



Milan

# SIGMAFINE SUMMIT 2019™

engineering data for digital transformation

## Engineering Data for Digital Transformation

How to transform process data & information into digital assets

**Bernard Morneau**  
Business Development Executive  
Pimsoft Inc.

**PiMSOFT**  
INNOVATIVE SOLUTIONS

# Does Data Quality matters in what we do?



**Let's pretend it does, just for today...**

# Our “Common Interest” has a name



# Our “Common Interest” has a name



Milan

**SIGMAFINE SUMMIT 2019**™  
engineering data for digital transformation





# Data Engineering

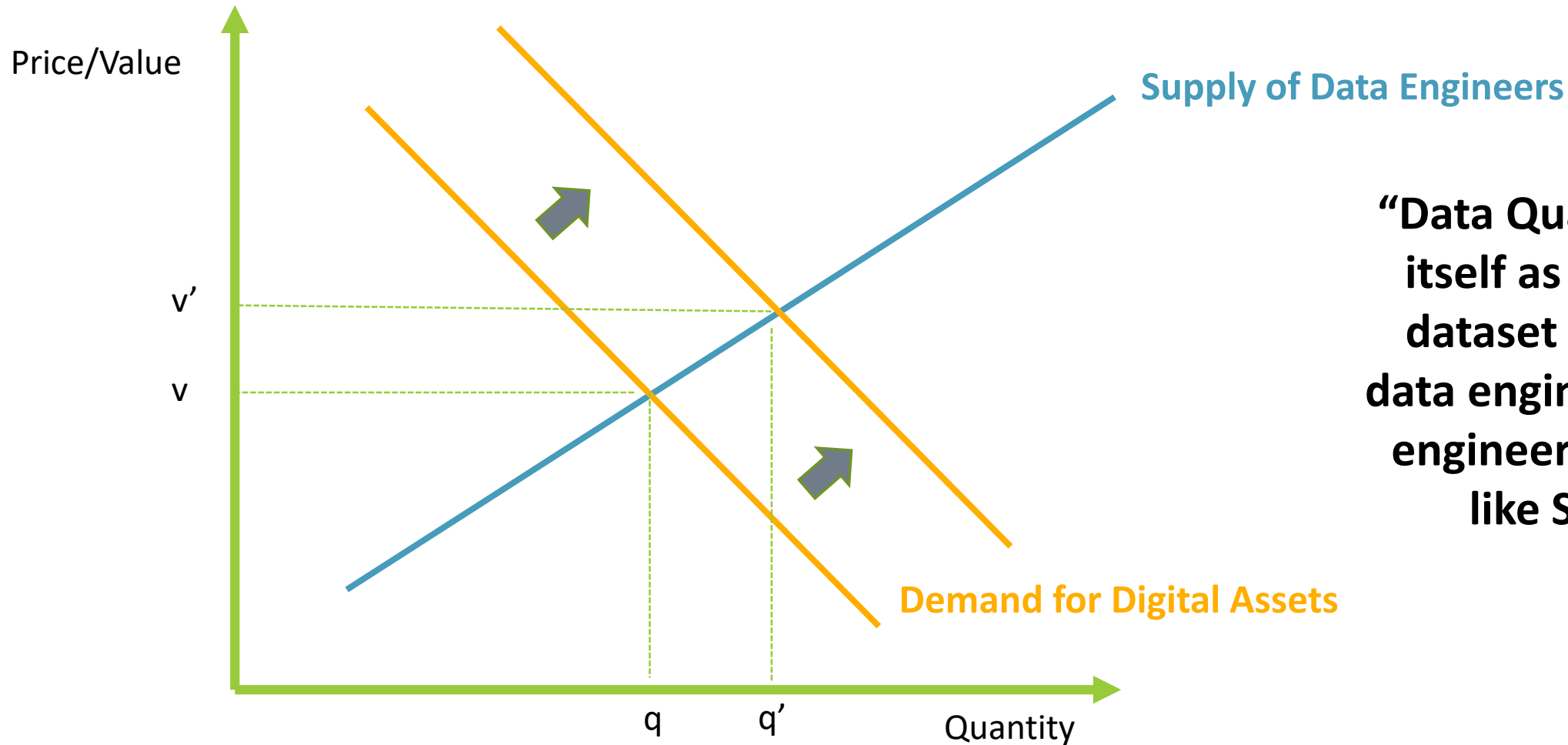
“...all the aspects of collecting, validating, reconciling, organizing, modeling, conditioning & shaping data into usable datasets

...implementing this practice as an enterprise-level service carrying out this task as autonomously as possible,

...ensuring sustainable data quality to users, applications, business processes & advanced analytics & simultaneously providing prescriptive KPI's to improve data quality over time.

