Semantic Data Management

Actors and their Requirements

Different actors have different assumptions, goals and challenges when working with data

Data Provider

- Knows his data and use cases
- **Provides data** so others can work with them
- Needs a way to s**hare his knowledge**

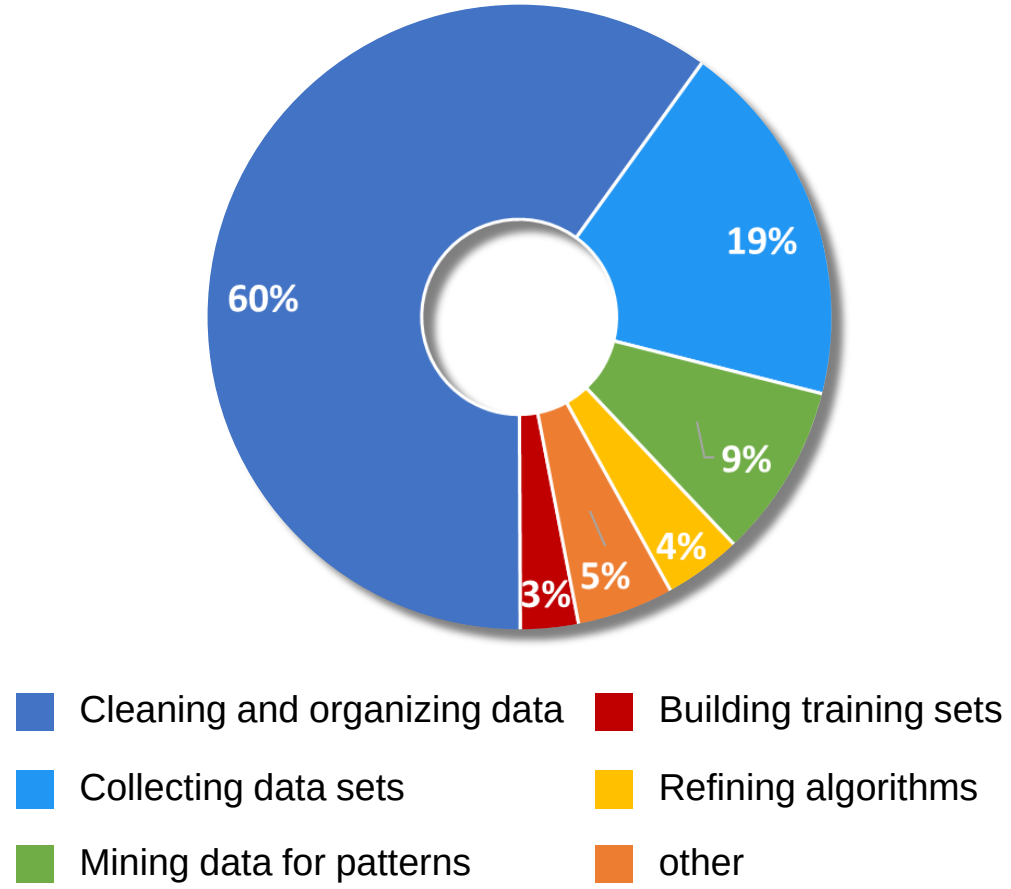
Data Consumer

- Wants to work with data
- Needs to **find the right data** for his use case
- Needs to **access the data**
- Needs to **understand the data**

Current Situation

- Data scientists spend almost **80% of their time with collecting, cleaning and organizing data**
- Common challenges:
  - **Quality**, **comprehensiveness, availability**, **retrievability**, and **expandability** of data
- To **make the most of your data**, they need to be available in a way that not only allows but supports further processing and future use cases
- Semantic data management effectively reduces the **time-to-application** and prepares your data for **future use cases**

