From big data to useful knowledge: challenges and opportunities in smart factories

Grzegorz J. Nalepa and **Edyta Brzychczy**

AGH University of Science and Technology

CHIST-ERA/IBDSF, Kraków, Poland, June 22, 2017

`home.agh.edu.pl/gjn`, `geist.re`, `bigdata.agh.edu.pl`

Outline

Photo from Schuterstock

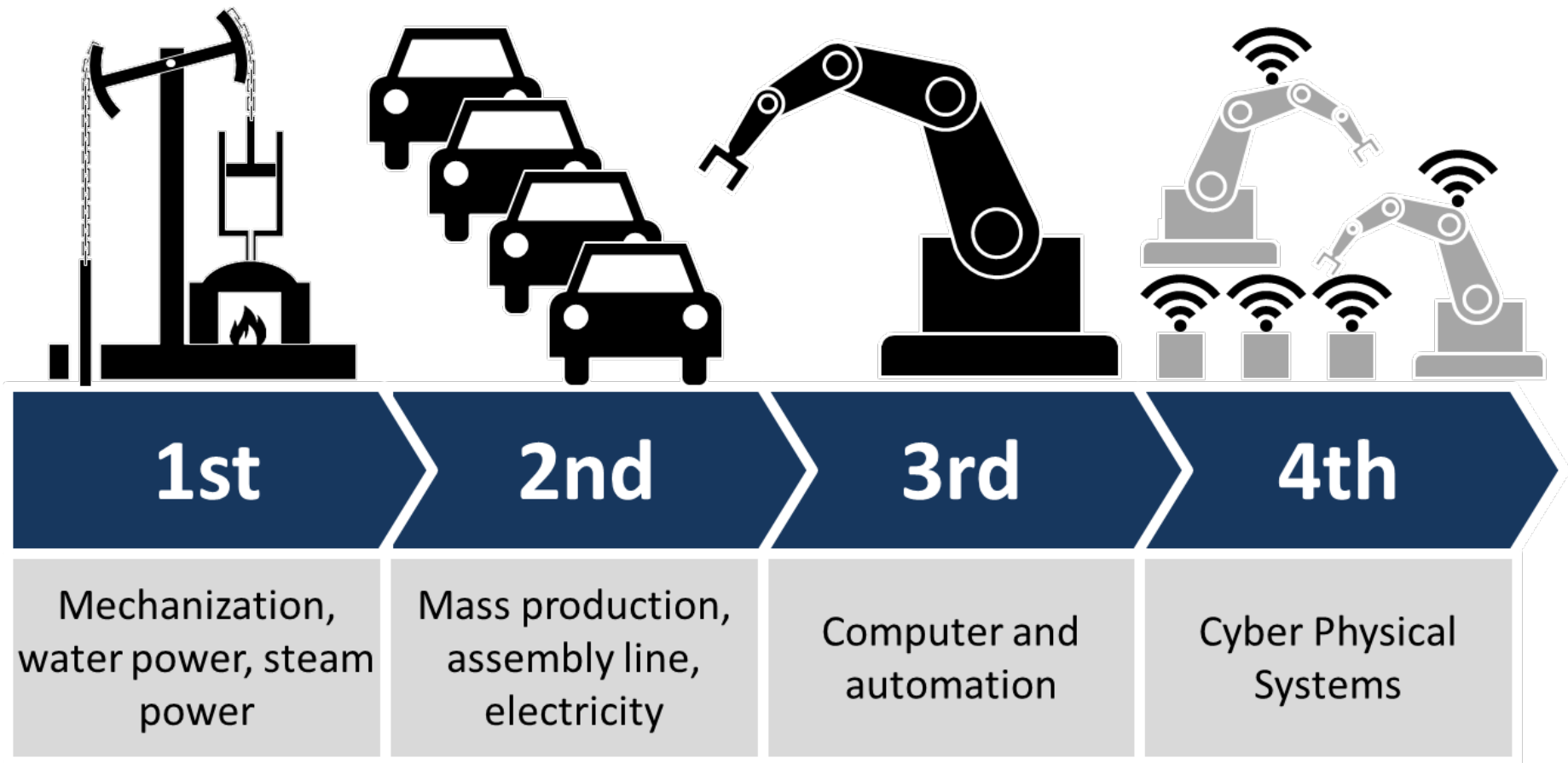


Photo from Wikipedia

Photo by ERIC LALMAND/AFP/Getty Images

Topics for this talk

1. learning
2. semantics
3. context
4. processes

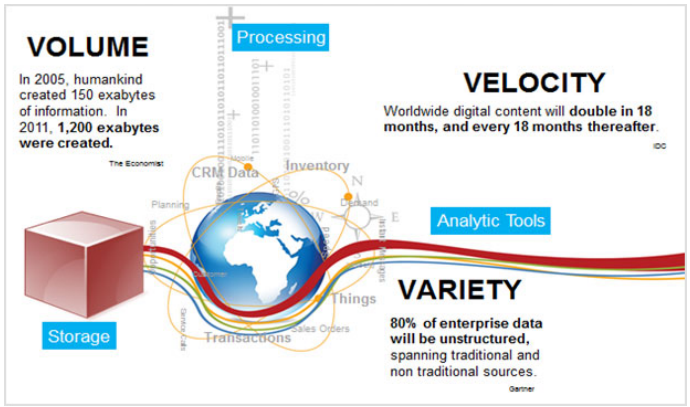
Outline


WHAT WE KNOW...
THE REST...
BIG DATA

Big Data may be Huge Problem

Big Data

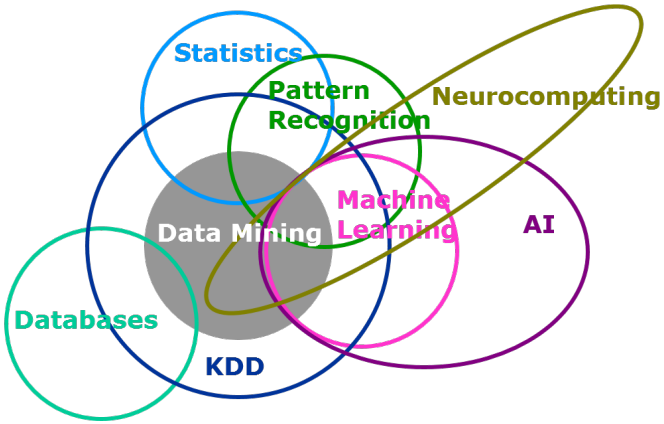
