

DH09

AI-supported and Demand-oriented Data Integration of Clinical Studies

HMS Analytical Software

Dr. David Hipp | david.hipp@analytical-software.de

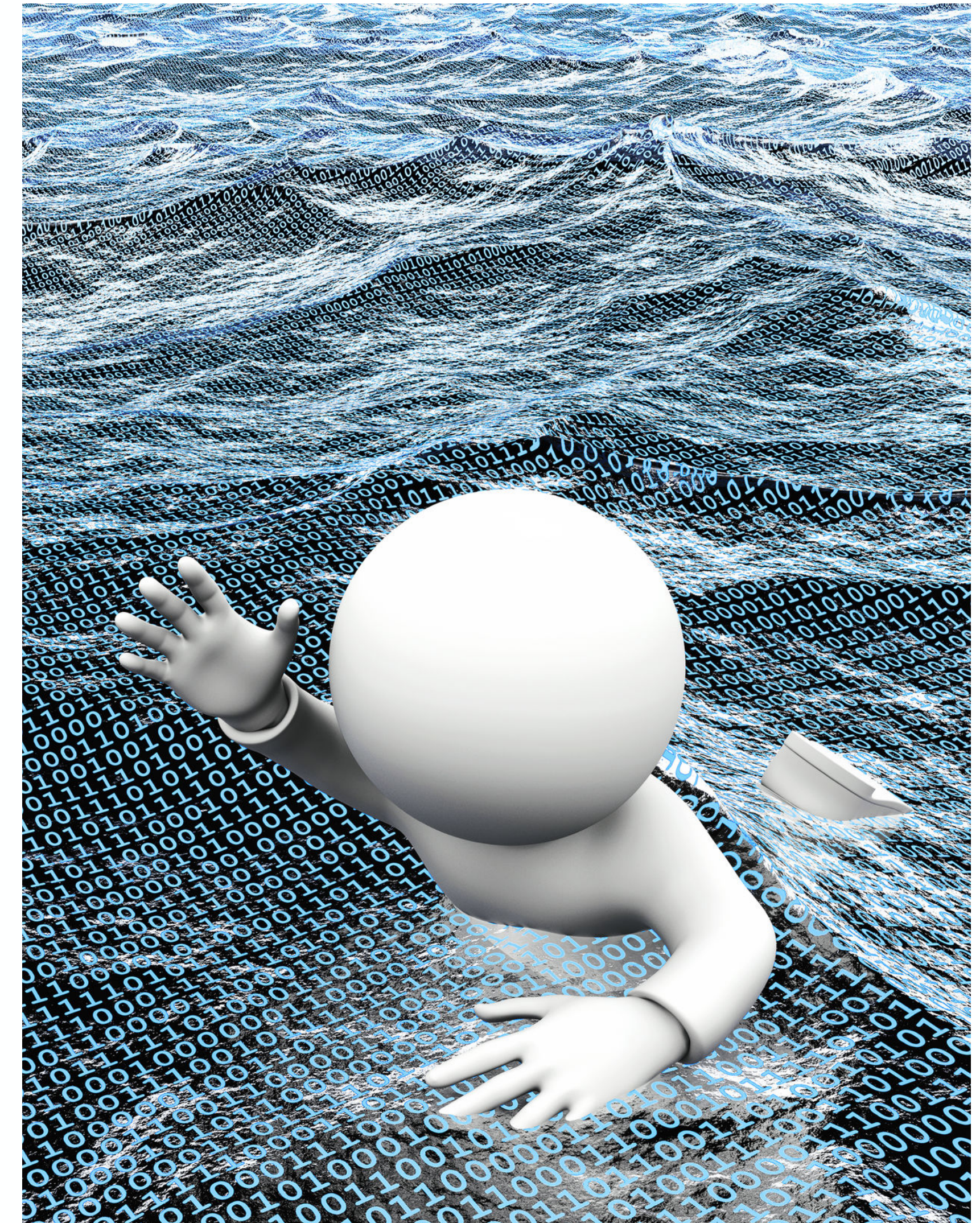
Kai Spielmann | kai.spielmann@analytical-software.de

Delivering the promises of advanced analytics

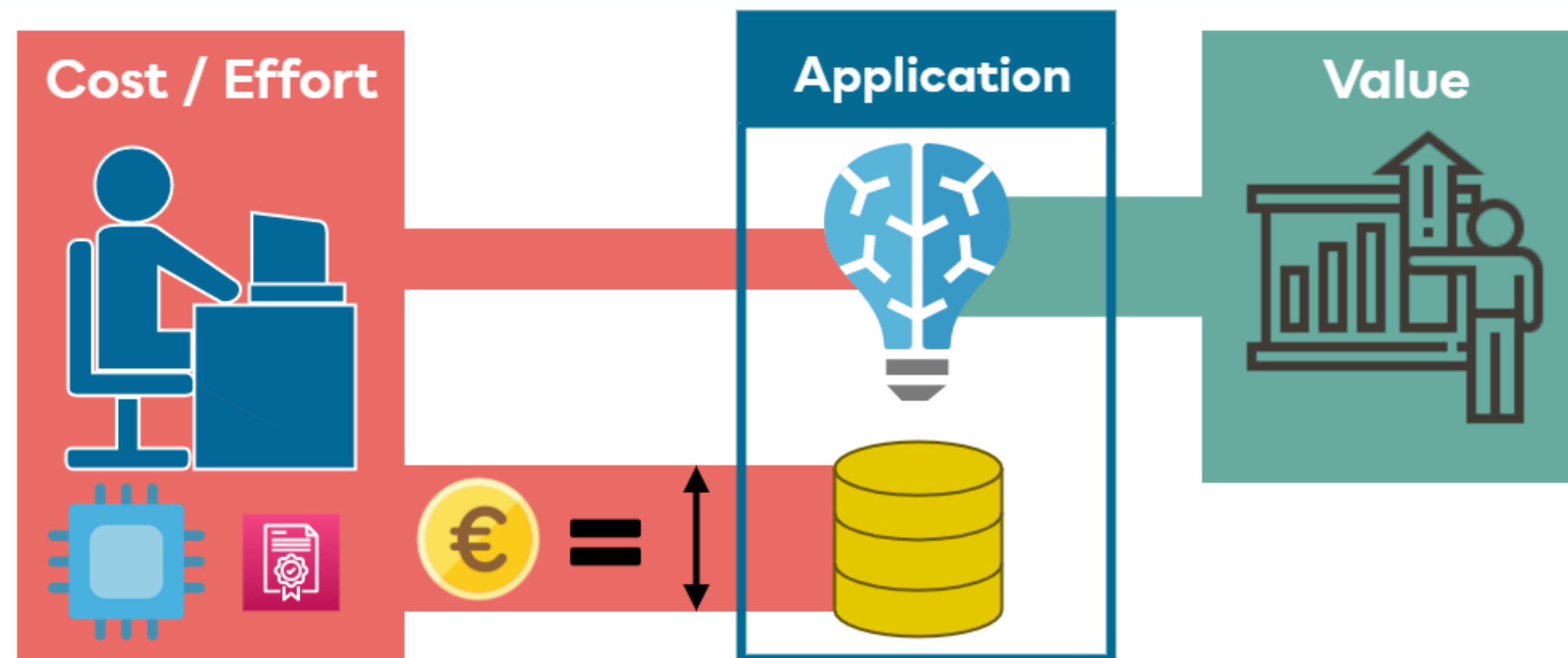
- High demand for advanced analytics for pharma, since many use cases promise considerable value
- Clinical study data of has been collected over decades and is an expensive, high quality asset