What data scientists spend the most time doing



Source: Forbes, 2016

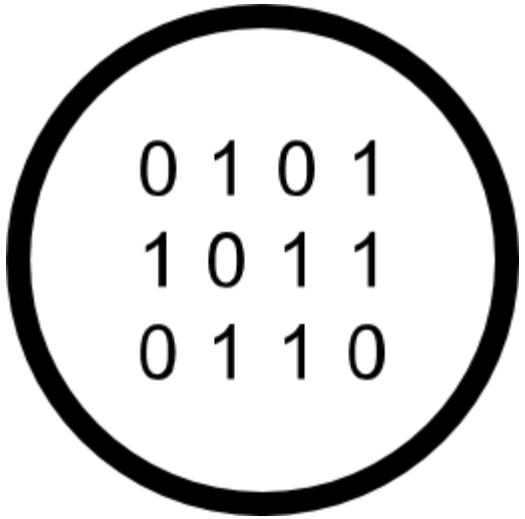
Challenges

Heterogeneous Provision

Heterogeneous Models

Relational

Hierachy



Heterogeneous Quality

Heterogeneous Formats

JSON

XML

CSV

Heterogeneous Use Cases

1

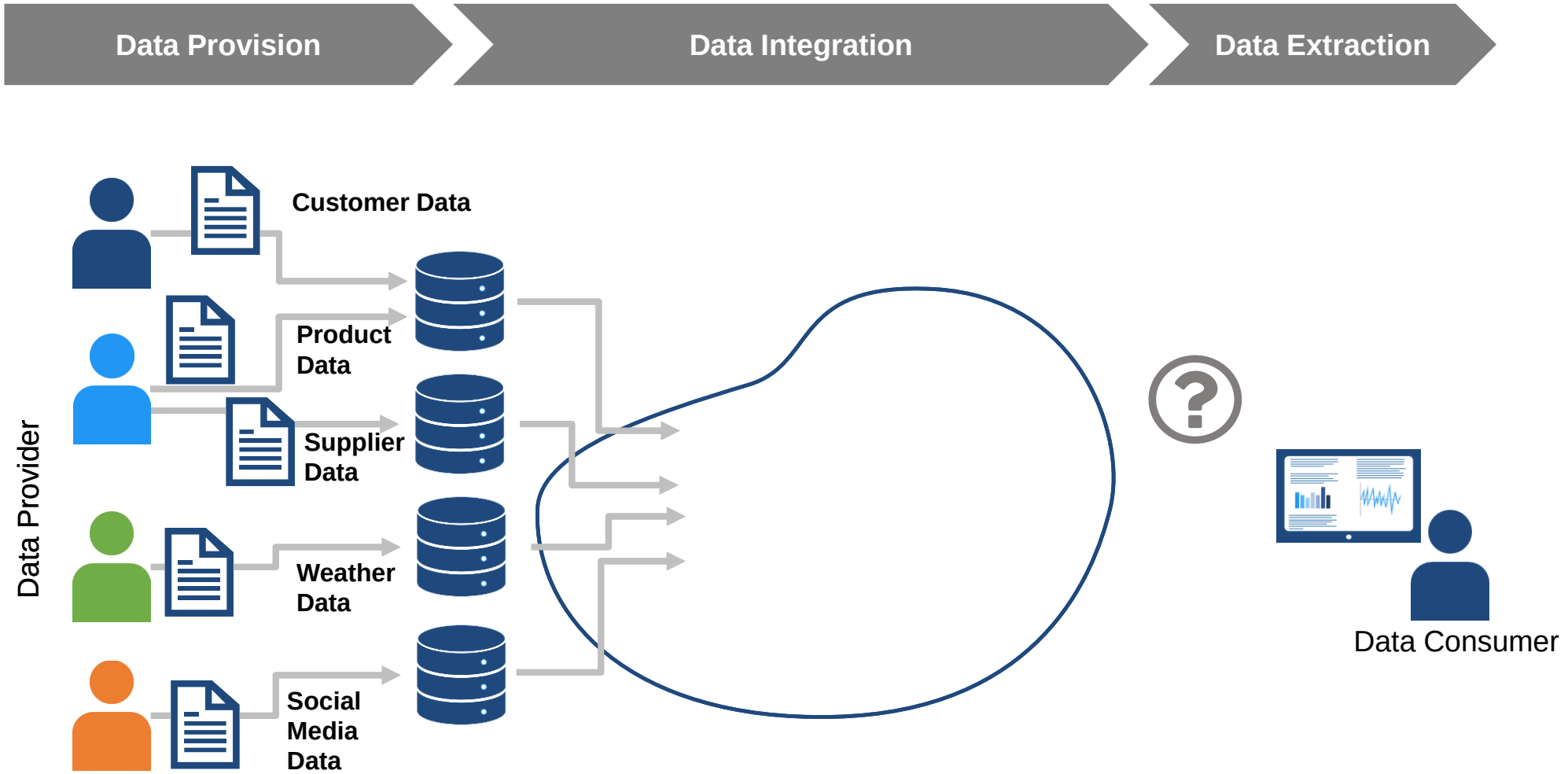
2

3

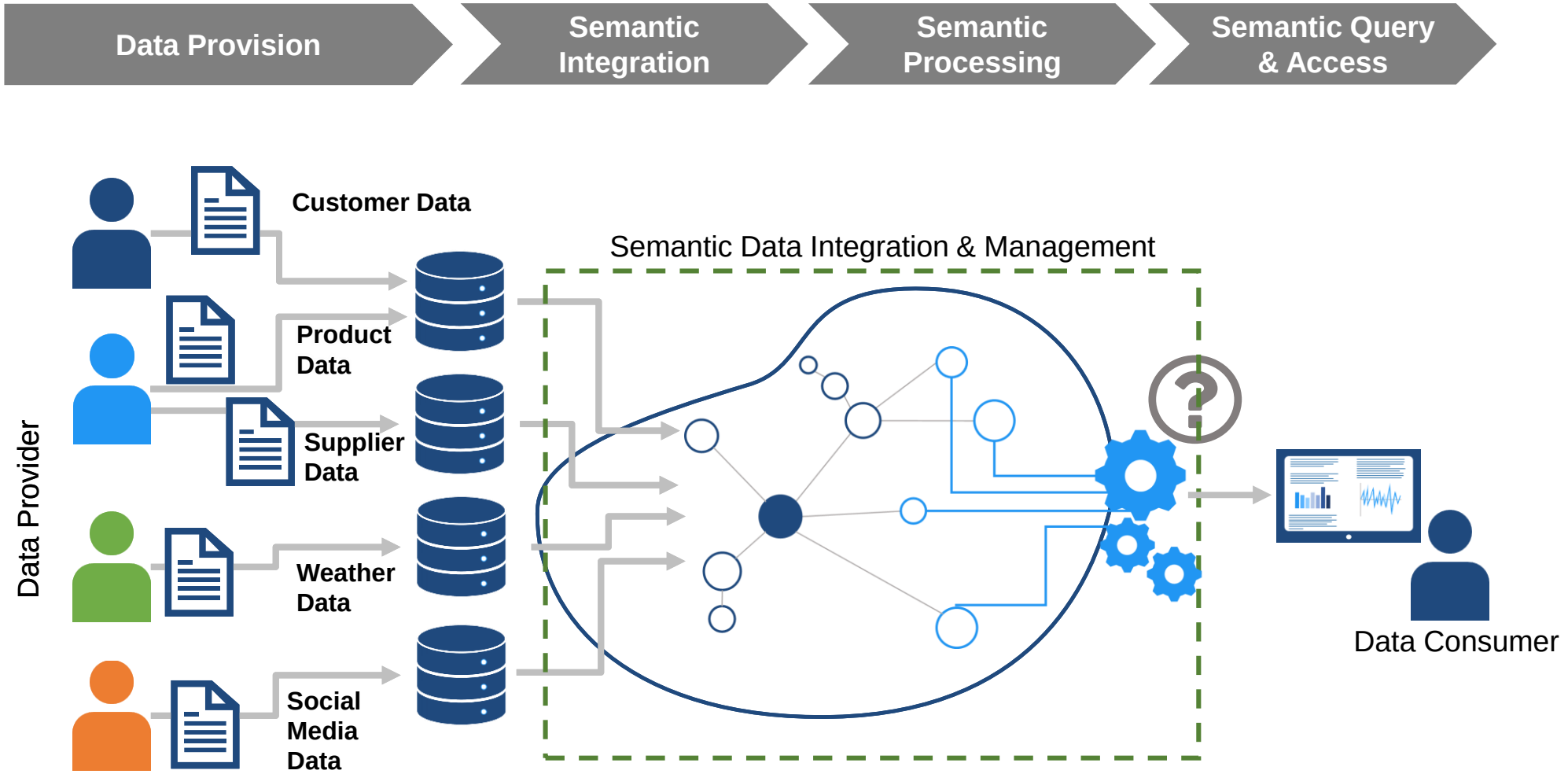
Heterogeneous Semantics

°C °F K

Schematic Representation of Initial Situation



Schematic Representation of Semantic Data Integration & Management



Our Mission

- Google **finds** and **"understands" websites without the need** for people to describe them explicitly
  - How can **enterprise data** be provided and found without complex and possibly incomprehensible descriptions?

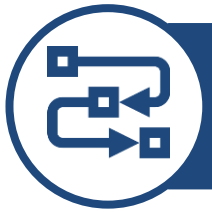
- Our aim is to reduce the **Time-to-Application** of enterprise data by supporting data scientist to answer pressing questions like:
  - Where to find the data needed for an application, analysis etc.?
  - What information is contained in the available data?
  - Are there other data sources that might be relevant to solve a challenge?