- we can record anything
- unlimited storage
- quite fast retrieval
- get it anywhere – cloud
- we can know more! (?)
- we can do more! (?)



Challenges

- *data* storage is not a solution
- distribution is nice but. . .
- goodbye ACID, welcome CAP
- no relationships, no *knowledge*
- statistics is not enough!
- to understand, we need to know what (how) questions to ask!

Machine Learning (and data mining)

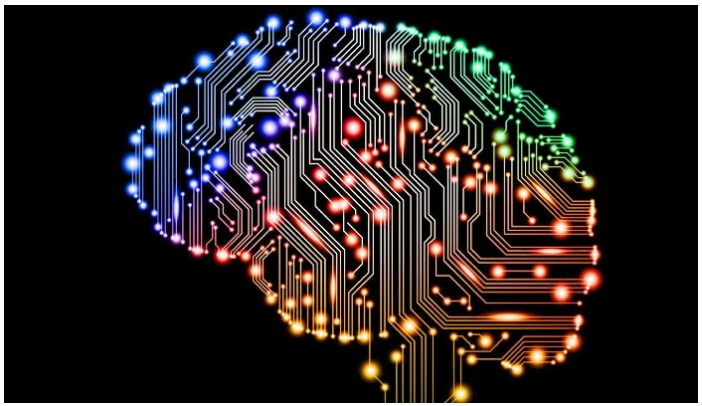


Benefits

- *automated* training
- *concept* learning
- object *classification*
- pattern *recognition*
- (new!) knowledge *discovery*
- *massive* parallelism (use of)
- ***incremental* models** (storage!)