For **cross-study analytics** those promises are hard to deliver, since ...

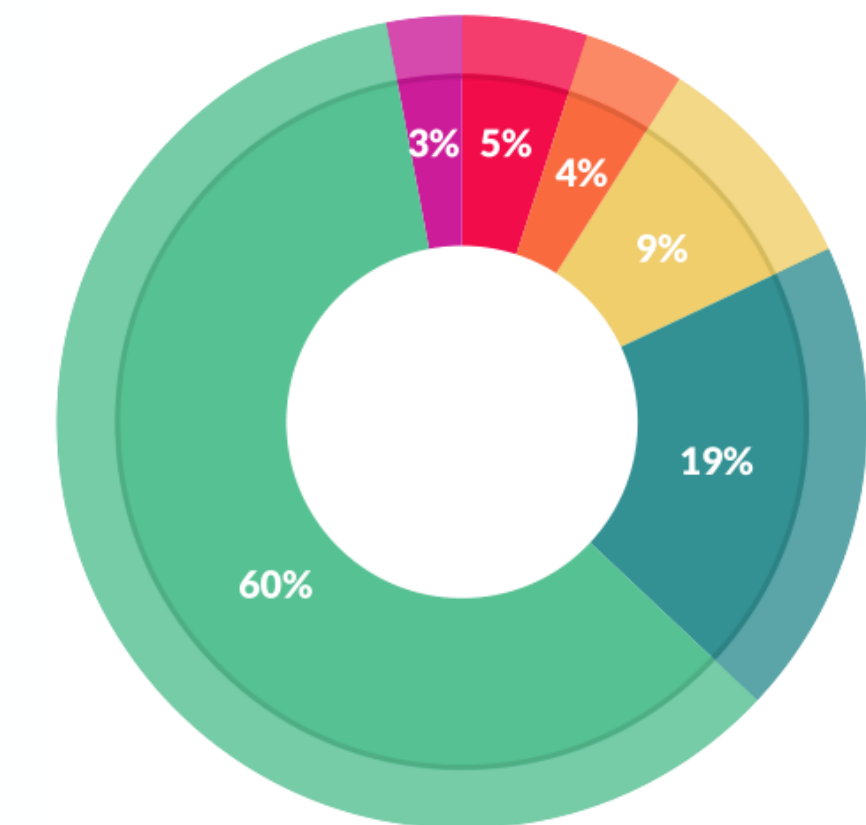
- **Lack of comparability**
Clinical study data is often heterogenous
- **Difficult access**
Legacy trial data is hidden away in data silos
- **Complex task**
Data integration is very time-consuming



Typical cost distribution for single analytics use case



What data scientists spend the most time doing



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



First-mover **dis**advantage

Integration of data prohibitively expensive for single use case

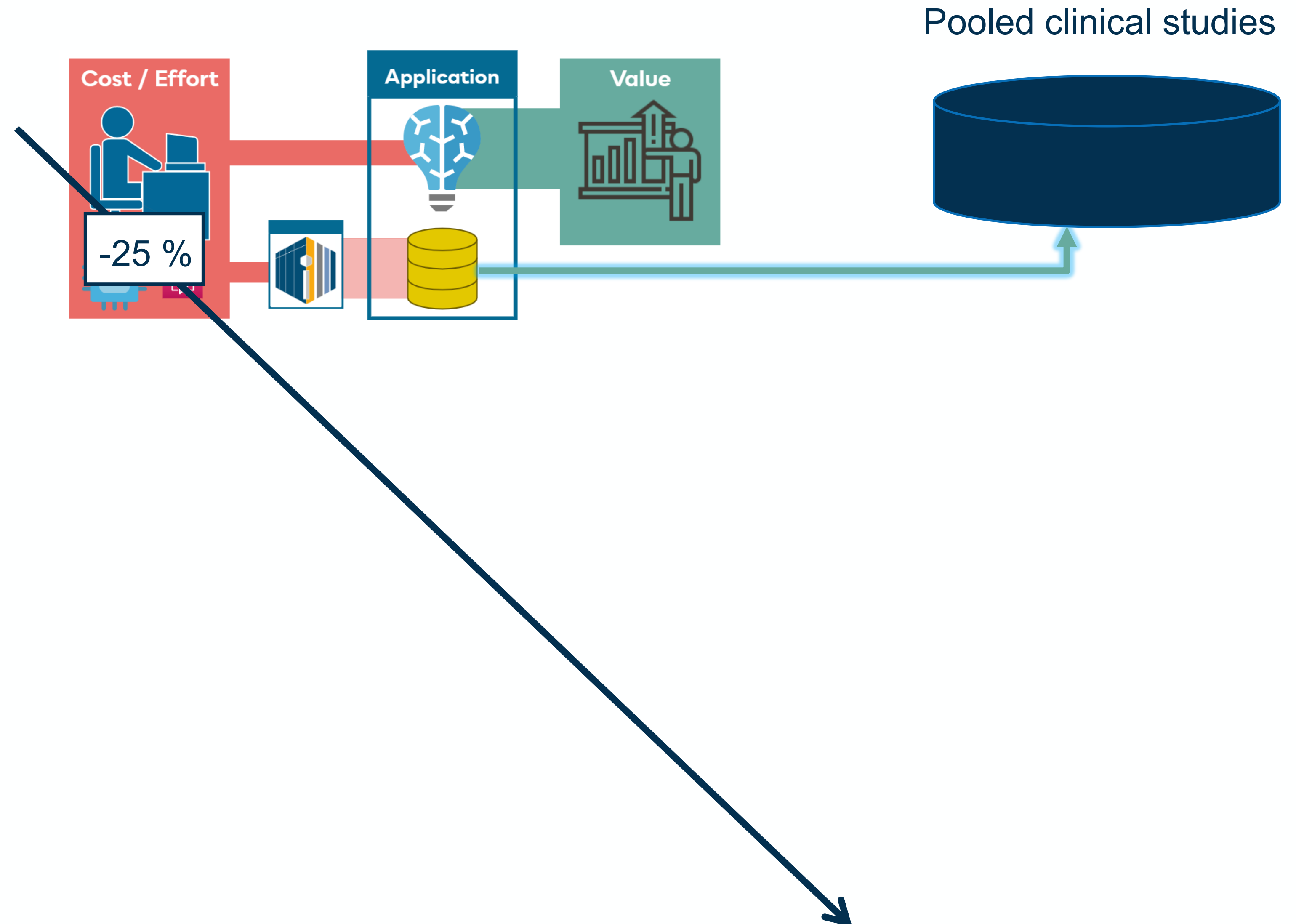
Lowering the cost barrier of entry via automation