Research Goals

Bridge the gap between physical & virtual world by developing intuitive concepts for Big Data collection

Develop new approaches for constructing and evolving Knowledge Graphs

Develop semantic mapping approaches that leverage machine learning & external knowledge

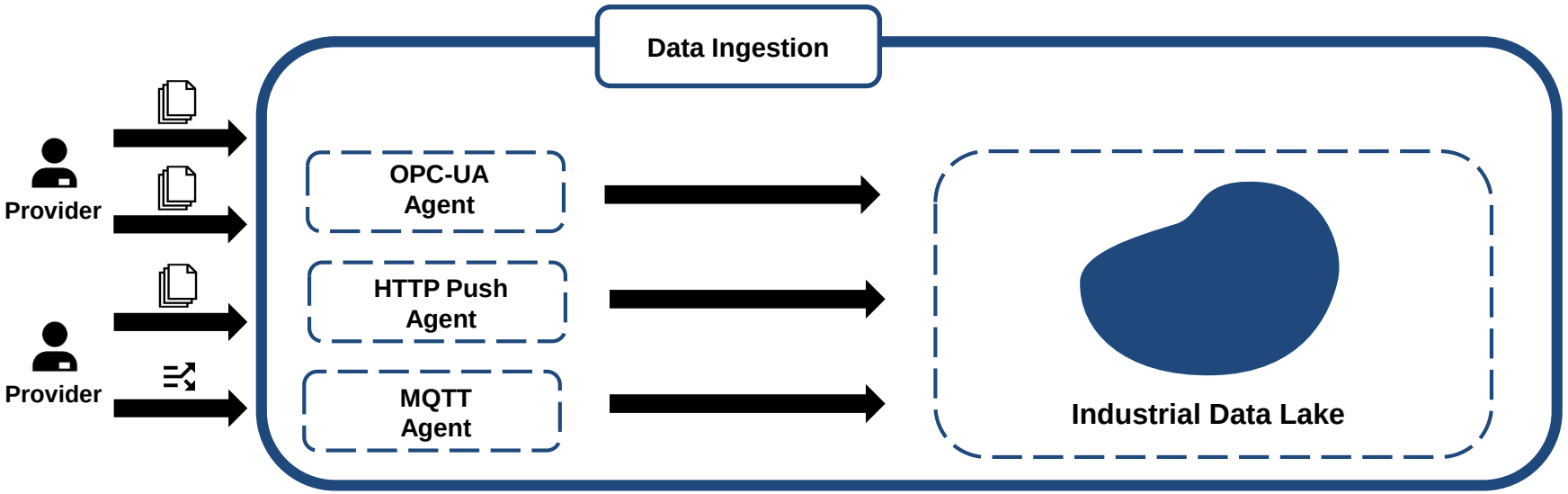


Enable semantic processing for data based on evolving Knowledge Graphs

Improve the usability & accessibility of data with the help of semantic querying

Big Data Collection

Bridge the gap between physical & virtual world by developing intuitive concepts for Big Data collection



- Plug & Ingest: Autonomous agents monitor data sources and collect the data in heterogeneous company-wide environments
- Agents proactively react to changing data sources and interfaces
- Data is ingested into a central location, e.g., into a data lake

Continuous Ontology & Knowledge Graph Creation

Develop new approaches for constructing and evolving Knowledge Graphs

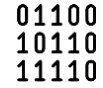
Knowledge graph learns continuously from interacting with its environment

