Big Data Pipelines on the Computing Continuum:
Ecosystem and Use Cases Overview

DataCloud H2020 project

https://datacloudproject.eu

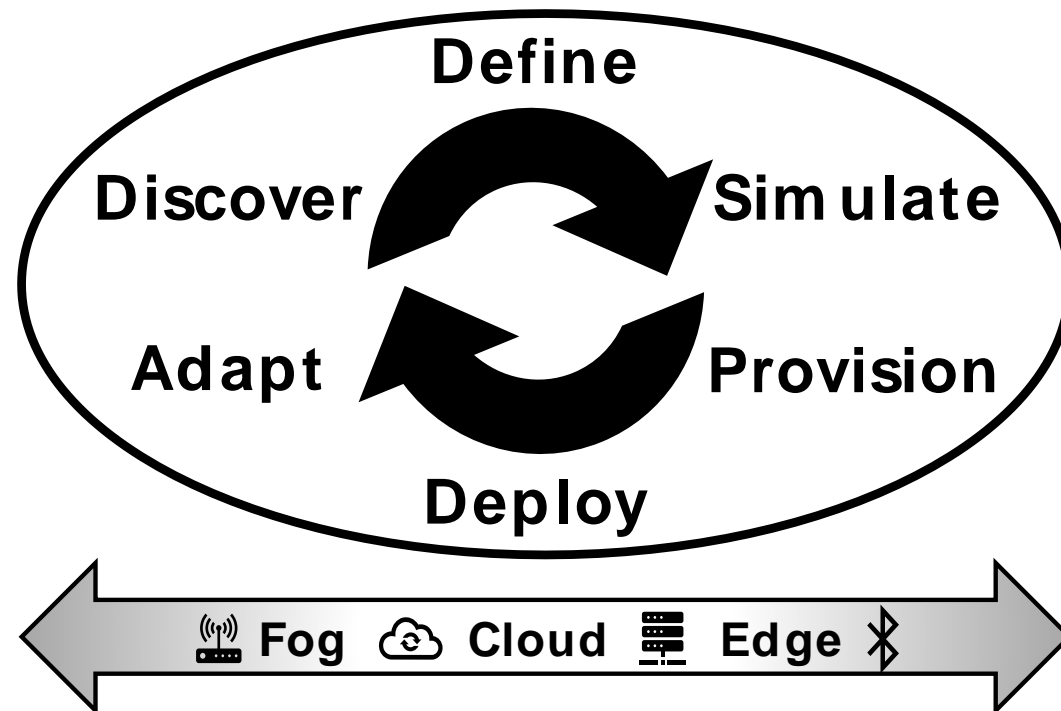Giannis Ledakis, UBITECH
gledakis@ubitech.eu

Outline

- DataCloud Big Data Pipeline Ecosystem

- Big data pipeline lifecycle on the Computing Continuum

- DataCloud Use Cases
  - Smart Mobile Marketing Campaigns Use Case
  - Automatic Live Sports Content Annotation Use Case
  - Digital Health System Use Case
  - Analytics of Manufacturing Assets Use Case
  - Analytics of Manufacturing Assets Use Case