- Idea: Use AI to reduce repetitive manual efforts
- HMS calls this concept

IntelligentDI



- Reduce the of costs of data integration
- Beneficial cost-benefit ratio for (smaller) single use cases
- **Scalable approach with repeatable savings**



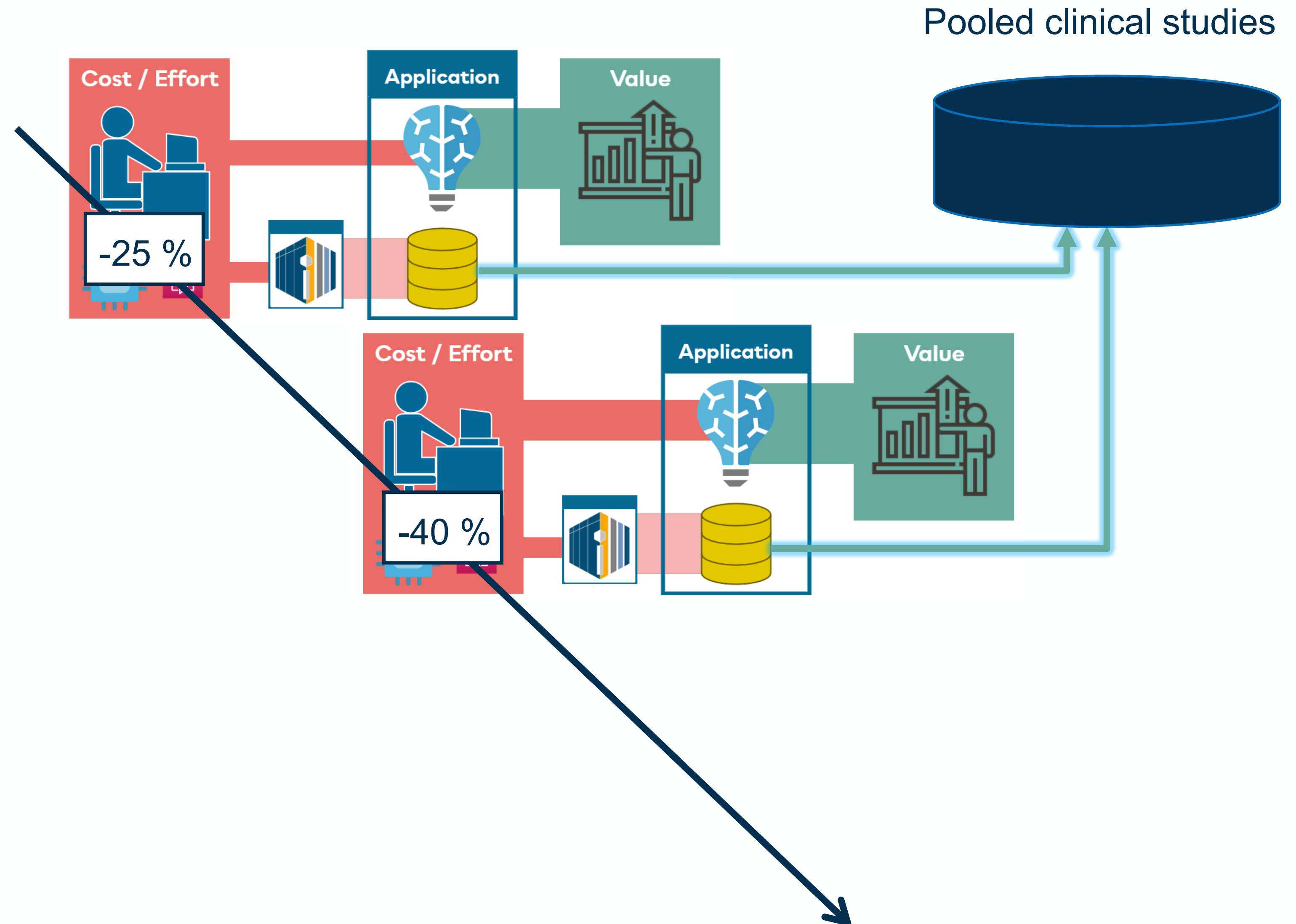
Lowering the cost barrier of entry via automation

- Idea: Use AI to reduce repetitive manual efforts
- HMS calls this concept

IntelligentDI



- Reduce the of costs of data integration
- Beneficial cost-benefit ratio for (smaller) single use cases
- **Scalable approach with repeatable savings**



