Technical Language Processing: Putting technical data to work in industrial applications

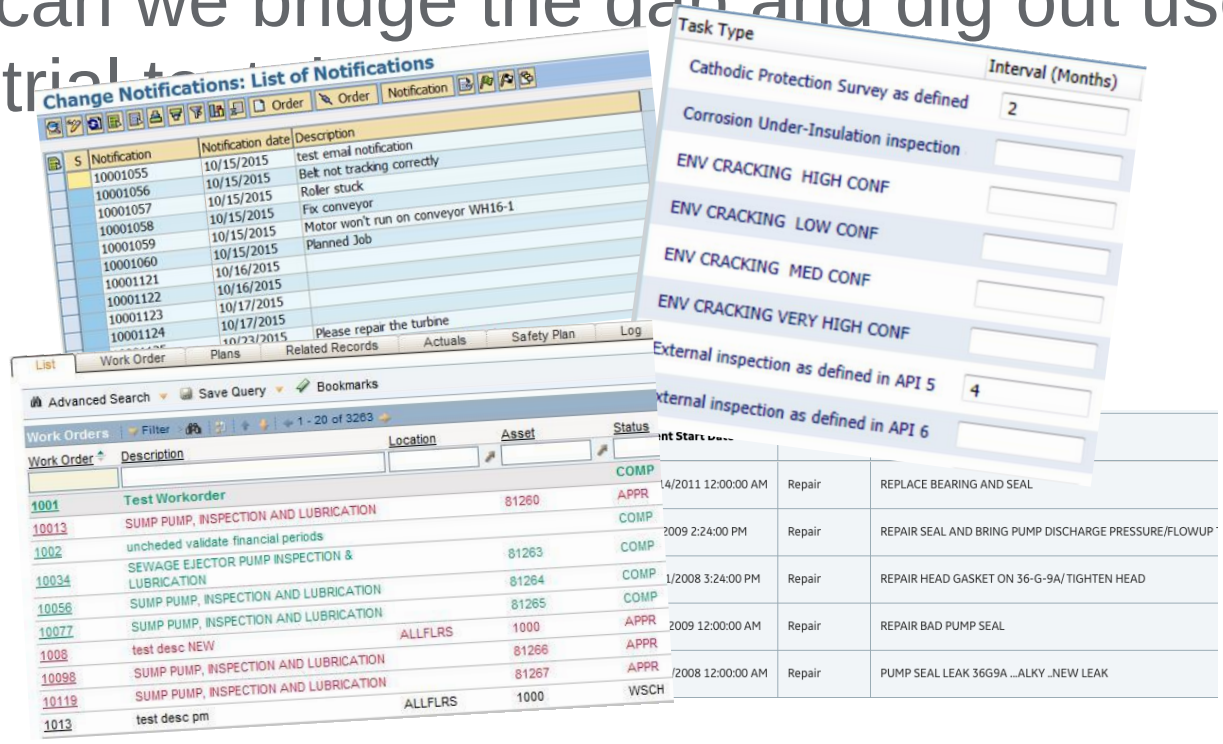
Sarah Lukens, Ph.D.
GE Digital

15 September 2021

Under-valued, under-utilized data today in industrial digital transformation

Large amount of text data which cannot be directly used by most analytics

How can we bridge the gap and dig out useful information from industrial te[illegible]



