

Gefördert durch:



Bundesministerium  
für Wirtschaft  
und Energie

aufgrund eines Beschlusses  
des Deutschen Bundestages



Smart Data



# SIDAP

Emanuel Trunzer, TU München

Peter Schleinitz, IBM

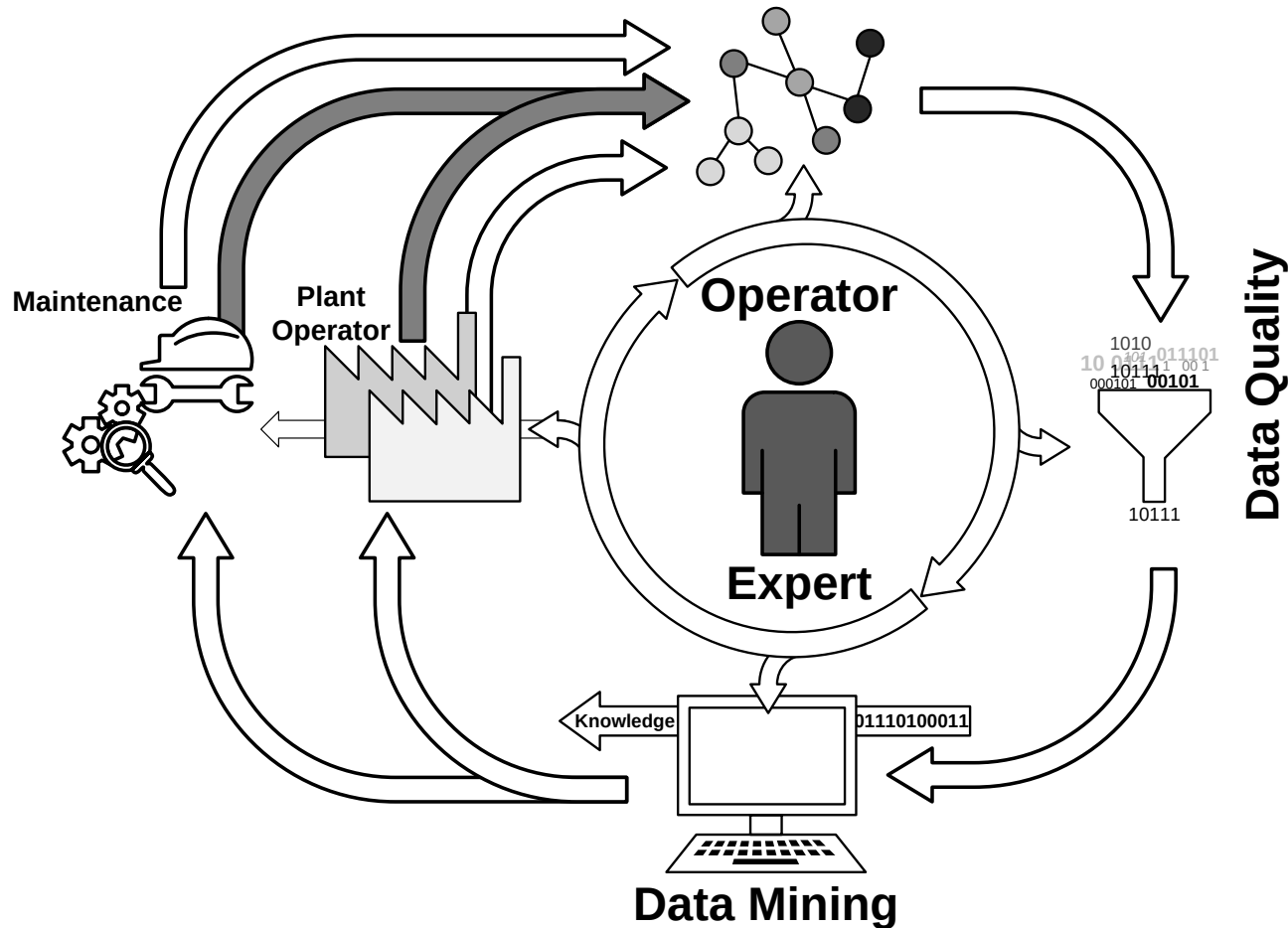
Skalierbares Integrationskonzept zur  
Datenaggregation, -analyse, -aufbereitung  
von großen Datenmengen in der Prozessindustrie



# Projektkonzept

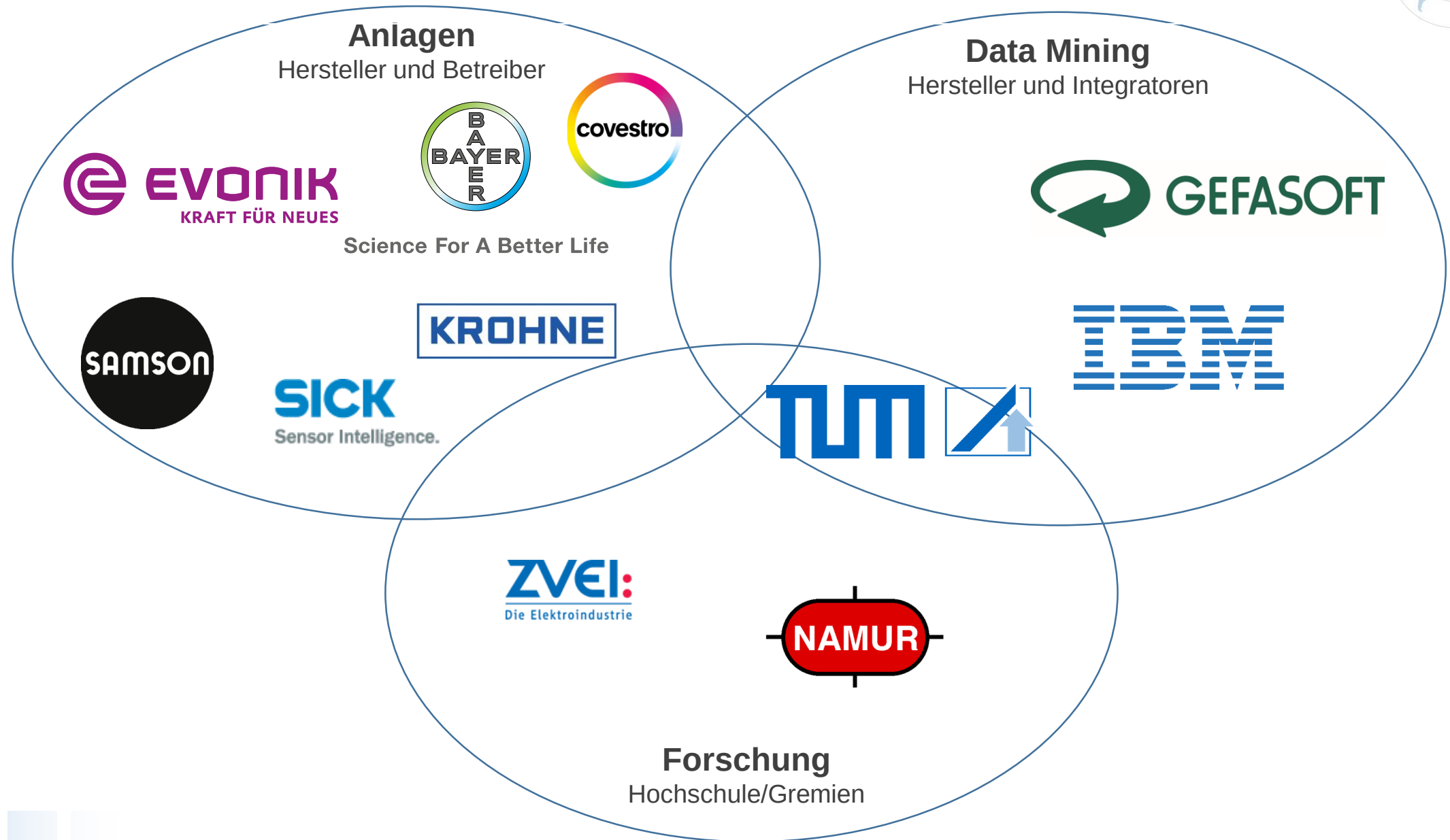
## Unternehmensübergreifende Datenanalyse

### Data Integration



## Forschungskonzept

- Ziel: Erhöhung der OEE (Overall Equipment Effectiveness)
- Herausforderungen: “Lernen” von verschiedenen existierenden Datenquellen (Entwicklung, Betrieb, Wartung), verschiedenen Werken und zwischen Hersteller, Engineering, Betrieb und Wartung
- Ansätze: Data Mining Algorithmen, Daten- und Systemarchitekturen, Cloud Technologien
- Ergebnisse: Voraussage fehlerhaften Verhaltens durch signal- und model-basierten Diagnosemethoden



Die Motivation des Projektes liegt in der **optimierten Informationsgewinnung**

## Chemieindustrie

Die Chemieindustrie ist einer der bedeutendsten Industriezweige in Deutschland. Sie erwirtschaftete 2012 mehr als **186 Mrd. Euro**. Damit liegt sie hinter dem Fahrzeugbau und dem Maschinenbau **auf dem dritten Platz**

Quelle: VCI, Initiative „Industrieland Deutschland stärken“



## STATUS QUO

### Komplexität

- Prozesse, weltweite Anlagen, 10.000 Geräte pro Anlage

### Sehr viele und heterogene Daten

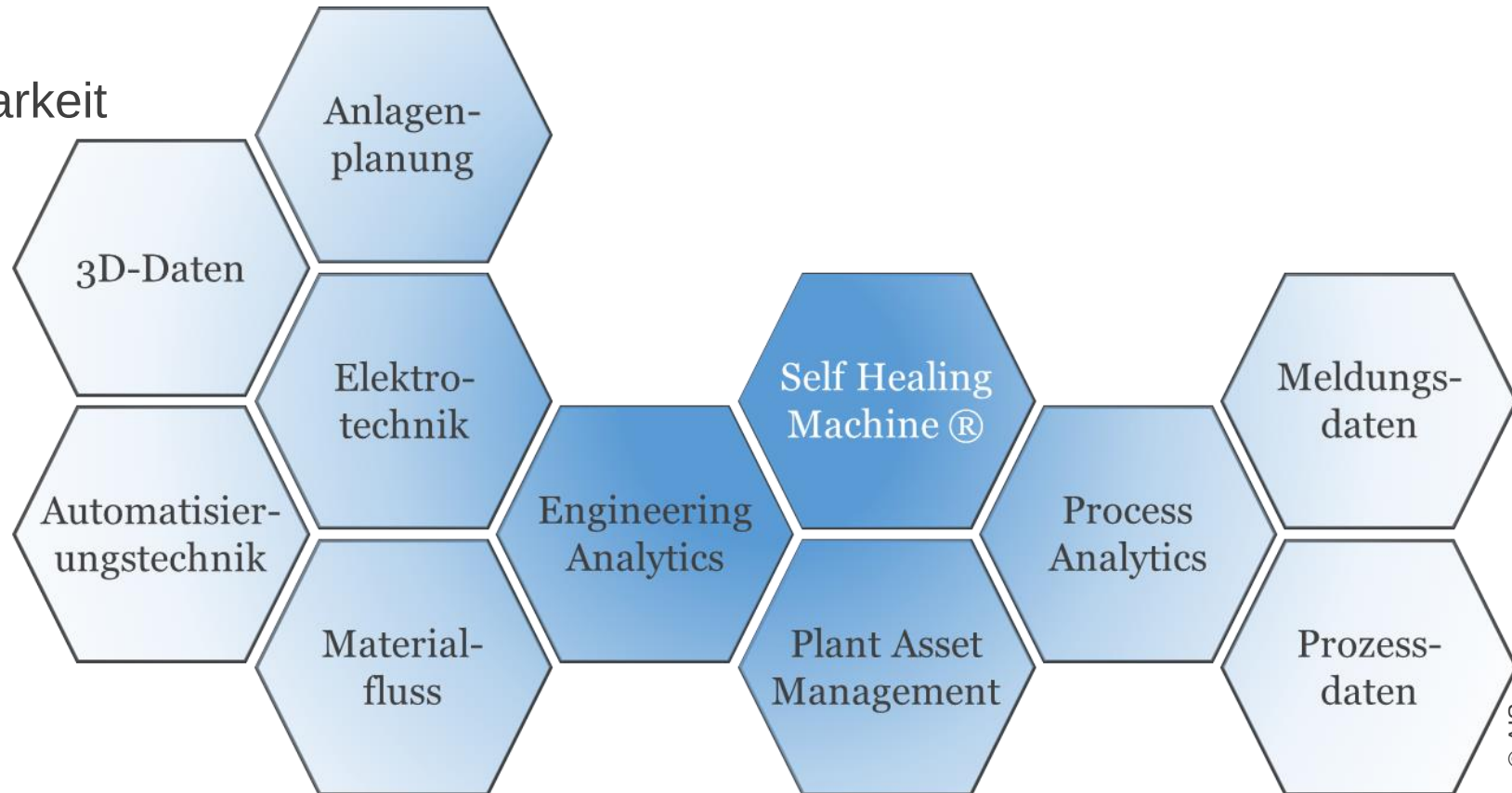
- strukturierte und unstrukturierte Daten