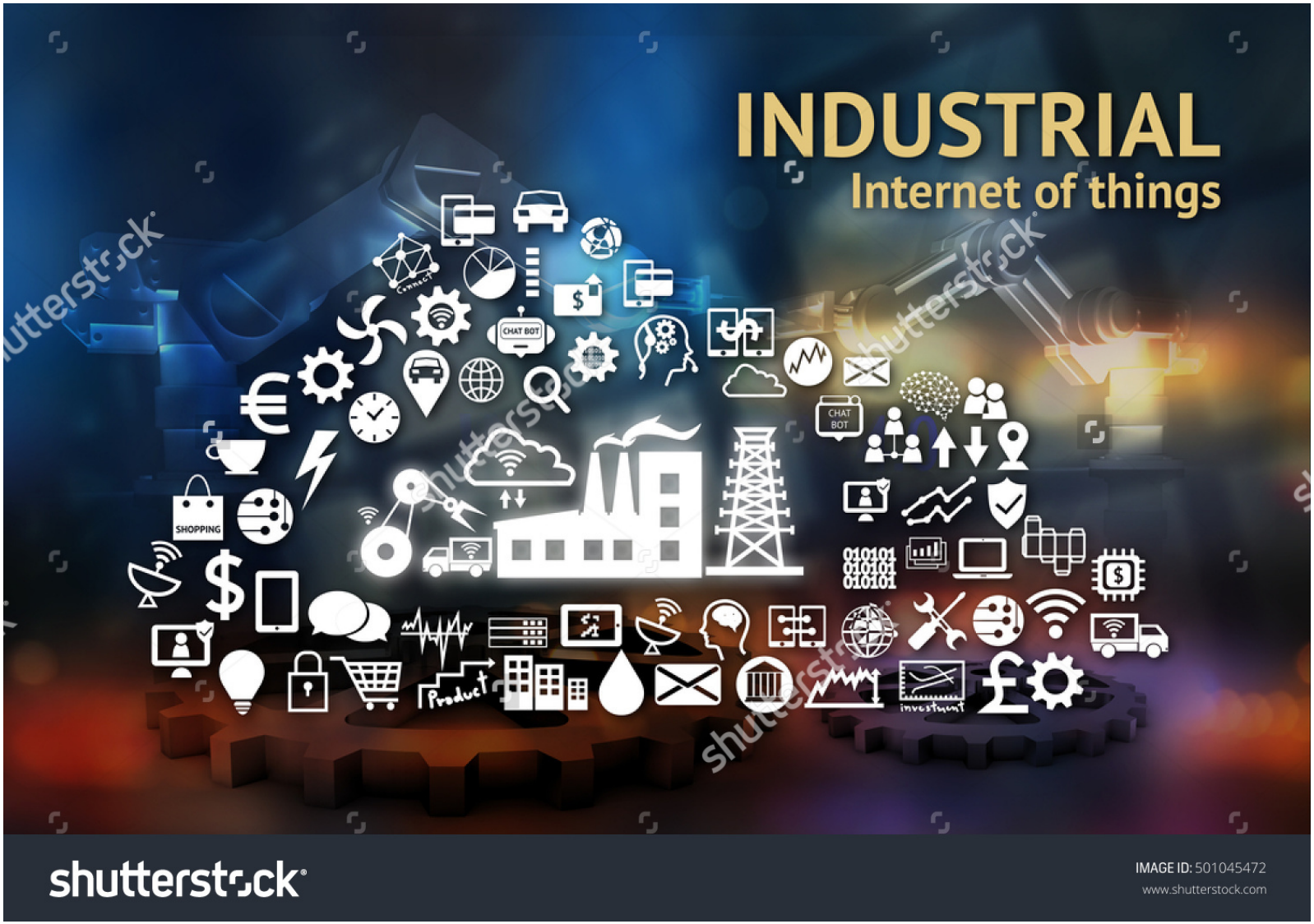
Methods

- learning from *examples*
- *unsupervised* learning
- *clusterization*
- deep learning
- **learning from *data streams***



Outline


50B CONNECTED THINGS BY 2020


INDUSTRIAL
Internet of things
CHAT BOT
SHOPPING
Product
investment
shutterstock
IMAGE ID: 501045472
www.shutterstock.com

Internet of Things! (or World Wide Mess?)

IoT

- *everything* gets connected
- accessible from *anywhere*
- sensor networks
- M2M communication
- „smart" X (house, car, city)
- *panopticon*-like monitoring

Challenges

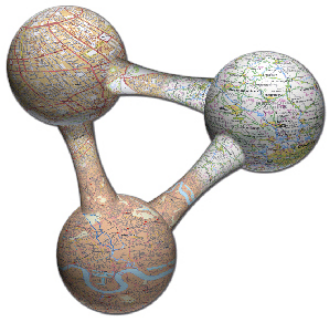
- huge data! maybe our *storage* is not unlimited after all?
- *protocols* (opportunistic)
- energy efficiency
- can machines *understand* each other?
- can they communicate with us with *our concepts*?