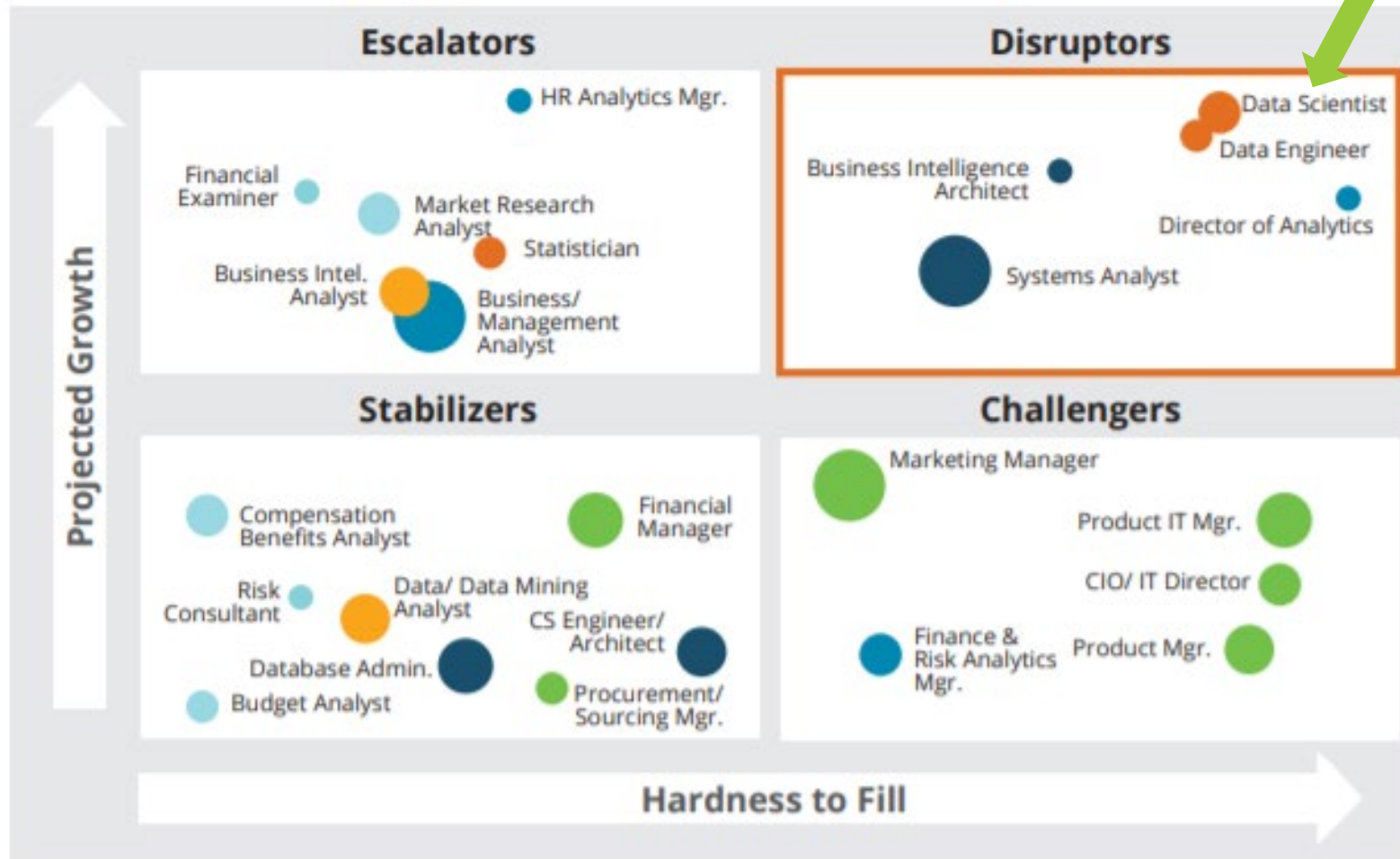
This service must also be able adapt to the evolving analytics mandates of the enterprise...”