### Silos/eingeschränkte Datenanalysen

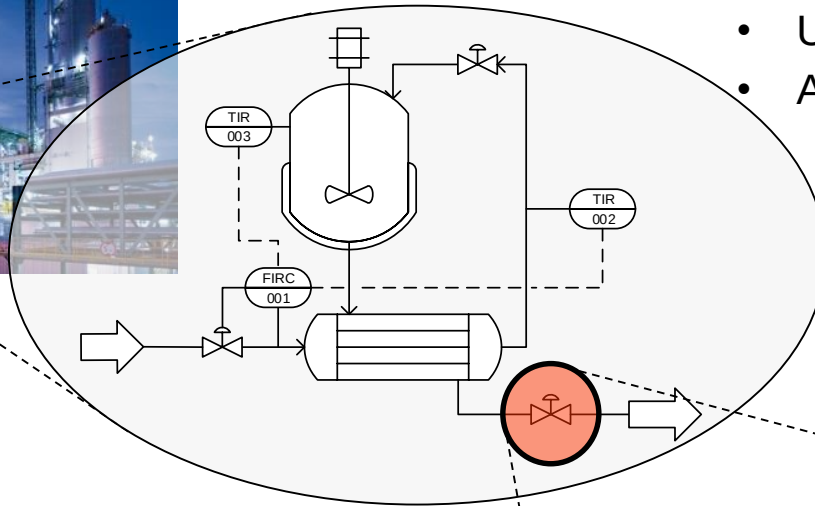
- Sehr hohe Aufwände bei der Datensammlung
- Verknüpfung von anlagen- / prozess- / produktspezifischem Wissen erforderlich

# Effizienzpotentiale

- Intelligentes und wissensbasiertes Asset Management System / Condition Monitoring System
- Steigerung der Anlagenverfügbarkeit und Produktqualität
- Integration der **Geschäftspartner** in Verfügbarkeitssteigerung
- Sicherung des Anlagenengineerings und -betriebs in **Deutschland**



# Anwendungsbeispiel *Ventildiagnose*

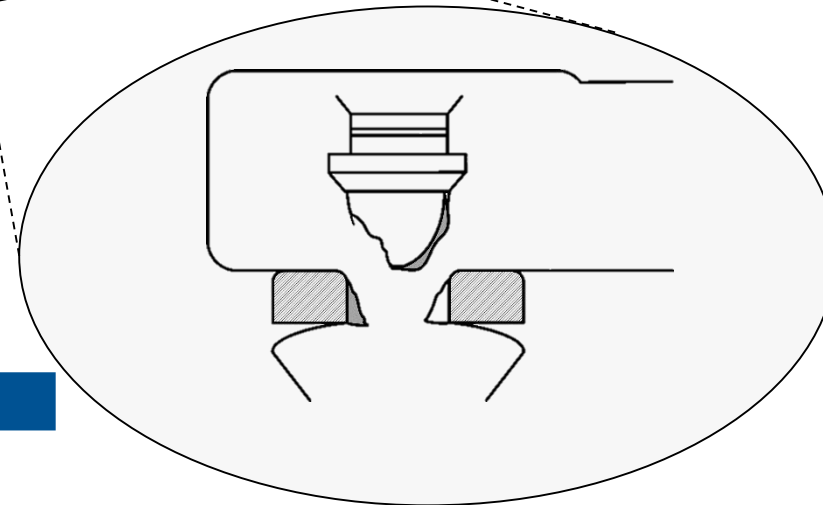


## Ventilfehler führen zu Anlagenstillständen

- Ungeplante Produktionsausfälle
- Austausch des Ventils

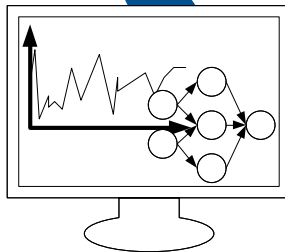
## Typische Ventilfehler

- Verschlissen
- Fouling/Blockage



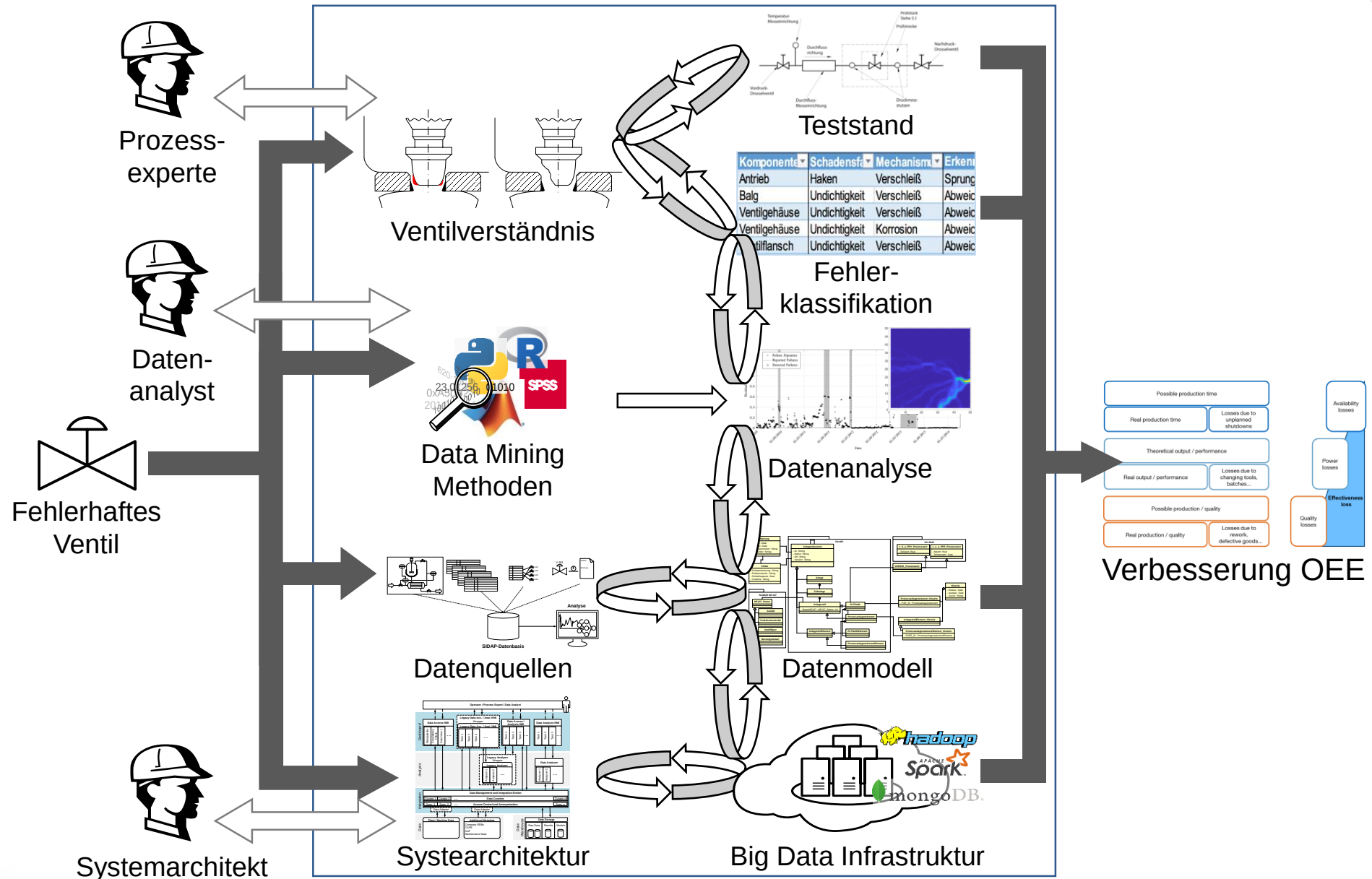
## Datenanalyse

- Verhindern von Produktionsausfällen
- Designempfehlungen
- Ventilverbesserungen