Metadata, Ontologies, *Semantics*

Methods

- symbolic knowledge *representation*
- data vs. *knowledge*
- numbers vs. *concepts*
- automated *reasoning*
- inference with *rules*
- *Semantic* (social) Web

http://business.data.gov.uk/companies



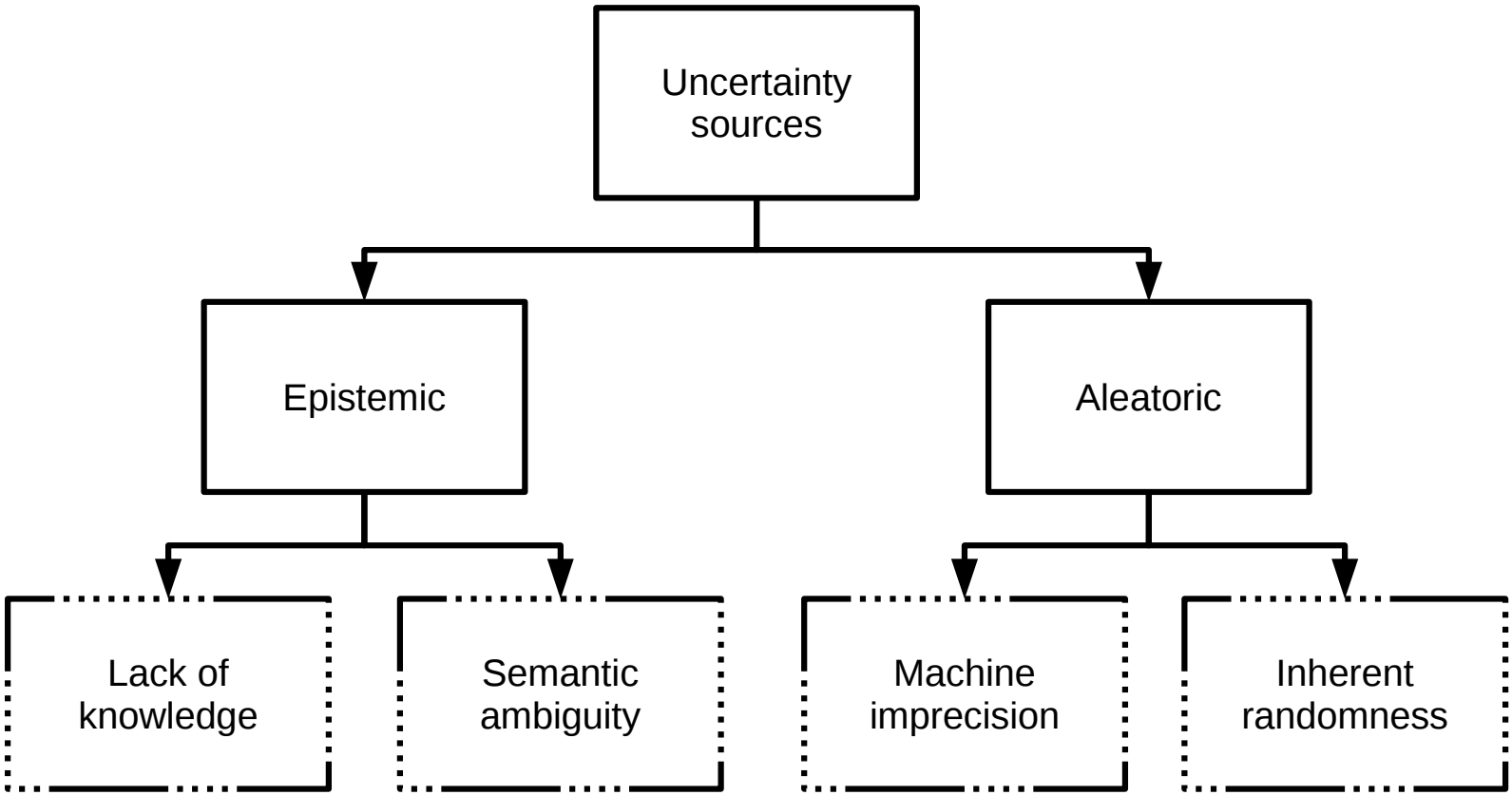
http://geovation.uk/linked-data-opportunity-for-developers

