POLITECNICO
MILANO 1863

Quality Data Analysis
aa. 2023-2024

Bianca Maria Colosimo
Department of Mechanical Engineering
biancamaria.colosimo@polimi.it

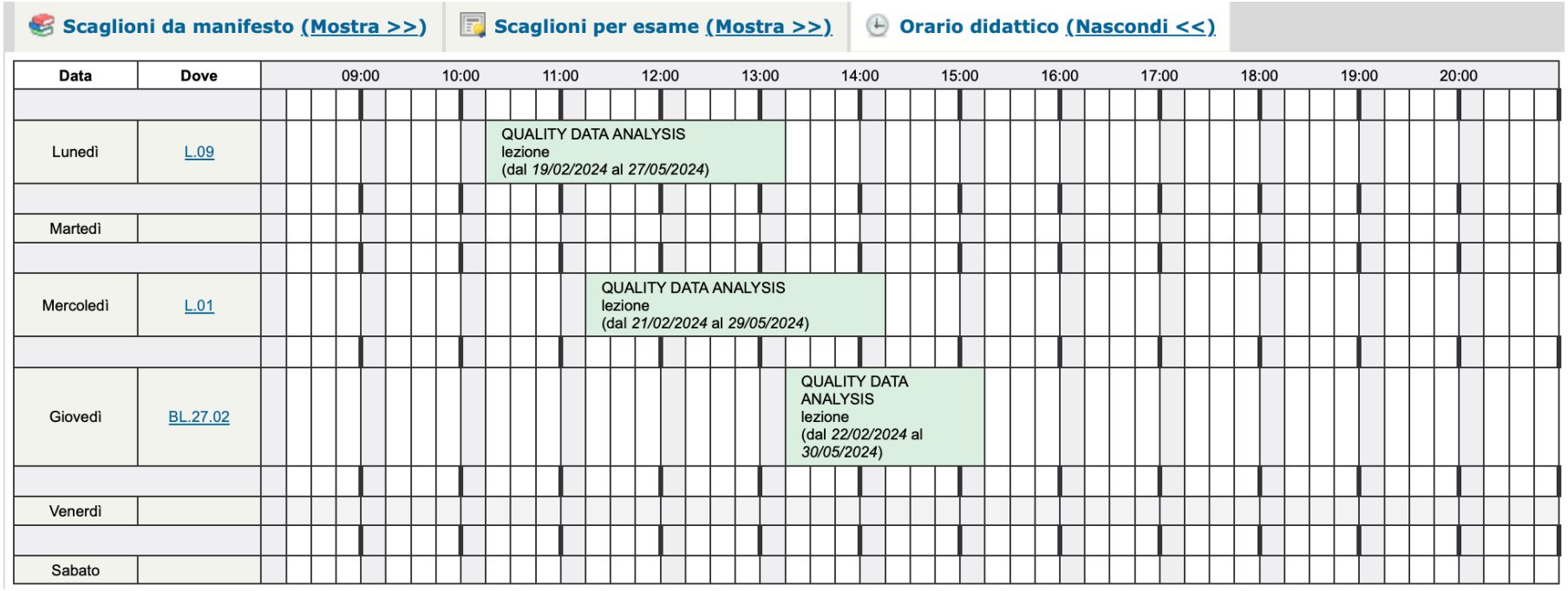
mecc

DIPARTIMENTO DI ECCELLENZA
MIUR 2018-2022


USEFUL INFORMATION

Useful information

1. Timetable:



2. Slides and data set: https://webeep.polimi.it/

Quality Data Analysis – THE TEAM

Lectures

Prof. Bianca Maria Colosimo
(biancamaria.colosimo@polimi.it)
Full Professor

Recitations (esercitazioni)

Matteo Bugatti
(matteo.bugatti@polimi.it)
Assistant Professor

Teaching assistants (lab projects)

- Egon Prioglio, PhD Candidate
- Patrizia Gironi, PhD Candidate
- Alessandro Margarita, PhD Candidate

Full Professor -
Co-founder of the **AddMe Lab**, **IC Labs and 3D cell Lab**

Polimi – Penn State

Member of the European Commission's platform **Manufuture**
Council members of **ASQ**, **Informs QSR** and **Enbis**

Senior Editor- Department Editor:
Progress in Additive Manufacturing- Additive Manufacturing Letters
Informs Journal of Data Science – IISE Transactions
Journal of Quality Technology

2023 Awards:

- **2023 Royal Swedish Academic of Engineering Sciences**
- **2023 EWF – Expert in Additive Manufacturing**
- **2023 ASQ Brumbaugh Award**
- **2023 ENBIS Box Medal Award**

Included among the top 100 Italian woman scientists in **STEM** (https://100esperte.it/)

La nostra rete

End users

Systems producers

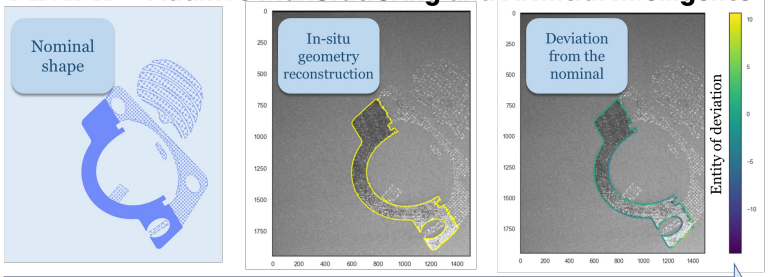
Platforms/networks

Research institutions

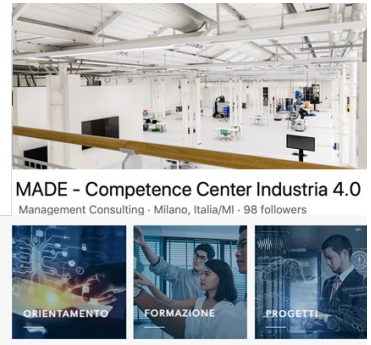
Some selected projects

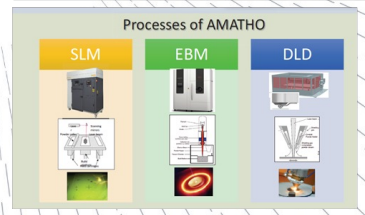
ESA benchmarck center – on structural integrity and in-situ monitoring
Polimi prime contractor – overall budet 500 KE
Polimi budget 300 KE

AMAI – Additive Manufacturing and Artificial Intelligence



Automated and robust in-situ detection of geometrical errors





Funded project (H2020- CleanSky 2)
Total budget POLIMI 1.250.000

MICS
Made in Italy
Circolare e Sostenibile

bio pros

BioProS

Biointelligent Production Sensor to Measure Viral Activity

Some of our current research partners (running research contracts):

SIEMENS

eMOTION of MOTION.
Avio Aero
A GE Aviation Business

Baker Hughes

INGERSOLL
Machine Tools

Awards

premi UCIMU (best thesis award)

Syllabus of the course

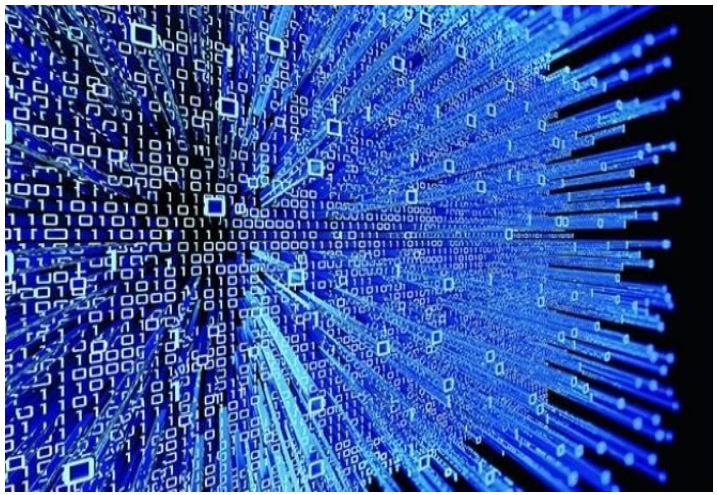
Objectives

Nowadays, an impressive amount of data can be collected in real industrial scenarios (Industry 4.0).
The course presents a set of **quantitative tools and methods** for **managing, modeling, monitoring data** in industrial and business scenarios.
Specific attention is given to **quality** data, i.e., all the key indicators of products and processes which play a relevant role in creating added value for the company.

Target

After successfully completing the course, students should be able to do the following:
1. Understand the philosophy and basic concepts of **quality data modelling**, monitoring and improvement.
2. **Extract relevant information from complex, high-dimensional data**
3. Identify **models to predict the expected pattern** of quality and performance indicators
4. Design and use **appropriate tools** to design and manage data in industrial and service scenarios.

Expected Results



Lectures will provide the basic tools to understand functions and processes, using appropriate quantitative tools for data analysis.

Recitations in computer labs will show how the learned tools can be effectively used to design new solutions using a scientific approach to face the problems at hand (applying knowledge and understanding).

The **lab project** will foster an additional insight to **develop** new ideas and solutions in business and industrial scenarios (making judgements and learning skills).

Requirements

A five-credit course in Statistics is strongly suggested/required.