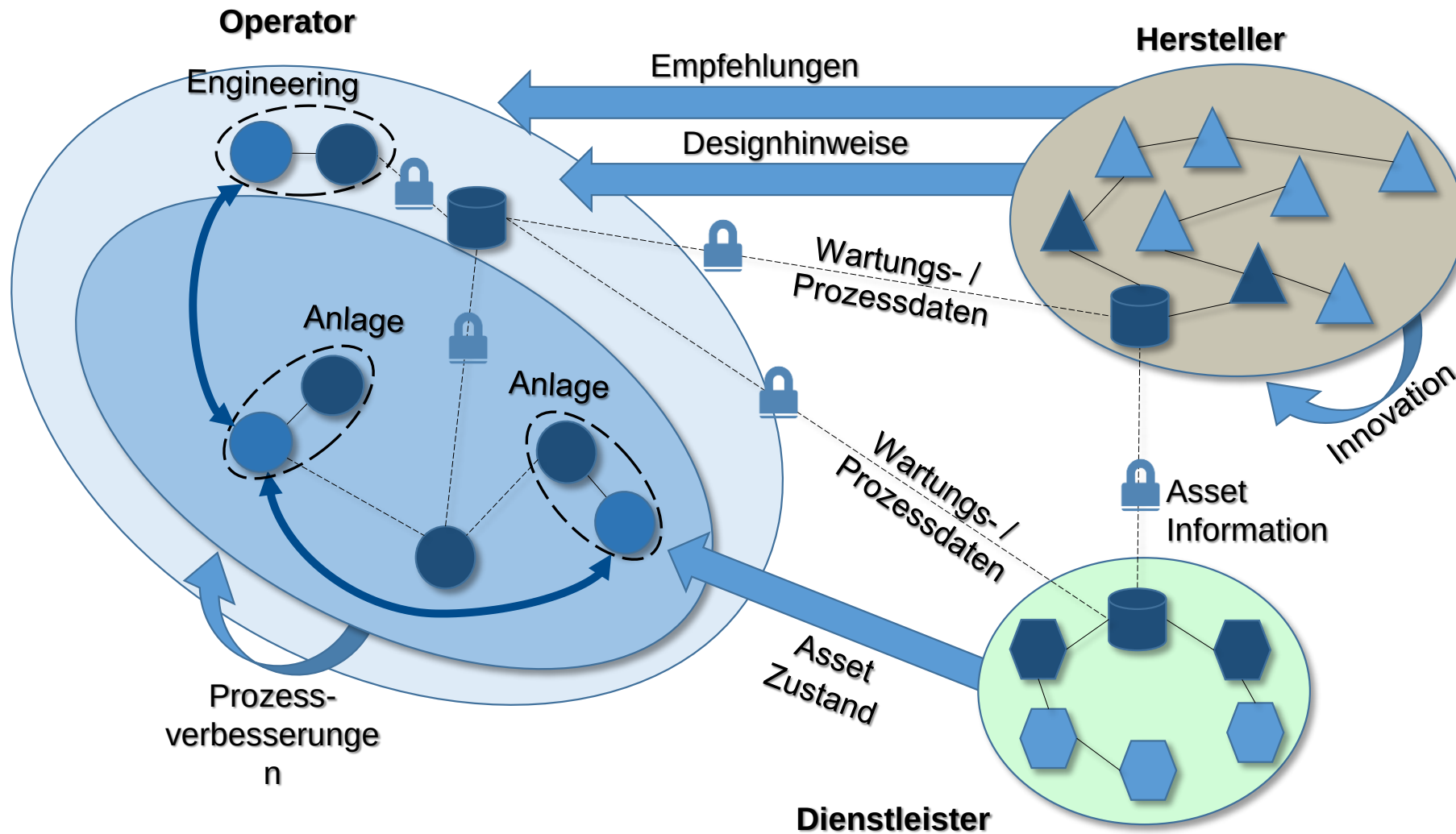




# Smart Data Prinzipien zur Analyse von Regelventilen

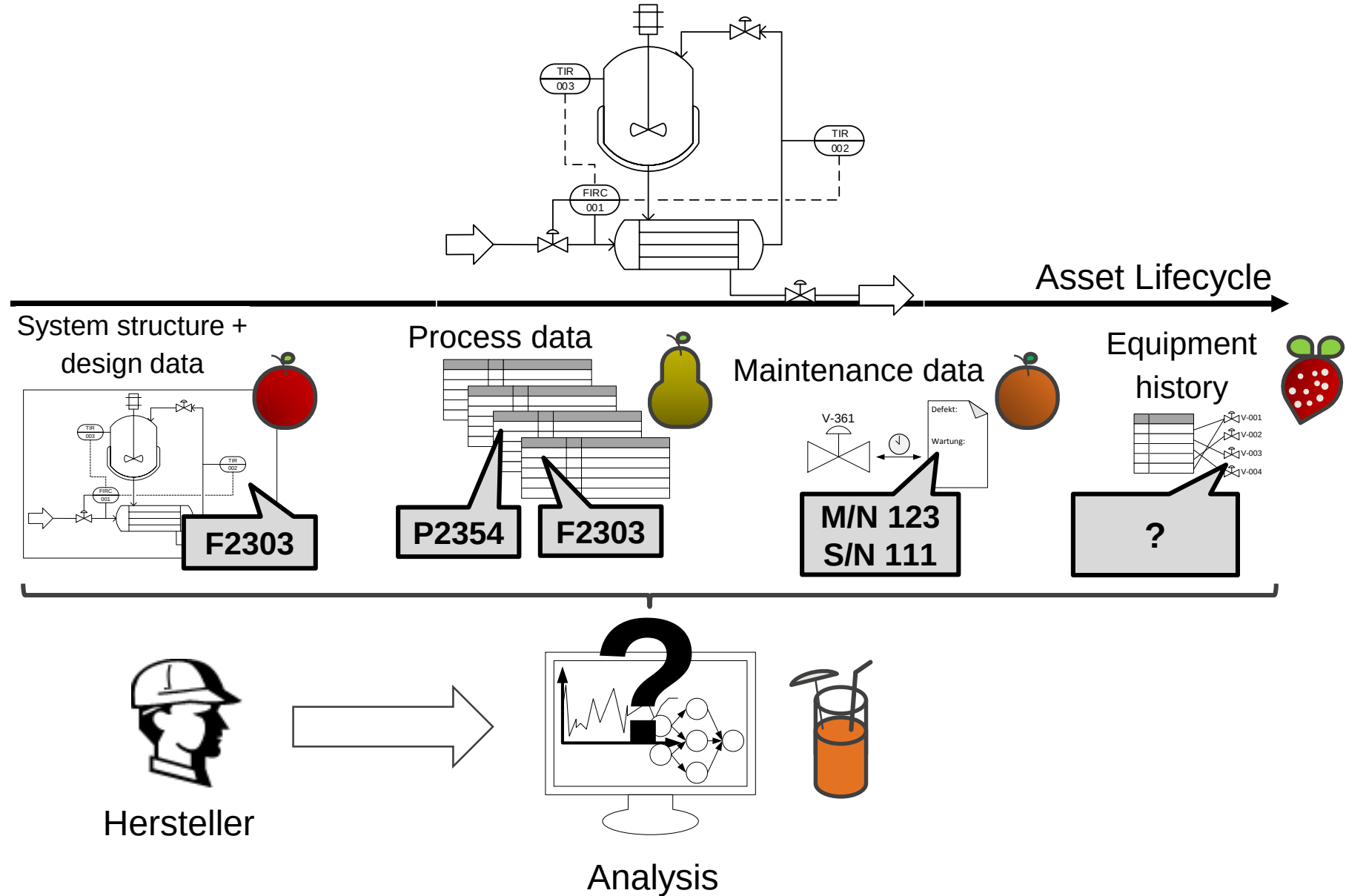


# Innovationspotentiale: Detailliertes Prozess- und Anlagenwissen bündeln

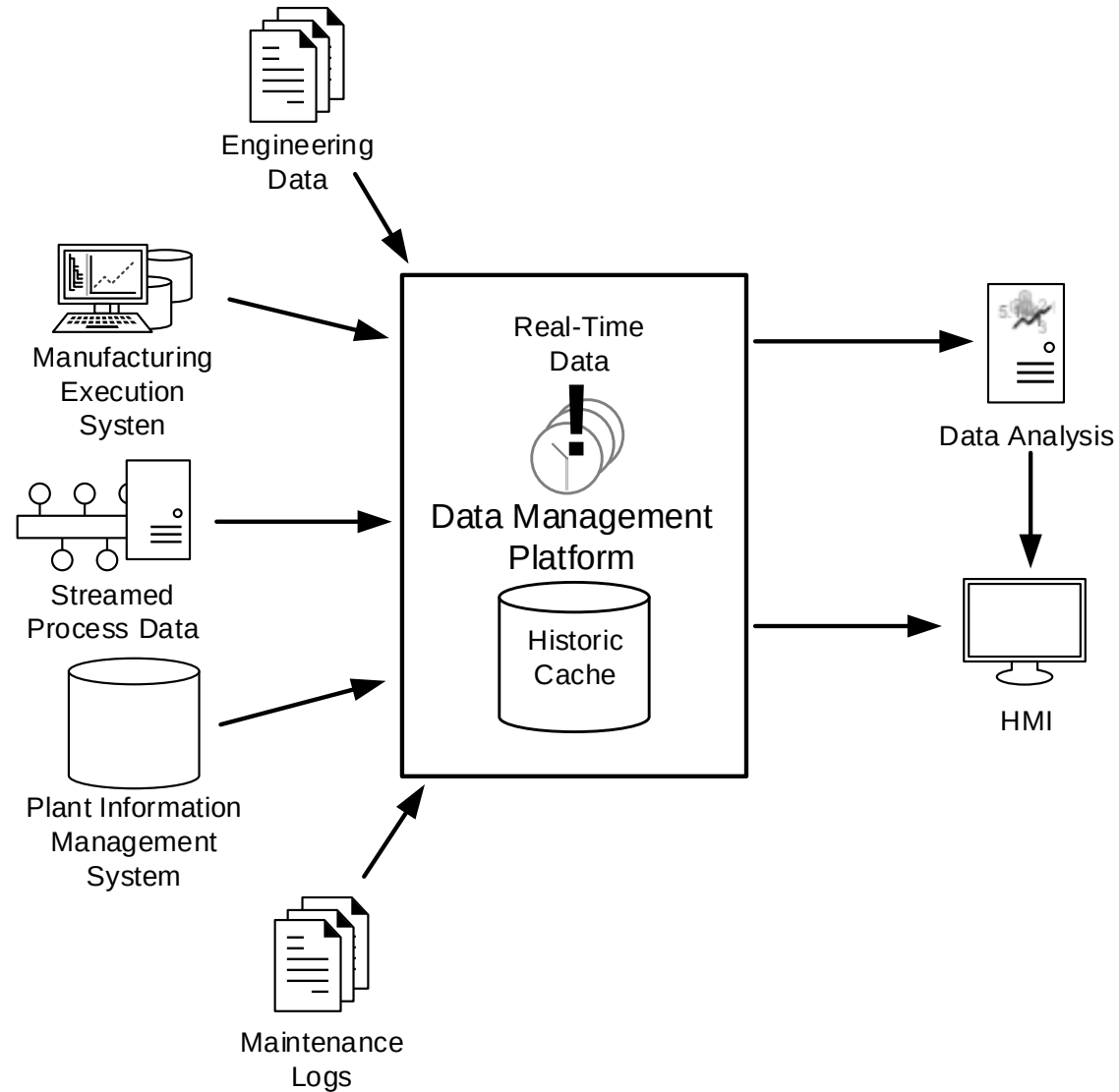




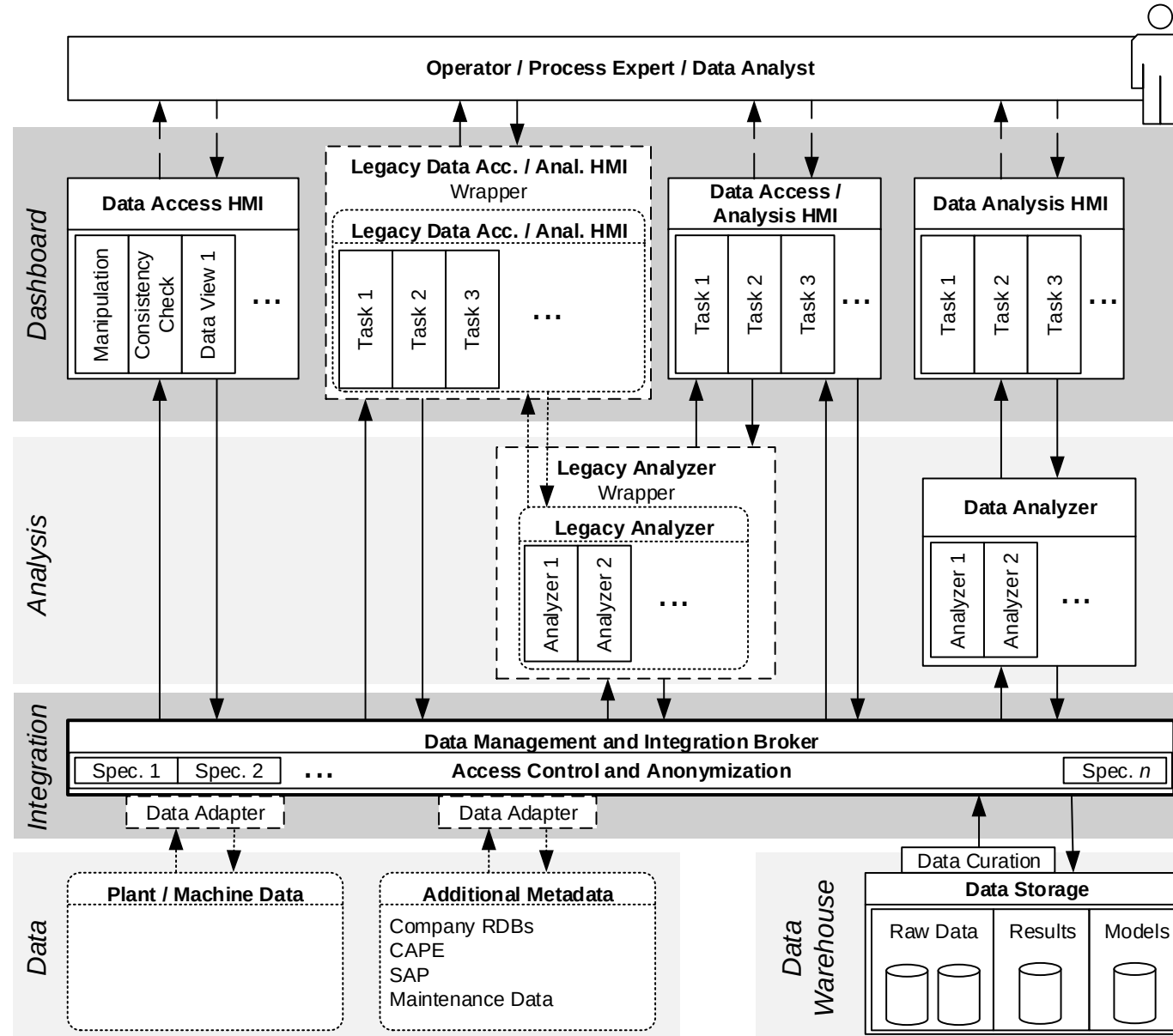
# Mehrwert durch Bündelung der Datenquellen



