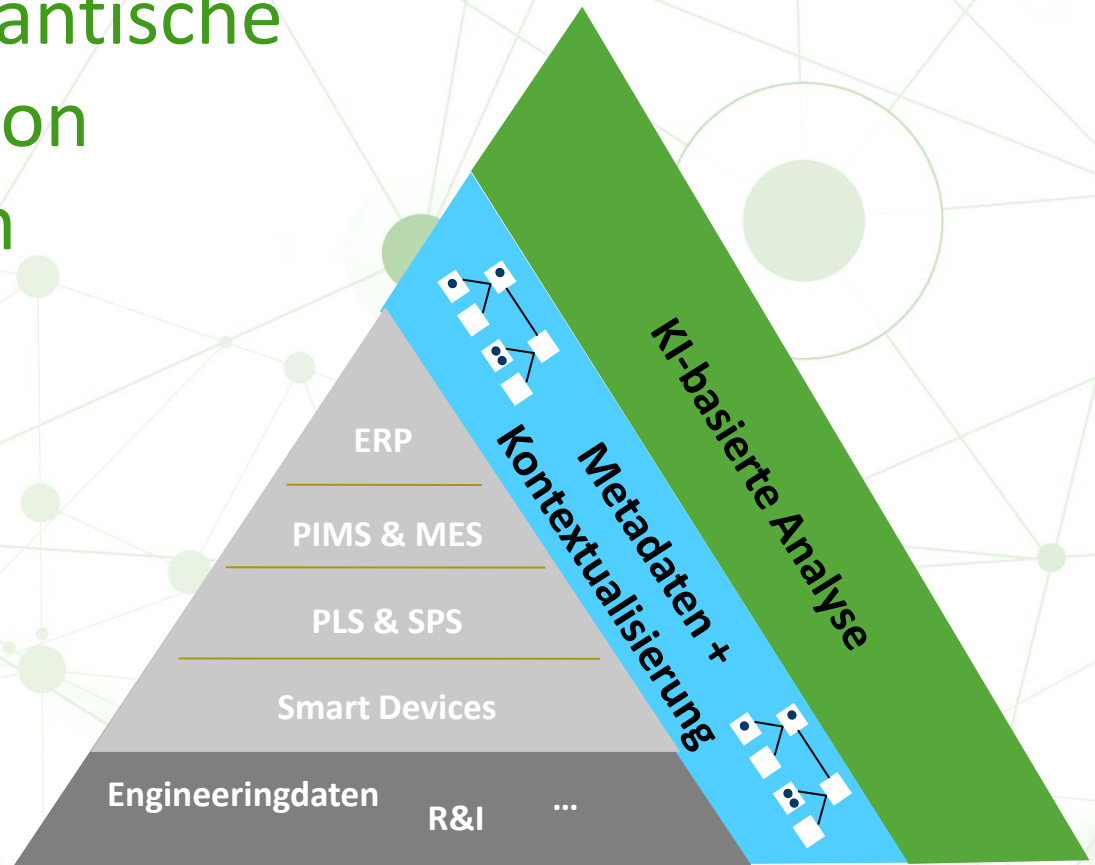
ML Model

The
Dataverse[®]
Project

Integration und semantische Aufbereitung von Daten

Integration und semantische Aufbereitung von Prozessdaten

- Daten aus unterschiedlichsten Systemen auslesen und semantisch verknüpfen
- Informationsgewinn für KI-basierte Analysen