- Initial Analysis of Scenarios

- Conclusions

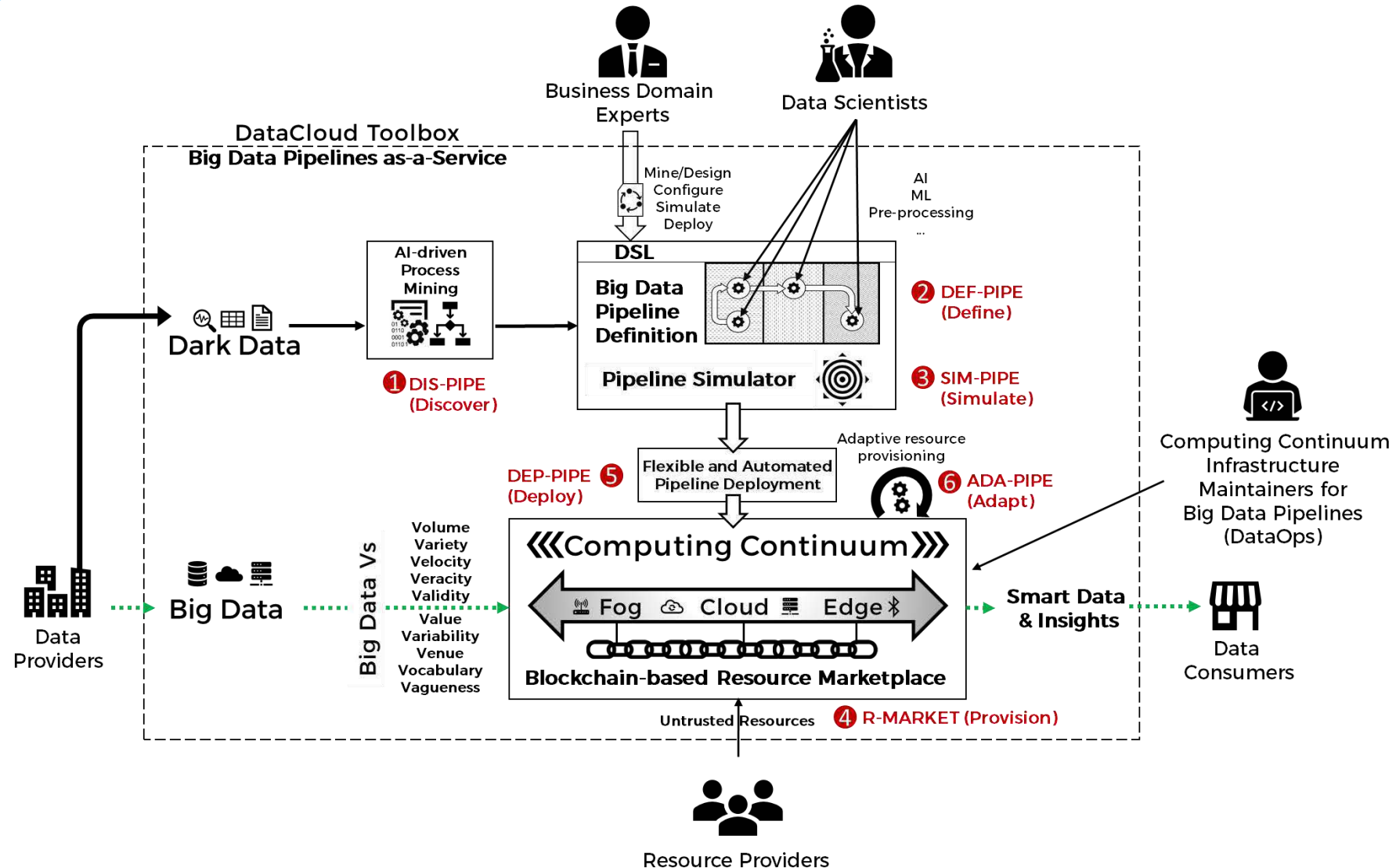# Challenge: Big Data Pipeline Lifecycle on Computing Continuum