Data Provider

Defines Semantic Models

Data Consumer

Queries Data Sets



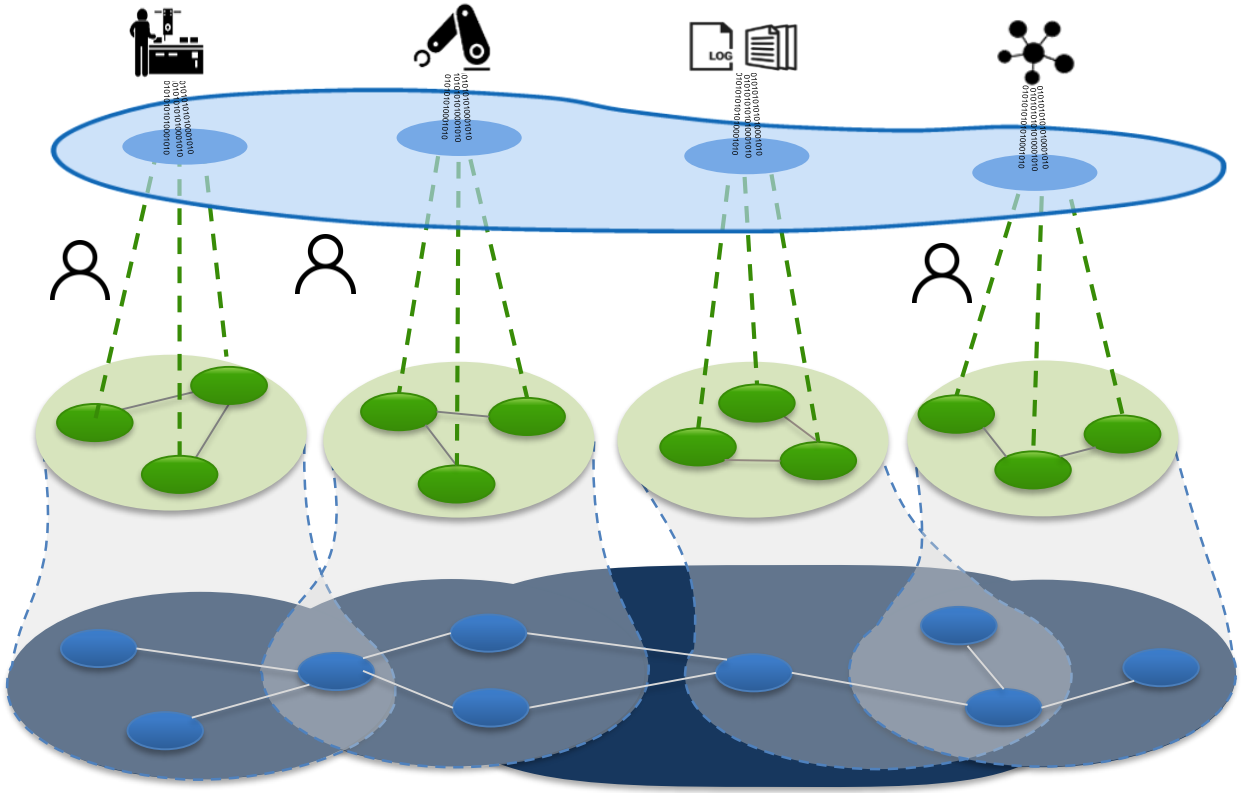
System

Autonomous Evolution

- Knowledge Graph **grows and strengthens continuously**
- Knowledge Graph allows **statements about relationships** of data sets
- **Advanced AI algorithms** monitor and support the evolution of the knowledge graph

Continuous Ontology & Knowledge Graph Creation

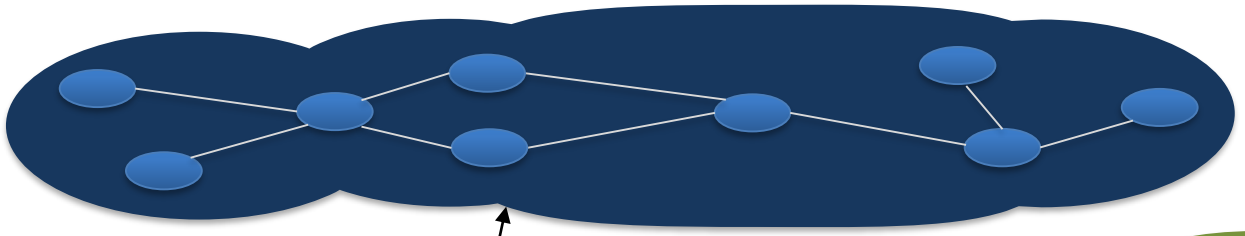
Develop new approaches for constructing and evolving Knowledge Graphs



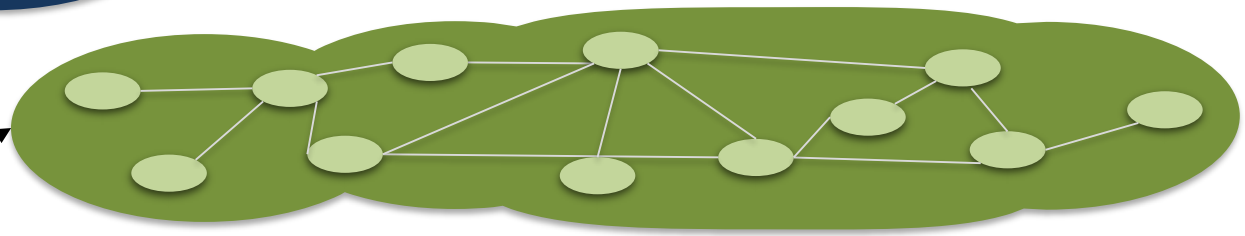
- **Data collection and access**
  Construction Industrial Data Lake
- **Data Description**
  Local Knowledge Representation
- **Learning from Data**
  Global Knowledge Graph

