A Data Quality Scorecard for Assessing the Suitability of Asset Condition Data for Prognostics Modeling

Sarah Lukens, Damon Rousis, Dominic Thomas, Travis Baer, Michael Lujan and Marshall Smith

PHM Society Conference 2022

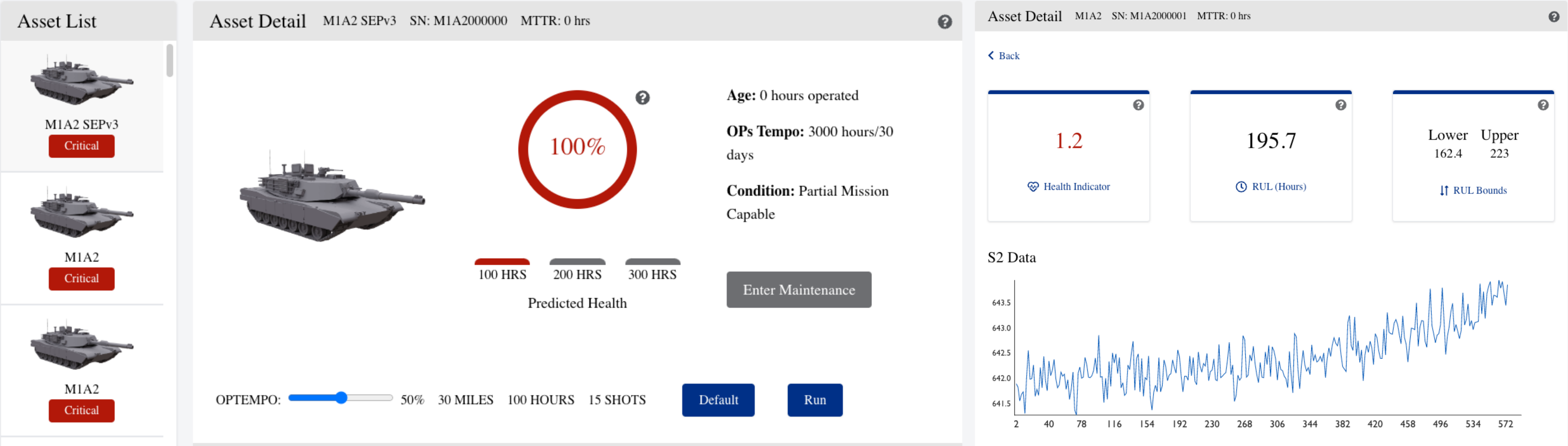
Link to paper: https://papers.phmsociety.org/index.php/phmconf/article/view/3188

LMI

What do I do at my day job?

LMI has over 60 years of experience delivering logistics solutions to DoD entities, including data analytics supporting:

- Reliability and sensor-based approaches for predictive maintenance & fleet management
- Supply chain & demand forecasting