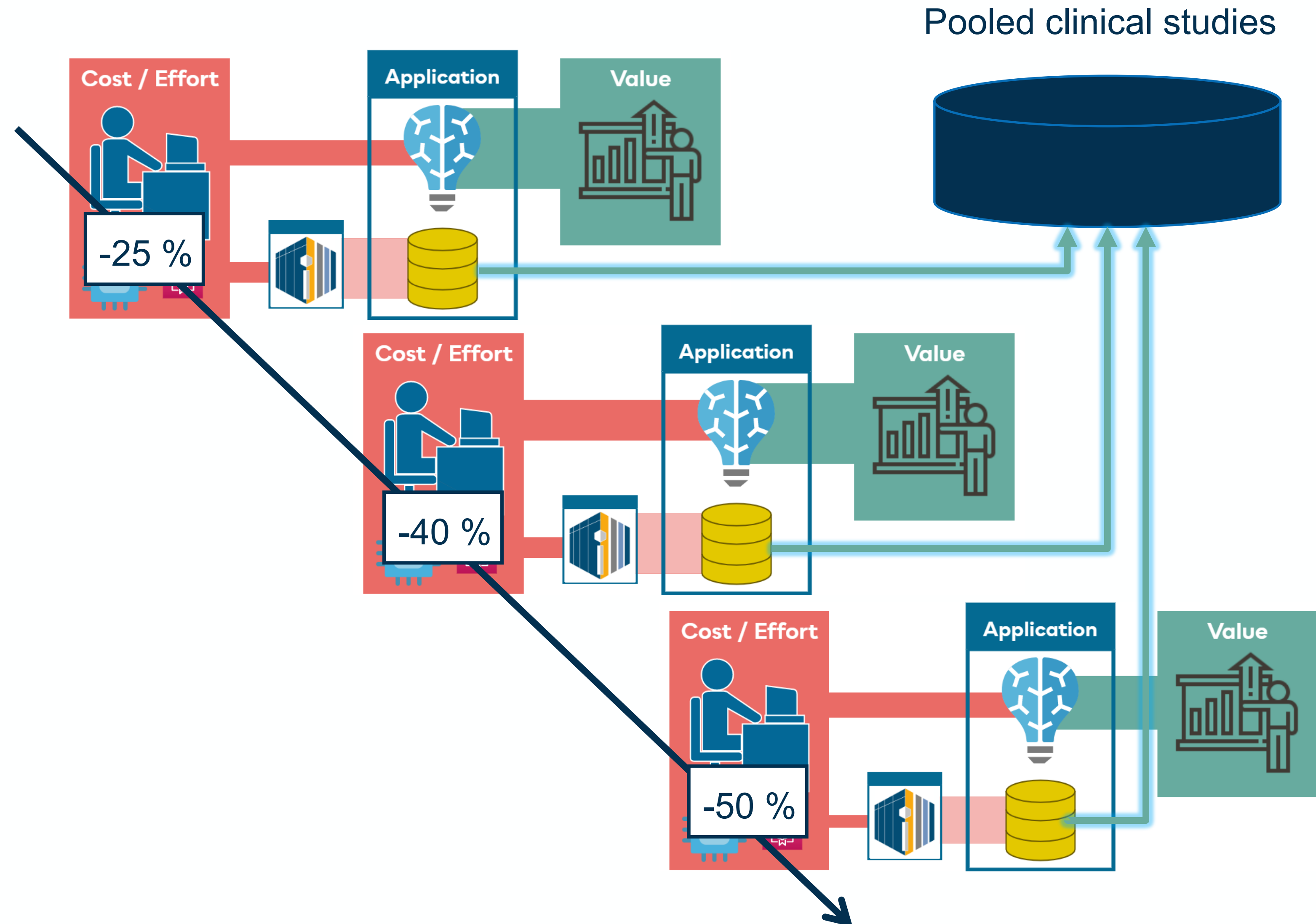
Lowering the cost barrier of entry via automation

- Idea: Use AI to reduce repetitive manual efforts
- HMS calls this concept

IntelligentDI



- Reduce the of costs of data integration
- Beneficial cost-benefit ratio for (smaller) single use cases
- **Scalable approach with repeatable savings**