Semantic Mapping: External Knowledge Recommendation

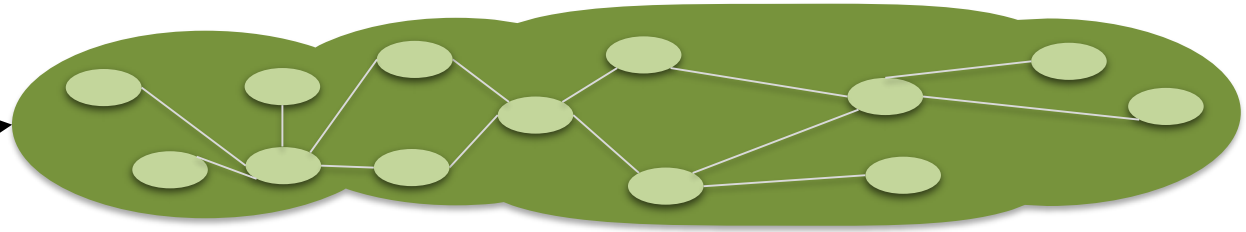
Develop semantic mapping approaches that leverage machine learning & external knowledge



Knowledge Graph



External Knowledge 1



External Knowledge 2

Timestamp (92,2%)

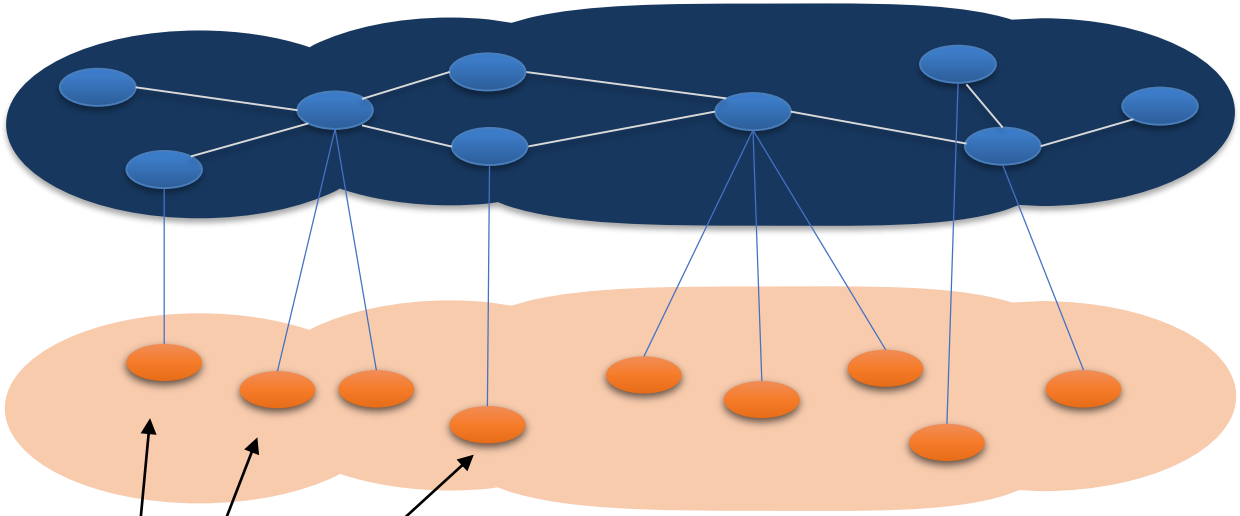
Identifier (4,2%)

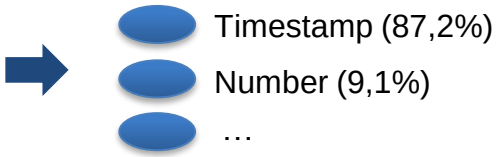
…

TS	TEMP	POS_Y	POS_X	ALT
1476835200	3	50.780296	6.104511	12
1476835200	4	50.774423	6.085856	4
1476835200	10	50.776057	6.043272	11