Technical language is characterized by…

domain-specific jargon

context

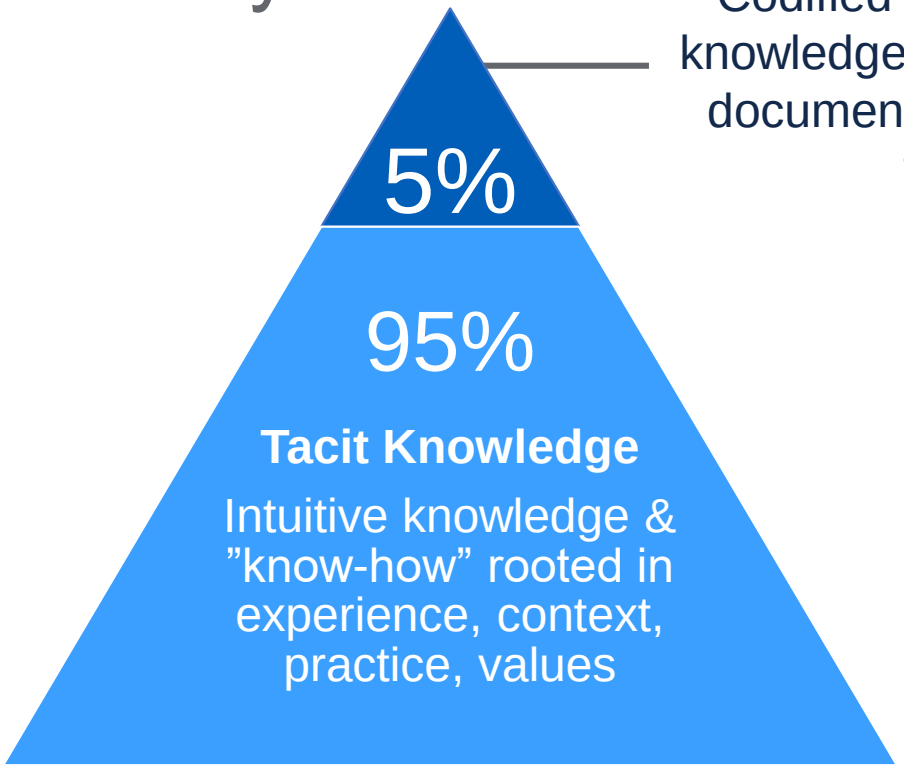
implied knowledge

non-standardized language

missing/inaccurate/ambiguous data

lack of labels

“Small data”

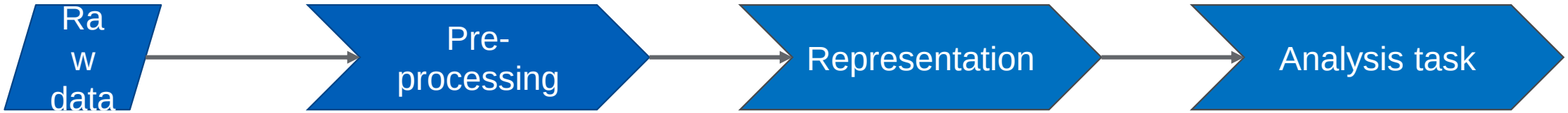


Explicit Knowledge
Codified or digitalized knowledge found in data, documents, records or files

#	Work Request Description
1	CL3-012 bfp not workin g also bad seal
2	block valve leaking
3	Monthly inspection by Joe

How do models pre-trained on Wikipedia, news articles, or Tweets behave when applied to work order descriptions?

Core NLP concepts – NLP pipelines and out-of-the-box challenges



Cleans up text such as lowercase, punctuation removal, etc.

Converts text to form which can be used by a numerical algorithm

Numerical machine learning algorithm such as neural networks or random forest

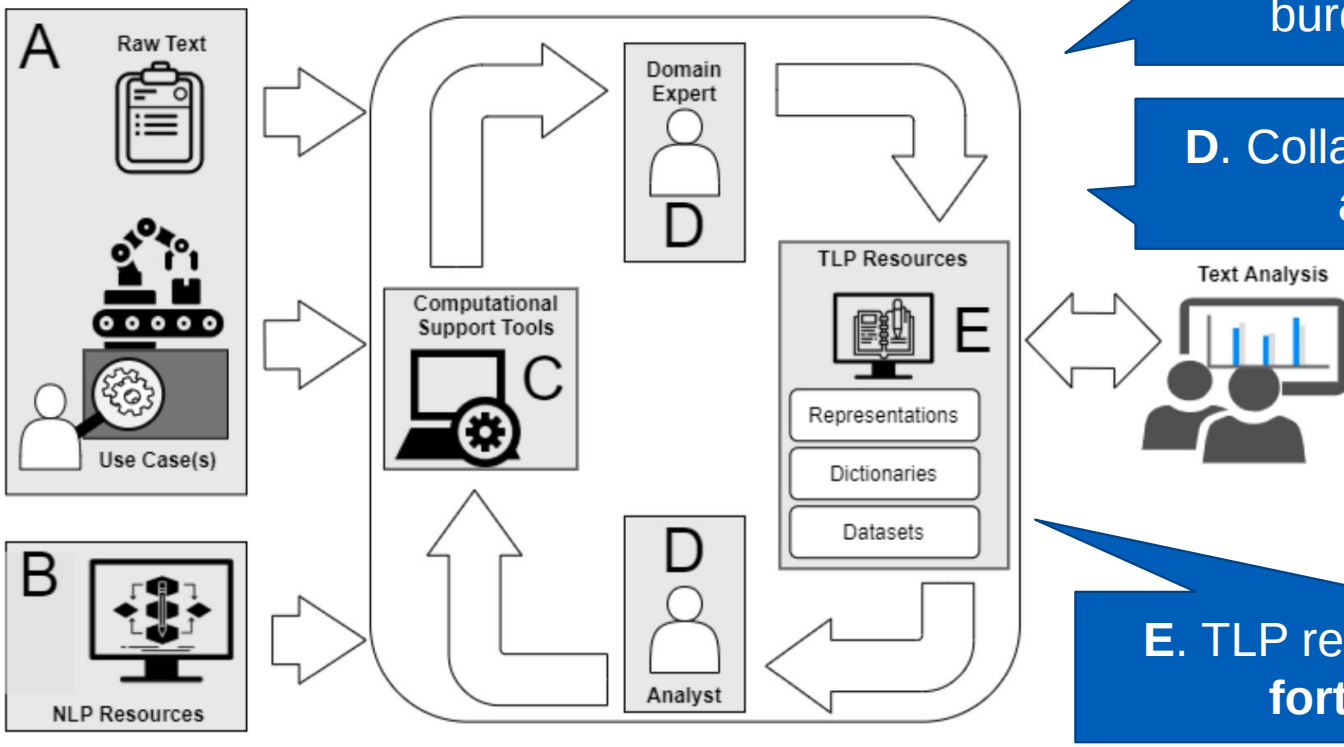
Examples:

Description	Challenge
Pump **not** workin g	**Pre-processing**: Stop word removal - OOTB tools may remove the word “not”
leakage in the **CO2** vlv	**Pre-processing**: Character removal may remove information such as technical abbreviations
leakage in the CO2 **vlv**	**Semantic meaning**: OOTB may not link technical concepts such as “valve” and ”vlv”
Pmp-01 Work Request per **proc 343**	**Missing context**: meaning dependent on knowing procedure 343

Making TLP a reality

A. Use case is explicitly considered as an input

B. Use NLP resources where it makes sense



C. Use computer tools to alleviate burden on domain expert

D. Collaboration between analyst and domain expert

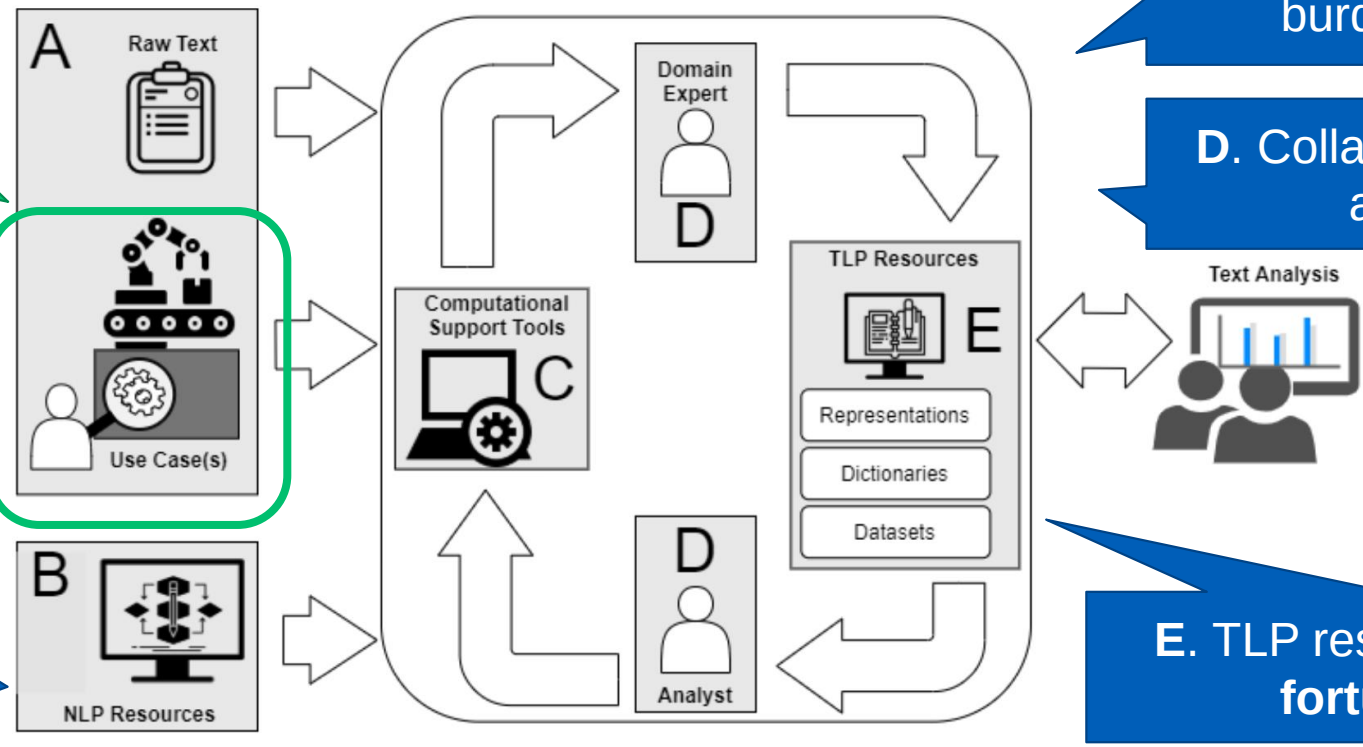
E. TLP resources – such as **fortuitous data**

Brundage, Sexton, Hodkiewicz, Dima & Lukens (2021). Technical language processing: Unlocking maintenance knowledge. *Manufacturing Letters*, *27*, 42-46.