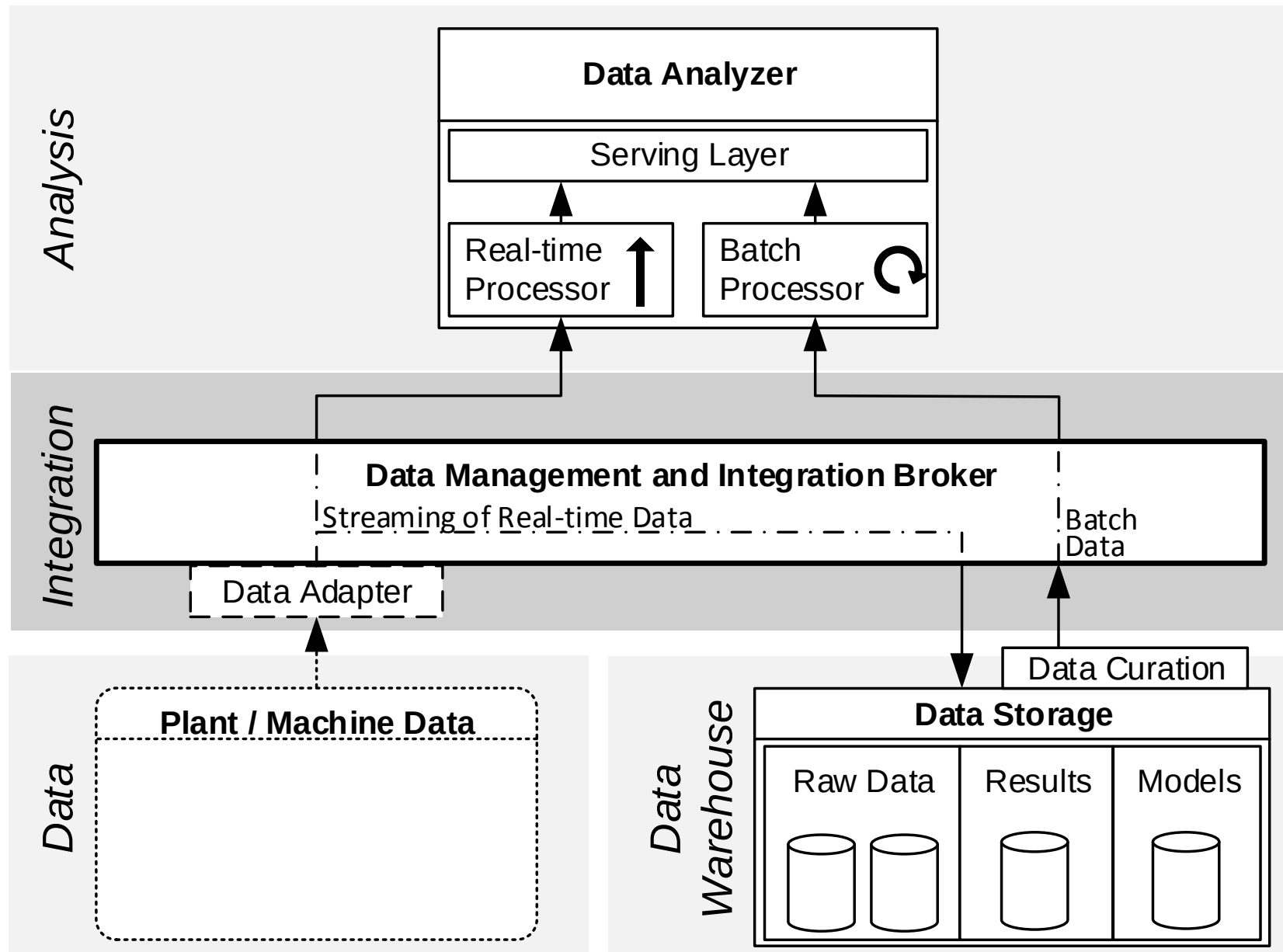
# Übersicht Gesamtarchitektur und Datenquellen



# Systemarchitektur / Datenquellen

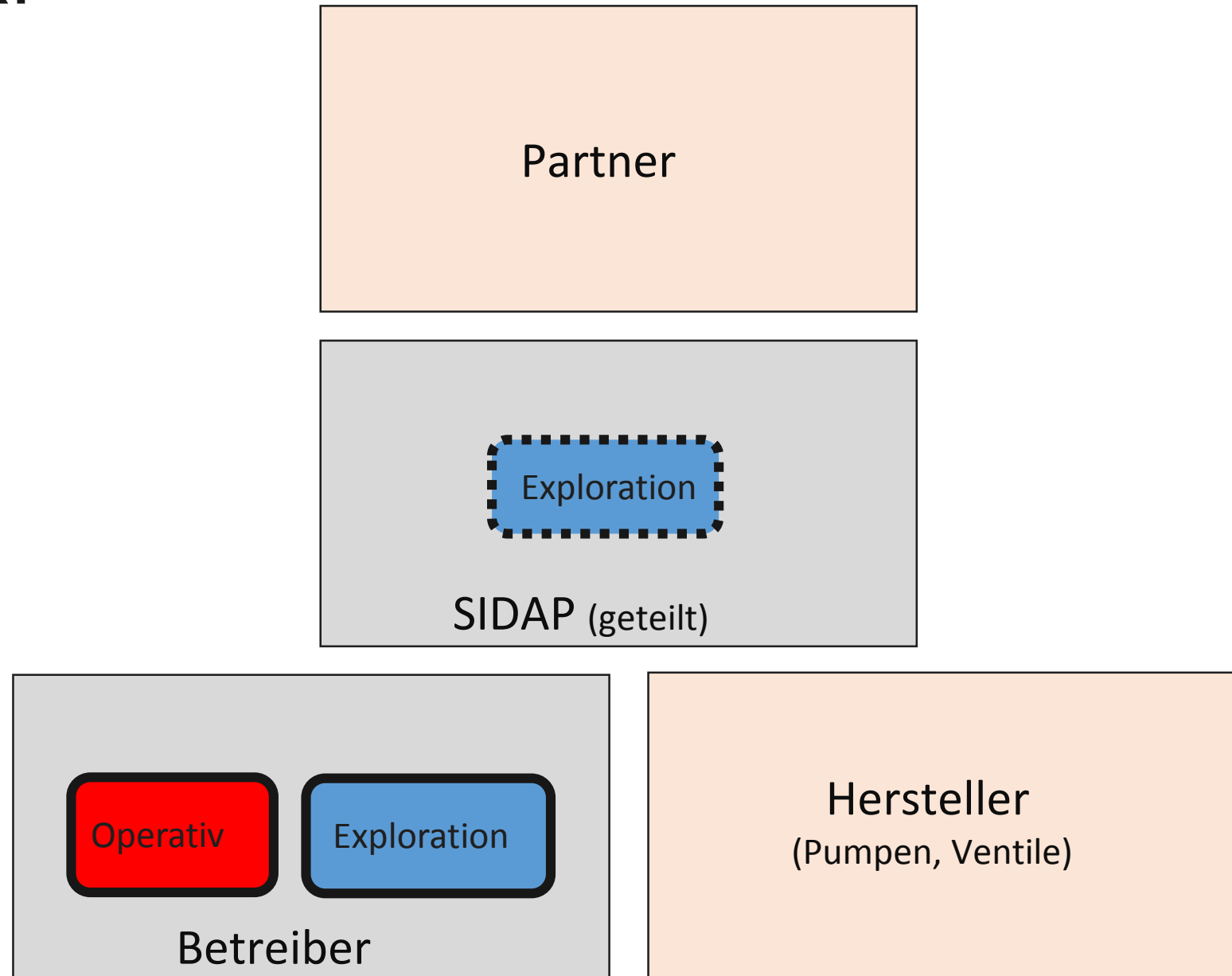


Emanuel Trunzer, Iris Kirchen, Jens Folmer, Gennadiy Koltun, Birgit Vogel-Heuser: A Flexible Architecture for Data Mining from Heterogeneous Data Sources in Automated Production Systems, Proceedings of the 2017 IEEE International Conference on Industrial Technology (ICIT), 2017



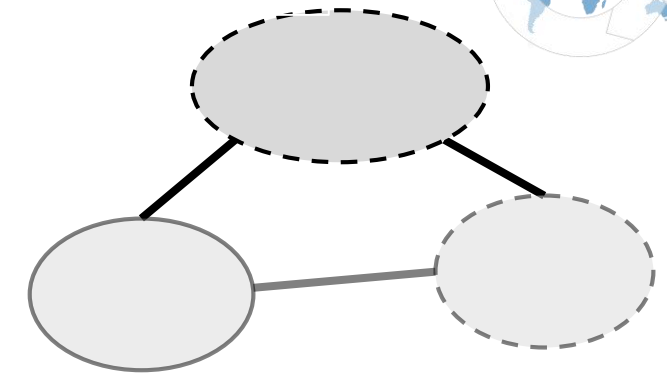
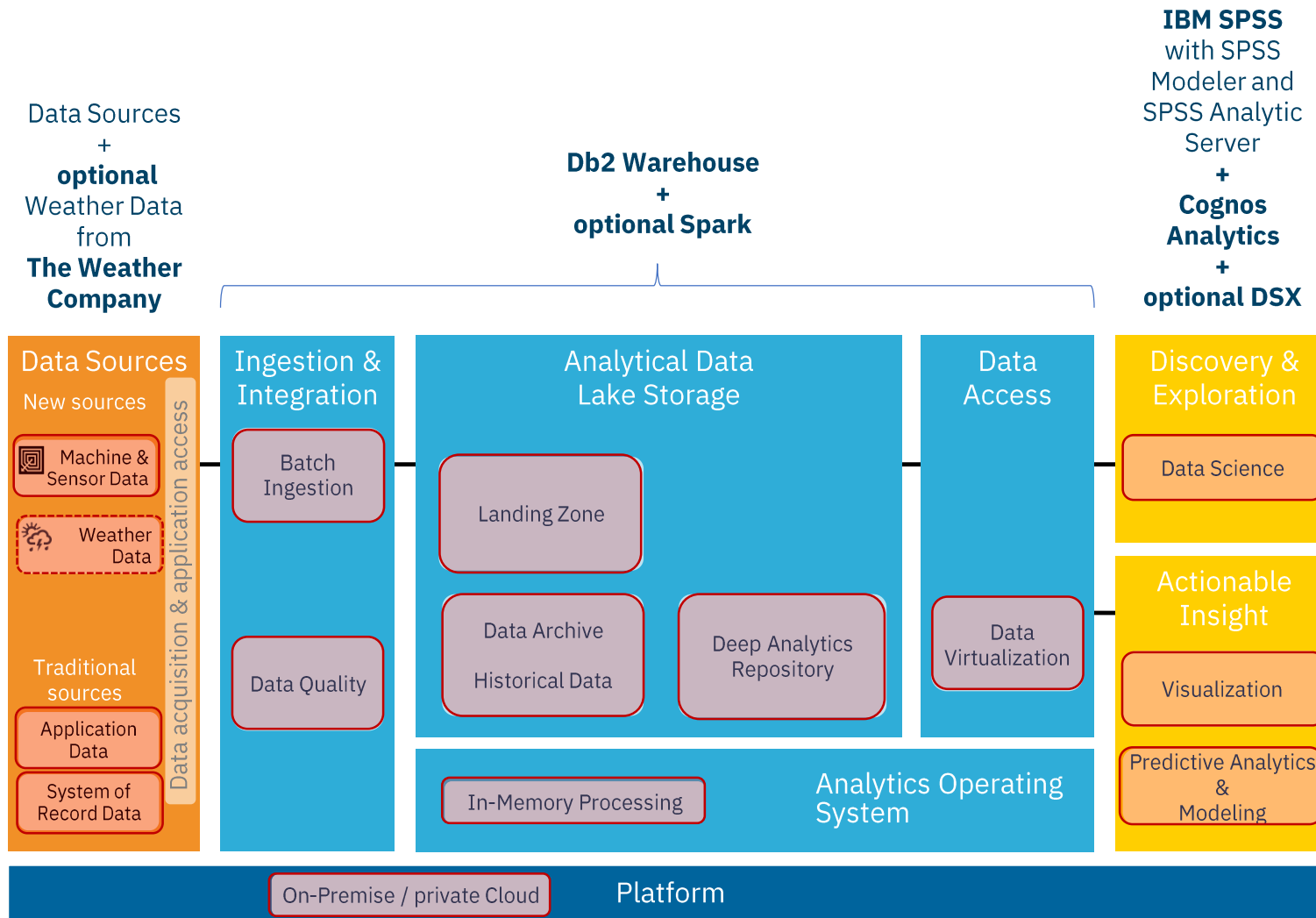
# SYSTEM CONTEXT