Semantic Mapping: Machine Learning Recommendation

Develop semantic mapping approaches that leverage machine learning & external knowledge





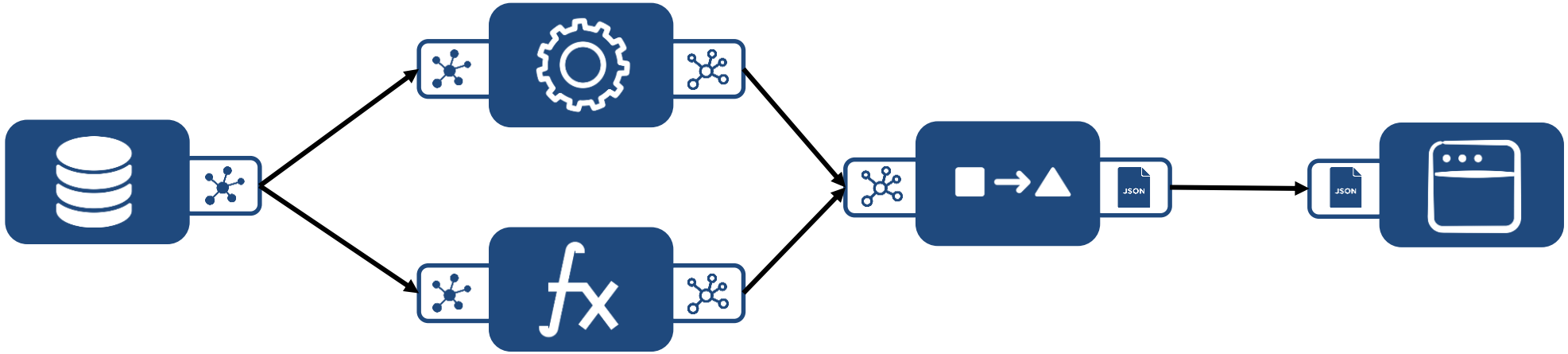
…

TS	TEMP	POS_Y	POS_X	ALT
1476835200	3	50.780296	6.104511	12
1476835200	4	50.774423	6.085856	4
1476835200	10	50.776057	6.043272	11

- **Knowledge Graph**
- **Machine Learning Layer**
- **Recommendation**

Semantic Processing

Enable semantic processing for data based on evolving Knowledge Graphs

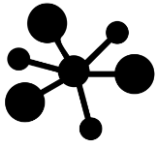


- Consider the continuously changing knowledge graph and enable the processing of data based on the provided semantics
- Prepare the data so that they are directly useable for analytics or applications

Developing Intuitive Concepts for Big Data Collection

Improve the usability & accessibility of data with the help of semantic querying

Natural Language & Visual Queries



Expand queries

Personalize the search experience

- Enable data set queries based on natural language and visual interfaces
- Leverage external knowledge databases to expand queries
- Learn from users queries in order to improve the personalized search experience

Selected Research Projects and Applications

Cross-site Enterprise Data Lake in Automotive Industry

Cross-site Enterprise Data Lake

Connection of **five plants** on **two continents**

Logistics
Construction
Paint Shop
Assembly
Rework

Live operation for more than 2 years
currently more than 6 TB/year

Configurable, intelligent **software connectors**

Optimization of throughput times on the conveyor belt

Goal

- **Construction** of a **cross-site enterprise data lake** that continuously collects and **integrates data from multiple factories**

Approach

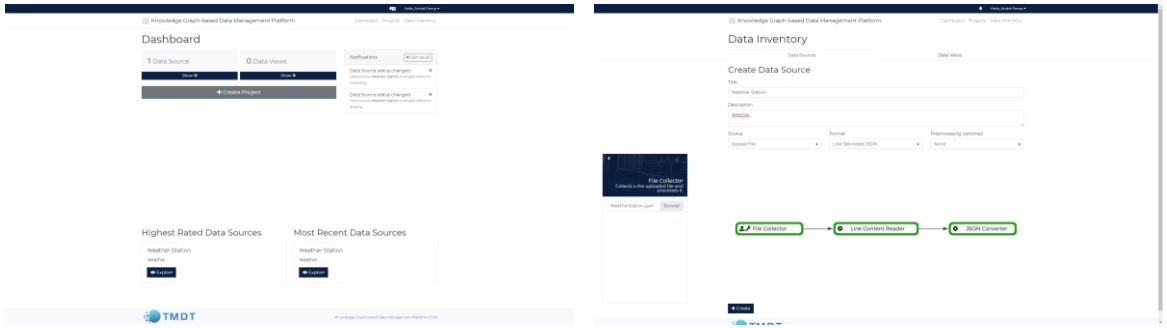
- Conceptual **design of cross-site, scalable, non-invasive data acquisition and -management**
- **Integration of heterogeneous data sources** in a common long-term data storage for each site through **configurable connector modules**
- **Automatic metadata enrichment**
- **Cross-site transparent consolidation of local data lakes** based on **information models**
- Provision of consolidated data through **common methods and familiar tools** (Tableau, Power BI, RapidMiner,…)