Making TLP a reality

A. Use case is explicitly considered as an input

B. Use NLP resources where it makes sense



C. Use computer tools to alleviate burden on domain expert

D. Collaboration between analyst and domain expert

E. TLP resources – such as **fortuitous data**

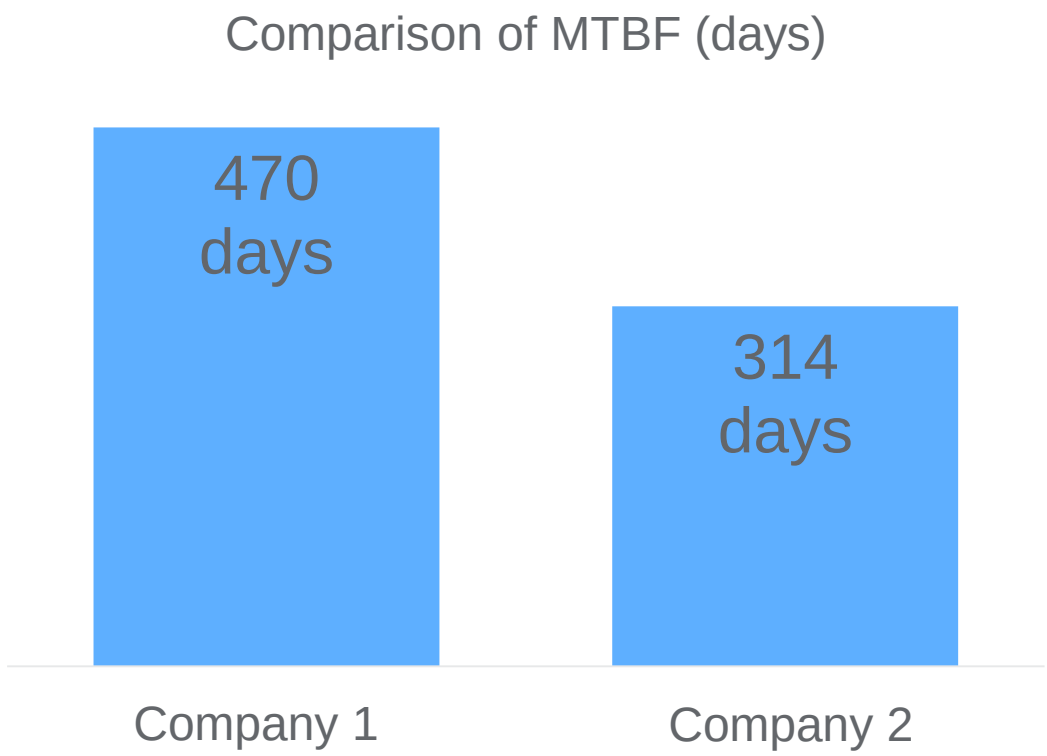
Brundage, Sexton, Hodkiewicz, Dima & Lukens (2021). Technical language processing: Unlocking maintenance knowledge. *Manufacturing Letters*, *27*, 42-46.

Use case #1: Enables consistent reliability metrics

Before: Inability to calculate Mean Time Between Failure (MTBF)