Students should mainly know:
- basics of statistical distributions (normal, poisson, binomial, Chi-square, F)
- Confidence intervals and Hypothesis tests

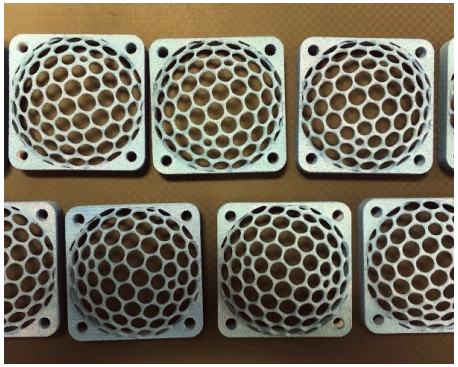
Project activity (last year example)

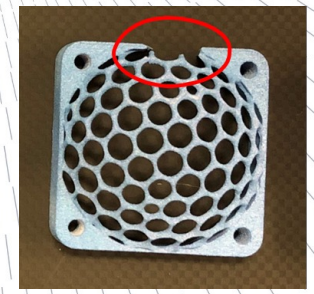
[MADE - Competence Center Industria 4.0 | Milano, Italia](https://www.made-cc.eu)
https://www.made-cc.eu

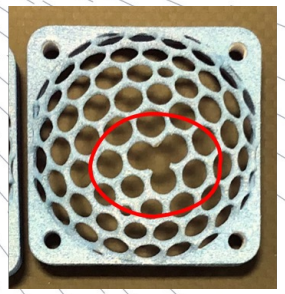
In-line vision-based quality analysis for automated defect detection (last year)

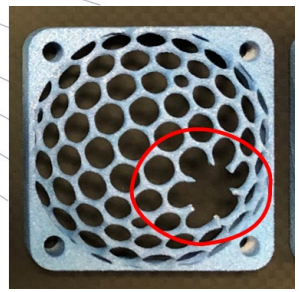
(in collaboration with Rockwell Automation and Siemens)

Test case: 3D printed fan covers (with HP Multijet-Fusion)









Problem: design and test your own statistical method to detect quality errors and anomalies affecting the 3D printed parts based on in-line gathered images

Topics

1. Quality engineering & Industry 4.0: data analysis as a basic tool for modeling, monitoring, control and improve.
2. Quality data modeling:
   - Standard assumptions and related tests;
   - Modeling patterns via linear models;
   - Modeling autocorrelated data via time series analysis;
   - Modeling survey data: Categorical and ordinal data
3. Quality monitoring of continuous variables
   - Traditional statistical process control (SPC) for the mean and the variance
   - SPC for autocorrelated data: Problems of traditional control charts for autocorrelated data;
   - Model based and model-free approaches for quality control of autocorrelated data.
4. Quality modelling and monitoring for "big" data streams: multivariate data
   - Modeling multivariate data
   - Dimensional reduction via Principal Component Analysis
   - Control chart for multivariate data - controlling the mean and the covariance
5. Toward zero-defect manufacturing: process quality and product specifications. Capability analysis. Univariate and multivariate control charts for small shifts (EWMA, CUSUM).
6. Modelling and monitoring attribute and qualitative data: control chart for defective rates and survey data
7. Quality Improvement – The role of improvement in the six-sigma roadmap. Quality improvement via empirical model building (for management engineering only- hints).

- 5 credit students: cover topics 1 – 2 (end of classes ~middle April)
- 5 credit students in Eng. Math can choose 3-4 (beginning Mid April – contact the prof.)
- 8 credit students: cover topics 1 – 3 (end of classes ~middle May)
- 10 credit students: cover all topics during the whole semester

Exam

- **Textbooks:**
- **D.C. Montgomery: Introduction to Statistical Quality Control – Wiley**
- **Additional textbooks:**
- L. C. Alwan “Statistical Process Analysis” – Irwin Mc Graw Hill
- E. del Castillo: “Statistical Process Adjustment for Quality Control” – Wiley
- Additional material will be provided

Written exam using appropriate SW (minitab, excel, …)
Project work – optional but highly suggested

- The project score sums up to the exam score only if written exam >=18)
- maximum increase of the final grade 3/30
- groups made of min 1, max 4 students
- The project scores are valid only for the sessions of this academic year

Oral exam – optional (only if written exam >=18)

- maximum increase of the final grade 3/30

Contacts

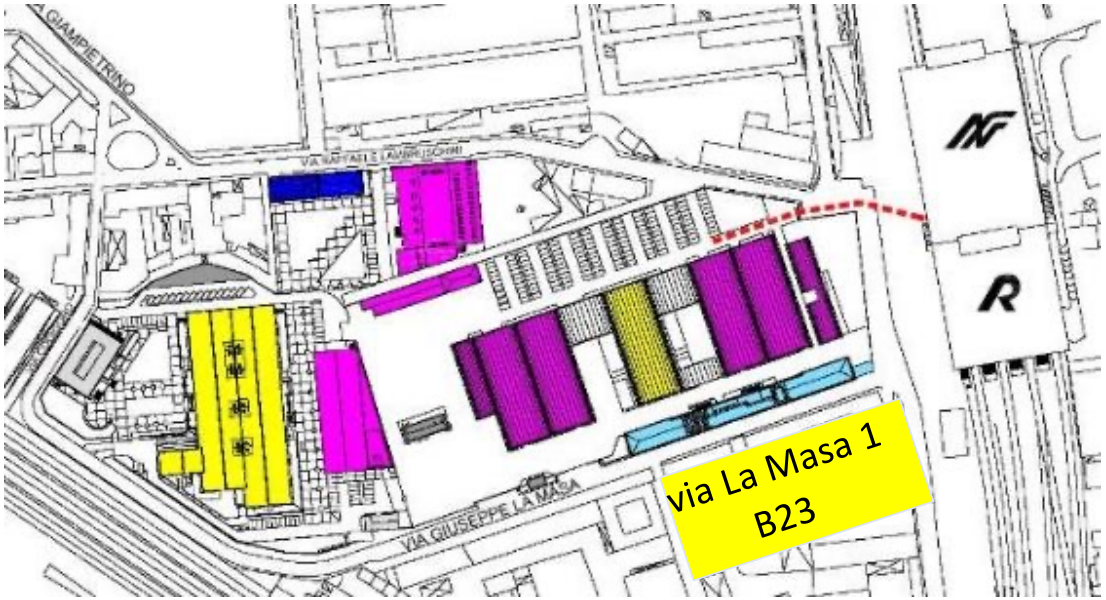
Bianca Maria Colosimo
Dipartimento di Meccanica
(Sesini Building, B23, via La Masa 1)

E-mail: biancamaria.colosimo@polimi.it

Monday 14:00-15:30
(please send an email to book an appointment)

Matteo Bugatti
Dipartimento di Meccanica
(Sesini Building, B23, via La Masa 1)

E-mail: matteo.bugatti@polimi.it



Recitations

We kindly ask you to:

- Bring your laptop computer (with electrical socket if required)
- All the instructions for software upload will be provided via webeep this week


INTRODUCTION

What is this course about?

This course deals with **Quality Data Analysis**
with special attention to
Quality Data Modelling Monitoring and Control

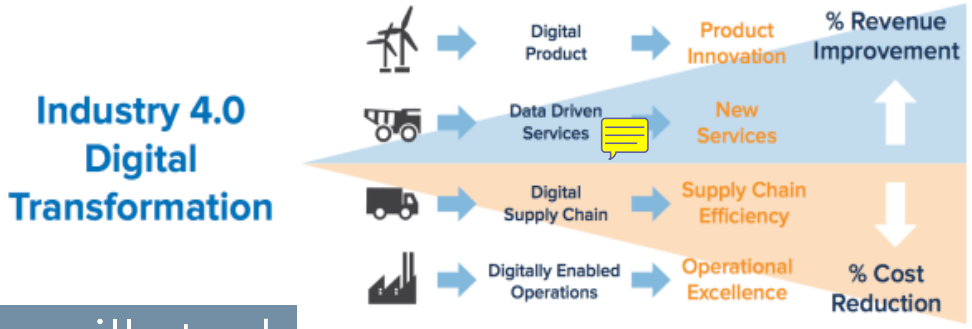
It trains engineers to handle and design quantitative approach for industrial data “science”

Quality Data Analysis

Quality

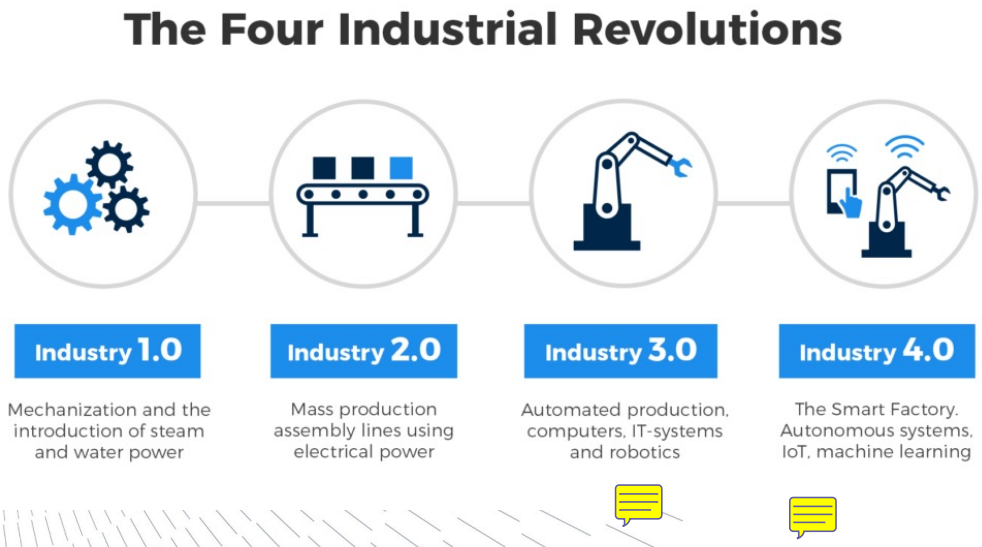
- capability to satisfy (expressed and unexpressed) customer needs - Deming
- All the information affecting the customer satisfaction

Data analysis – modelling and monitoring



All the approaches that we will study can be applicable to other data types, too

Industry 4.0 Is Enabling A New Era Of Manufacturing Intelligence And Analytics



- 72% of manufacturing enterprises predict their use of data analytics will substantially improve customer relationships and customer intelligence along the product life cycle.
- 86% of manufacturers surveyed expect to secure simultaneous gains from both lower costs and added revenue in the next five years.
- 35% of companies adopting Industry 4.0 expect revenue gains over 20% over the next five years.
- Japan and Germany are the furthest along in digitizing internal operations and partnering across their value chains.
- Data analytics and digital trust are the foundations of Industry 4.0.

