

# Data (space) related research @ Software AG

Data Week 2023, Luleå, Sweden

Dr. Harald Schöning. Vice President Research, June 13, 2023



# Various projects to create data spaces

## RESOURCE-EFFICIENT, ECONOMIC AND INTELLIGENT FOODCHAIN

<https://ki-reif.de/>



Supported by:



Federal Ministry  
for Economic Affairs  
and Climate Action

Reduce food waste along the value chain

on the basis of a decision  
by the German Bundestag

# Various projects to create data spaces



## EuProGigant

<https://euprogigant.com/en/>

The goal of the project is the demonstration and scaling of a multi-location, digitally networked production ecosystem with resilient, data-driven and sustainable value creation to strengthen the pioneering role of European industry.



Supported by:



on the basis of a decision  
by the German Bundestag

# Various projects to create data spaces



## Intelligent Empowerment of Construction Industry

<https://ieco-gaiax.de/>

Create a Gaia-X based data space for the Construction Industry to close the productivity gap → distributed digital twin



Supported by:



on the basis of a decision  
by the German Bundestag

# Various projects to create data spaces

Funded by:



Federal Ministry  
of Education  
and Research



## Datenraum Industrie 4.0 für die zerspanende Produktion



Create a Gaia-X based Value creation ecosystem for the machining sector



# Various projects to create data spaces



## GAIX-X for Advanced Mobility Services

The project provides data and services related to automated driving and cooperative system alliances based on connected and automated vehicles as well as intelligent transportation infrastructures

# Gaia-X 4 AMS

Supported by:



on the basis of a decision  
by the German Bundestag

# Various projects to create data spaces – commonalities ?



Gaia-X 4 AMS

Application  
specifics

Domain  
specific tooling

General Data  
space tooling

GAIA-X

SIMPL (?)



# Business model?



Plattformen

providing data from various sources

~~and (AI) services~~

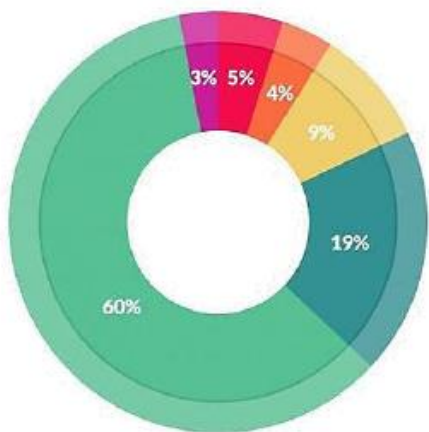
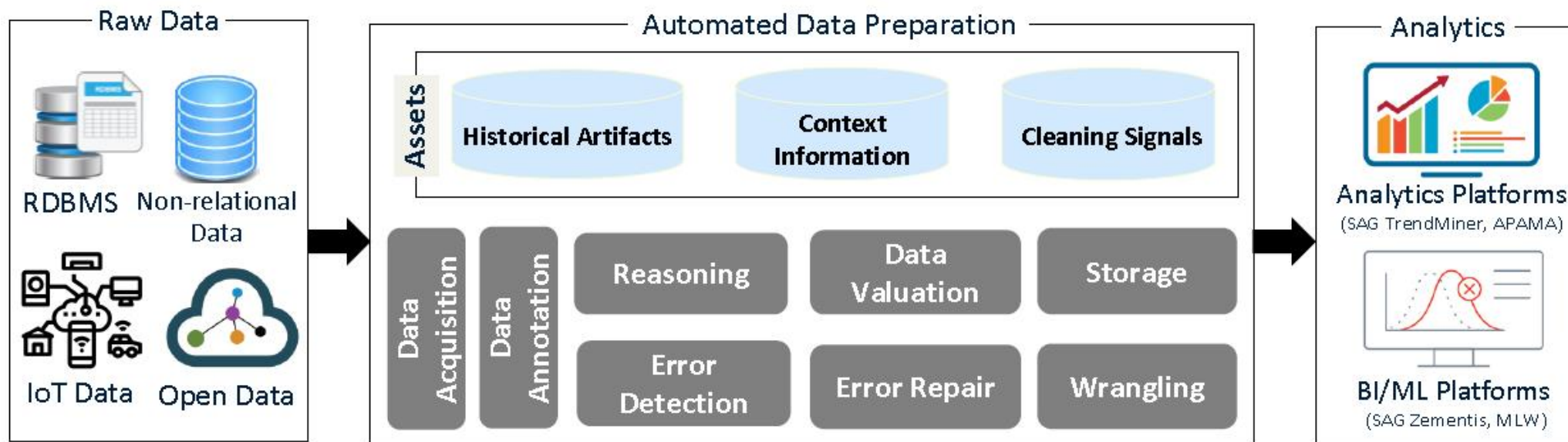