# Increasing demand for “Digital Assets”



**“Data Quality manifest itself as conditioned dataset prepared by data engineers and data engineering solutions like Sigmafine**

# The Quant Crunch (IBM)



## Key:

- Data Driven Decision Makers
- Functional Analysts
- Data System Developers
- Data Analysts
- Data Scientists & Adv. Analysts
- Analytics Managers

## Postings:

- Less than 5,000
- 5,000 - 10,000
- 10,000 - 50,000
- 50,000 - 150,000
- +150,000

Miller, S. et al., The quant crunch, IBM, 2017 (<https://www.ibm.com/downloads/cas/3RL3VXGA>) / <https://www.burning-glass.com/research-project/quant-crunch-data-science-job-market/>

## Top Trending Job Titles (U.S., Y/Y)

JOB	INCREASE Y/Y (%)
Data Engineer	88
Actuary	75
Computer Scientist	62
Video Game Designer	58
Geographer/GIS Specialist	57
Business Continuity Planner/Analyst	49
Data Scientist	49
Biostatistician	47
Security Management Specialist	47
Multimedia Designer/Animator	46
UI / UX Designer/Developer	45
Mathematician	41
Web Designer	40
Network/Systems Support Specialist	39
Database Architect	35

<https://insights.dice.com/2019/06/04/data-engineer-remains-top-demand-job/>

Statistics from [www.burningglass.com](http://www.burningglass.com)

### 25 Jobs with the Best Career Opportunities in the U.S.

Rank	Job Title	Career Opportunities Rating	Median Base Salary	Job Openings on Glassdoor
1	Tax Manager	4.1	\$112,021	4,803
2	Salesforce Developer	4.1	\$81,721	3,193
3	Product Designer	4.1	\$102,054	2,045
4	Strategy Manager	4.0	\$142,328	3,131
5	HR Manager	4.0	\$84,700	4,351
6	Audit Manager	4.0	\$102,521	3,050
7	Data Scientist	3.9	\$110,160	6,789
8	Business Development Manager	3.9	\$80,341	6,982
9	Java Developer	3.9	\$82,087	9,367
10	Marketing Manager	3.9	\$83,513	8,194
11	Product Marketing Manager	3.9	\$121,780	2,319
12	Mobile Developer	3.8	\$94,555	2,097
13	Devops Engineer	3.8	\$107,619	5,505
14	Communications Manager	3.8	\$80,188	2,434
15	Accounting Manager	3.8	\$86,562	3,784
16	Data Engineer	3.8	\$101,527	5,489
17	Scrum Master	3.8	\$97,771	2,580
18	Dentist	3.8	\$140,298	3,339
19	Physician Assistant	3.7	\$107,356	10,061
20	Product Manager	3.7	\$117,113	12,865
21	Compliance Manager	3.7	\$96,580	3,030
22	Nursing Manager	3.7	\$85,181	14,755
23	Solutions Architect	3.7	\$130,595	7,850
24	QA Manager	3.7	\$91,660	2,010
25	Software Development Engineer	3.6	\$117,250	9,659