Forbes August 7th, 2017

Big data - Volume

From Harvard Business Review, 2012:

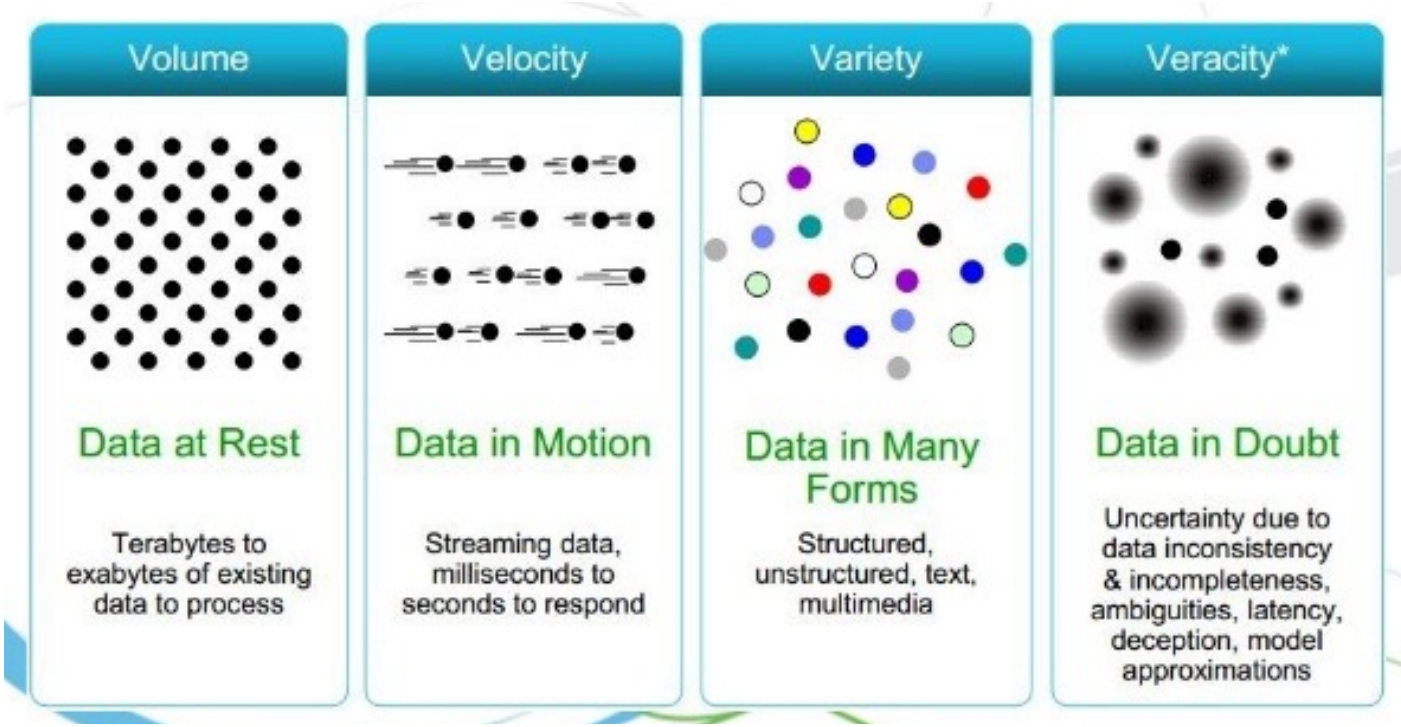
The shortage of data scientists is becoming a serious constraint in some sectors.

DATA

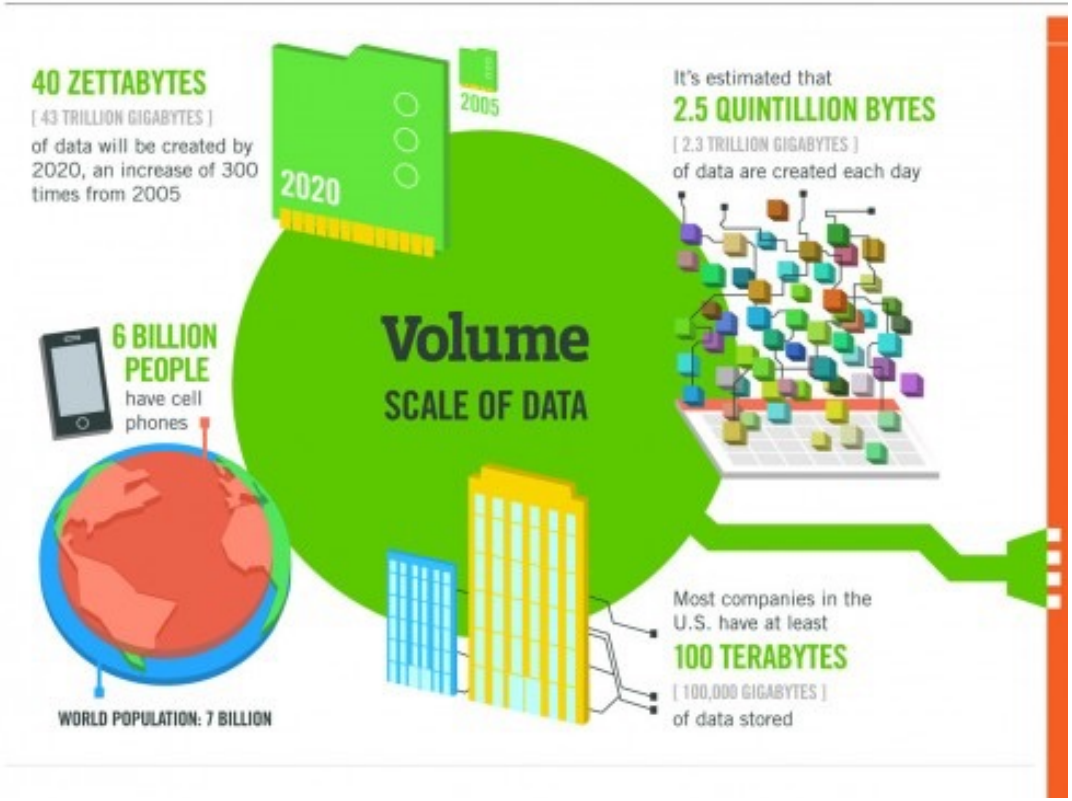
Data Scientist: The Sexiest Job of the 21st Century

by Thomas H. Davenport and D.J. Patil

FROM THE OCTOBER 2012 ISSUE



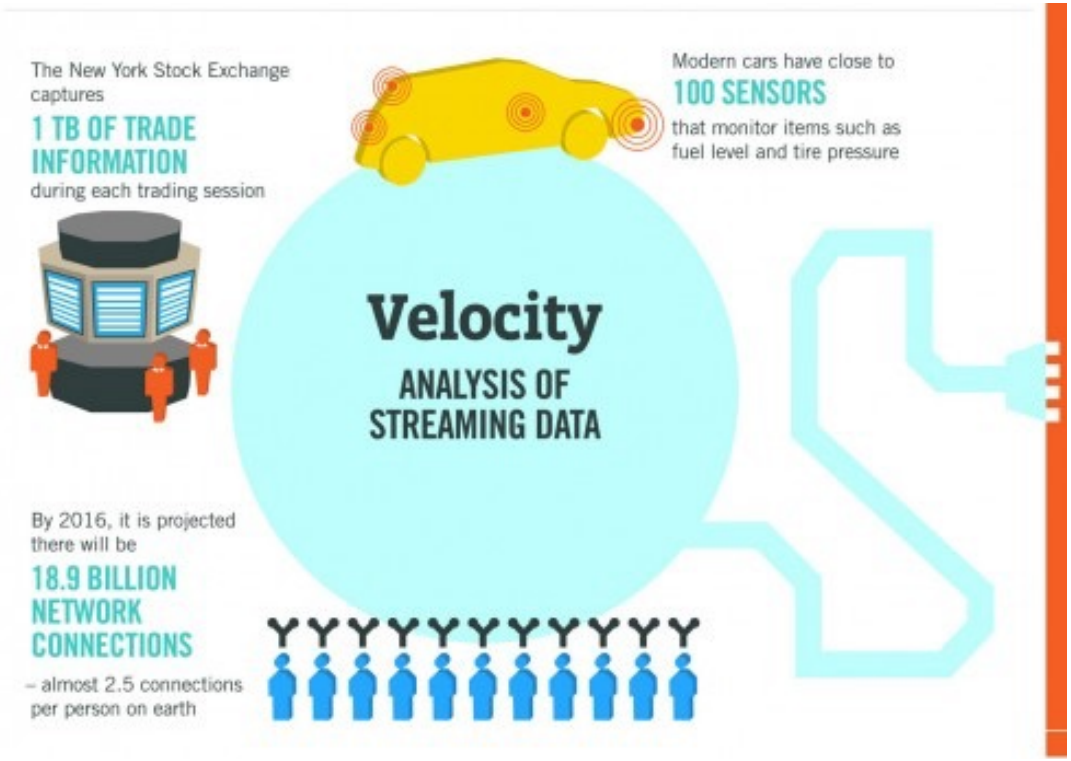
Data in the modern age



Volume refers to the **vast amounts of data** generated every second. We are not talking Terabytes but Zettabytes or Brontobytes. If we take all the data generated in the world between the beginning of time and 2008, the same amount of data will soon be generated every minute! This increasingly makes data sets too large to store and analyse using traditional database technology. With big data technology we can now store and use these data sets with the help of distributed systems, where parts of the data is stored in different locations and brought together by software.