Benefits

- metadata – data interpretation
- processing with human concepts
- *shared* conceptualizations
- (semantically) Linked Open Data
- semantic *interoperability*
- *query*, not search

Outline

Data in dynamic environments is often uncertain and ambiguous



Context-Aware Systems

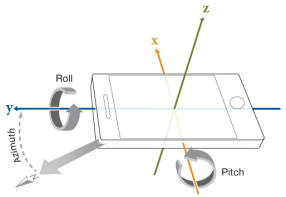
context = any information that can be used to characterize the situation of an entity

Hardware

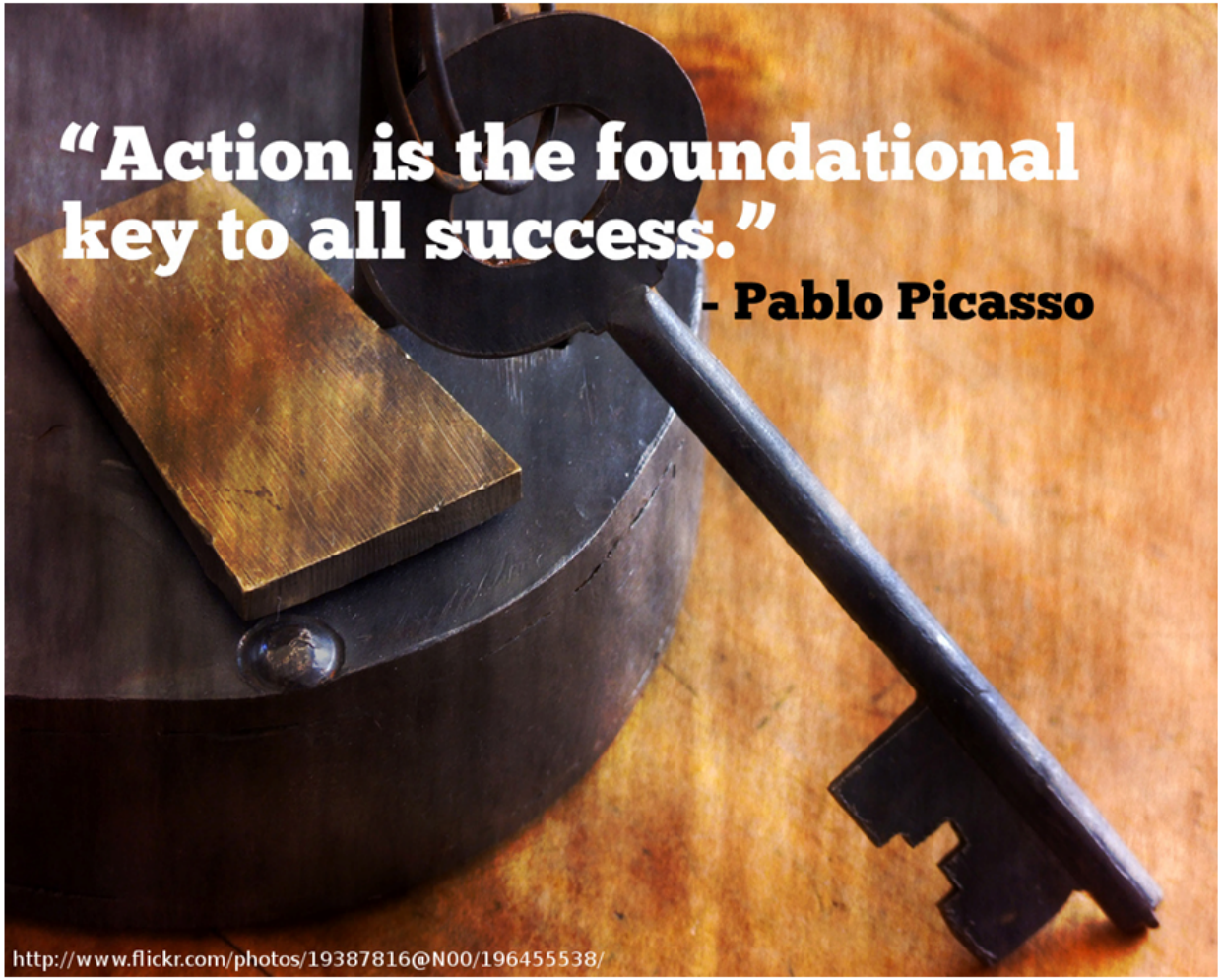
- mobile, wearable
- versatile sensors
- environment monitoring
- new interfaces
- ambient intelligence