Big Data pipelines are composite pipelines for processing data with non-trivial properties and characteristics, commonly referred to as the *Vs of Big Data.*
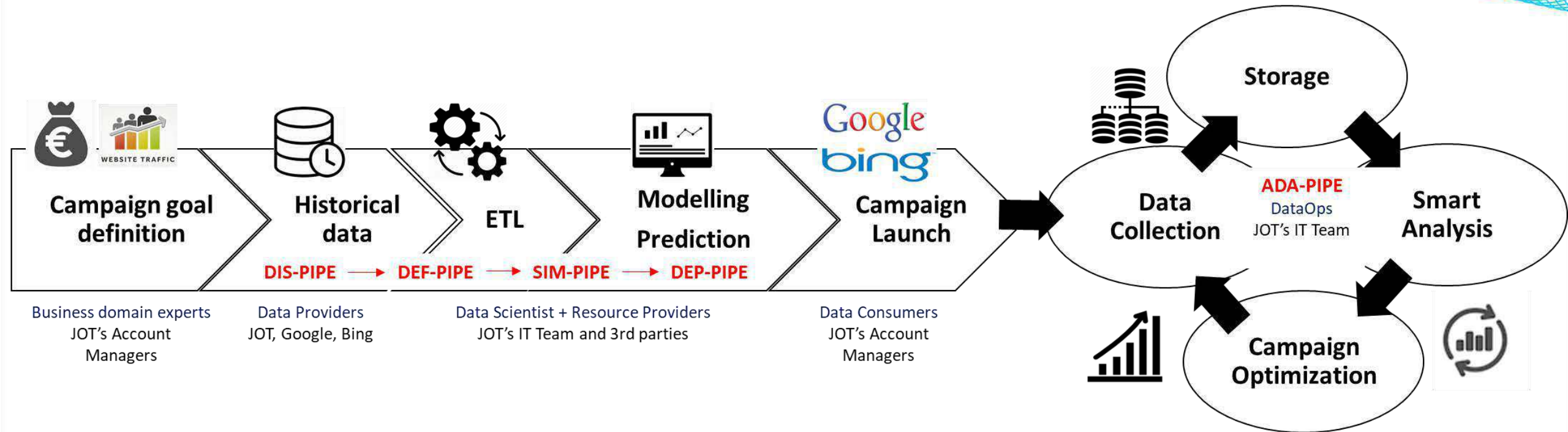
Example: 30GB file

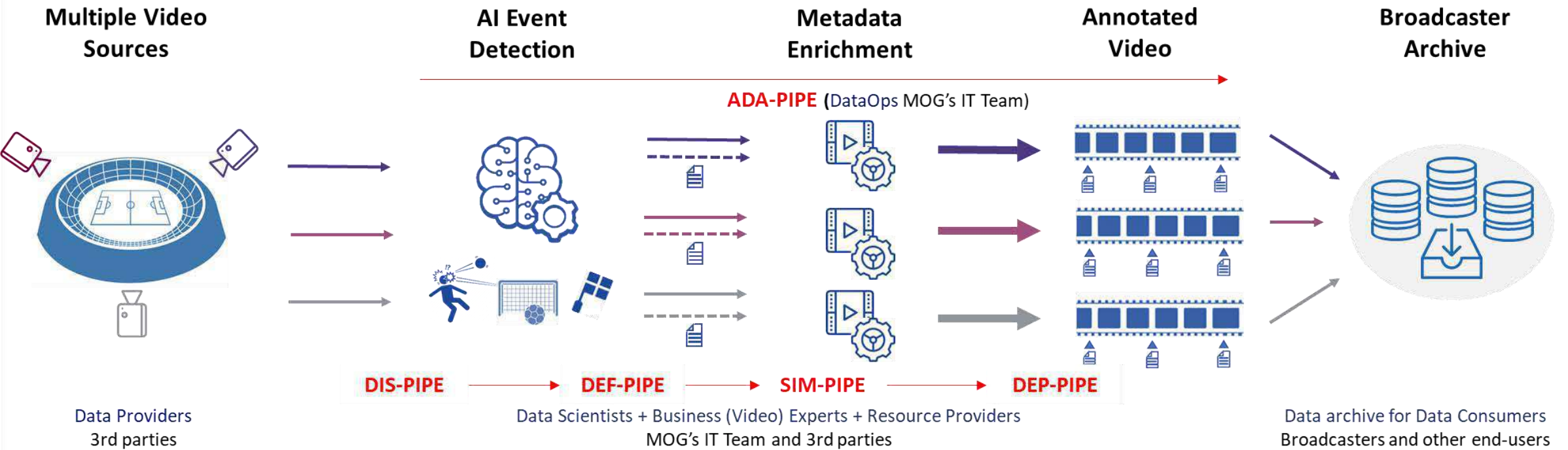| |
|---|
| Unzip |
| TSV to CSV |
| Split |
| Transform |
| Store |