After: Benchmarking comparison of MTBF is possible

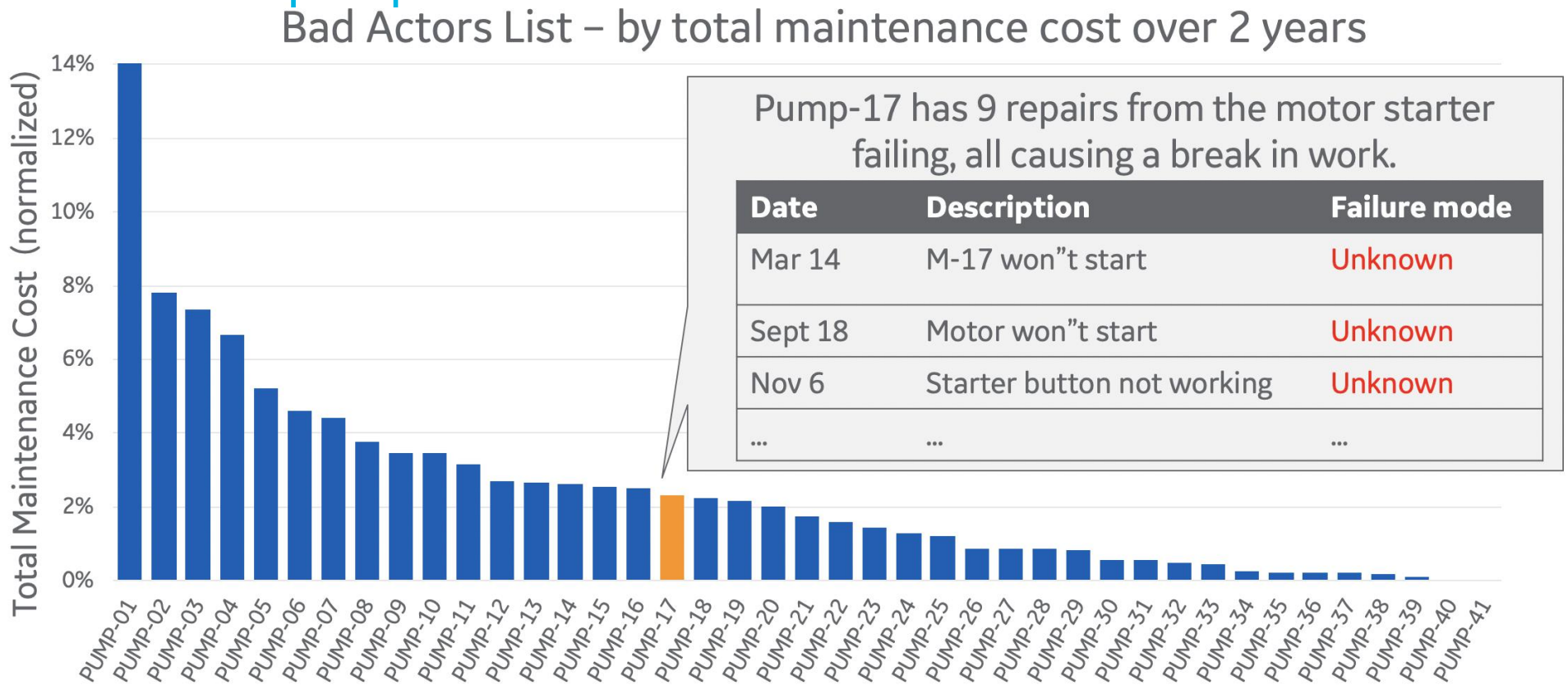
Description	Before: Breakdown Indicator	After: Is a Failure event?
Seal is leaking badly	FALSE	True
Block valve is broken open and inoperable	FALSE	True
00120-Pump 1 work request	FALSE	False
Check impeller size	FALSE	False



Lukens, S., Markham, M., Naik, M., & Laplante, M. (2019). Data-driven approach to estimate maintenance life cycle cost of assets. *Model-Based Enterprise Summit (MBE 2019)*

Use case #2: Finding hidden bad actors

Root cause analysis revealed no proper procedures in place for motor maintenance. Once remedied, failures ceased with estimating annual savings of $54,000 for the one pump alone