Big data - https://www.linkedin.com/pulse/20140306073407-64875646-big-data-the-5-vs-everyone-must-know/

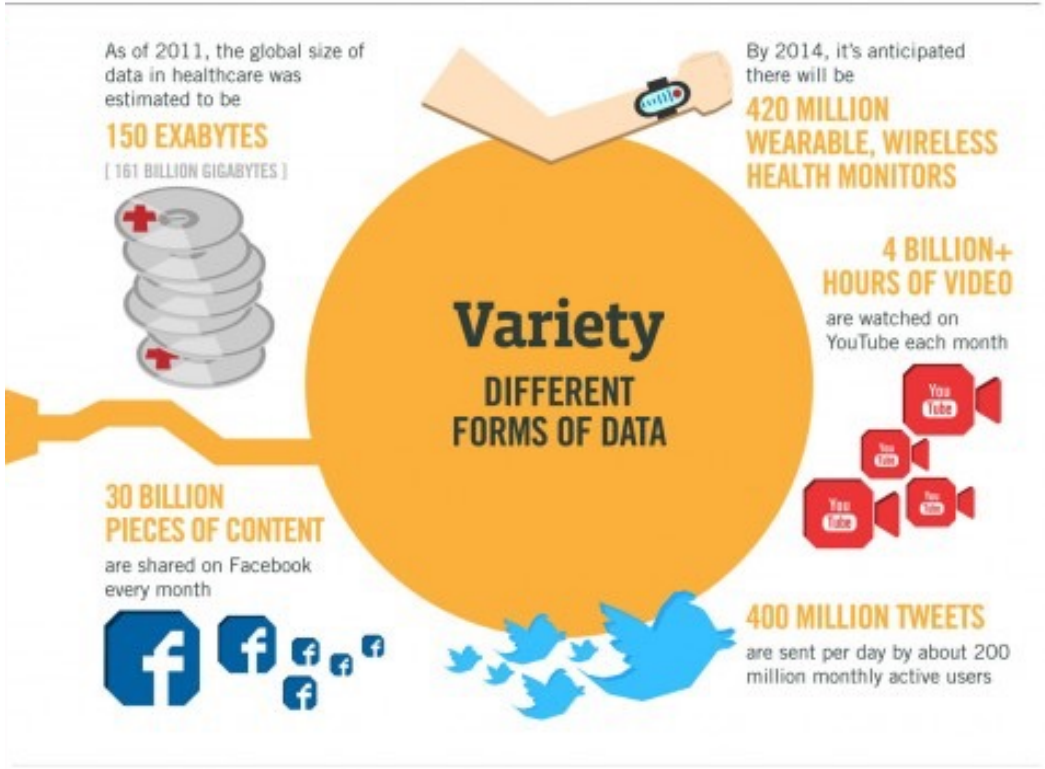
Data in the modern age



Velocity refers to the **speed** at which new data is generated and the speed at which data moves around. Just think of social media messages going viral in seconds, the speed at which credit card transactions are checked for fraudulent activities, or the milliseconds it takes trading systems to analyse social media networks to pick up signals that trigger decisions to buy or sell shares. Big data technology allows us now to analyse the data while it is being generated, without ever putting it into databases.

Big data - https://www.linkedin.com/pulse/20140306073407-64875646-big-data-the-5-vs-everyone-must-know/

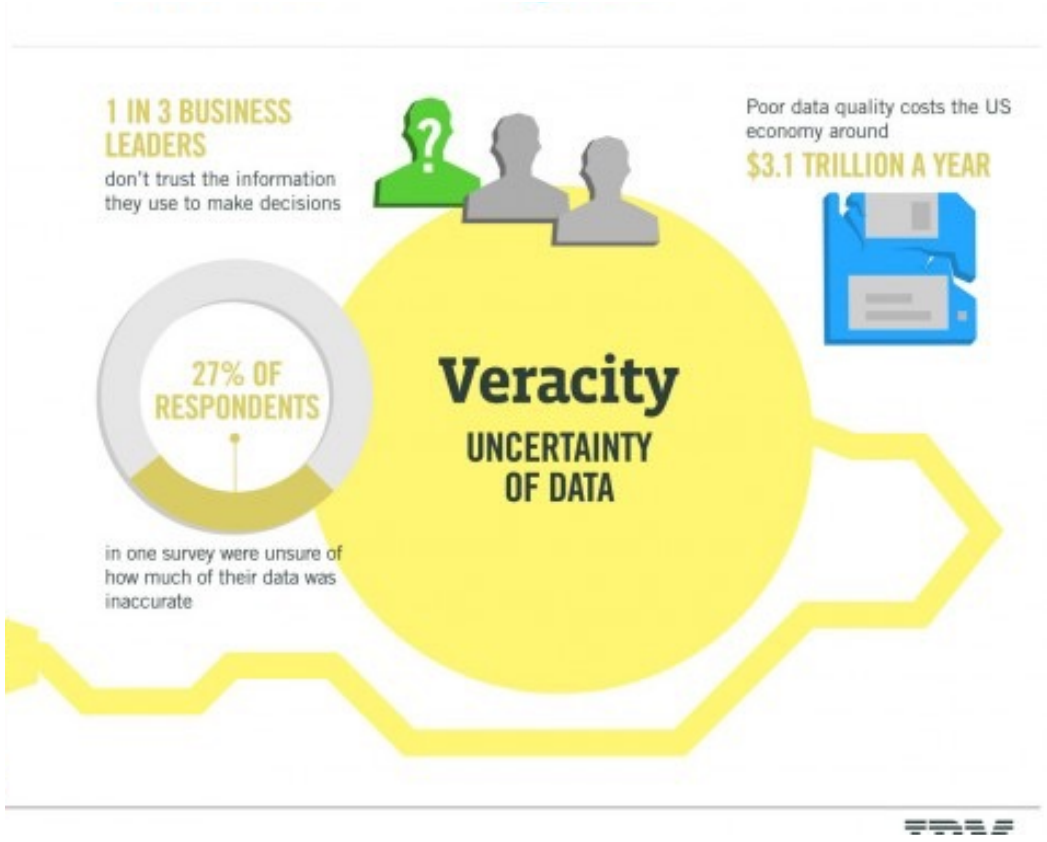
Data in the modern age



Variety refers to the **different types of data** we can now use. In the past we focused on structured data that neatly fits into tables or relational databases, such as financial data (e.g. sales by product or region). In fact, 80% of the world's data is now unstructured, and therefore can't easily be put into tables (think of photos, video sequences or social media updates).

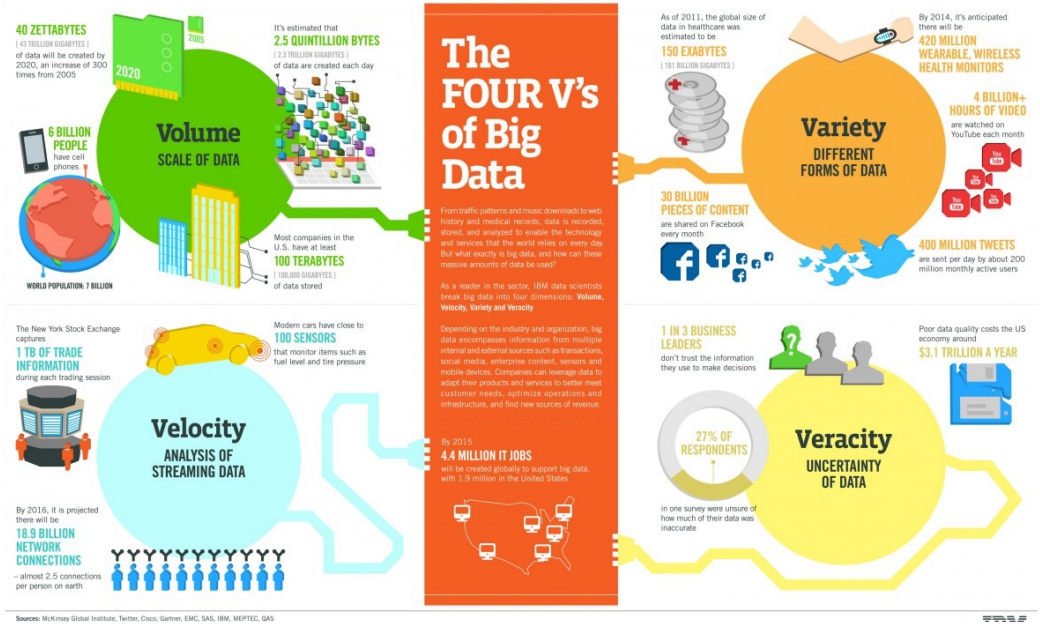
With big data technology we can now harness differed types of data (**structured and unstructured**) including messages, social media conversations, photos, sensor data, video or voice recordings and bring them together with more traditional, structured data.

Data in the modern age



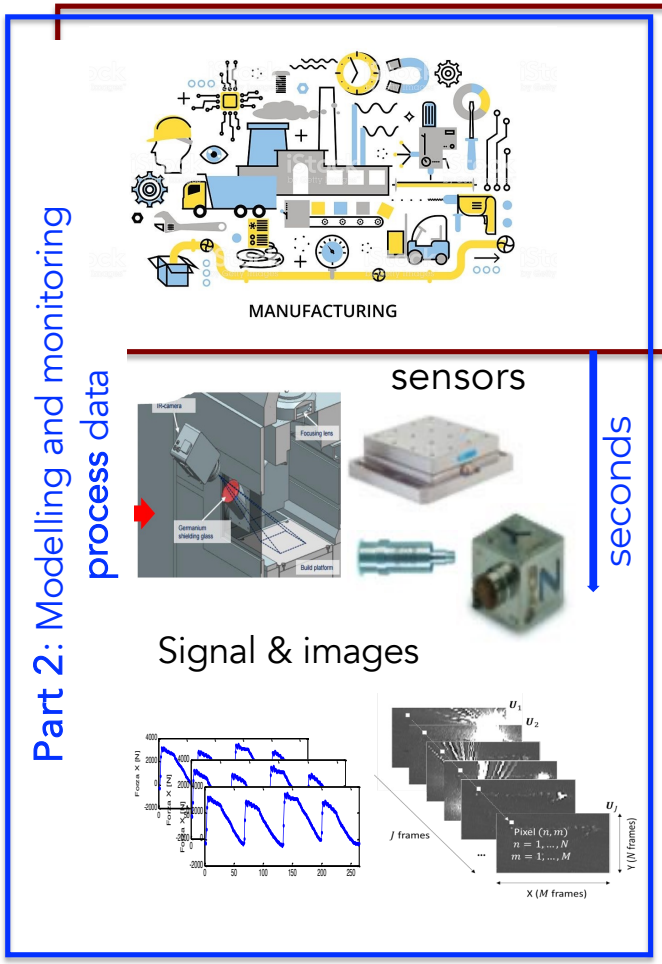
Veracity refers to the messiness or trustworthiness of the data. With many forms of big data, quality and accuracy are less controllable (just think of Twitter posts with hash tags, abbreviations, typos and colloquial speech as well as the reliability and accuracy of content) but big data and analytics technology now allows us to work with these type of data. The volumes often make up for the lack of quality or accuracy.