UX for AI-supported clinical study integration

Legacy

DEMO
PATIENT NO
START DATE
START TIME
AGE

PATIENT NO	START DATE	START TIME	AGE
1	12.07.2003	10:35	45 Y
2	12.07.2003	13:15	33 Y

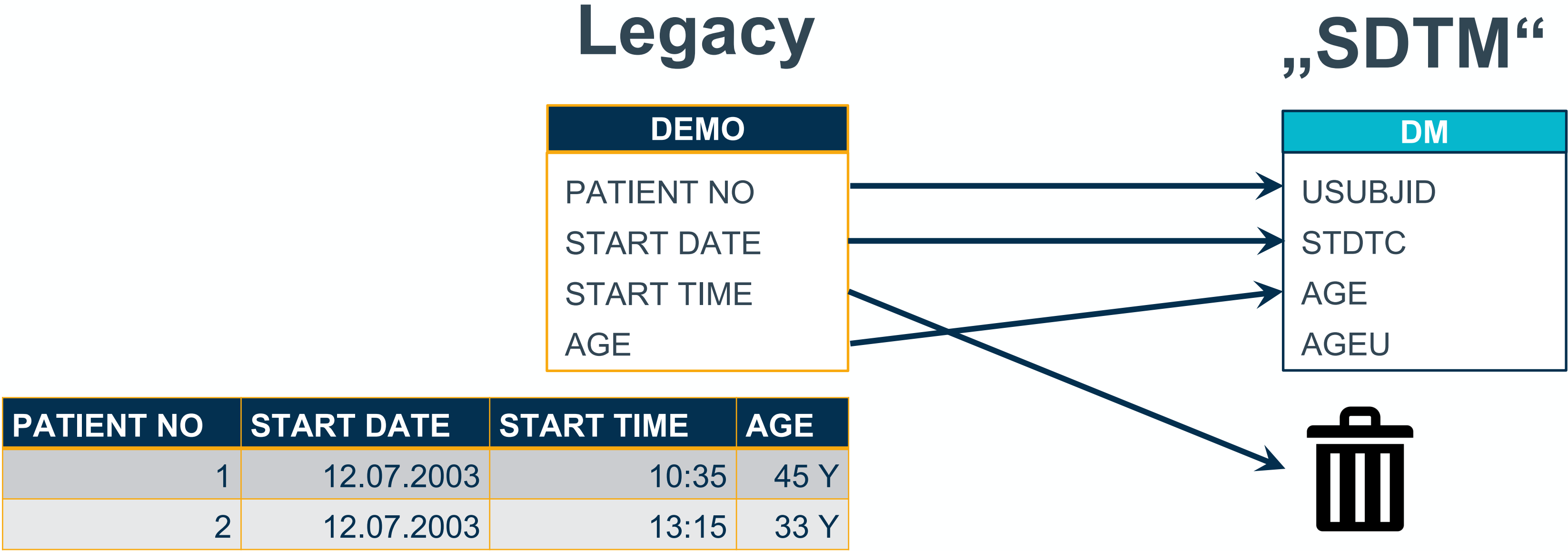


„SDTM“

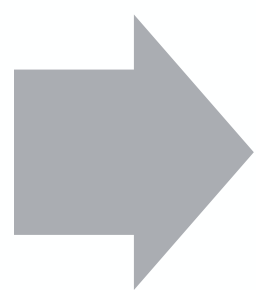
DM
USUBJID
STDTC
AGE
AGEU



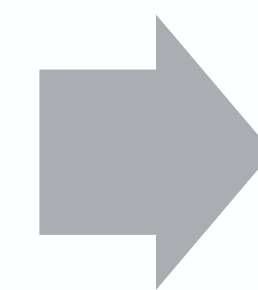
UX for AI-supported clinical study integration



Prediction



Validation



Enhancement

UX for AI-supported clinical study integration

Legacy

DEMO
PATIENT NO
START DATE
START TIME
AGE

PATIENT NO	START DATE	START TIME	AGE
1	12.07.2003	10:35	45 Y
2	12.07.2003	13:15	33 Y

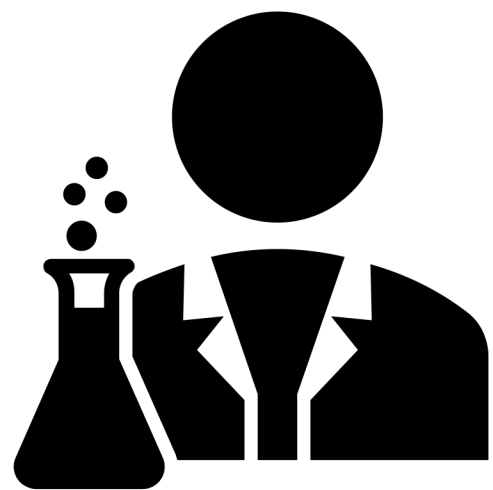
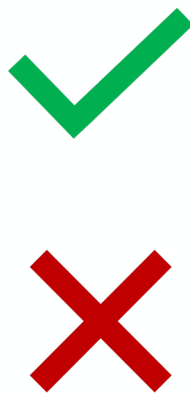
„SDTM“

DM
USUBJID
STDTC
AGE
AGEU

USUBJID	STDTC	AGE	AGEU
1	12.07.2003 + 10:35 = ?	45 Y	45 Y
2	12.07.2003 + 13:15 = ?	33 Y	33 Y



Prediction



Validation

Enhancement

UX for AI-supported clinical study integration

Legacy

DEMO
PATIENT NO
START DATE
START TIME
AGE

PATIENT NO	START DATE	START TIME	AGE
1	12.07.2003	10:35	45 Y
2	12.07.2003	13:15	33 Y

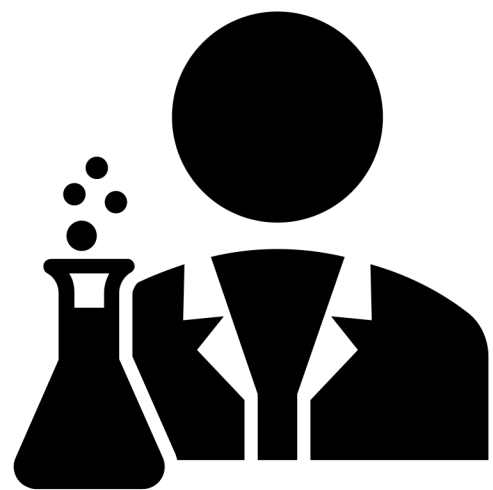
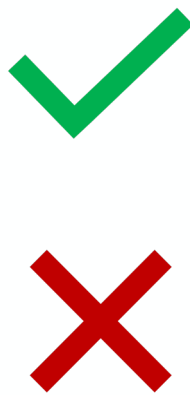
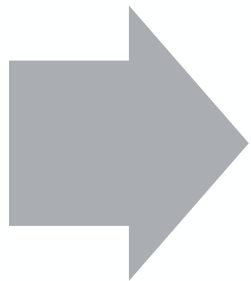
„SDTM“

DM
USUBJID
STDTC
AGE
AGEU

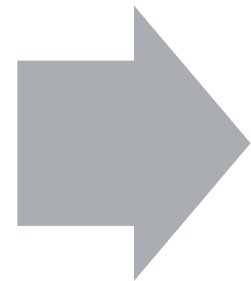
USUBJID	STDTC	AGE	AGEU
magicPot3147_1	2003-07-12T10:35:00	45	Y
magicPot3147_2	2003-07-12T13:15:00	33	Y