## 3 UMGEBUNGEN



# Architectural Overview

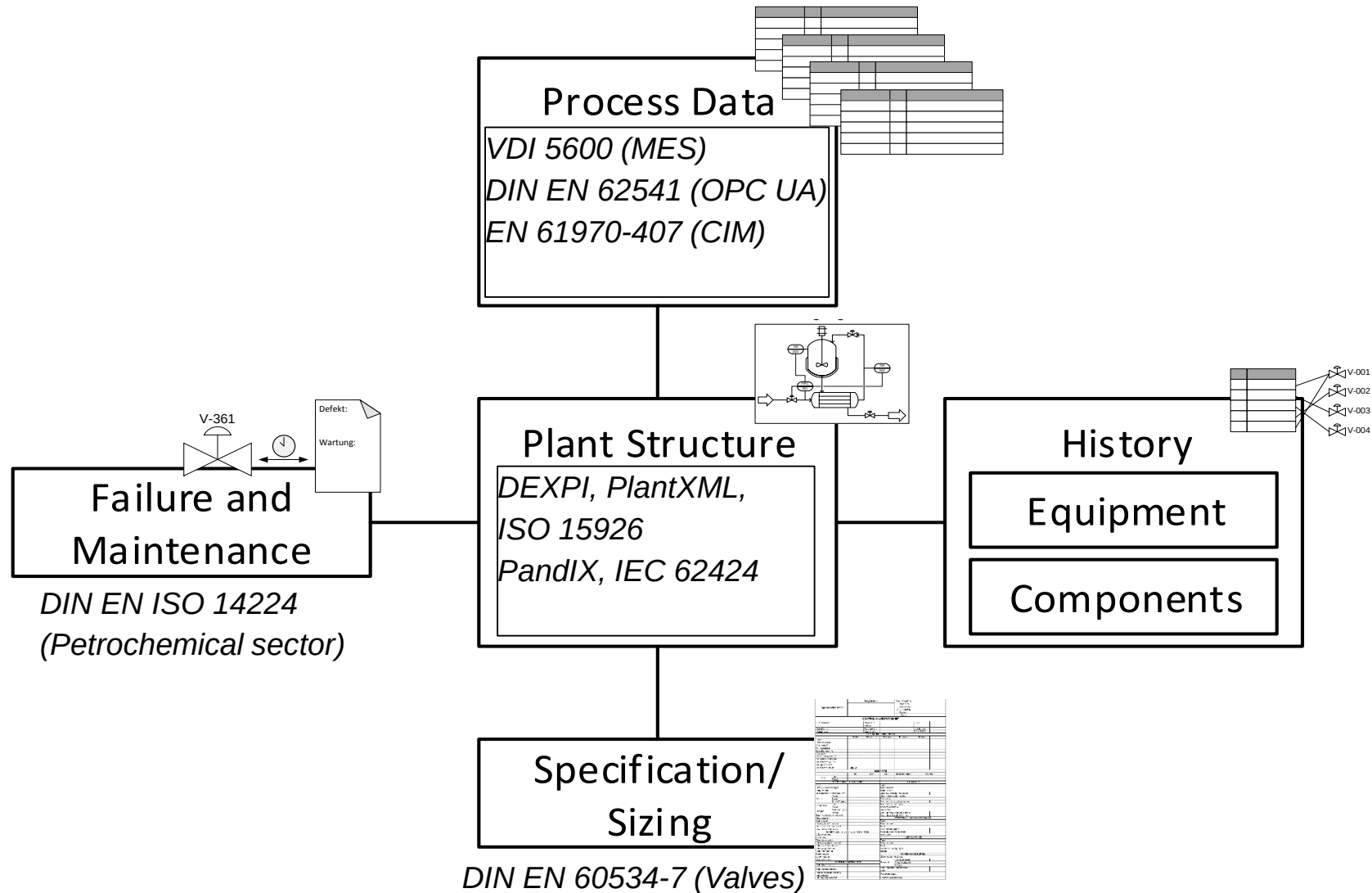
## 3 UMGEBUNGEN– SIDAP-Cloud / shared environment for exploration



- Batch ingestion of anonymized data using loadfiles with predefined XML format
- Storing of data in relational tables (Db2)
- Explore and analyze with IBM Cognos and SPSS
- Generating of predictive models from anonymized data sets
- Optional enhancements with Spark analytics engine and Data Science Experience (DSX) platform for exploration
- Private on-premise / private Cloud installation shared by all members of consortium



# SIDAP Datenmodell – Teilmodelle und Standards



# TEXT ANALYTICS , COGNITIVE

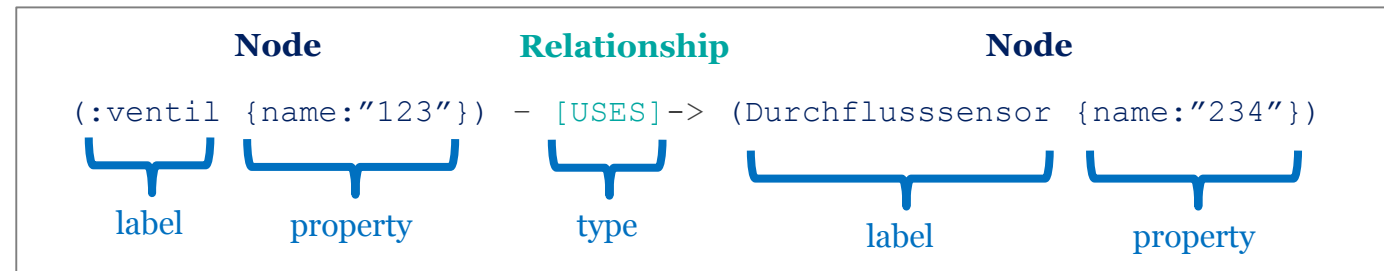
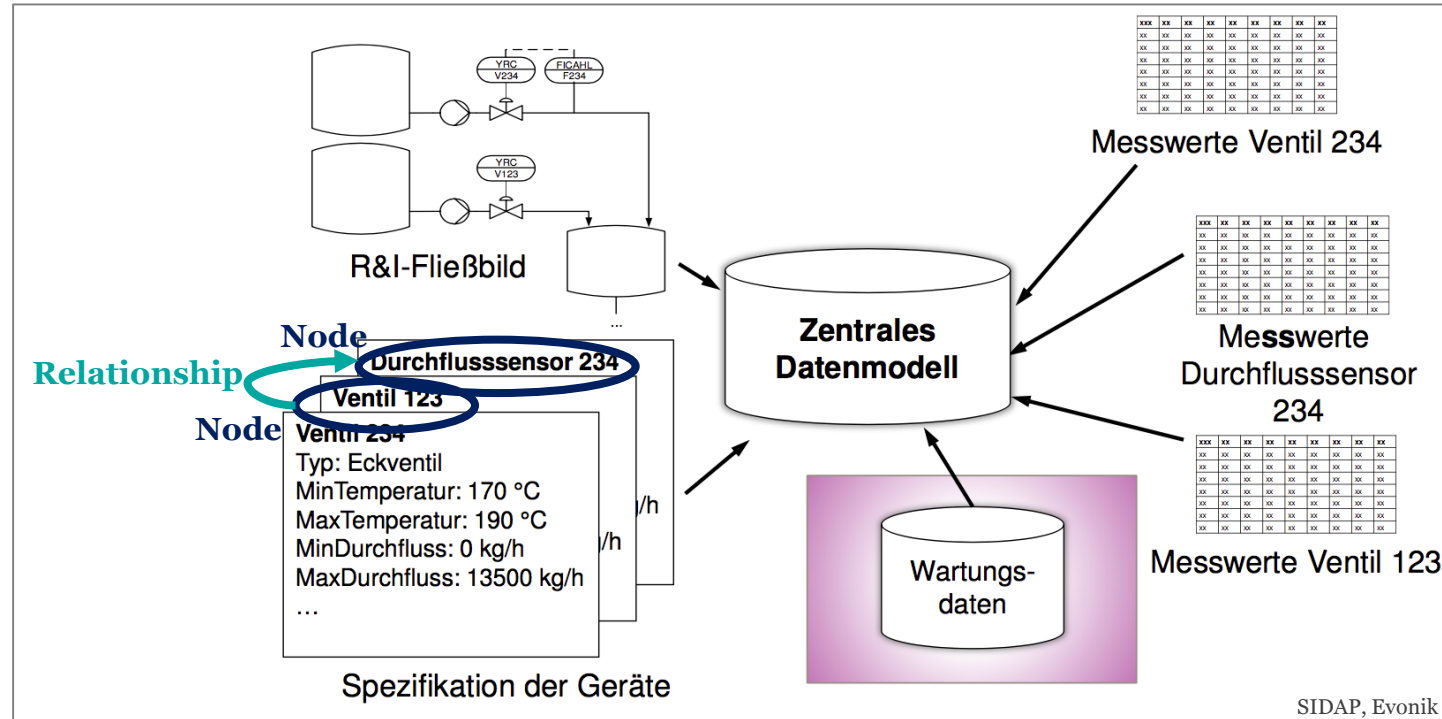
## WOODSIDE: Saving and sharing tribal knowledge with Watson



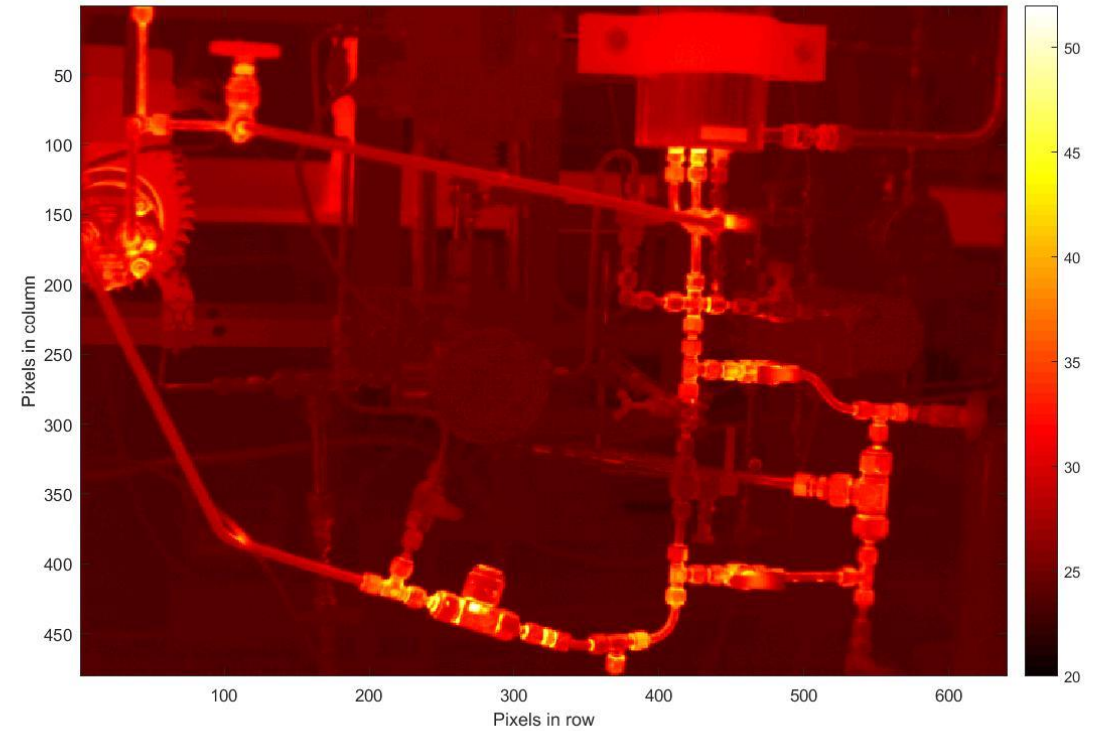
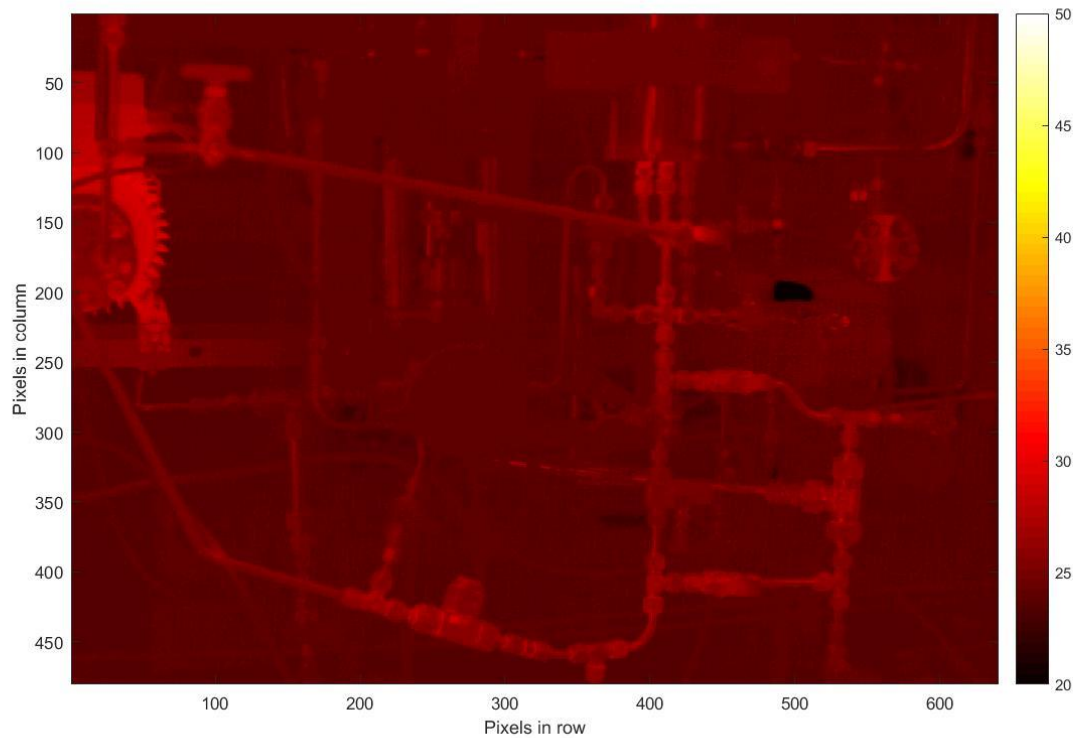
<https://www.ibm.com/blogs/cloud-computing/2017/04/woodside-energy-watson/>