Software

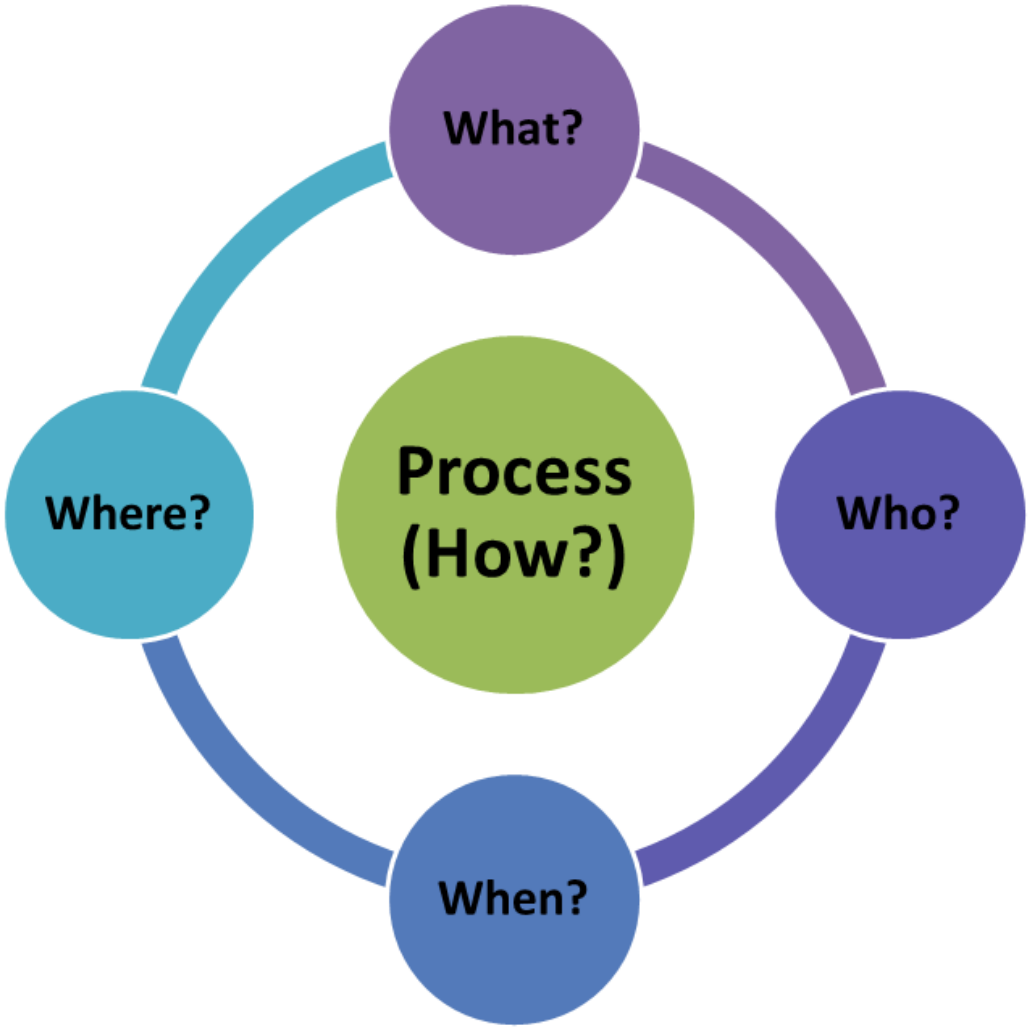
- *adaptation* of functionality
- learning of user *habits*
- reasoning with concepts
- *recommendation (relevance)*
- decision support
- *mediation* of context
- continuous *assistance*