<https://www.glassdoor.com/research/studies/best-career-opportunities/>

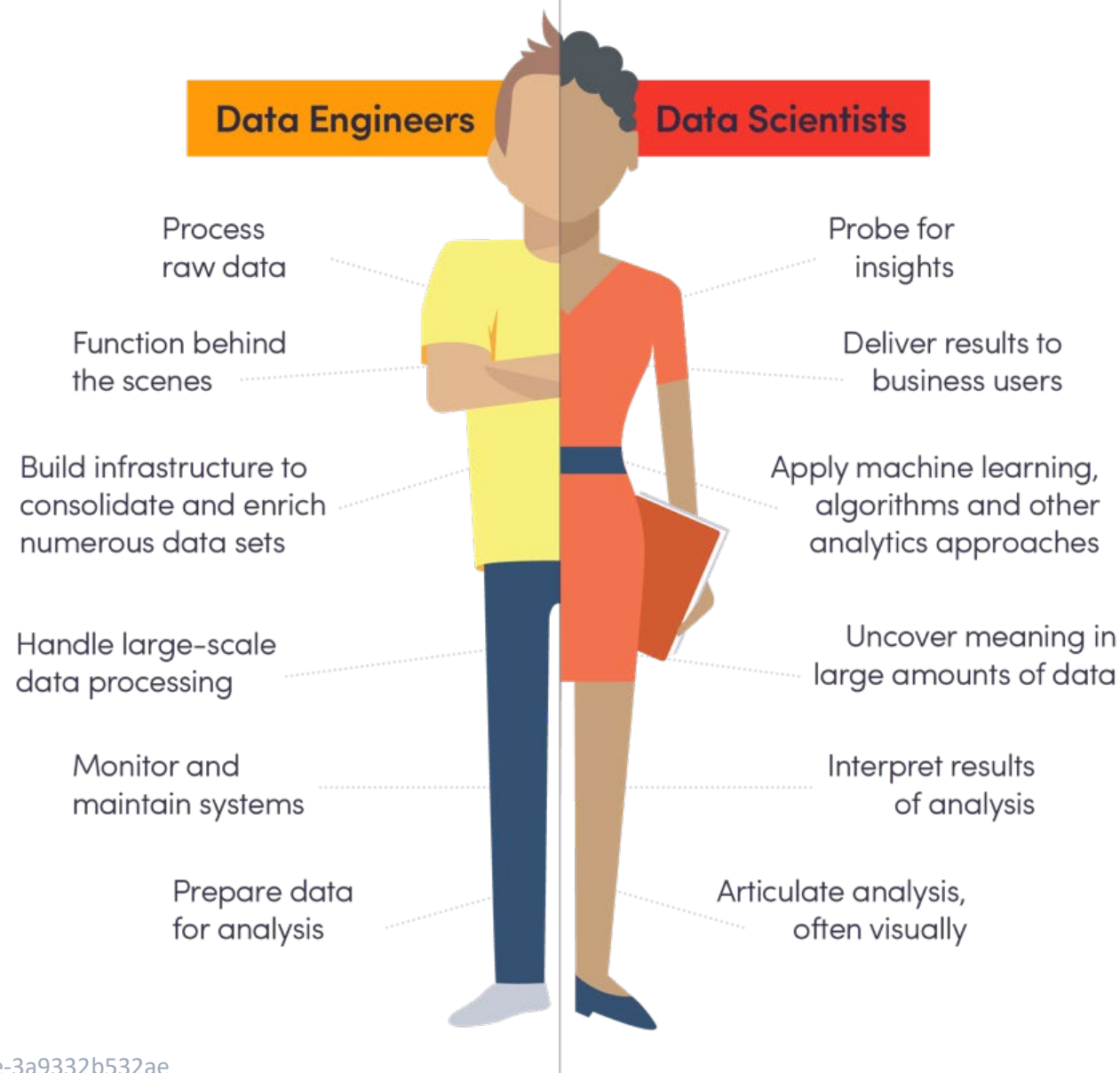




**“...data engineers are actually holding the gateway of innovation...”**

*Lissie Mei, Analytics Consultant at Hilti*

<https://towardsdatascience.com/data-engineering-the-cousin-of-data-science-is-troublesome-3a9332b532ae>



# Passion for Data Quality Recommended Resources

## DATA DOC (TOM REDMAN)



[You tube series on data quality and impact on machine learning](#)

## DAN MYERS



[Conformed dimensions of Data Quality](http://dqmatters.com) (<http://dqmatters.com>)

## IQ INTERNATIONAL

The logo for IQ International, featuring the letters 'iQ' in a stylized blue font followed by the word 'International' in a grey sans-serif font.