# GRAPHEN-DATENBANKEN

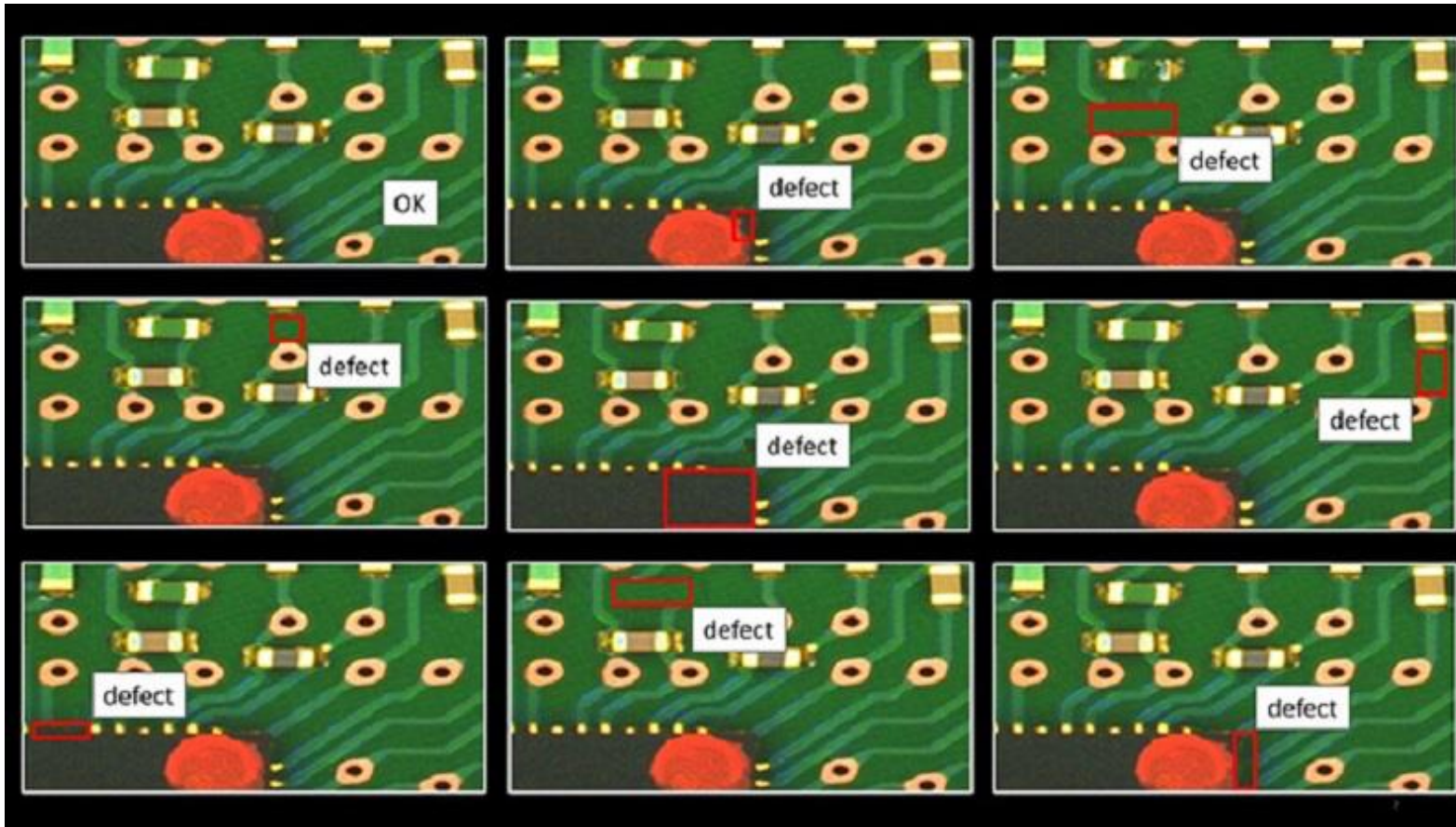
## FLEXIBLERES DATENMODELL



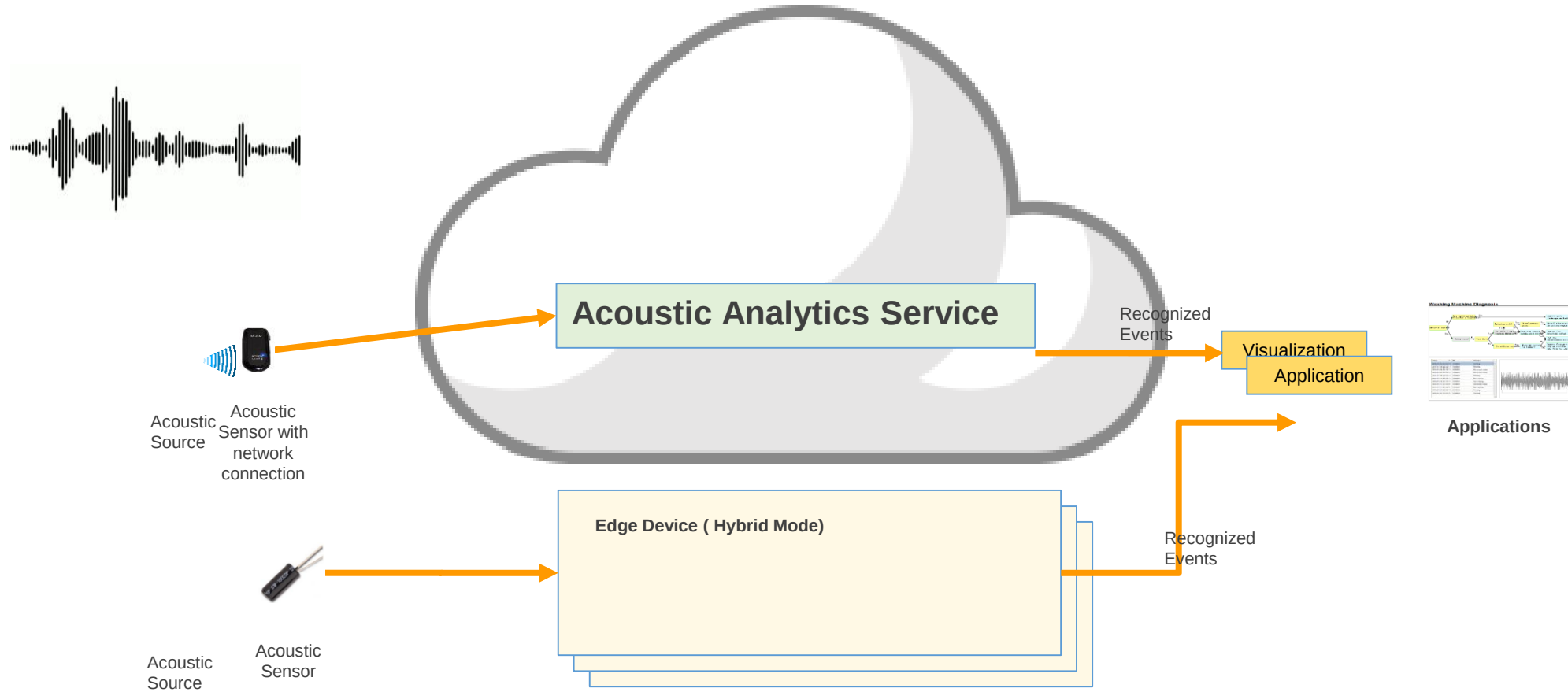
## Analyse thermographischer Bildinformationen zur Leckageerkennung