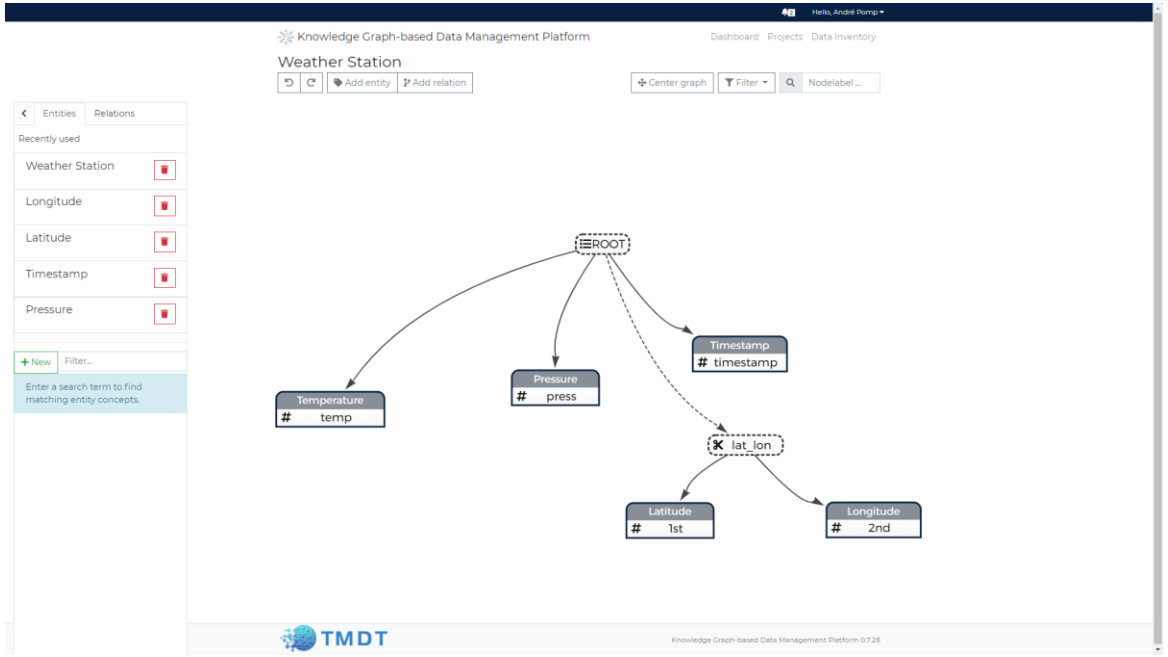
Results

- Functional cross-site data acquisition and management infrastructure for five plants on two continents
- Over 1000 different data sources integrated

Knowledge Graph-based Data Management Platform (KGBDM Platform)

KGBDM Platform





Goal

- **Reduce the time-to-application** by handling **heterogeneous data sources** on an information level

Approach

- Provide **different connector modules** for adding data sources
- Data providers **create semantic models for their data sets**
- Semantic model creation is **supported with the help of recommendation systems and artificial intelligence**
- **Semantic Data Integration** based on the provided semantic models
- Data access via the **semantic search**
- **No need to define ontologies**. The KGBDM platform **learns a knowledge graph** based on the user provided semantic models and queries

Results

- **Available state of the art semantic data management platform** that is **used in various research and development projects**

Your Contact Person:

André Pomp, M.Sc.
Tel: +49 (0)202 439 1153
pomp@uni-wuppertal.de

Chair for Technologies and Management of Digital Transformation
Univ. Prof. Dr. Ing. Tobias Meisen
https://www.tmdt.uni-wuppertal.de/
Campus Freudenberg
Rainer-Gruenter-Str. 21
D-42119 Wuppertal
Germany

University of Wuppertal
School of Electrical, Information and Media Engineering

TMDT

BERGISCHE UNIVERSITÄT WUPPERTAL