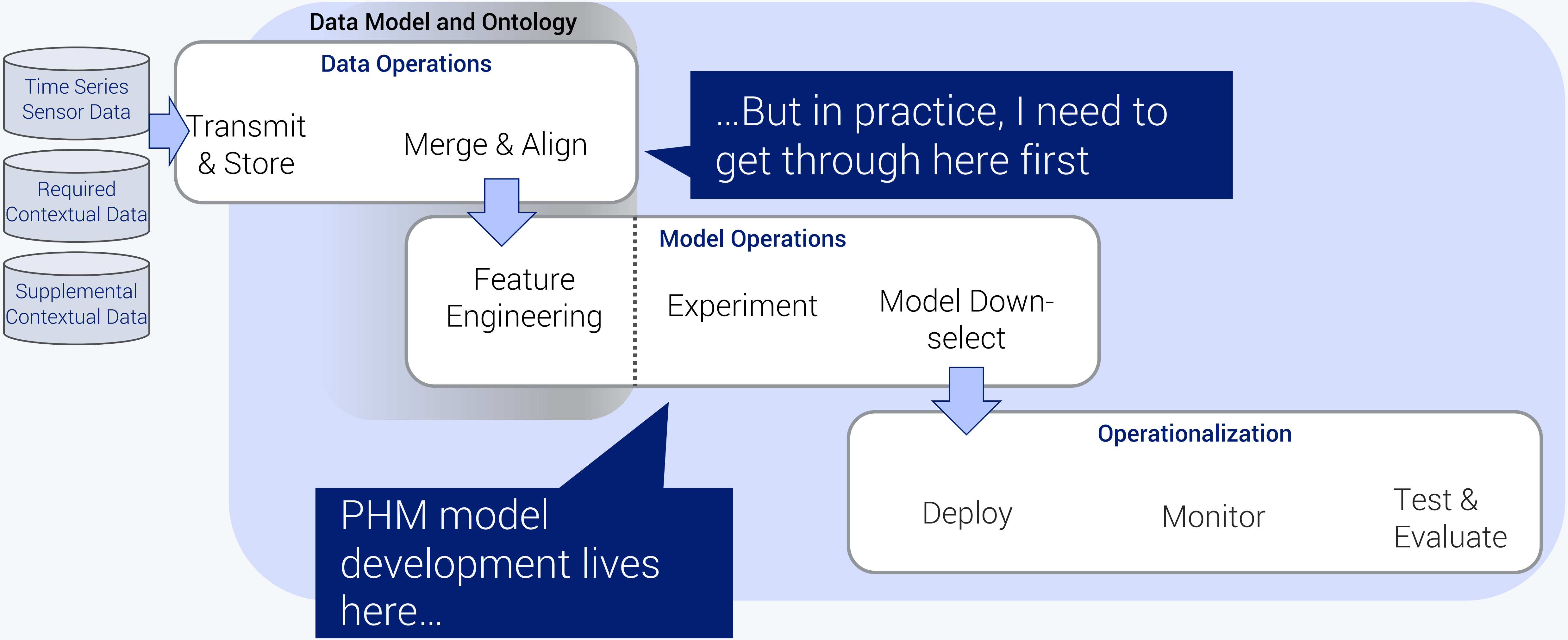
LMI

Background

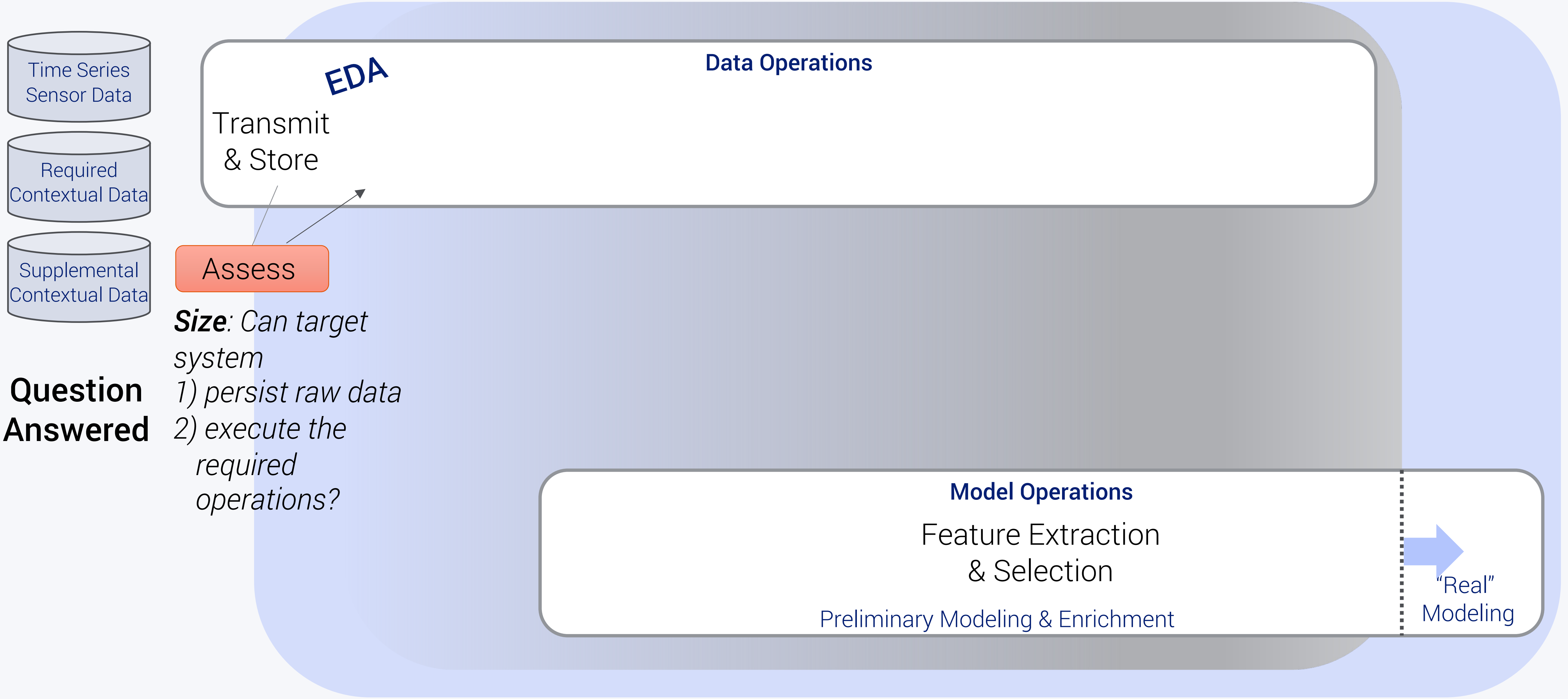
Have you ever seen this happen?