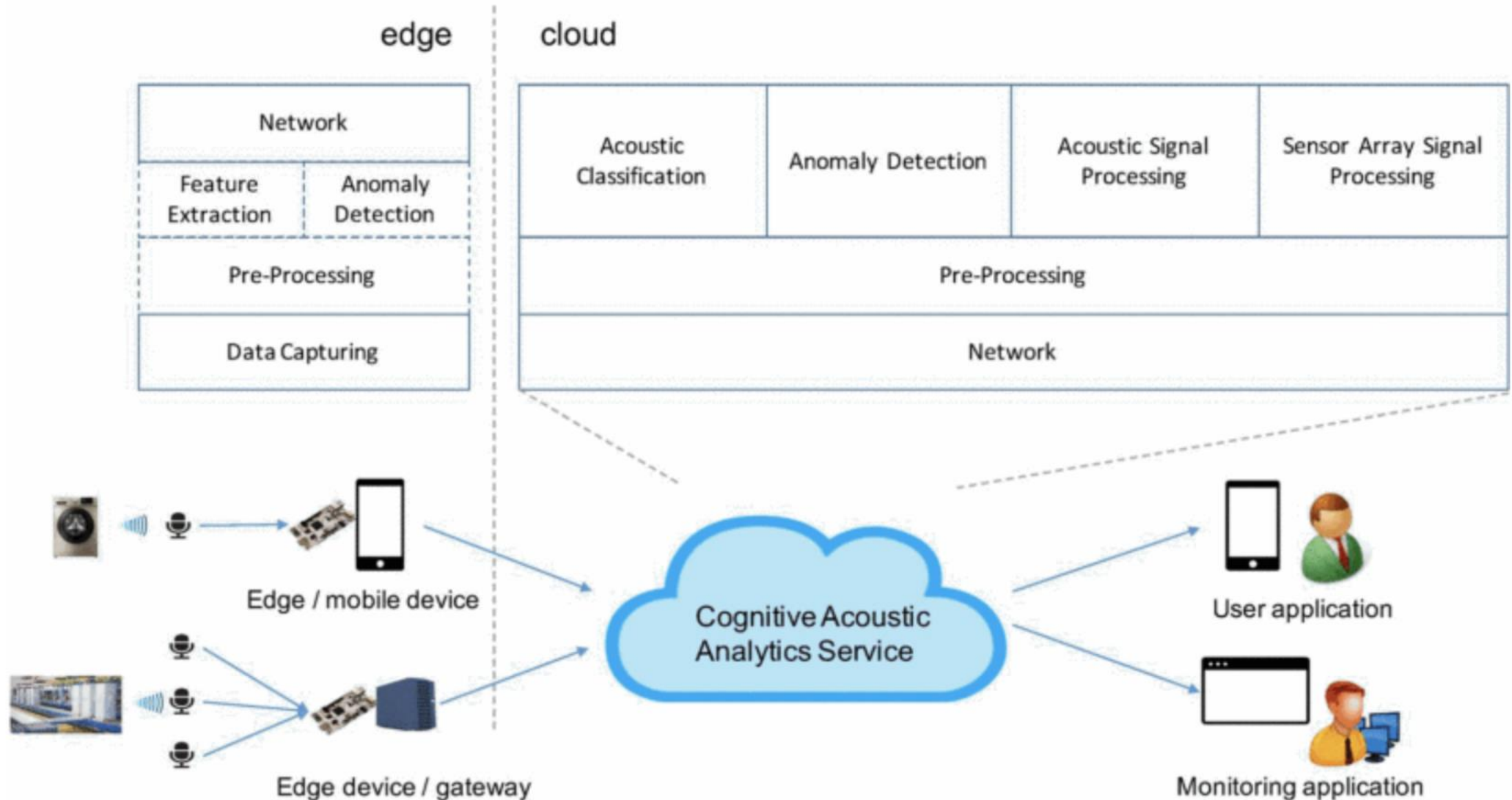


<https://www.ibm.com/bs-en/marketplace/visual-inspection-for-quality>



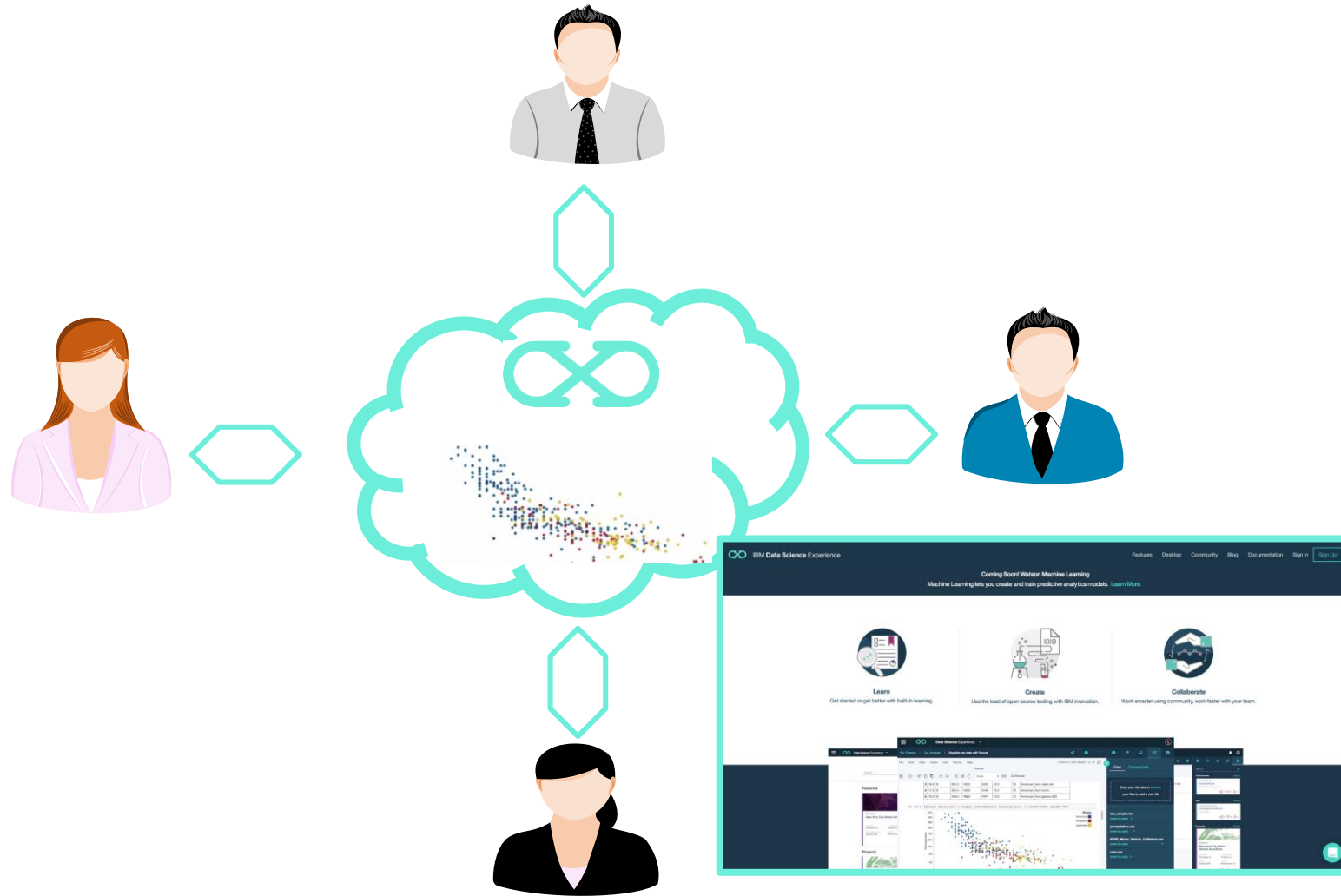
draft

# Cognitive Acoustic Analytics Service for Internet of Things



# COLLABORATION / DATA SCIENCE

## JUPYTER/IBM Data Science Experience





## Environments



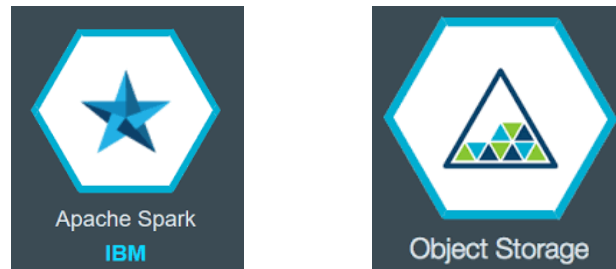
- Integration of RStudio IDE
- Shiny apps
- Work with Jupyter Notebooks

## Coding



- Code in Scala/Python/R/SQL
- Add libraries

## Powered by



- Fully managed Spark environment
- 5GB Object Storage

## Collaborate



- Projects, Users, Roles
- Share notebooks & results
- Versioning & history

<http://datascience.ibm.com/>

# Data Science Experience

- **Data Science as a Team Sport**  
Lets *data scientists/engineers, analysts, stakeholders collaborate* to collect, share, explore, *analyze data* in order to *derive insights and train models, and share or deploy resulting assets*
- **Projects** - collaborate as team or work individually
- **Jupyter Notebooks + IBM value add**
  - Integrated in Projects with access control
  - *Spark integration with R/Python/Scala kernels*
  - *Versions, comments, share link, publish to GitHub*
  - *PixieDust, Brunel, ...*
- **Machine Learning integrated in Projects:**  
*Use ML Wizard and Flows to train Models*
- **RStudio integrated with Spark**
- **DSX Integrates with Data in many places**
  - Object Storage (SWIFT now, new Cloud Object Storage soon)
  - Watson Data Platform Services and WDP Catalog
  - Message Hub and IBM Streaming Analytics
  - Can call any IBM service, e.g. Watson, Quantum, etc
  - Third party data services on premise or on other clouds
- **Built on the IBM Cloud platform**

 IBM Data Science Experience

Features Blog Documentation Sign In [Sign Up](#)

## Master the art of data science

Solve your toughest data challenges with the best tools and the latest expertise in a social environment built by data scientists.

 **Watch Now (2:31)**

[Sign Up For a Free Trial](#)

**Coming Soon! Watson Machine Learning**  
Machine Learning lets you create and train predictive analytics models. [Learn More](#)



### Learn

Get started or get better with built-in learning.



### Create

Use the best of open source tooling with IBM innovation.