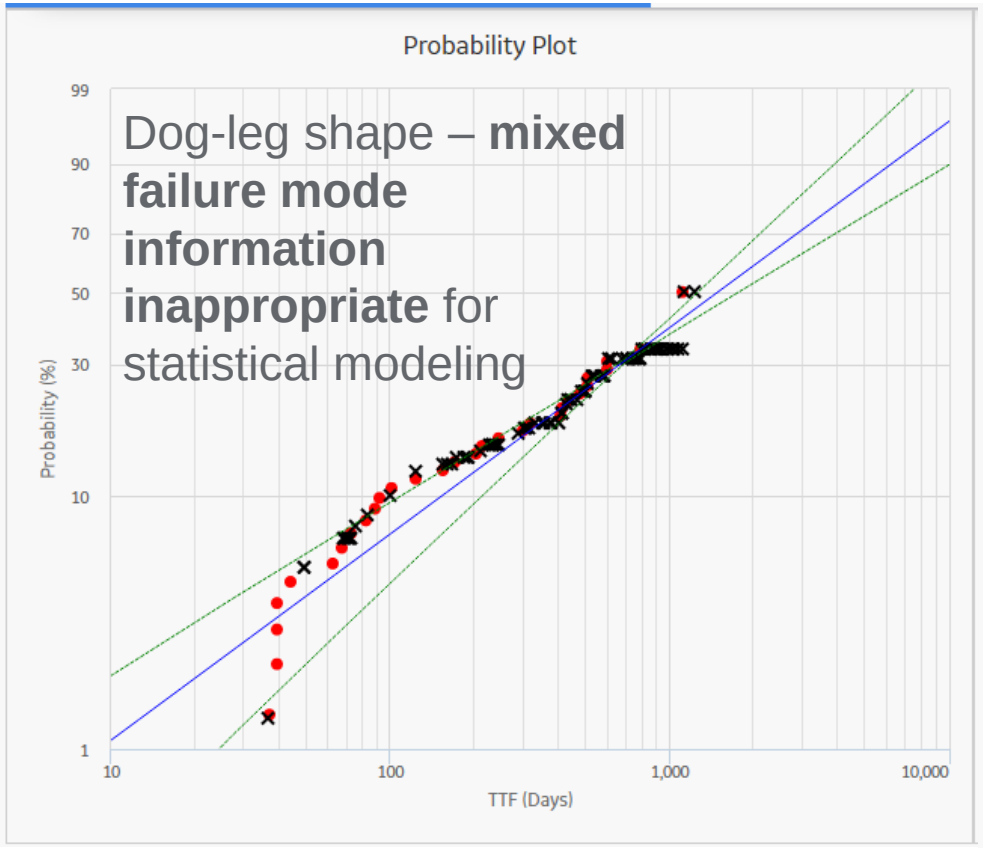


https://www.ge.com/digital/blog/technical-language-processing-how-put-industrial-data-work

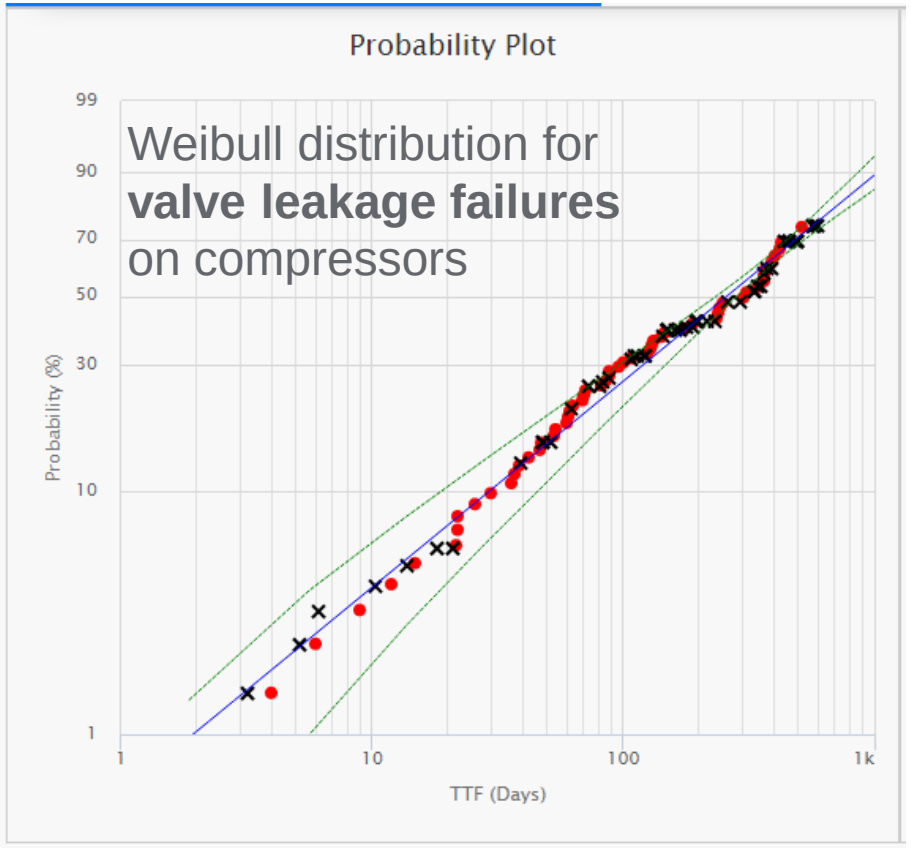
Use case #3: Enables reliability distribution fitting

Failure mode characterization can be used for reliability-data based survival models such as Weibull analysis