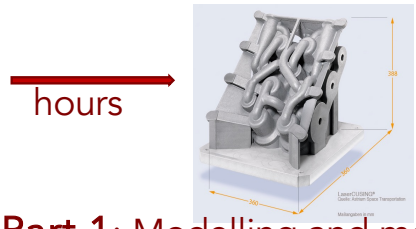
Data in the modern age



- **Value**: Then there is another V to take into account when looking at Big Data: Value! It is all well and good having access to big data but unless we can turn it into value it is useless. So you can safely argue that 'value' is the most important V of Big Data. It is important that businesses make a business case for any attempt to collect and leverage big data. It is so easy to fall into the buzz trap and embark on big data initiatives without a clear understanding of costs and benefits.

Quality Data Analysis

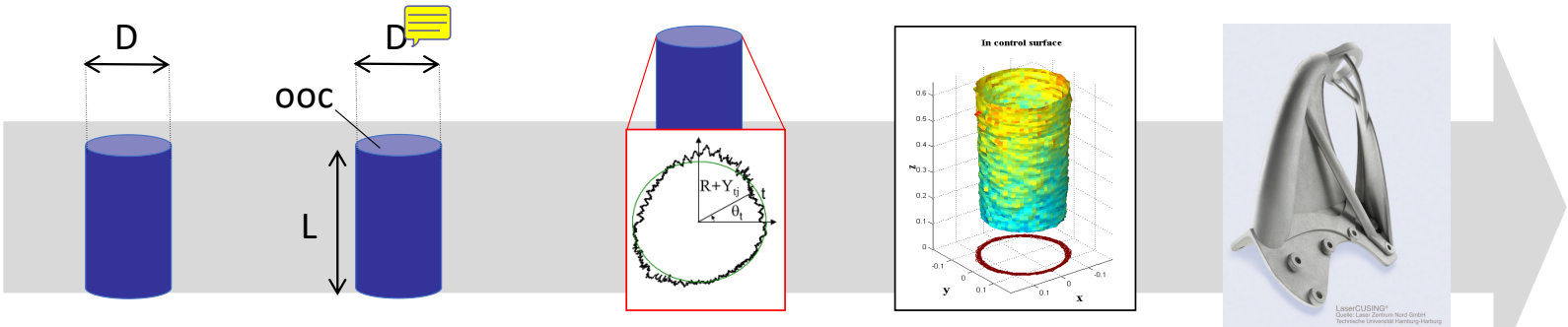




We are assisting to a paradigm shift in product and process data

- from simple to complex data (image, point clouds, voxels, network data)
- from normal size to massive data
- from slow to fast data
- from single to multiple streams

Product data



$X = D \in \mathbb{R}$

One quality characteristic

X=[D,L,OOC,…]'

$\mathbf{X} \in \mathbb{R}^p$

p quality characteristics

$X = f(\theta) + \varepsilon$

Quality profiles

$X = f(\theta, z) + \varepsilon$

Quality Surfaces

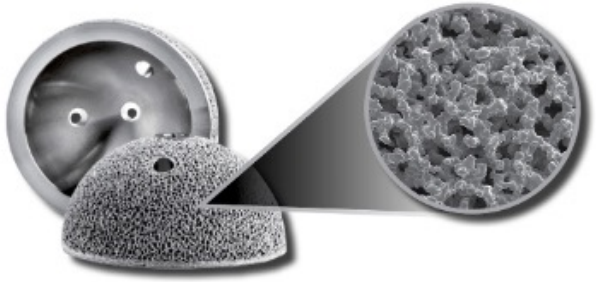
?

Complex shapes

Lightweight structures

Complex shapes (topological optimization)

Internal channels and features

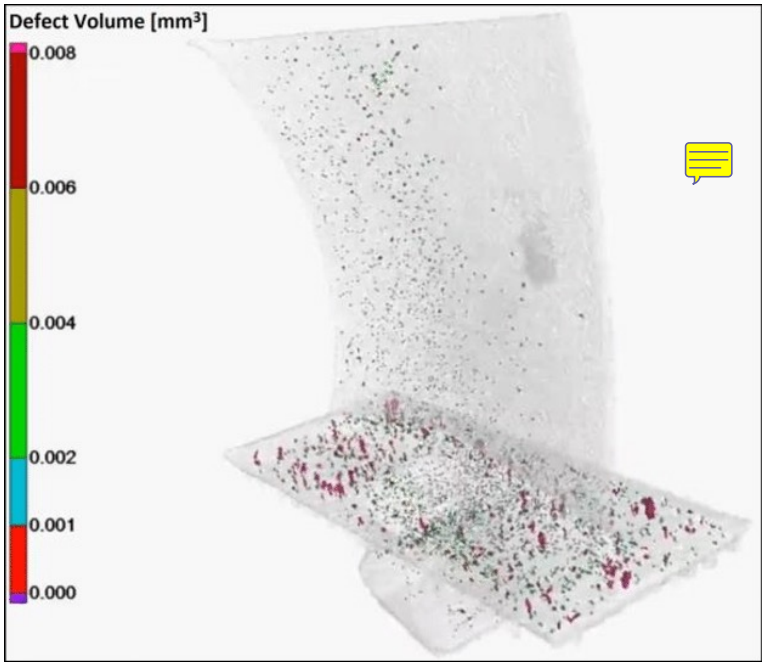


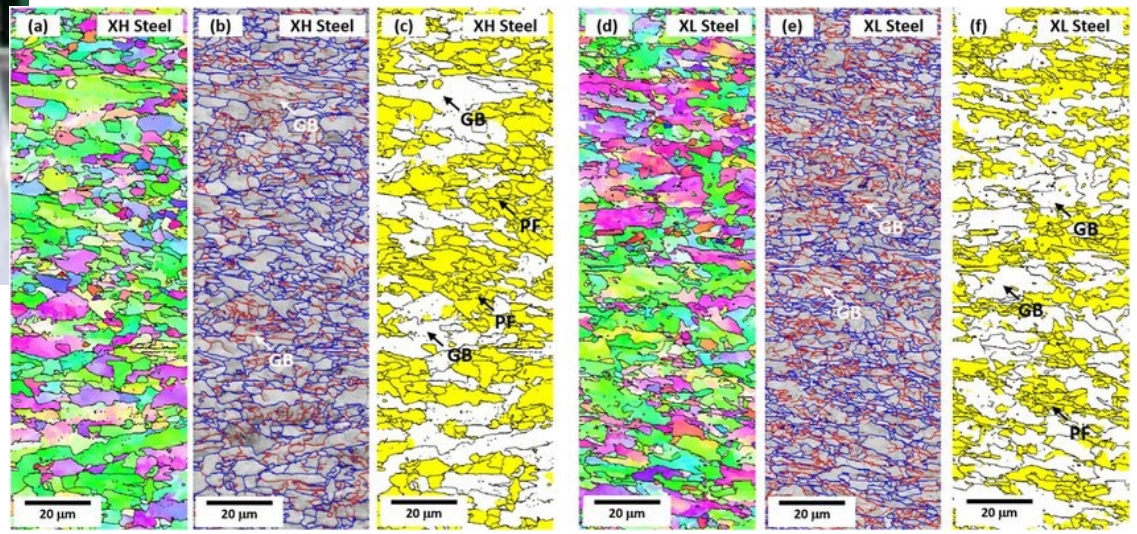
Surface and functional patterns

Airbus A380 bracket

Source: Concept

Product data

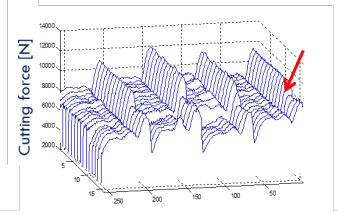




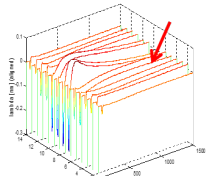
Data 4.0

signals

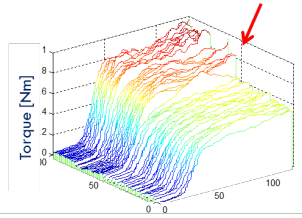
Milling hard-to-cut materials



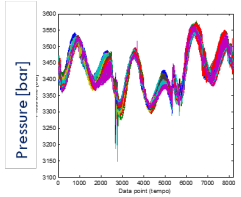
Drilling of multi-layer materials (aerospace)



Tapping monitoring



Waterjet and Abrasive Waterjet cutting

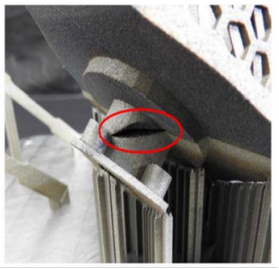


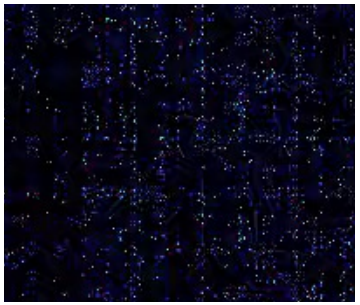
Images /video

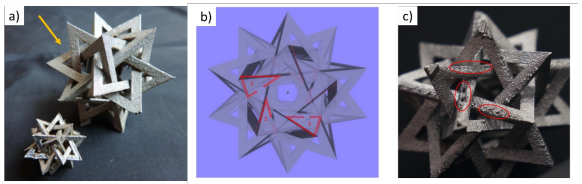
POST-SCAN

Red line: nominal contour
Blue area: detected departure between in-situ reconstruction and nominal contour