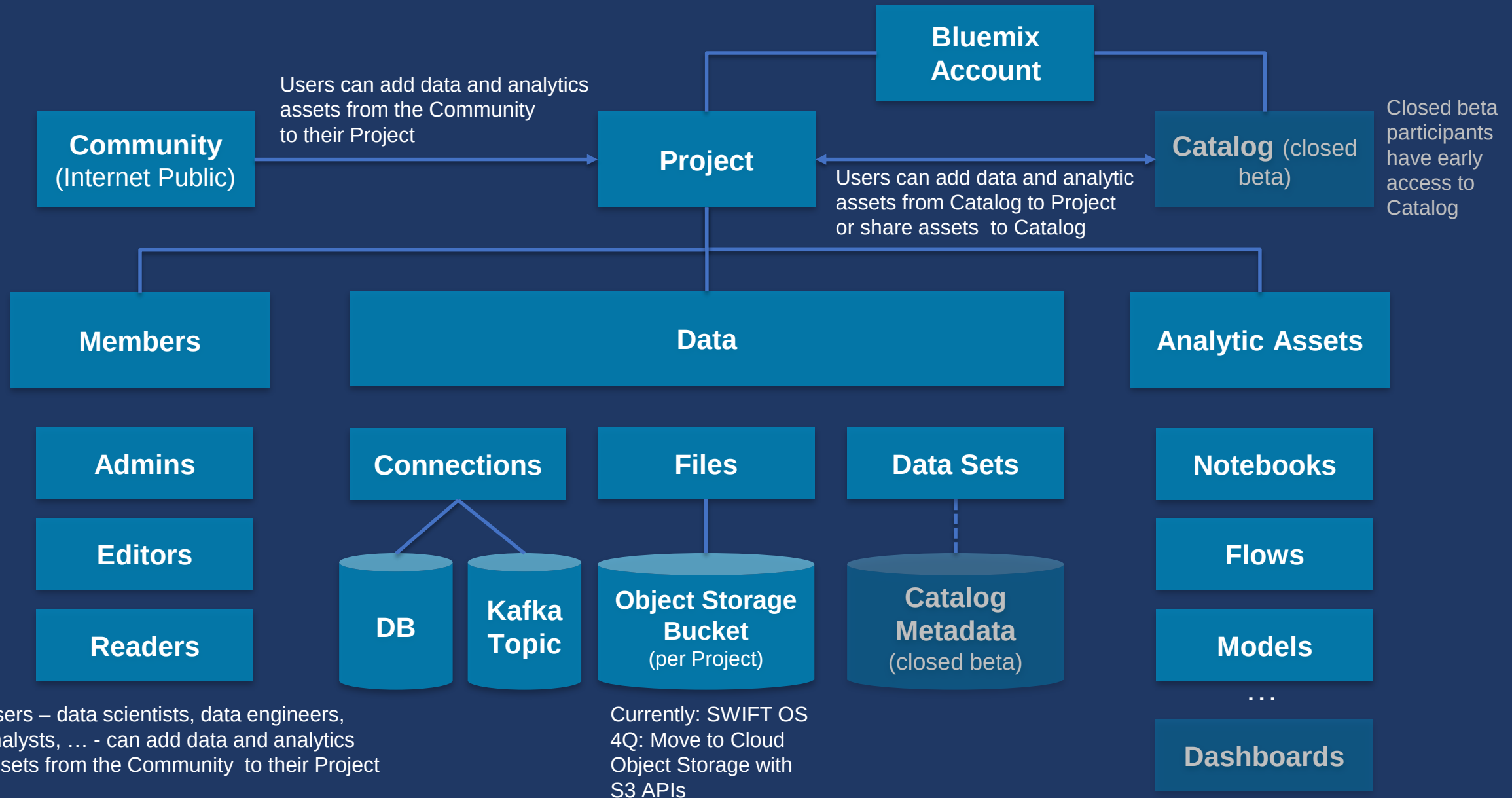


### Collaborate

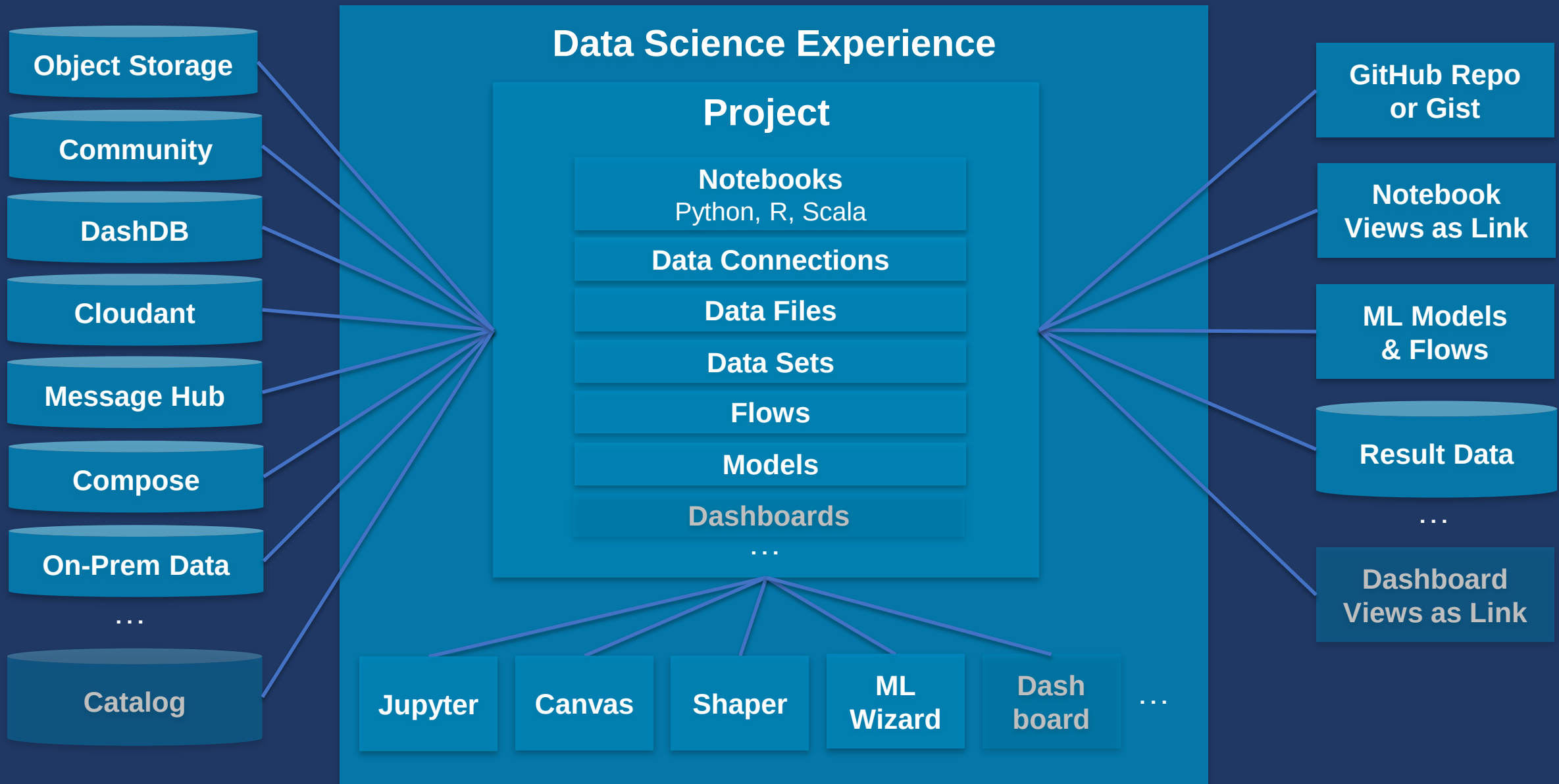
Work smarter using community, work faster with your team.

Try it yourself at <https://datascience.ibm.com>

# Projects allow users to work and collaborate

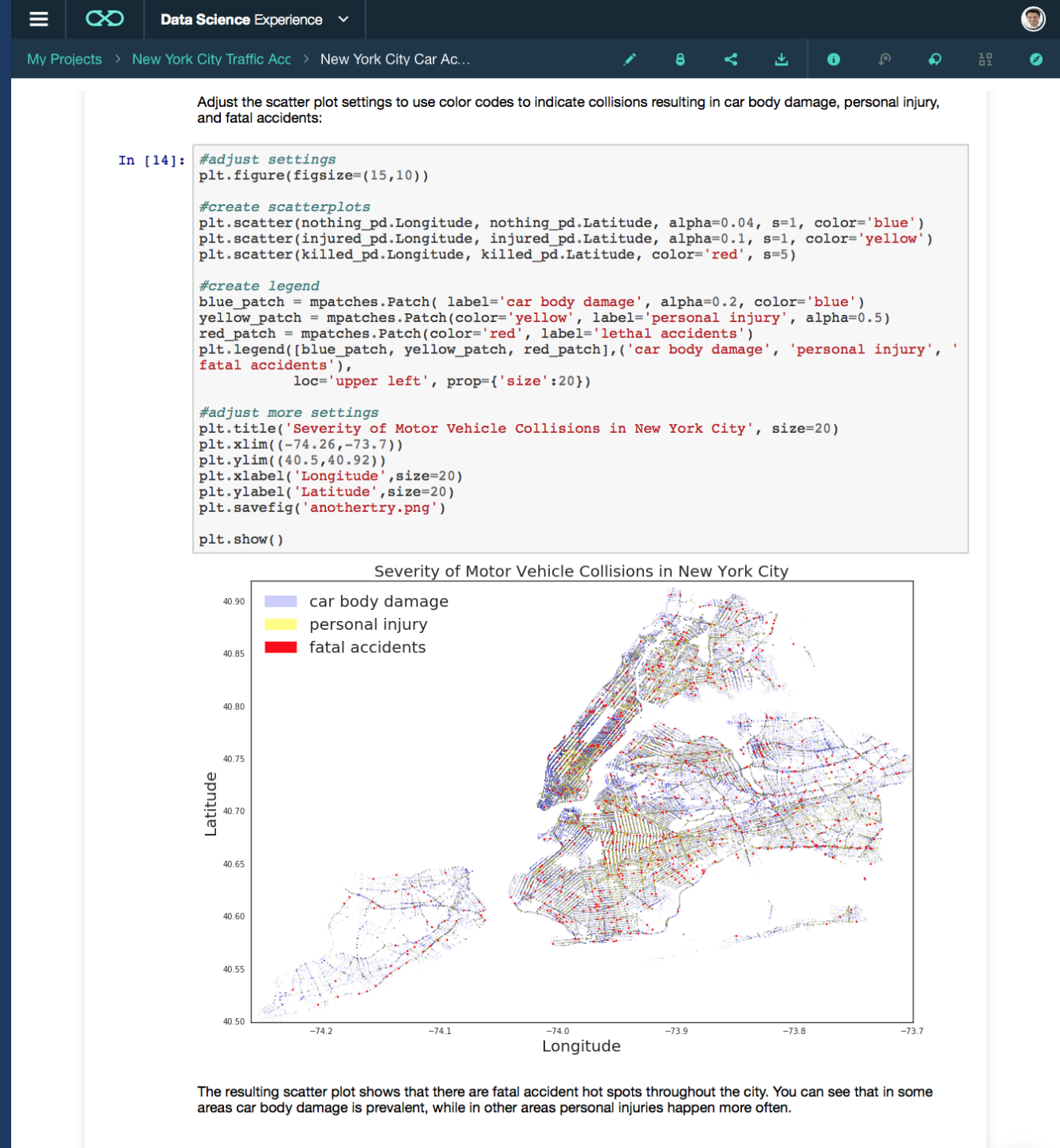


# DSX - Analyze Data from many Sources to generate Insights



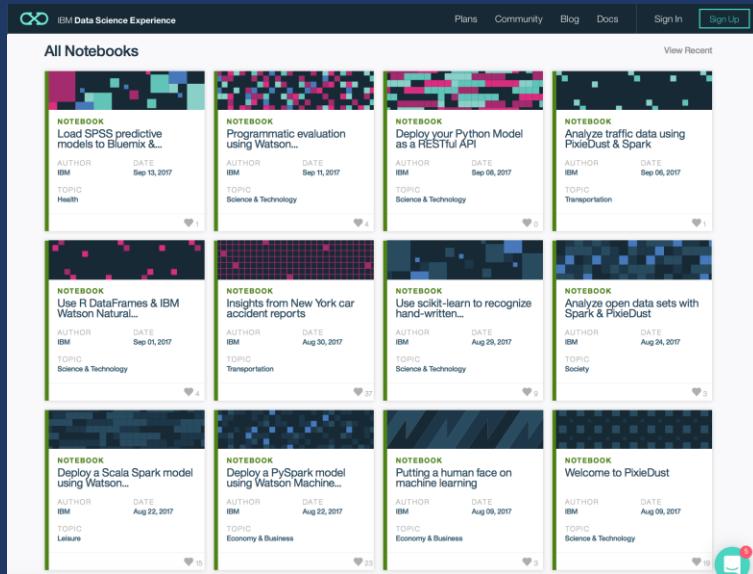
# Jupyter Notebooks

- **Interactive Documents** that contain insights and the code to achieve them
  - Text to explain what the Notebook does and document insights
  - Code cells for accessing, processing, and analyzing data
  - Result cells with output and visualizations
- **Jupyter** <http://jupyter.org> is the most popular open source project and de facto standard
  - Allows to create, run, edit notebooks in web browsers
  - Can run with Python, R, Scala, and many other kernels
  - Notebooks are stored as .ipynb JSON objects
- **Not only for Data Science**, also useful for other applications that benefit from text + code + insights + viz in interactive docs
- **Popular for publishing interactive tutorials & demos** with runnable code inside, and reproducible research



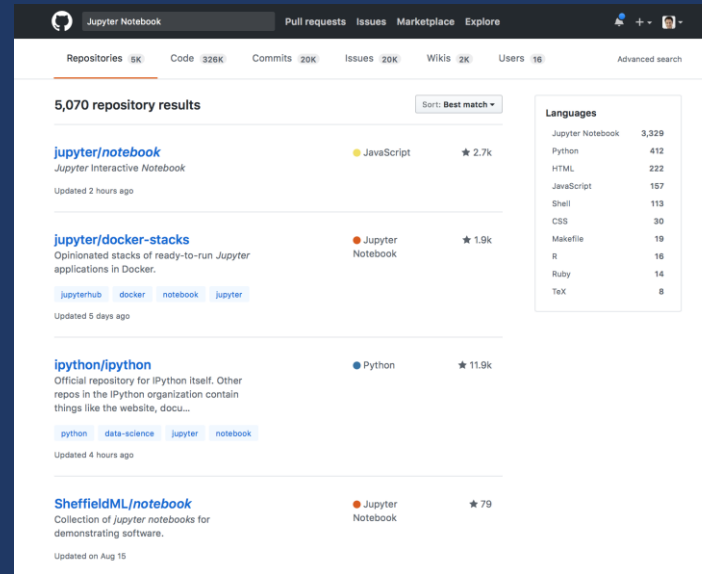


# Jupyter Notebooks – Where to find Examples



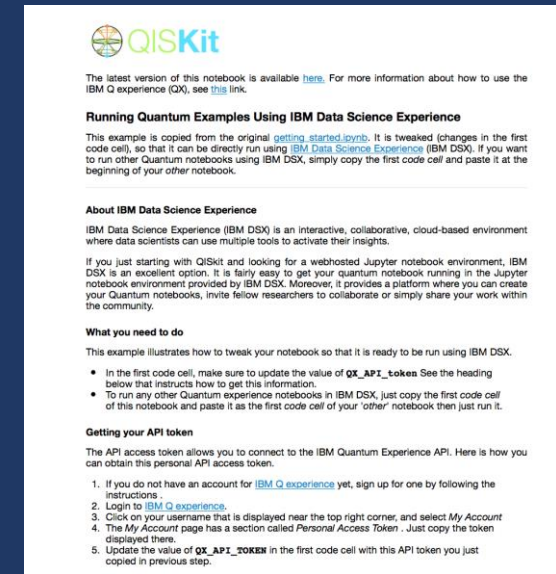
<https://datascience.ibm.com/community> --> Notebooks

Curated example notebooks and tutorials in the Data Science Experience Community



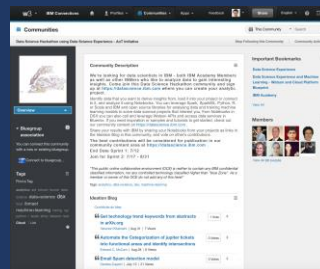
<https://github.com/search?utf8=✓&q=Jupyter+Notebook&type=Repositories>

Hundreds of thousands of Jupyter Notebooks on GitHub



[https://github.com/Qiskit/qiskit-tutorial/blob/master/1\\_introduction/running\\_on\\_IBM\\_DSX.ipynb?cm\\_mc\\_uid=24574902395214519070223&cm\\_mc\\_sid\\_50200000](https://github.com/Qiskit/qiskit-tutorial/blob/master/1_introduction/running_on_IBM_DSX.ipynb?cm_mc_uid=24574902395214519070223&cm_mc_sid_50200000)

An example of a notebook that can run in a DSX Project and call the IBM quantum computing service



Recent AoT Data Science Hackathon on Data Science Experience  
<https://w3-connections.ibm.com/communities/service/html/communitystart?communityU uid=0c2c13a6-8260-4dc1-bd83-efbf6b534004>