Outline

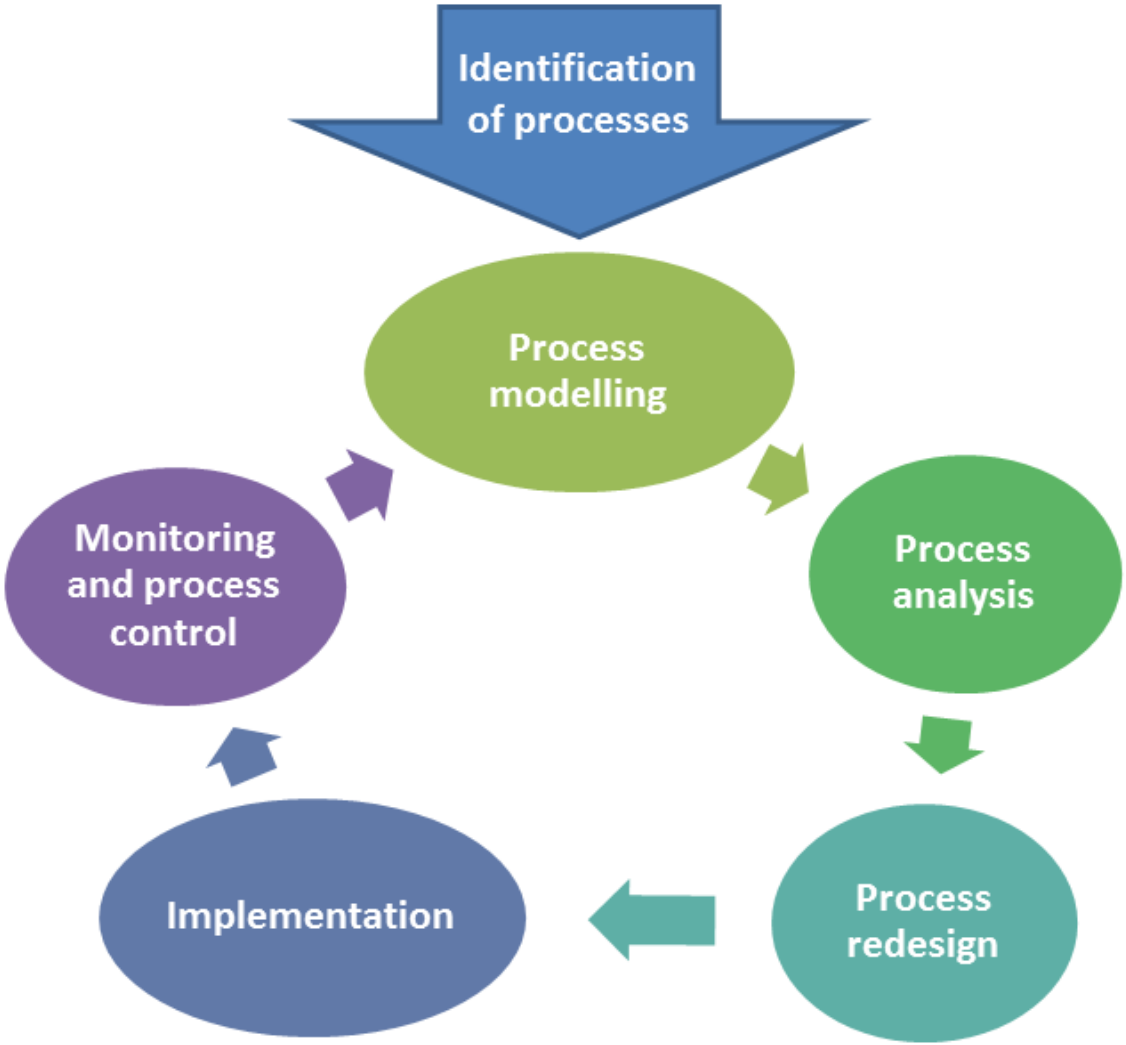

“Action is the foundational key to all success.”
- Pablo Picasso
http://www.flickr.com/photos/19387816@N00/196455538/

WHY?


What?
Where?
Process
(How?)
Who?
When?

Marketing and customer services	Design process of products and technical production preparation	Manufacturing	Distribution
market analysis, include forecasting i.e. prices	demand forecasting	process analysis and forecasting	pre-trading analysis
customer profiling	input requirements analysis and forecasting	internal transport (route) optimisation	forecasting costs
feedback and complaints analysis	predicting production outputs	quality analysis and forecasting	route/fleet optimisation
churn prediction	operation scheduling	predictive maintenance	credit risk analysis