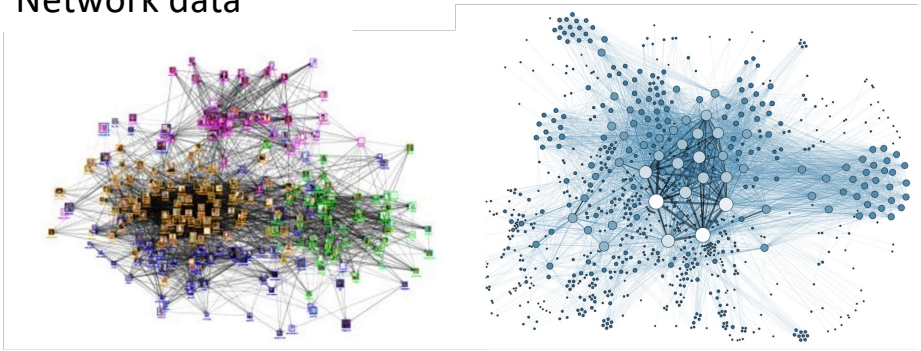
DEFECT







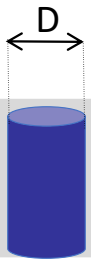
Network data



Social network data

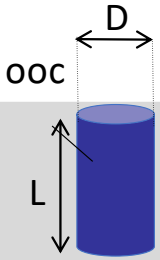
From product to process data

PRODUCT DATA



$X = D \in \mathbb{R}$

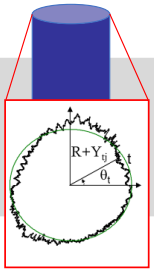
univariate



X=[D,L,OOC,...]'

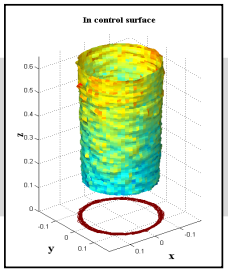
$\mathbf{X} \in \mathbb{R}^p$

multivariate



$X = f(\theta) + \varepsilon$

Profiles/
functional data

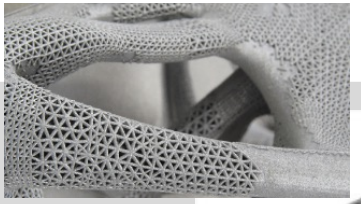


$X = f(\theta, z) + \varepsilon$

Surfaces

?

Manifold



?

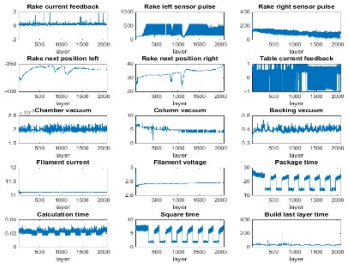
Lattice

CHALLENGES

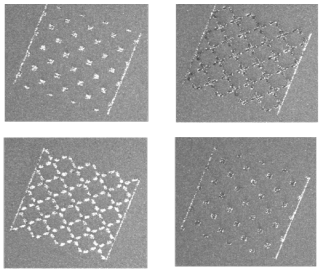
OPPORTUNITY

PROCESS DATA

signals



Image

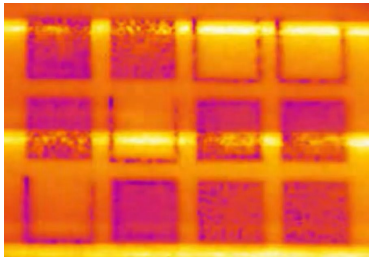


5 - 10 Gbyte

High-speed videos

5 – 10 Tbyte

IR videos



50 – 100 Tbyte

Why data analysis is so important?

✓Having data is not enough if we don't know how to let them speak.

✓"Data analysis" does not mean "Complex analysis"

✓**Very often data are not appropriately analyzed because appropriate tools and methods are not known.**

- "One important idea is that science is a means whereby learning is achieved, not by mere theoretical speculation on the one hand, nor by the undirected accumulation of practical facts on the other, but rather by a motivated iteration between theory and practice."

- George E. P. Box (1976)

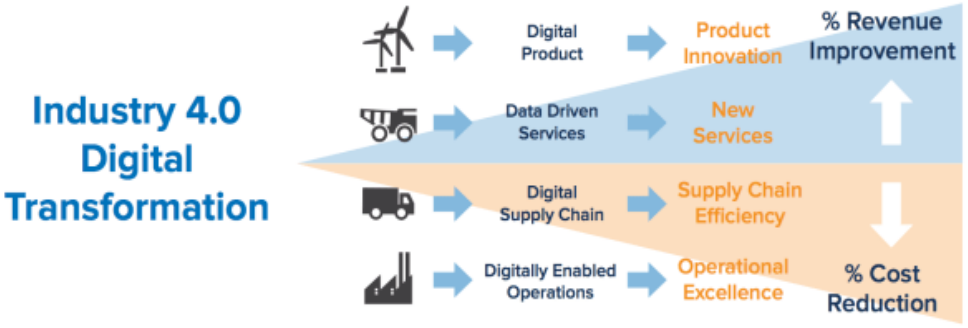
✓"*Essentially, all models are wrong, but some are useful*." Box, G. E. P., and Draper, N. R., (1987)

Quality Data Analysis

Quality

- capability to satisfy (expressed and unexpressed) customer needs - Deming
- All the information affecting the customer satisfaction

Data analysis – modelling and monitoring

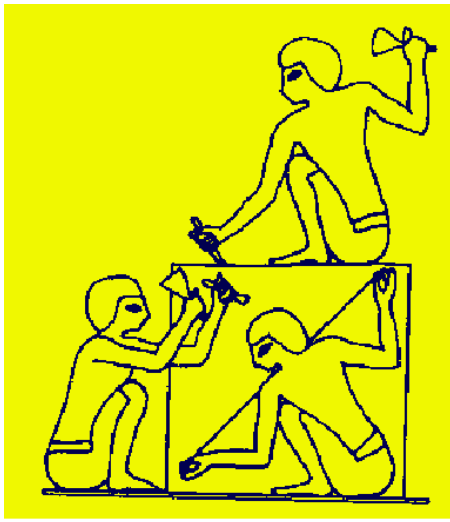


Quality & Customer satisfaction

> Quality—you know what it is, yet you don't know what it is. But that's self-contradictory. But some things are better than others, that is, they have more quality. But when you try to say what the quality is, apart from the things that have it, it all goes poof! There's nothing to talk about. But if you can't say what Quality is, how do you know what it is, or how do you know that it even exists? If no one knows what it is, then for all practical purposes it doesn't exist at all. But for all practical purposes it really does exist.

Quality: a never ending problem

- **XX b.C.**
  "When standards were repeatedly violated, *Phoenicians inspectors* cut the hands of people who did the work."



- **1450 b.C. Egyptian tomb**: with the help of a chord, the inspector is checking whether the stone faces are perpendicular or not

- **50 a.C. roman** water main (aqueduct): built to supp
  the wind at a speed of 215 Km/h
  (max wind speed measured at that time 100 Km/h
  standard were the same for all the regions in the empire

Quality and its relevance: Food industry

The secret of Coca Cola:
is Statistical Process Control!

Importance of Quality Control in Food Industry

- **Reduced production cost**
- **Better goodwill (reputation)**
- **Increase sales (social media!)**
- **Facilitates Pricing**
- **Improved techniques of production**
- **Higher employee morale**

Quality and its relevance: aerospace 💬

Airlines flying Boeing's 787-10 Dreamliner are complaining about quality they say is 'way below acceptable standards'

Issues at the North Charleston plant were reported in April in a comprehensive **New York Times** investigation, which found evidence of shoddy production, poor oversight, and a culture that "made speed a priority over safety."

The **report came a month after Boeing's 737 Max jet was grounded worldwide after the second fatal crash in five months**.
The Department of Justice expanded an inquiry into the 737 Max to include issues at the North Charleston factory in June.

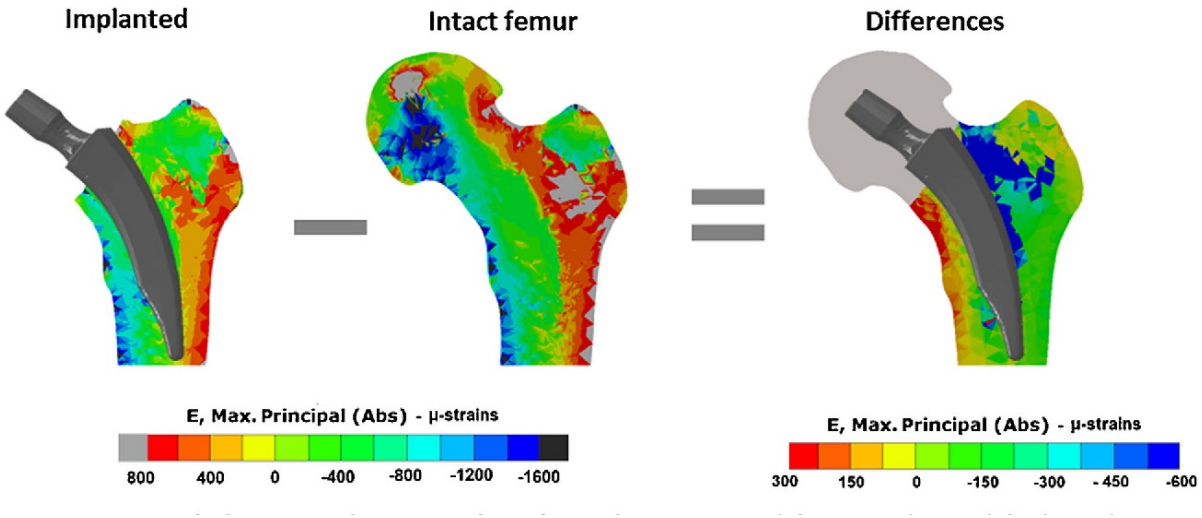
http://Qualityassurancemag.com

Boeing overhauls quality controls: more high-tech tracking but fewer inspectors
Originally published January 20, 2019 at 5:01 pm