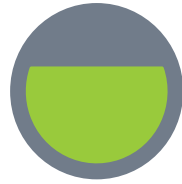
[Non-profit organization with low membership dues, for the passionate](#)

# Points of View on Data Quality



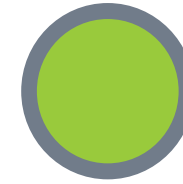
## Raw

- As measured
- Off the sensors



## Cleansed

- Processed at acquisition



## Conditioned

- Post processed
- Model based

**Data Quality Objective**  
➔ **Zero Defects**

**Data Quality Objective**  
➔ **Fitness for use**

# Life cycle of data vs. Data Engineering Activities

Stage 1

Raw Data

Real-Time Data Validation

Data  
Elements

Real-Time Data  
Cross Validation

Real-Time Data  
Aggregation

Real-Time Data  
Filtration

Stage 2

Cleansed Data

Datasets

Model-Based  
Data Validation

Model-Based  
Data Reconciliation

Stage 3

Conditioned Data



# Life cycle of data vs. Data Engineering Activities



**EMERSON™**

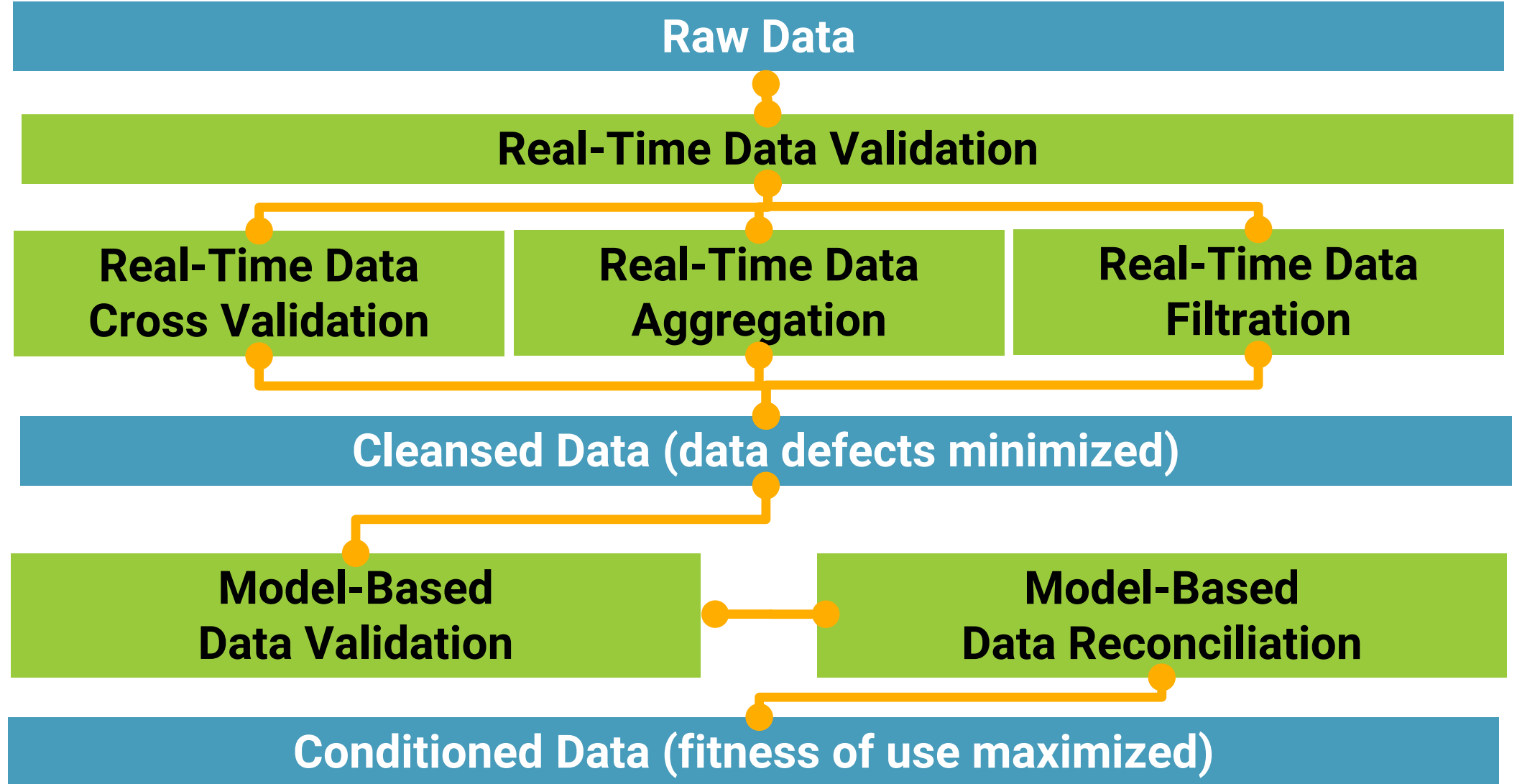
**At Data  
Generation**



**At Data  
Acquisition**



**Post Data  
Acquisition**



# The game we are playing...

## Avoidance game...

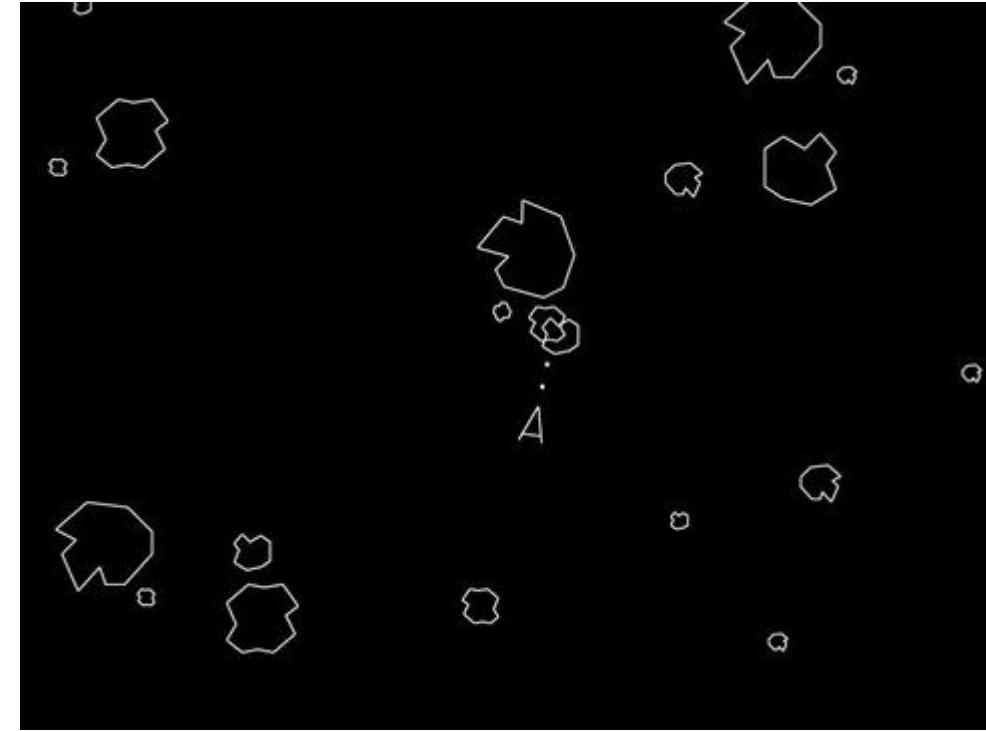
- Data streams & defects occur
- Human vigilance has its limit
- Perfect sensors not available

## Avoiding bad data (zero defect)

- Detect & Mitigate Data Defects
- Eliminate root causes
- Improve measurement systems