Prediction



Validation

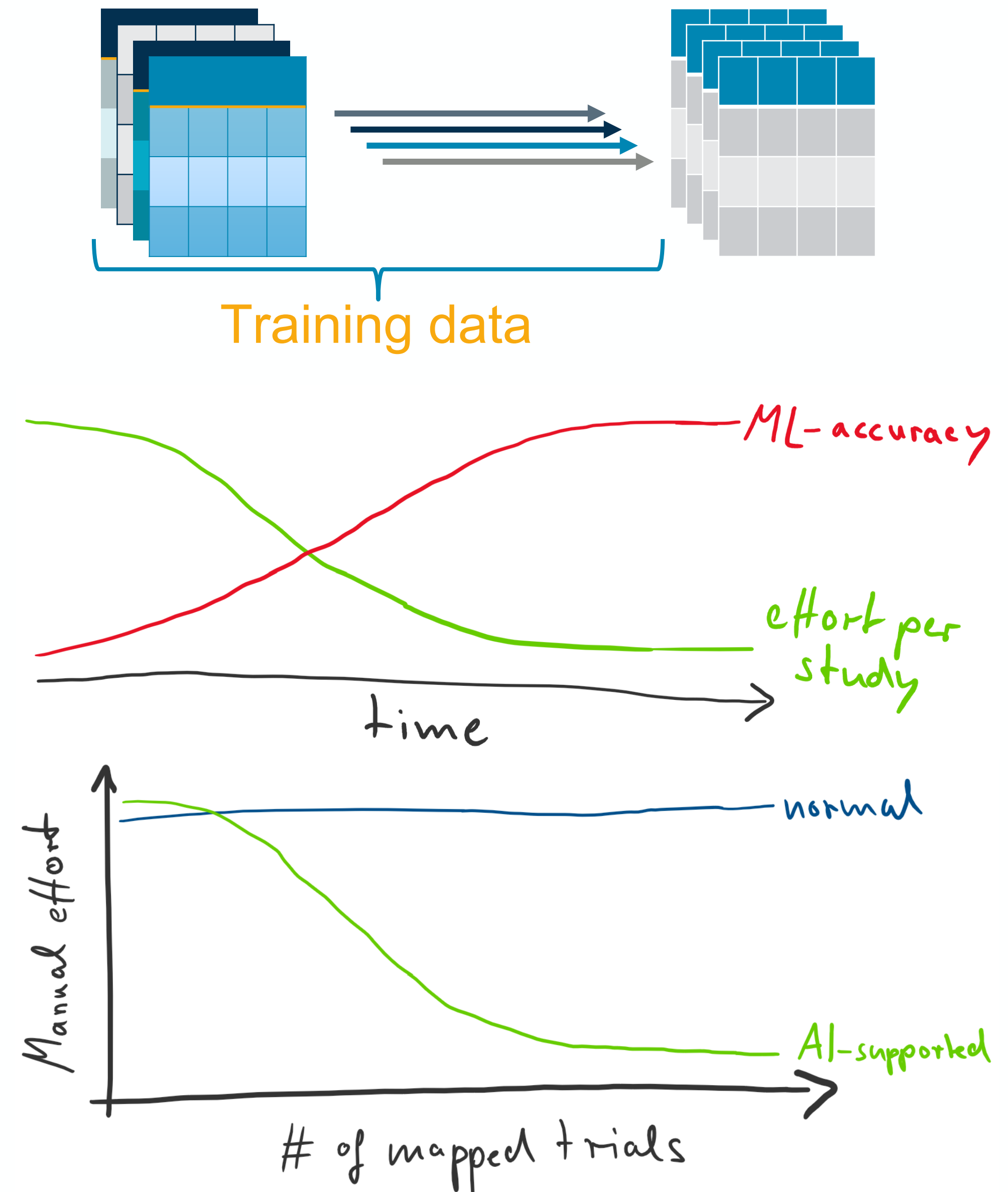


- `create_usubjid()`
- `concat_ISO_date()`
- `split(" ")`

Enhancement

Why AI-support pays off over time

- For this talk AI = Machine Learning (ML)
- ML = „Statistical pattern recognition“
- Initially **only little/no training data** (i.e. mapped trials), thus no AI-support
- Over time:
 - **More and more training data** from expert validations and mappings
 - Decreased manual efforts
- Feasible strategy: Step-wise scaling to new domains and schemas



So what has IntelligentDI to do with clinical study integration?

IntelligentDI by **HMS** ...

- Is an industry agnostic
 - Concept
 - Analytical paradigm
 - Code library
- Was first applied in reinsurance use case:
 - Data collection from primary insurers
 - Data in manually created Excel-sheets



IntelligentDI was applied for pharmaceutical company ...

- ≥ 250 clinical trials
- Strongly heterogeneous data formats due to
 - Legacy
 - Different CROs and data providers
 - Bought products
 - Evolving data standards
- Initiated by R&D analytics department which had requests for cross-study analysis
- Tool suite for integrating clinical data in an SDTM-inspired format incl. automatized enrichments to prepare the data for analytics



The IntelligentDI Python package Machine Learning kernel

Python package developed by **HMS** 

- More than 4 years of project experience

- `import pandas, sklearn, xgboost`

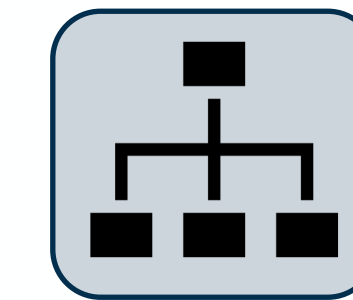


Goal: „Save 5 weeks of project work“ by using existing code

- **Framework** for predicting target columns
- **ML-features** that make the algorithm see the differentiating aspects of the data
- **Model selection routine** to choose the best ML model