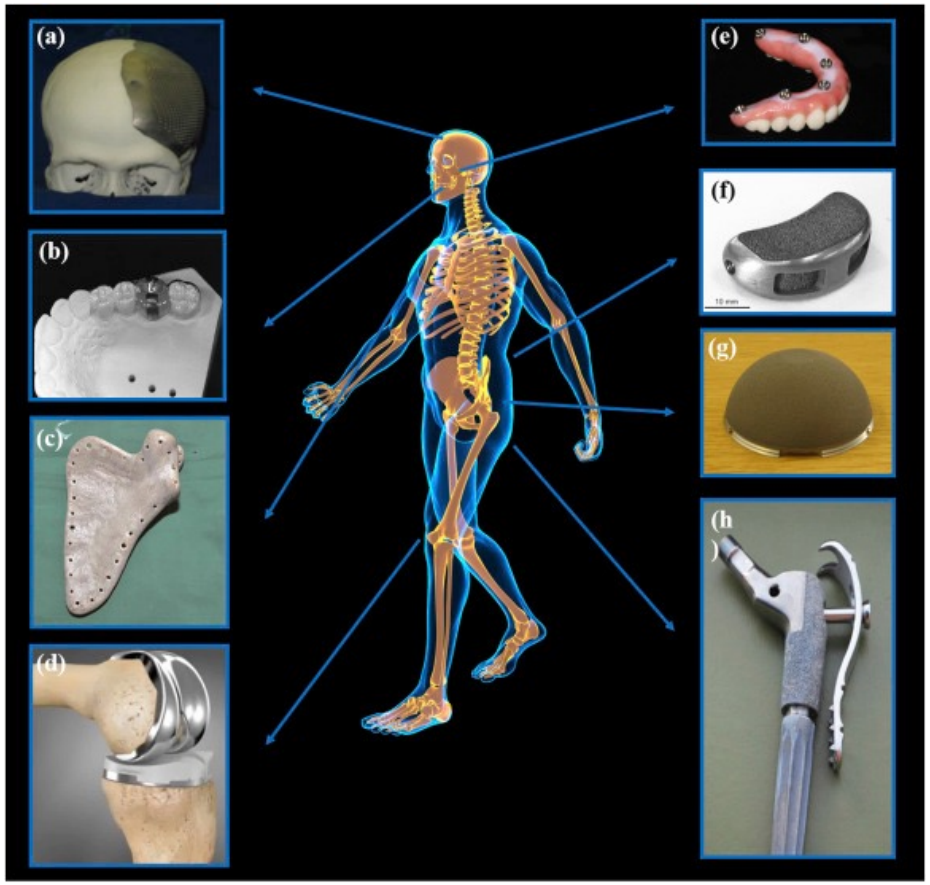
https://www.seattletimes.com/business/boeing-aerospace/boeing-overhauls-its-quality-controls-more-high-tech-tracking-but-fewer-inspectors/

Quality and its relevance: medical implants

Quality is a key element in biomedical applications

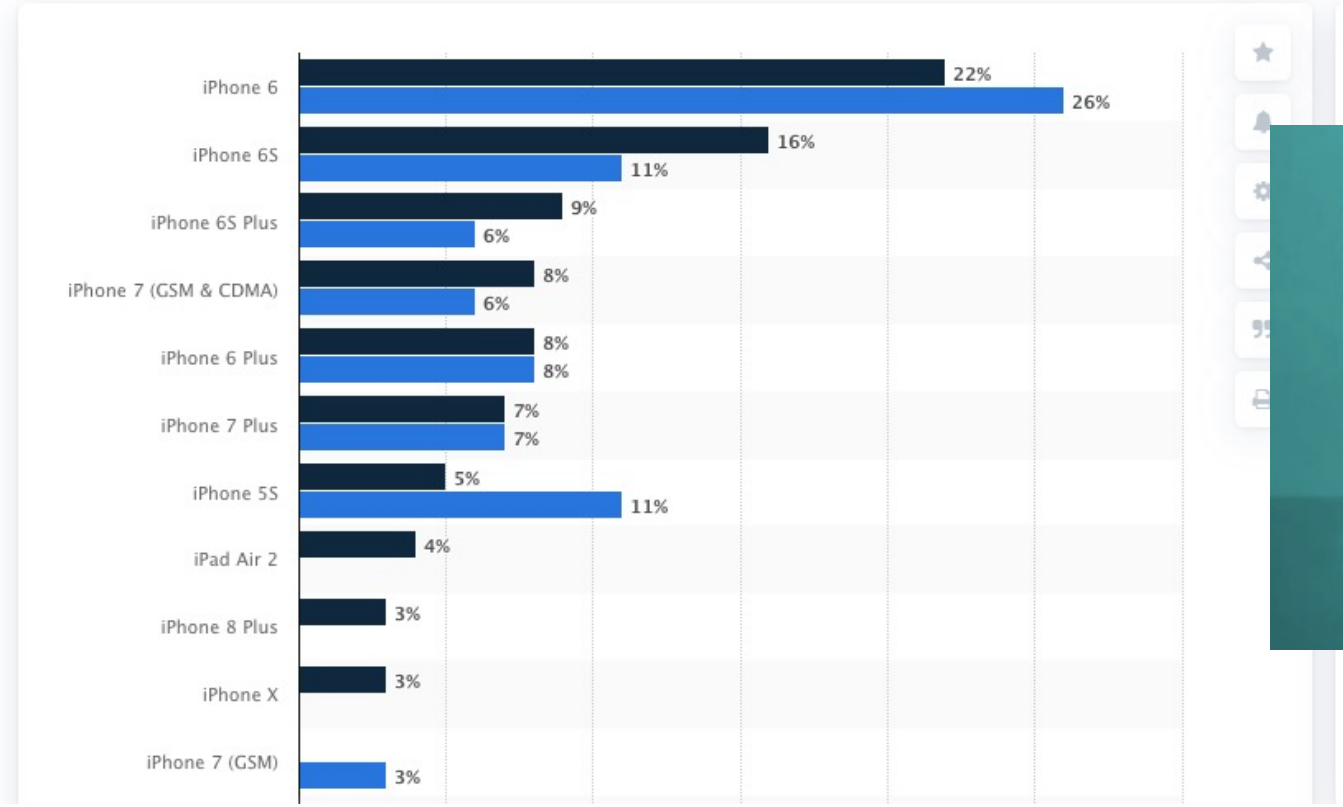


Product quality inspection via X-ray CT



Qualty and its relevance

iphone- failure rate by model worldwide



https://www.statista.com/statistics/804359/iphone-failure-rate-by-model-worldwide/

Quality in services: Quality and customer satisfaction

Examples
Customer care
Teaching evaluation

Examples
Car sharing
Railway/airplane
transportation

Salute

HOME ALIMENTAZIONE E FITNESS MEDICINA E RICERCA SALUTE SENO ONCOLINE

Le classifiche degli ospedali che funzionano meglio: il nord primeggia, il sud insegue

La cardiochirurgia è una specialità che si misura abbastanza bene, visto che gli interventi che vengono realizzati in questi reparti super specialistici sono soprattutto due, il bypass e la sostituzione delle valvole. Anche qui lavora meglio chi lavora di più e si valuta la riuscita dell'intervento andando a vedere la mortalità dopo 30 giorni dall'uscita dalla sala operatoria. Ecco quali sono le eccellenze nel nostro Paese.

Bypass aortocoronarico mortalità a 30 giorni (strutture con volumi più alti, almeno 250 casi rilevati da Agenas a biennio)

1 - San Raffaele, Milano - 0%
2 - Cliniche Gavazzeni, Bergamo - 0,37%
3 - Santa Chiara, Trento - 0,50%
4 - Hesperia Hospital, Modena - 0,55%
5 - Spedali Civili, Brescia - 0,62%
6 - Fondazione Giovanni Paolo II, Campobasso - 0,66%
7 - Ospedale di Lecco, Lecco - 0,73%
8 - Città di Lecce, Lecce - 0,74%
9 - Poliambulanza, Brescia - 0,80%
10 - Ismett, Palermo - 0,85%
Media italiana, 2,36%

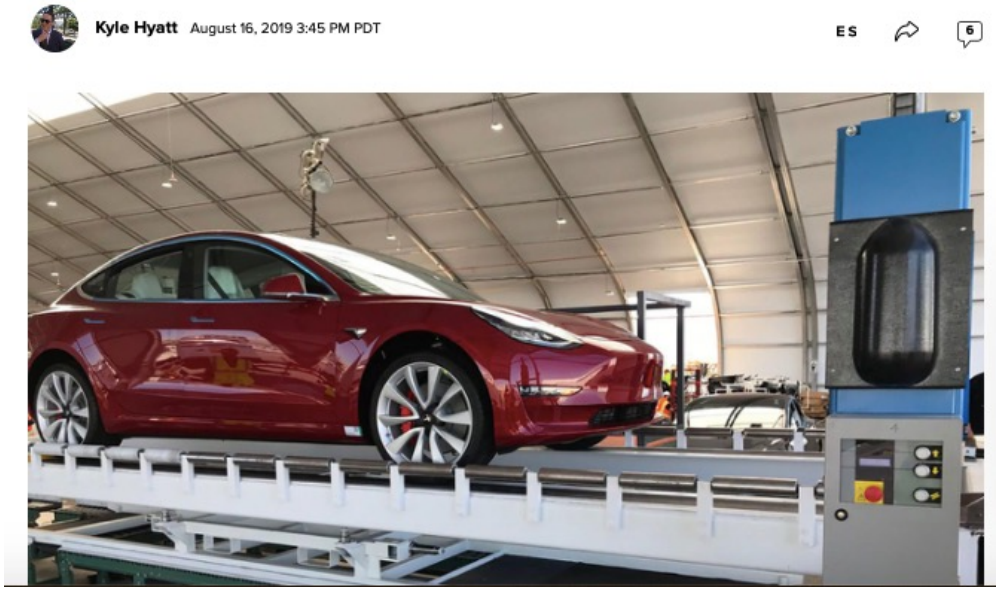
ht
/1
15

Examples
Hospitals

Quality and its relevance

Quality issues lead NextMove to bail on its massive Tesla Model 3 order

The German specialist EV rental company experienced a number of issues with the first 15 Model 3s of a 100-car order and canceled the remainder.