→ Data Governance Act

Gaia-X 4 AMS

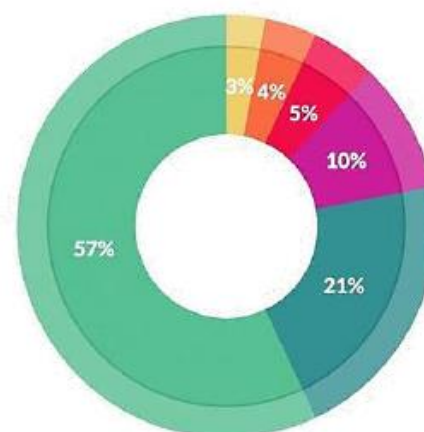


# Facilitating data preparation



What data scientists spend the most time doing

- Building training sets: 3%
- Cleaning and organizing data: 60%
- Collecting data sets: 19%
- Mining data for patterns: 9%
- Refining algorithms: 4%
- Other: 5%



What's the least enjoyable part of data science?

- Building training sets: 10%
- Cleaning and organizing data: 57%
- Collecting data sets: 21%
- Mining data for patterns: 3%
- Refining algorithms: 4%
- Other: 5%

Source: Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Gil Press, Forbes, March 23<sup>rd</sup>, 2016

# Facilitating data preparation - AI for data cleaning

Research is done in the following projects:

KompA+KI

<https://kompaki.de/>

IML4E

<https://iml4e.org/>



Funded by:



Federal Ministry  
of Education  
and Research

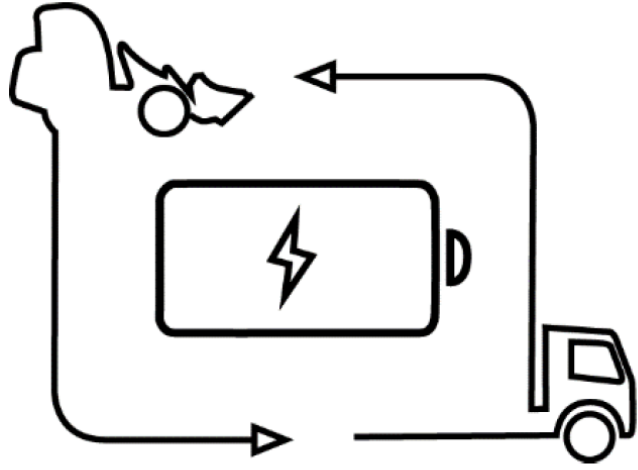
Several papers (see next slide)

Three patent applications

# AI for data cleaning - References

- Mohamed Abdelaal, Christian Hammacher, and Harald Schoening. “REIN: Comprehensive Benchmark Framework for Data Cleaning in ML Pipelines”. In: *26th International Conference on Extending Database Technology (EDBT)*. Mar. 2023.
- Mohamed Abdelaal, Rashmi Koparde, and Harald Schoening. “AutoCure: Automated Data Curation Framework for ML Pipelines”. In: *Sixth International Workshop on Exploiting Artificial Intelligence Techniques for Data Management (aiDM, SIGMOD workshop)*, June. 2023.
- Daniel Del Gaudio, Tim Schubert, and Mohamed Abdelaal. “RTClean: Context-aware Tabular Data Cleaning using Real-time OFDs”. In: *2023 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops)*. Under Review. Mar. 2023.
- Tobias Clement, Nils Kemmerzell, and Mohamed Abdelaal. “XAIR: A Systematic Meta-Review of Explainable AI aligned to the Software Development Process”. In: *Machine Learning and Knowledge Extraction, special issue on Advances in Explainable Artificial Intelligence (XAI) (2022)*. January 2023.
- Benjamin Hilprecht, Christian Hammacher, Eduardo Reis, Mohamed Abdelaal, and Carsten Binnig. “DiffML: End-to-end Differentiable ML Pipelines”. In: *7th Workshop on Data Management for End-to-End Machine Learning (DEEM, SIGMOD workshop)*, June 2023.

# Environmental (Life Cycle) Assessment



REVAMP

For batteries (e.g. in vehicles)



**CLICE - DiPP**

Climate-neutral circular economy  
enabled by  
Digital Product Carbon Pass