- I have sensor data!!
- I'm going to spend some $$$ so I can predict future failures on my industrial assets!!
- The ROI will be worth it!

PHM implementation has a lot of moving factors



Zoom into "Data Operations"



Question Answered

Size: *Can target system*
1) persist raw data
2) execute the required operations?

LMI

Data Survey – what could go wrong?

Example: ground vehicle data. Can I use all this data?

Meta-data checklist	Sensor data (provided)	Maintenance data (provided)
❑ Scope of asset & component coverage	~360 variables/tags on 5 ground vehicles	~250K work orders on ~500 vehicles
❑ Calendar period	May 2019 – August 2019 (4 months)	Jan 2011 – Dec 2021
❑ What variables?	6 tags with high number of quantitative distinct readings, including voltage, RPMs, etc. (Most tags are diagnostic)	Of the ~250K WO's, only ~200 on 3 of the 5 assets. These WO's span 2020-2021.
❑ How big is the data?	15 GB	1 GB

4 months insufficient period for baselining health

Data sources not compatible

- Standard process streamlines figuring this out very quickly
- Provides explicit requirements for data collection and gathering

LMI

Preliminary steps before PHM model building

Time Series Sensor Data

Required Contextual Data

Supplemental Contextual Data

Data Operations

EDA

Transmit & Store

Data Model Assessment

Assess

Assess

Question Answered

Size: *Can target system*
1) persist raw data
2) execute the required operations?

Structure: *Does the data model support the ability to provide* ***context*** *to raw data?*

Model Operations

Improve historical Data (NLP, imputation)