DataCloud: Approach and Toolbox

Smart Mobile Marketing Campaigns

| Campaign goal definition	Historical data	ETL	Modelling Prediction	Campaign Launch
DIS-PIPE → DEF-PIPE → SIM-PIPE → DEP-PIPE				
Business domain experts JOT's Account Managers	Data Providers JOT, Google, Bing	Data Scientist + Resource Providers JOT's IT Team and 3rd parties		Data Consumers JOT's Account Managers

Automatic Live Sports Content Annotation Use Case
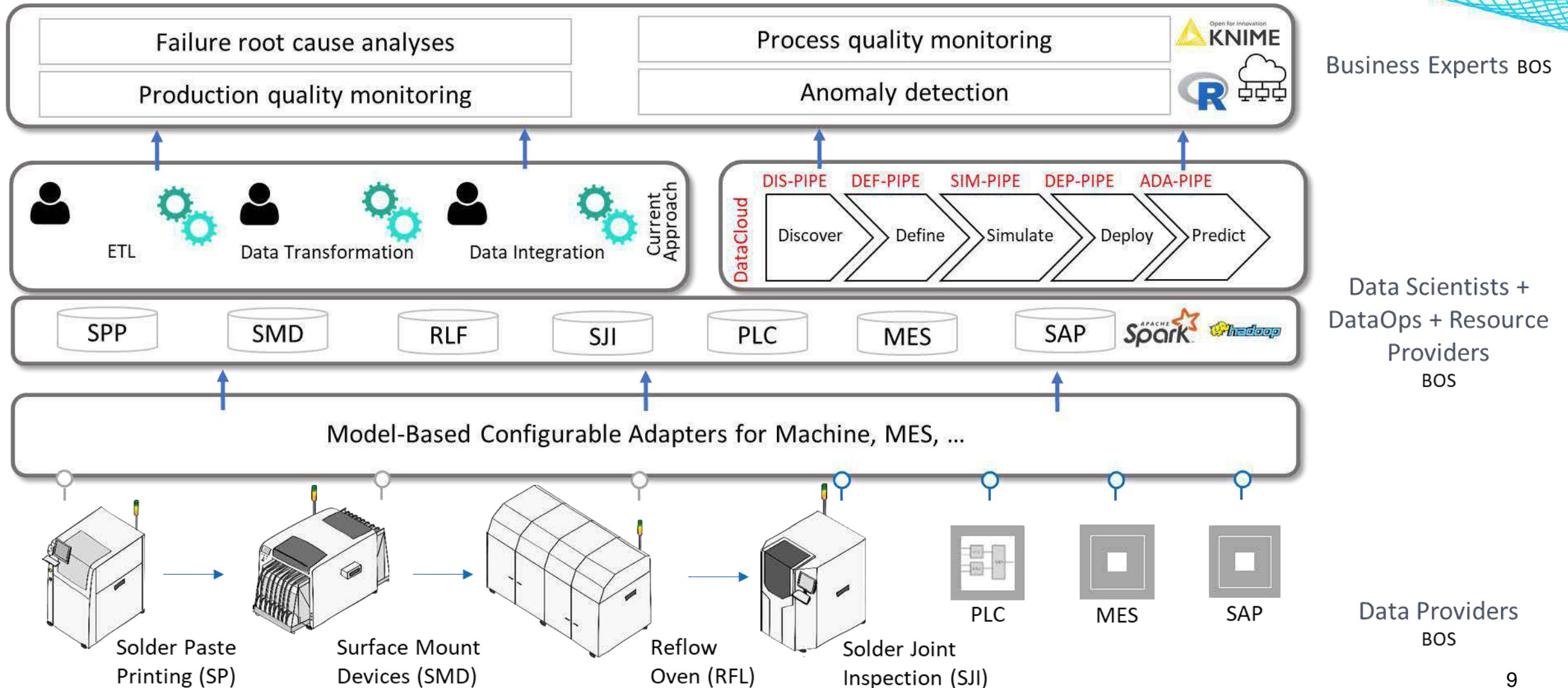
Digital Health System Use Case
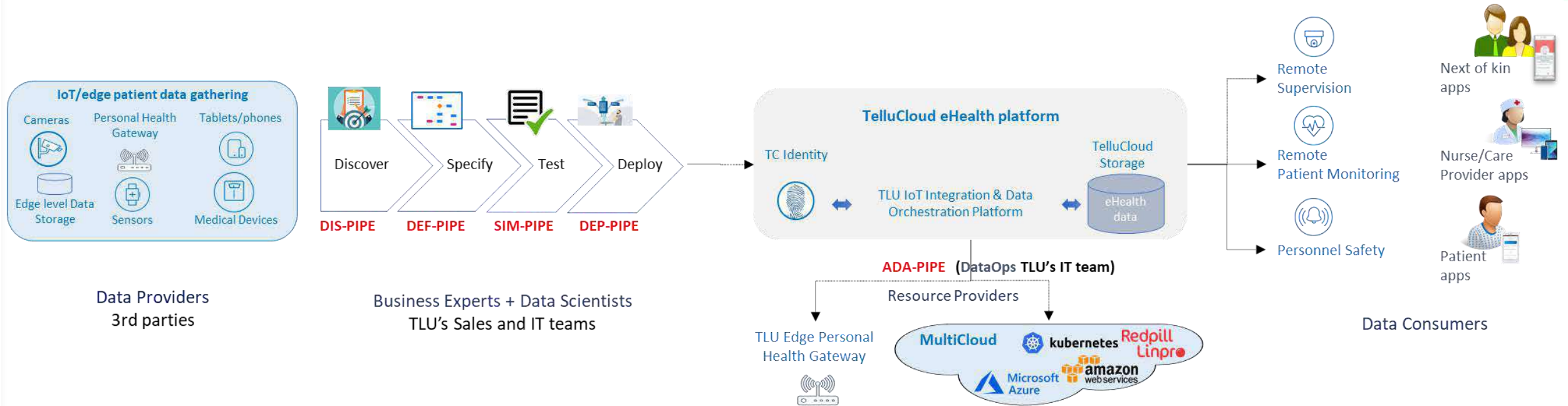
Predicting Deformations in Ceramics Manufacturing