Before

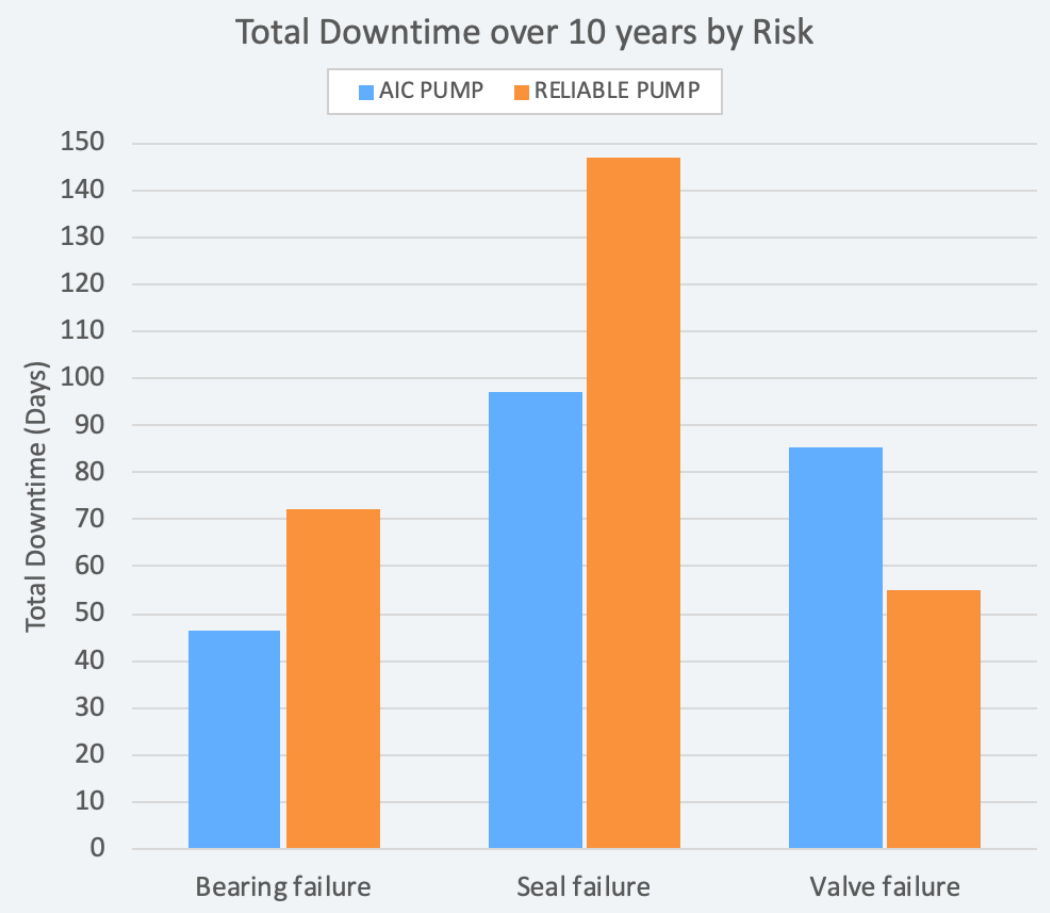


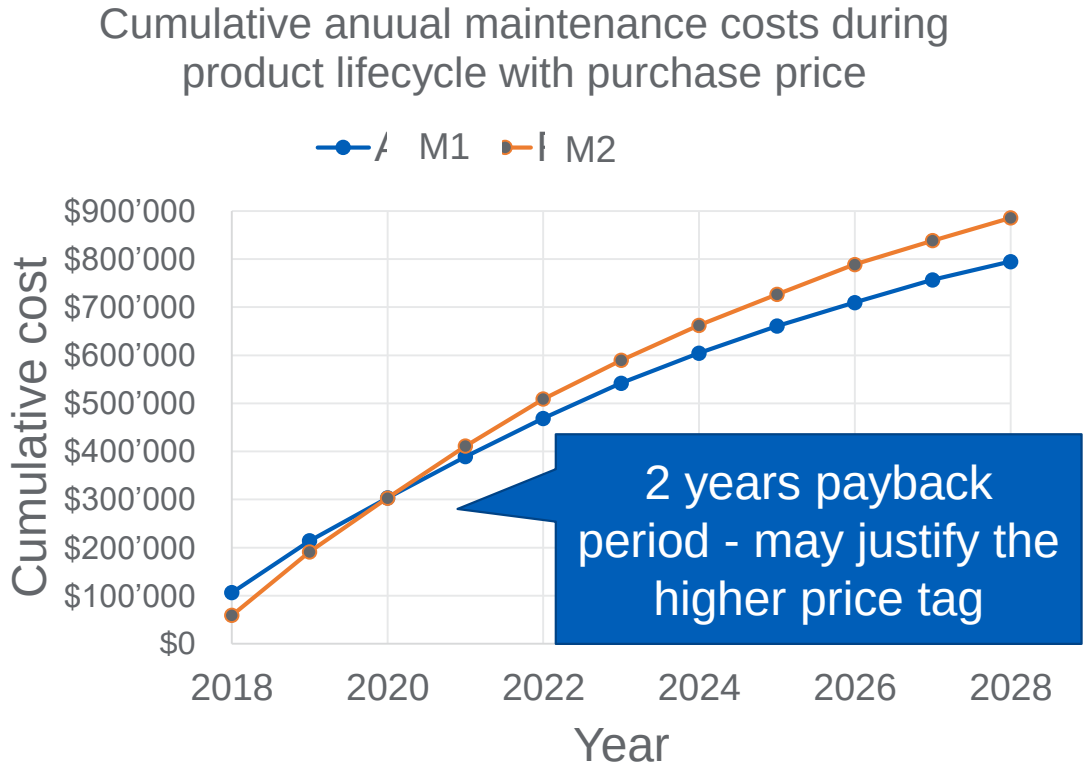
After



Reliability distribution fitting can be used for decision making

Repair or replace? Simulation model:





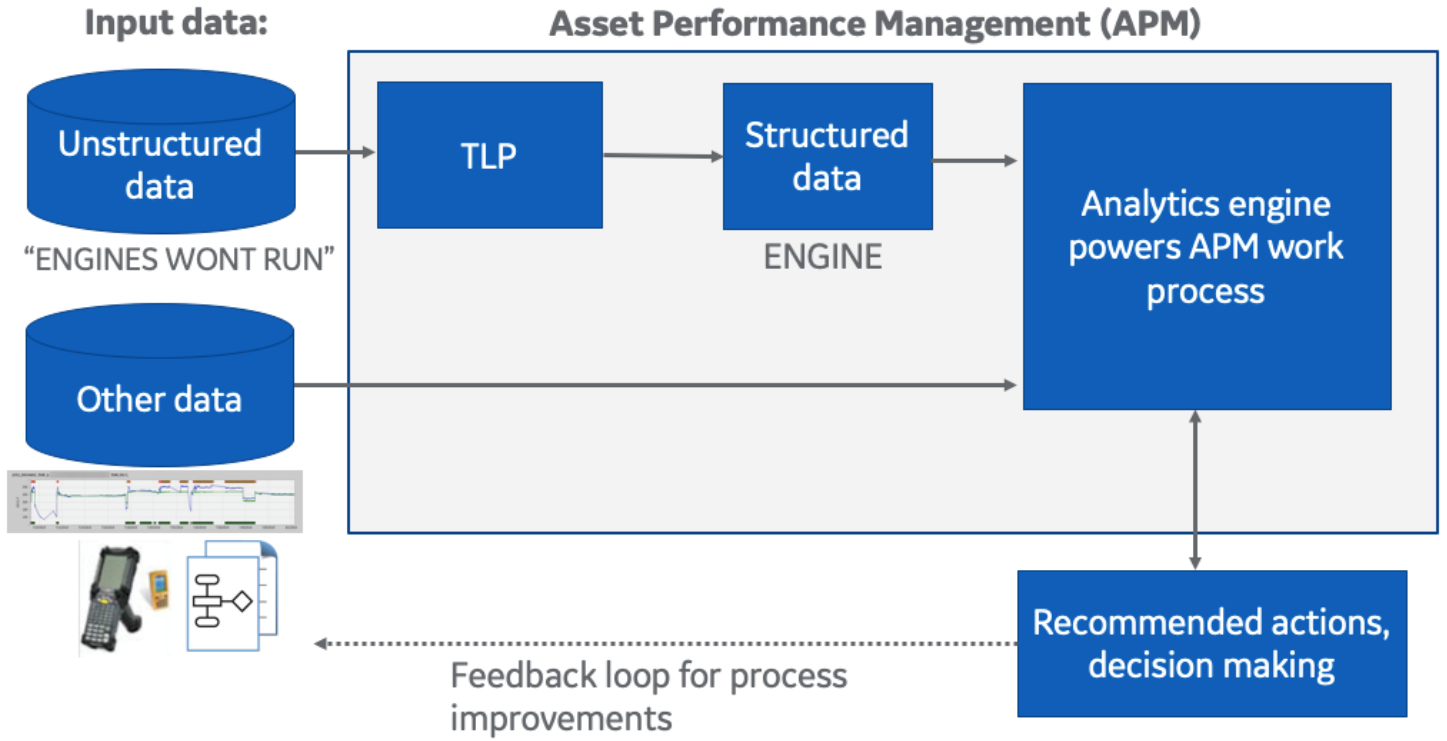
Model 1 is more reliability but costs more than Model 2

https://www.ge.com/digital/blog/digital-twin-approach-designing-cost-effective-maintenance-strategies

Putting industrial data to work with TLP

TLP enables industrial companies to make decisions using their historical data.

The value of TLP comes with integrating the data into business processes in ways that create value.



Thank you!

Contact me: sarah.lukens@ge.com

Making TLP a reality

Data Representations & Feature Engineering

- Need for use case-driven representations
- Need for computational support tools to assist in annotation of most-used concepts

Entity Types Definitions and Dictionaries

- No wide-spread community consensus or adoption exists for agreed on entity sets or hierarchies in maintenance.

Raw and Annotated datasets

- Industry perception that datasets are valuable intellectual property.

Verification & Validation approaches

- High fidelity requirements in industrial applications