Feature Extraction & Selection

Preliminary Modeling & Enrichment

"Real" Modeling

LMI

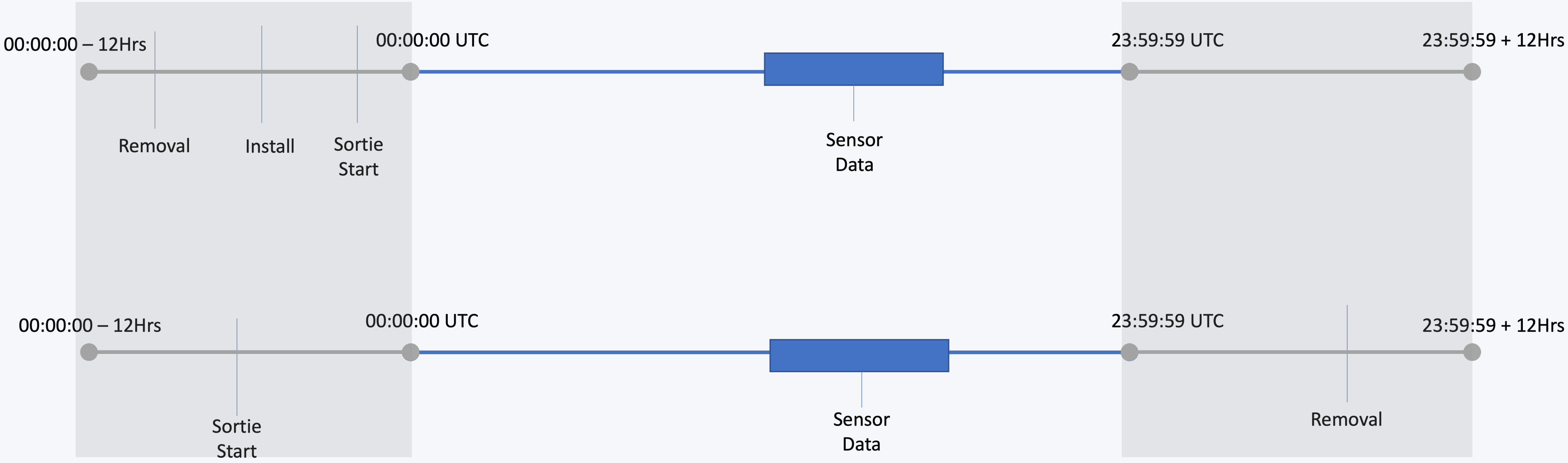
Data Model Assessment

	Scenario 1: Perfect	Scenario 2: Promising	Scenario 3: Insufficient
Sensor data is human readable			
Sensor data is machine readable			N/A
Sensor data join key/s exist			N/A
Sensor data join key/s unambiguous			N/A
Maintenance data is human readable			
Maintenance data is machine readable			
Maintenance data join key/s exist			
Maintenance data data join key/s unambiguous			
Operating time data is human readable			
Operating time data is machine readable			
Operating time join key/s exist			
Operating time join key/s unambiguous			
Failure cause is human readable			
Failure cause is codified			
Maintenance action human readable			
Maintenance action codified			

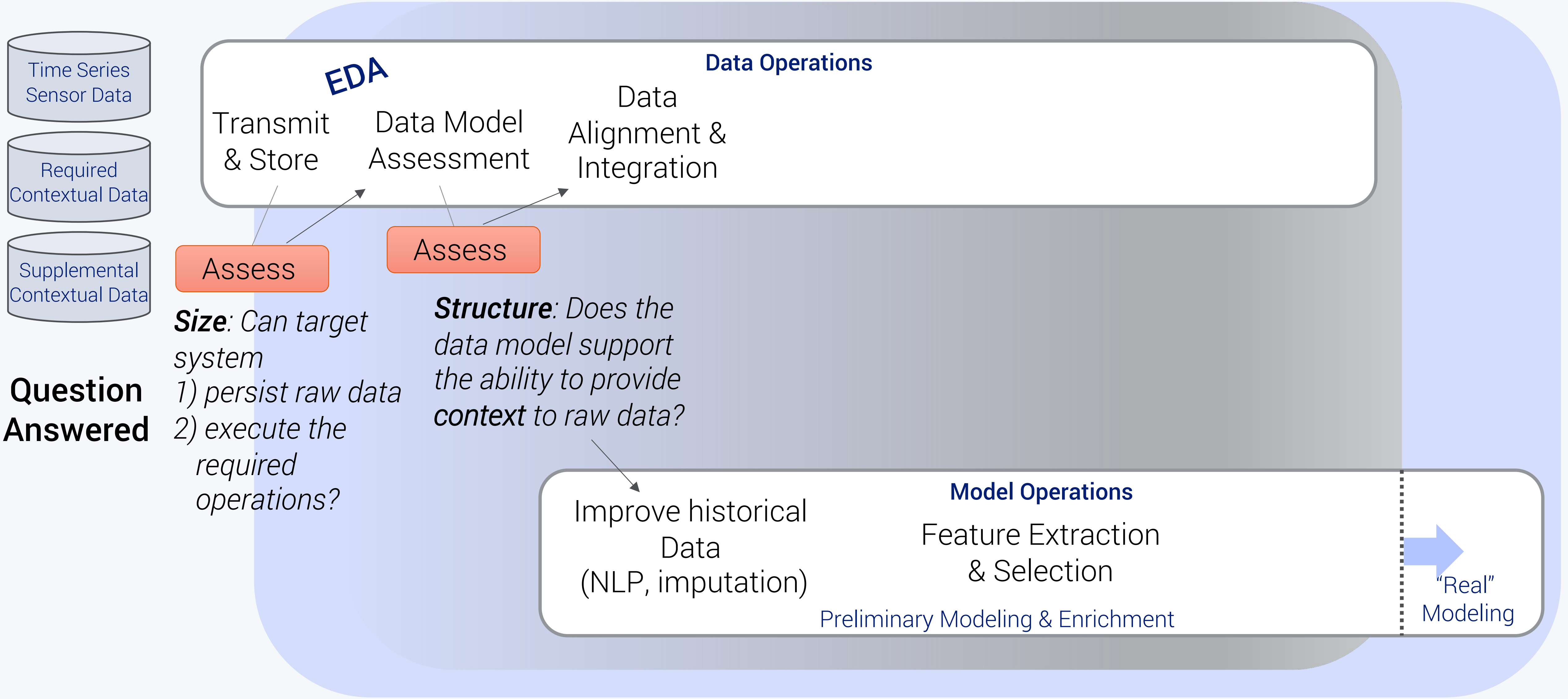
Data Model Assessment – what could go wrong?