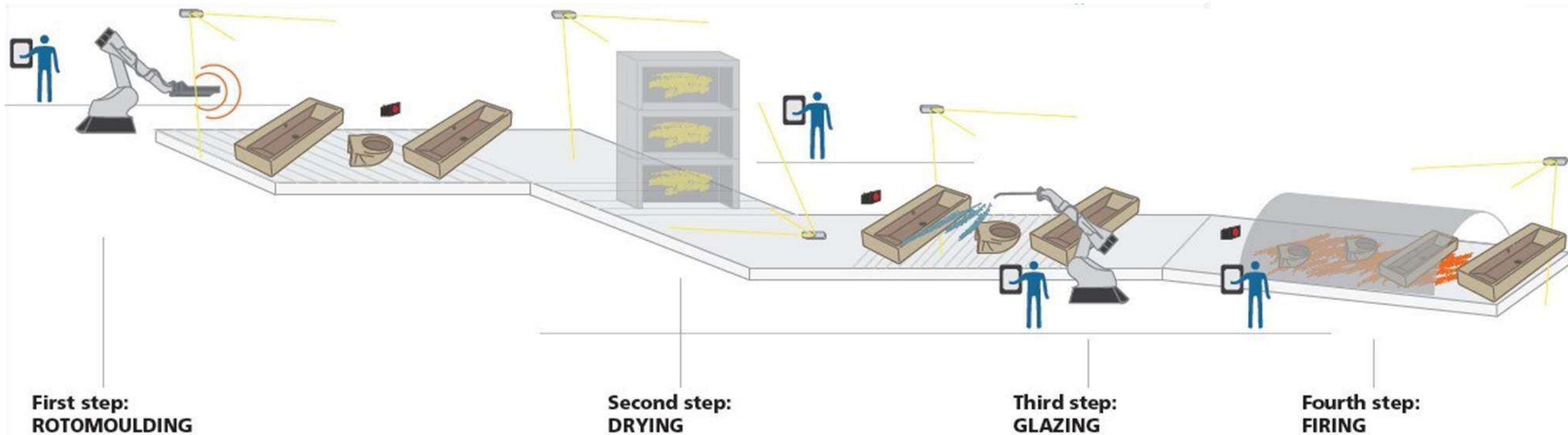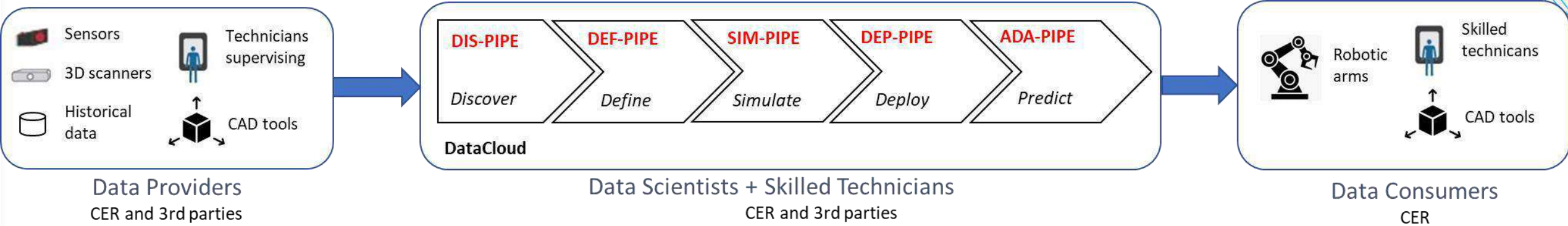
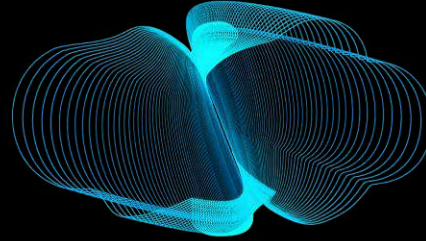# Analytics of Manufacturing Assets Use Case

Overview of lifecycle usage and stakeholders involved

TABLE I
USE CASES AND THEIR DEGREE OF COVERAGE FOR PIPELINE LIFECYCLE.

Case	Pipeline Discovery	Pipeline Definition	Pipeline Simulation	Resource Provisioning	Pipeline Deployment	Pipeline Adaptation
Marketing	Low	Moderate	High	Moderate	High	Low
Media	Moderate	Moderate	Moderate	High	High	High
Healthcare	Low	High	High	Moderate	High	Moderate
Manufacturing	High	High	Moderate	Moderate	Moderate	High
Industry 4.0	High	High	Moderate	High	Moderate	Low

TABLE II
STAKEHOLDERS AND THEIR DEGREE OF INVOLVEMENT IN EACH USE CASE.

Case	Data providers	Business domain experts	Data scientists	Resource providers	DataOps operators	Data consumers
Marketing	High	Moderate	High	Moderate	Moderate	High
Media	High	Low	Moderate	High	High	High
Healthcare	High	Moderate	Moderate	Moderate	High	High
Manufacturing	High	High	Low	Moderate	Moderate	High
Industry 4.0	High	High	High	Moderate	Moderate	High

THANK YOU!

https://datacloudproject.eu/