Between January 2018 and September 2019, new owners reported an average of 66 problems per 100 new vehicles.

Bloomberg noted that it's difficult to make direct comparisons between Tesla and the rest of the auto industry. J.D. Power's 2019 initial-quality survey found 91 problems per 100 vehicles across the auto industry, though J.D. Power measures problems over 90 days, rather than the 30-day period covered by Bloomberg, and Tesla is not included in J.D. Power's survey.

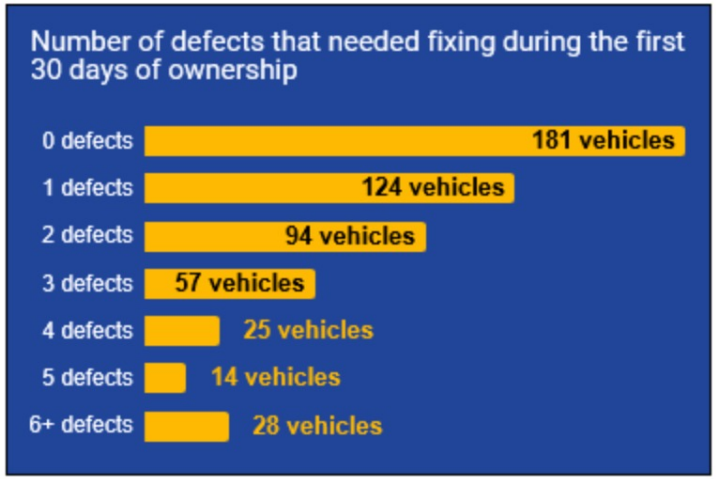
https://www.businessinsider.com/tesla-fixing-quality-issues-nearly-5000-model-3-owners-say-2019-10?IR=T

Tesla and the Quality costs

- It should not be news to many Tesla investors that **Model 3 quality** to date has been atrocious.
- Widespread stories of customers rejecting cars, cars failing first day, rapidly growing Lemon Law cases, etc. Model 3 public tracker continues to show quality levels that put the car in the pantheon of the lowest quality cars ever made.

https://www.patreon.com/posts/cost-of-quality-22258241?l=it

Consider the data from the public tracker (images below).



35% of Model 3s had no defects during the first 30 days of ownership.

Quality costs

A ridiculous 25% of Model 3s are sitting in Tesla service centers for AT LEAST one day for repairs.

About 10% of Model 3s delivered are spending 6+ days in Tesla service centers within the first month. No wonder the Lemon Law claims are growing.

All of this comes at a great cost to customers and Tesla.

Customer Costs Due To Poor Model 3 Quality

- Customers will have to spend considerable amount of time running around Tesla service centers
- Customers may have to take time off from employment or other activities to attend to the car.
- Customers may be stranded at inopportune times and places
- The "brand new" car that the customers purchased may go through extensive surgeries, and will never have factory quality, fit and finish (although with Tesla factory quality is not a quality anyone would want)

All of this, slowly but surely, damages Tesla brand.

Tesla Direct Costs Due To Poor Model 3 Quality

From an investor perspective, there are direct costs that Tesla will incur in rectifying the quality problems.
These costs include:

- Cost of customer service to log the problems and schedule an appointment.
- Cost of transporting the car to the service center when needed
- Problem diagnosis and repair.
- Cost of storage for the duration.
- Cost of any additional damage that occurs under Tesla care (this appears to be a rampant problem)
- Cost of customer loaner car for the duration of the repair.

We will not be able to estimate many of these costs without good breakdowns on specific kind of problems that Model 3s are experiencing. But, we will make an attempt. Note that, given the large number of cars that Tesla sells in its home state, California cost structure dominates the calculations. The labor and infrastructure costs of performing these repairs in California will be very expensive.

We modeled several likely scenarios and estimate that direct costs to Tesla are likely around $750 to $1,000 per each Model 3 sold. This is within the first month – excluding any service or warranty repairs beyond the 30 day period.
In other words, ***roughly 2% of Tesla Model 3 gross margins are likely going out as direct quality costsin the FIRST 30 days.***

Indirect Costs Are Even More Difficult To Quantify

The level of repair rates for Model 3 completely swamp the Tesla service center network. The more Model 3s are sold, the bigger the problem becomes. We can expect customer dissatisfaction leading to lower sales but there are no good models to quantify this effect.
On a more objective and measurable basis, multitude of comments on Twitter and TMC Forums show that Tesla is witnessing large delays in customer cars being serviced.
We seem to be close to a breaking point in terms of service infrastructure that may lead to massive customer backlash.

https://www.patreon.com/posts/cost-of-quality-22258241?l=it

Quality Costs

Quality costs– PAF method (Prevention – Appraisal – Failure)*:

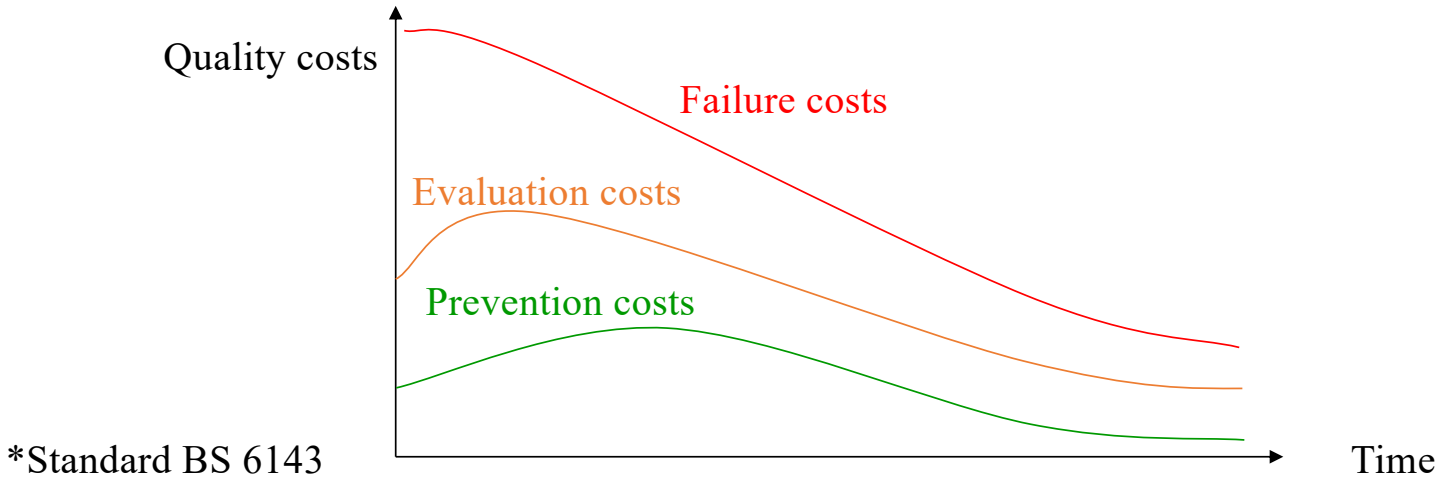
– Costs for prevention (costs to prevent non-conformities in products and services)

– Evaluation costs (associated to measures, inspections, audits, …)

– «Failure» costs (non-conformity):

  – Internal (remanufacturing, wastes, redesign, downtime)

  – External (claims, subsitutions, discounts for non-conforming, lost commissions, …)



*Standard BS 6143

Zero defects and zero-waste

Zero Defects

- **Improving** Product Quality and **Eliminating** Waste
- Zero defects means getting it right the first time, every time!

How much do quality failures cost your company?
Quality defects have significant costs associated with them – some of the most obvious being money, time, resources, and lost reputation. And programs to eliminate quality defects can be expensive and time consuming

https://www.mindtools.com/pages/article/newTMC_87.htm

Twin transition: green and digital

Green

- Zero-defect first-time right versus zero waste:
  - reduced time to market,
  - Material savings
  - Resource savings (energy)
  - Reduced scrapes and rework

Digital

- Massive data
- Multi-stream
- high velocity
- Big data

- Fast reaction
- Decision making

- (Edge computing
- Distributed data)

I4.0 in EU

- Ursula von der Leyen, President of the European Commission, said: “Europe's industry is the motor of growth and prosperity in Europe. And it is at its best when it draws on what makes it strong: its people and their ideas, talents, diversity and entrepreneurial spirit.
- This is more important than ever as Europe embarks on its ambitious **green and digital transitions** in a more unsettled and unpredictable world. Europe's industry has everything it takes to lead the way and we will do everything we can to support it.”
- Thierry Breton, Commissioner for Internal Market “Europe has the strongest industry in the world. Our companies - big and small - provide us with jobs, prosperity and strategic autonomy. Managing the **green and digital transitions** and avoiding external dependencies in a new geopolitical context requires radical change - and it needs to start now.”

Why data analysis is so important?

- Standard ISO: “Principle 7: Factual approach to decision making: Effective decisions are based on the analysis of data and information”
- Kume (leader in the Japanese quality movement) 1985: “Data is a guide for our actions: from data we learn pertinent facts and take appropriate actions based on these facts”
- Hoerl (1998) – General Electrics
    - …Similarly, vague statements as “We think we have that under control” or “Recent performance seems to be improving” are no longer accepted at any level. The typical response to these kinds of statements now is:
        - “**Show me the data**!”
- Quality professional (Alwan, 2000): “In God we trust, all others bring data”