## Avoiding the economic downside of bad data/datasets (fitness for use)

- Improve data/information redundancy
- Use computer, technology & Software



Asteroids<sup>®</sup> by Atari

# The game we are playing...

## Avoidance game...

- Data streams & defects occur
- Human vigilance has its limit
- Perfect sensors not available

## Avoiding bad data (zero defect)

- Detect & Mitigate Data Defects
- Eliminate root causes
- Improve measurement systems

## Avoiding the economic downside of bad data/datasets (fitness for use)

- Improve data/information redundancy
- Use computer, technology & Software



**Collision Avoidance Systems**



- 1. The only end goal of “Data Engineering” is “Data Quality”**
- 2. Sustainable “Data Quality” is not a task, it is a process & a multidisciplinary effort**
- 3. “Data Quality” has 2-sided benefit model**
  - Rewarded by good data, conditioned data
  - Rewarded by avoidance of economic downside resulting from bad data





# Our shared Mission



**Make Data Quality  
Great Again!**



Milan  
**SIGMAFINE SUMMIT 2019™**  
engineering data for digital transformation



**Bernard Morneau**

Business Development Executive

Pimsoft Inc.

[bernard.morneau@pimsoftinc.com](mailto:bernard.morneau@pimsoftinc.com)

+1 415 349 1281

**PIMSOFT**  
INNOVATIVE SOLUTIONS