Example: Ambiguous join keys

- Events could have occurred anywhere in a 24 hour window because only DATE was recorded
- Example of two possible timelines:



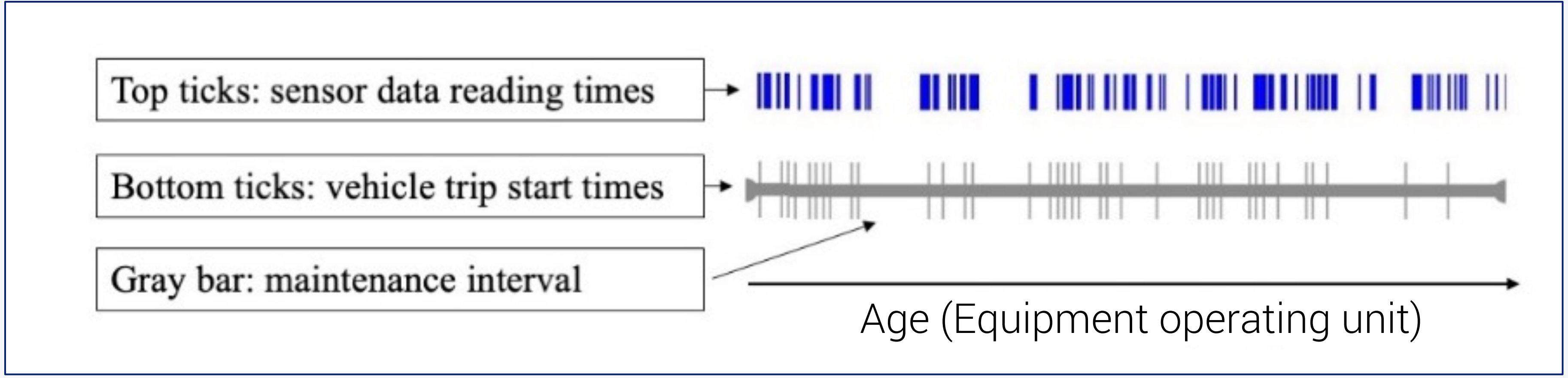
Preliminary steps before PHM model building