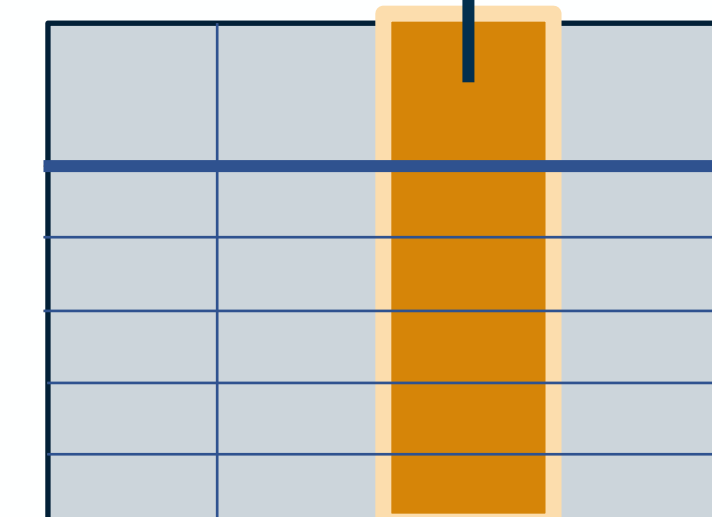


IntelligentDI ML-features



1. Metadata

2. Column name

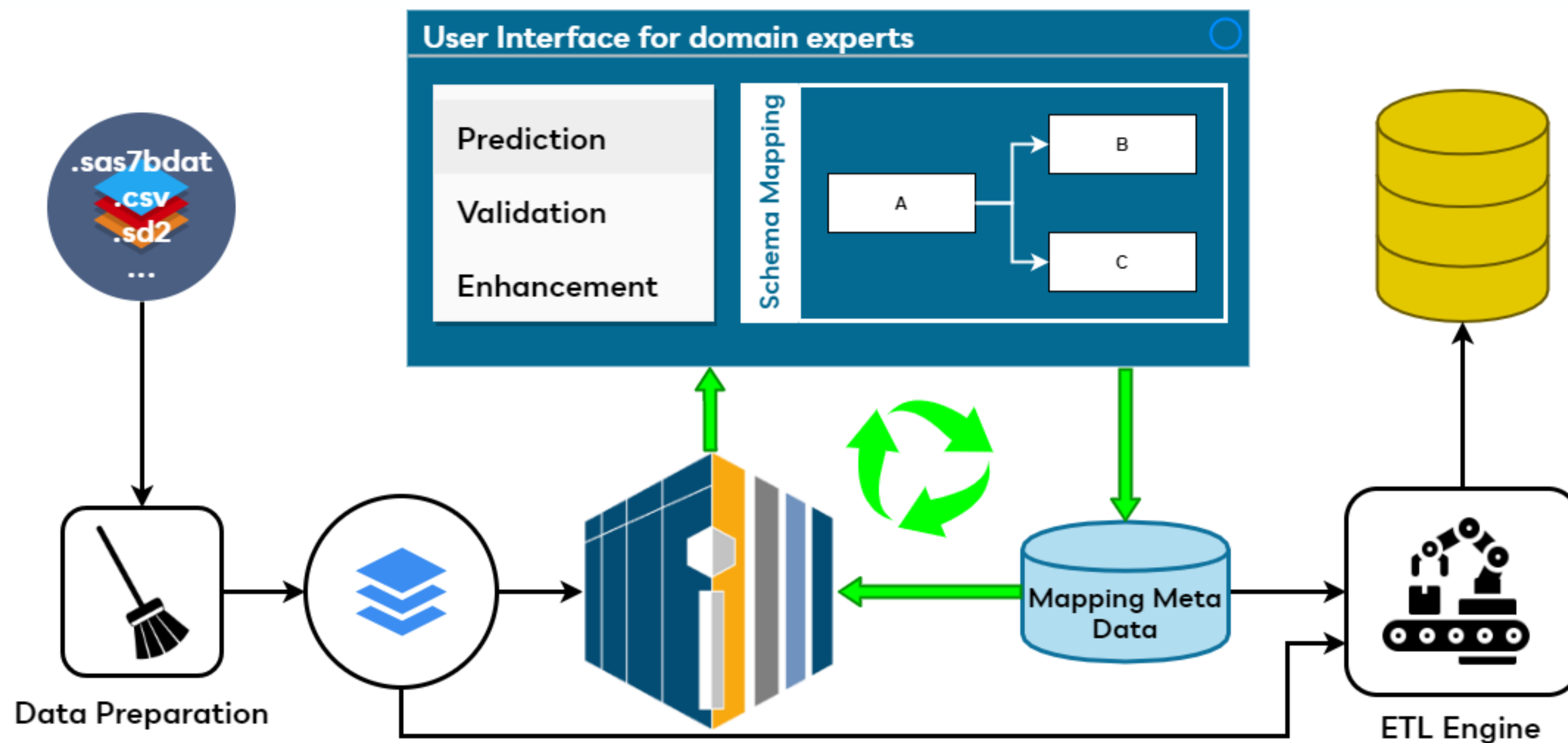


3. Content

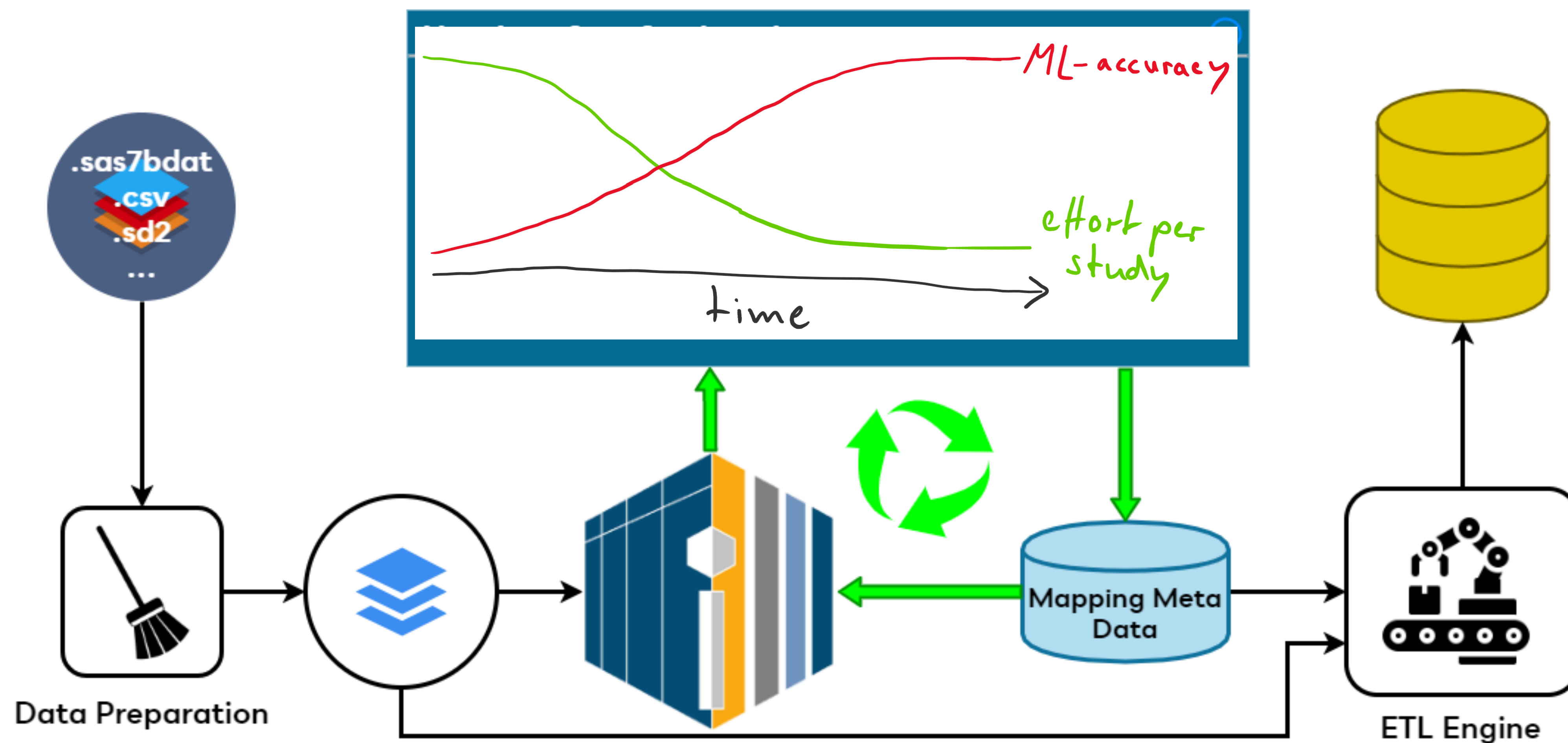


4. Context

Building an AI-supported clinical data integration tool



Building an AI-supported clinical data integration tool



Economical considerations for developing an AI-supported clinical data integration tool

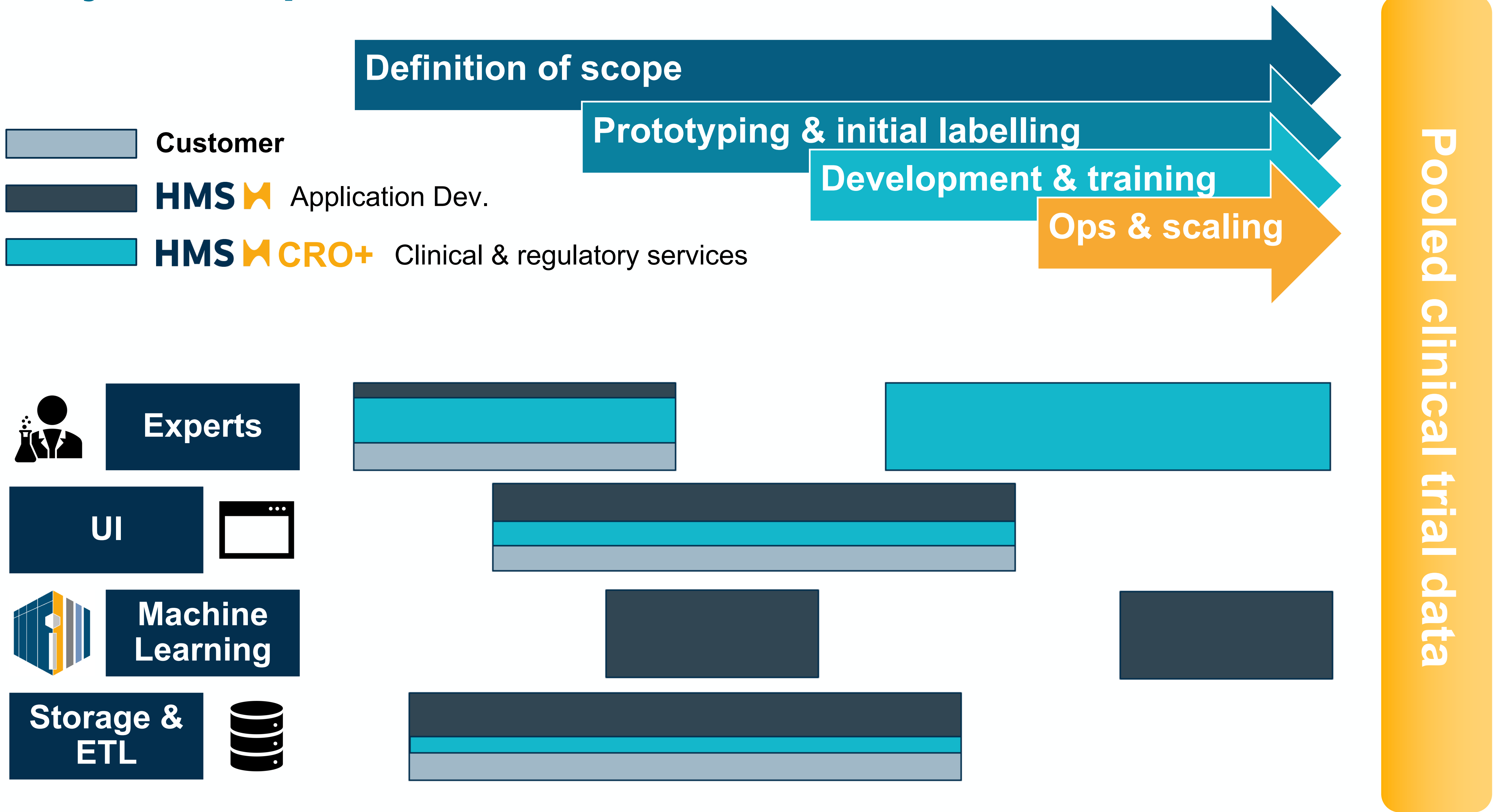
Investments need to be made for:

- Application development
- Manpower for labelling training data
- Manpower for integrating trials
- Application Ops and maintenance and
- IT infrastructure

Cost savings and added value is a matter of:

- **Scale** as # of clinical trials
- **Increased efficiency** by Standardization and an ergonomic UI
- **Availability** of on-demand study integration with little preliminary lead time
- **Increased automation** based on
 - AI-support (depends on ML-accuracy)
 - Automatic data enrichments

Project template



Any questions? We would be happy to hear from you



CONTACT INFORMATION

Kai Spielmann

Head of Life Science

kai.spielmann@analytical-software.de



David Hipp

Project Manager & Data Scientist

david.hipp@analytical-software.de