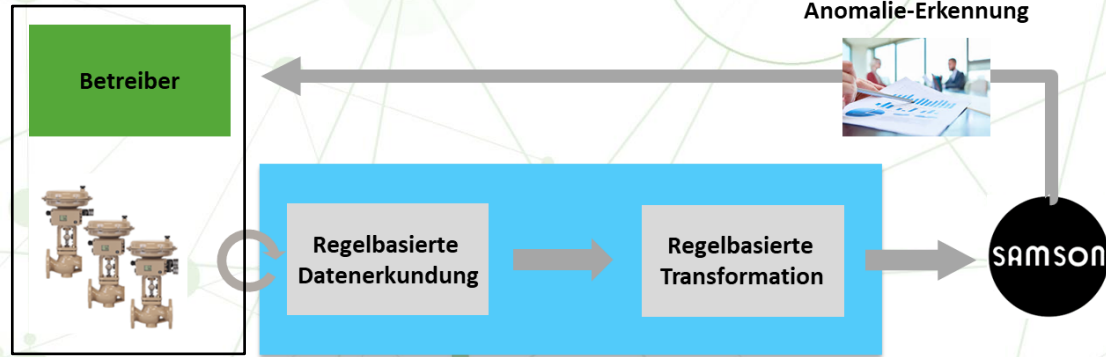




Use Case 1

KI-basierte Anomalie-Erkennung von Feldgeräten:

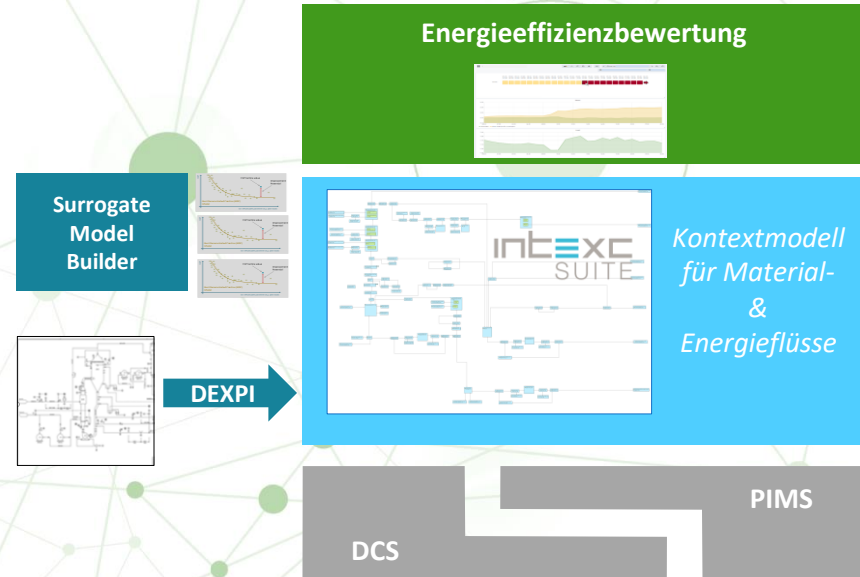
- Konfigurationsloses Einlesen von Gerätedaten
- Transformation in einheitliche Informationsmodelle
- KI-basierte Anomalie-Erkennung
- Mehrwert für Anlagenbetreiber:
 - Stillstandsvorbereitung
 - Condition Monitoring



Use Case 2

Kontextualisierung von Daten zur KI-basierten Energieeffizienzbewertung

- R&I-Diagramme (DEXPI) + Informationsmodell + Identifikation von Bilanzkreisen
- Verlinkung von Prozessdaten zum Informationsmodell
- Generierung von Surrogat-Modellen zur Energieeffizienzbewertung
- Pilot: Installation bei Evonik in Marl
- Publikation zur „Toolchain“:
DOI [10.1002/cite.202300004](https://doi.org/10.1002/cite.202300004)



Zusammenfassung

Ein **systematisches Datenmanagement** ist Grundlage für die Entwicklung und den Einsatz von KI rund um den Datenlebenszyklus.

Bausteine aus KEEN (Auswahl)

- Smart Hub - kabellose Integration von Sensoren
- Sammlung und (semantische) Integration von Metadaten – „Toolchain“
- Metadatenschema ProMetaS zur strukturieren Datendokumentation
- Plattform zum Daten teilen, archivieren und publizieren (Dataverse)