LMI

Data alignment & integration of contextual data sources

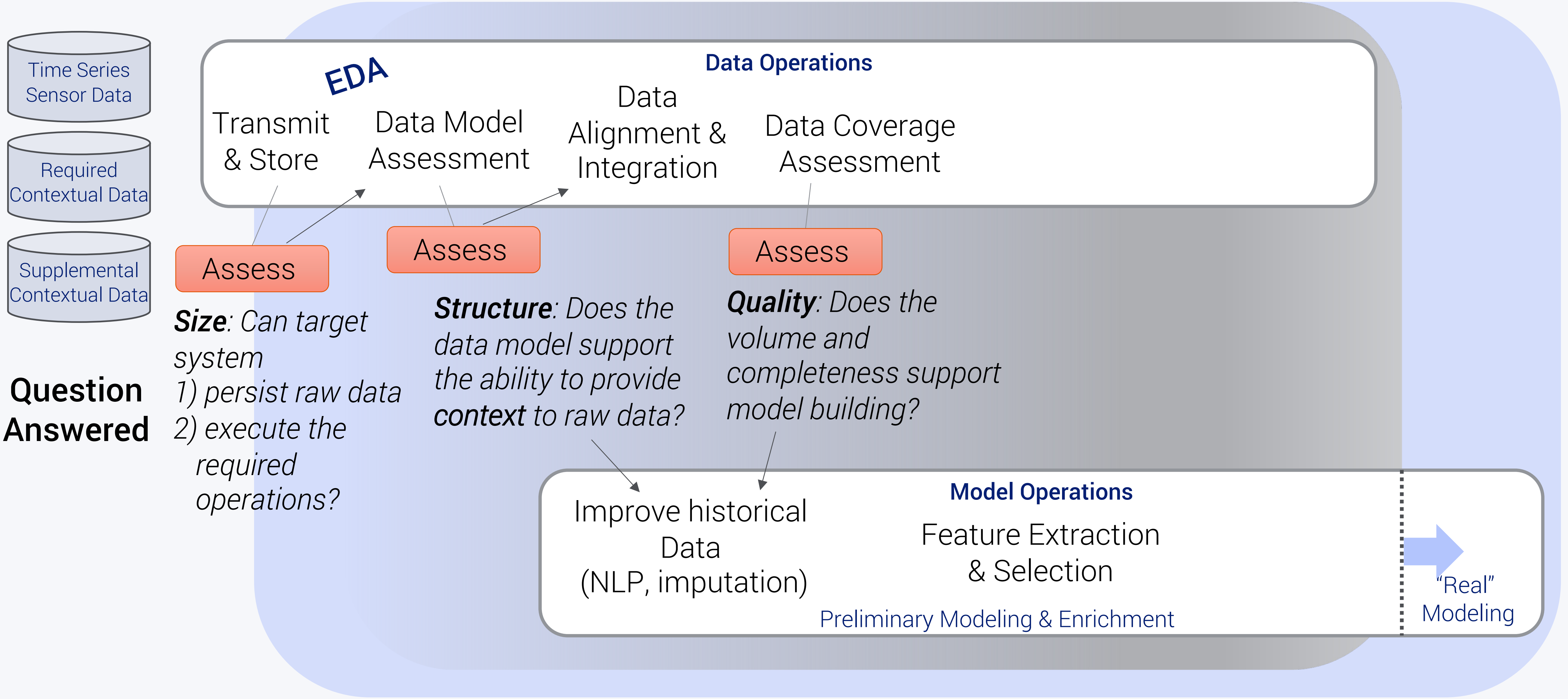
3 data sources:
1. Sensor readings
2. Maintenance intervals
3. Operating time



Sensor readings | Maintenance intervals | Operating time

Time	Sensor 1	Interval ID	Install date	Duration	Trip ID	Age
2022-5-03 01:00	641.82	1	2022-5-01	31	1	0
2022-5-03 01:05	642.15	1	2022-5-01	31	1	5
2022-5-03 01:10	642.35	1	2022-5-01	31	1	10
2022-5-04 03:45	642.35	1	2022-5-01	31	2	15
2022-5-04 03:50	642.37	1	2022-5-01	31	2	20

Preliminary steps before PHM model building

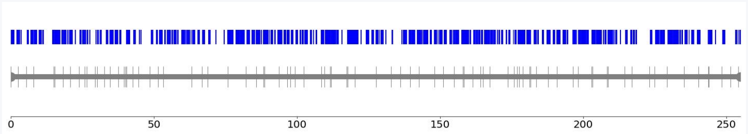
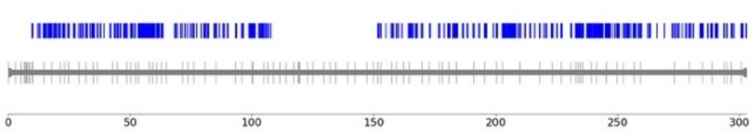
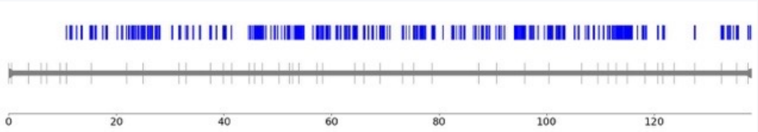
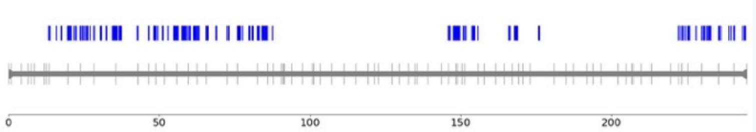
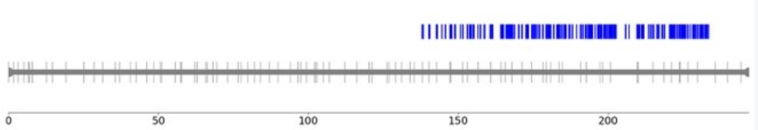
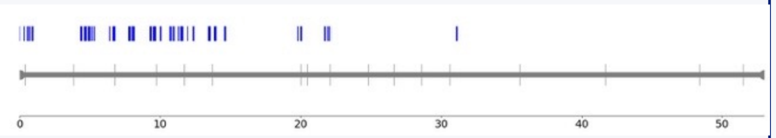


Integration of contextual data sources: what could go wrong?

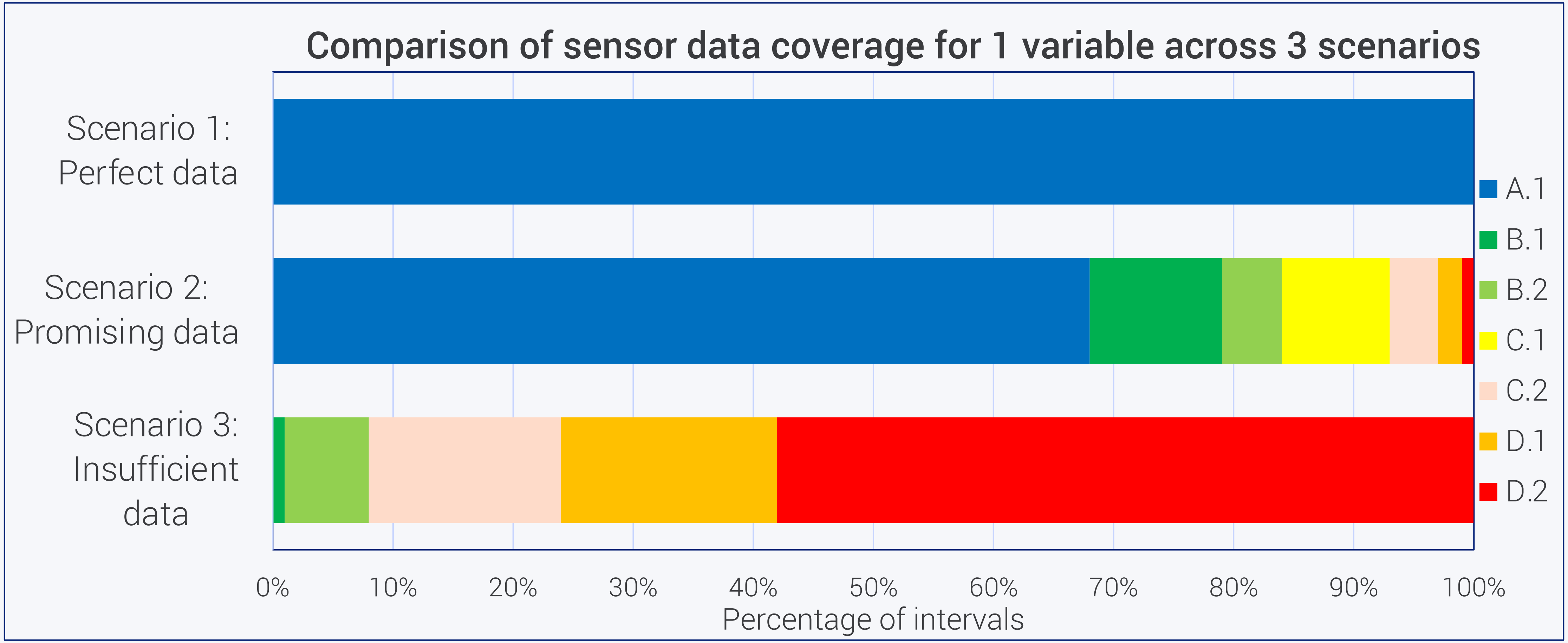
Issue	Impact on equipment lifetime model	Scenario 1: Perfect	Scenario 2: Promising	Scenario 3: Insufficient
No operating data or gaps in operating data	Component aging may not match operating time, contributing to uncertainty in lifetime modeling Without remediation, these intervals are not usable for training	0%	32%	60%
Low component aging (<10 hours)	Inclusion/Exclusion to model may introduce bias and change training data size	0%	3%	12%
No sensor data	Cannot build sensor-based prognostics model	0%	2%	19%
Measurement ambiguity	May result in prognostics algorithm trained on mislabeled data	0%	0%	1%

Measures expressed as a percentage of maintenance intervals in the dataset containing the issue.

Sensor Data Coverage Assessment

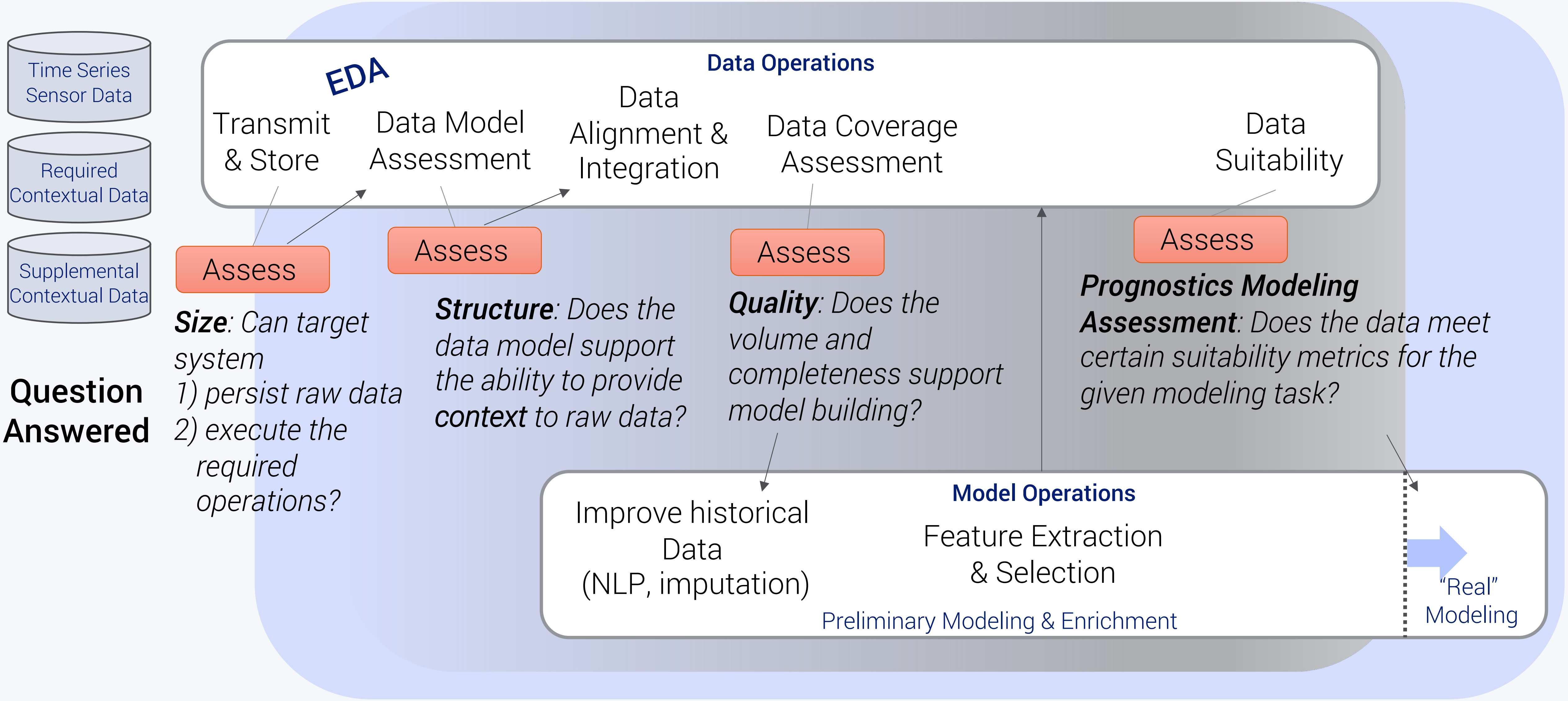
Type	Description	Example
A.1	Sensor data is consistent from interval start to end	
B.1	Small gap in sensor data in the middle of the interval*	
B.2	Small gap in sensor data at beginning or end of interval	
C.1	Large gap in sensor data in the middle of the interval†	
C.2	Large gap in sensor data at the beginning or end of interval	
D.1	Sensor data very sparse throughout the interval	
D.2	Sensor data does not exist in the interval	

Example: Sensor Data Coverage Assessment



LMI

Preliminary steps before PHM model building



Soo… given my existing data, can I build PHM models?

If you have the data, it is worth investigating if the investment into PHM is possible..

But in a cost- and effort-effective fashion!

We've shown:

- Formal, repeatable approach for evaluating PHM modeling suitability which …
- Brings scientific rigor to EDA for PHM
- Suggests measures to articulate, quantify and communicate the suitability of entering into the model building phase of PHM

And now we drink

Thank you

Sarah Lukens

sarah.lukens@lmi.org

LMI