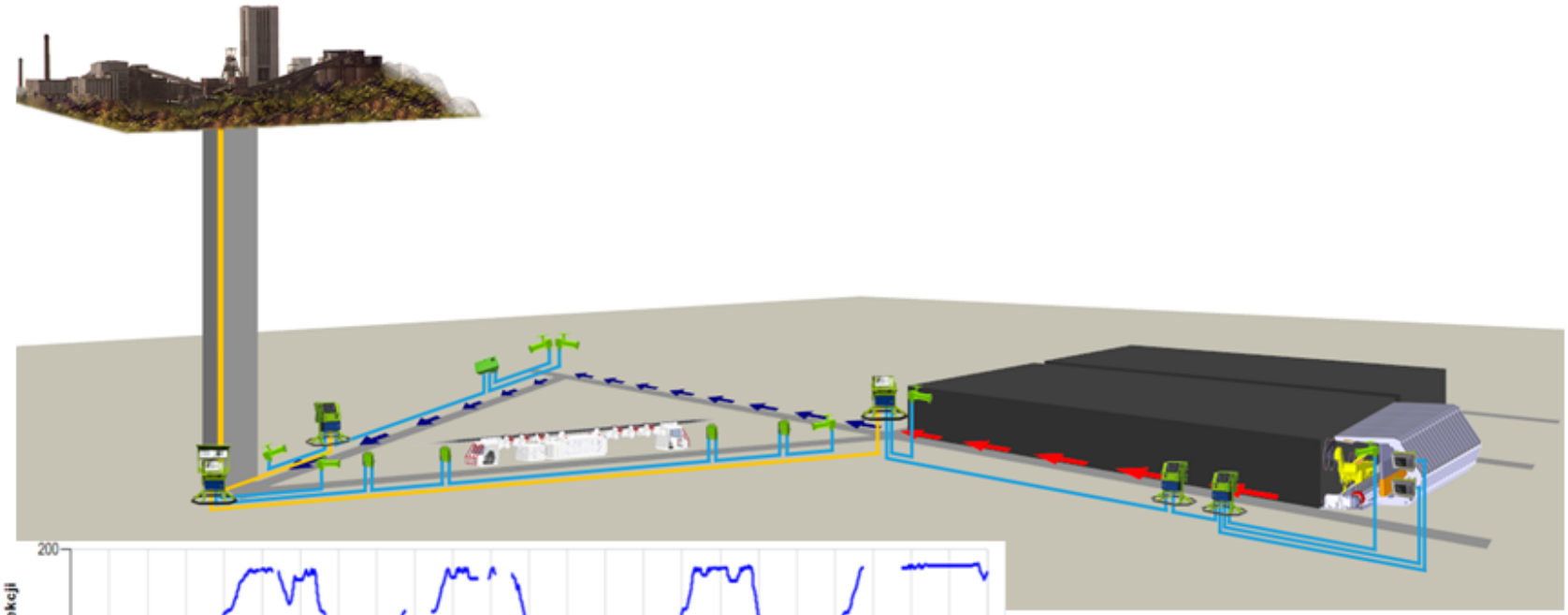
BPM Cycle

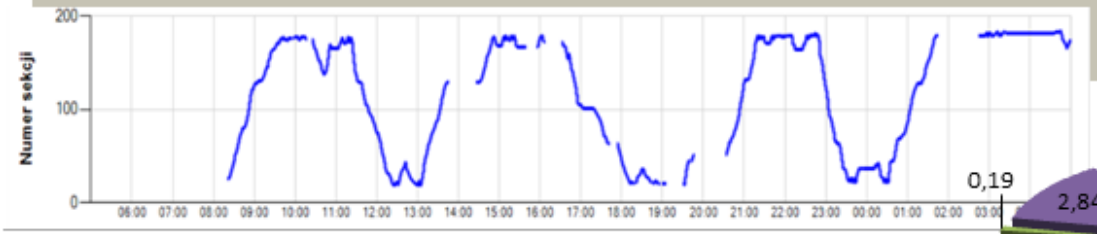

Identification of processes
Process modelling
Process analysis
Process redesign
Implementation
Monitoring and process control

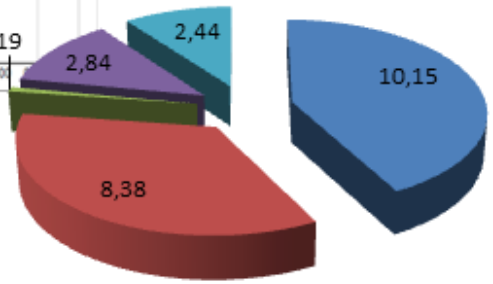
Process mining vs. mining process

Source: Korski. J, FAMUR S.A.

Process monitoring



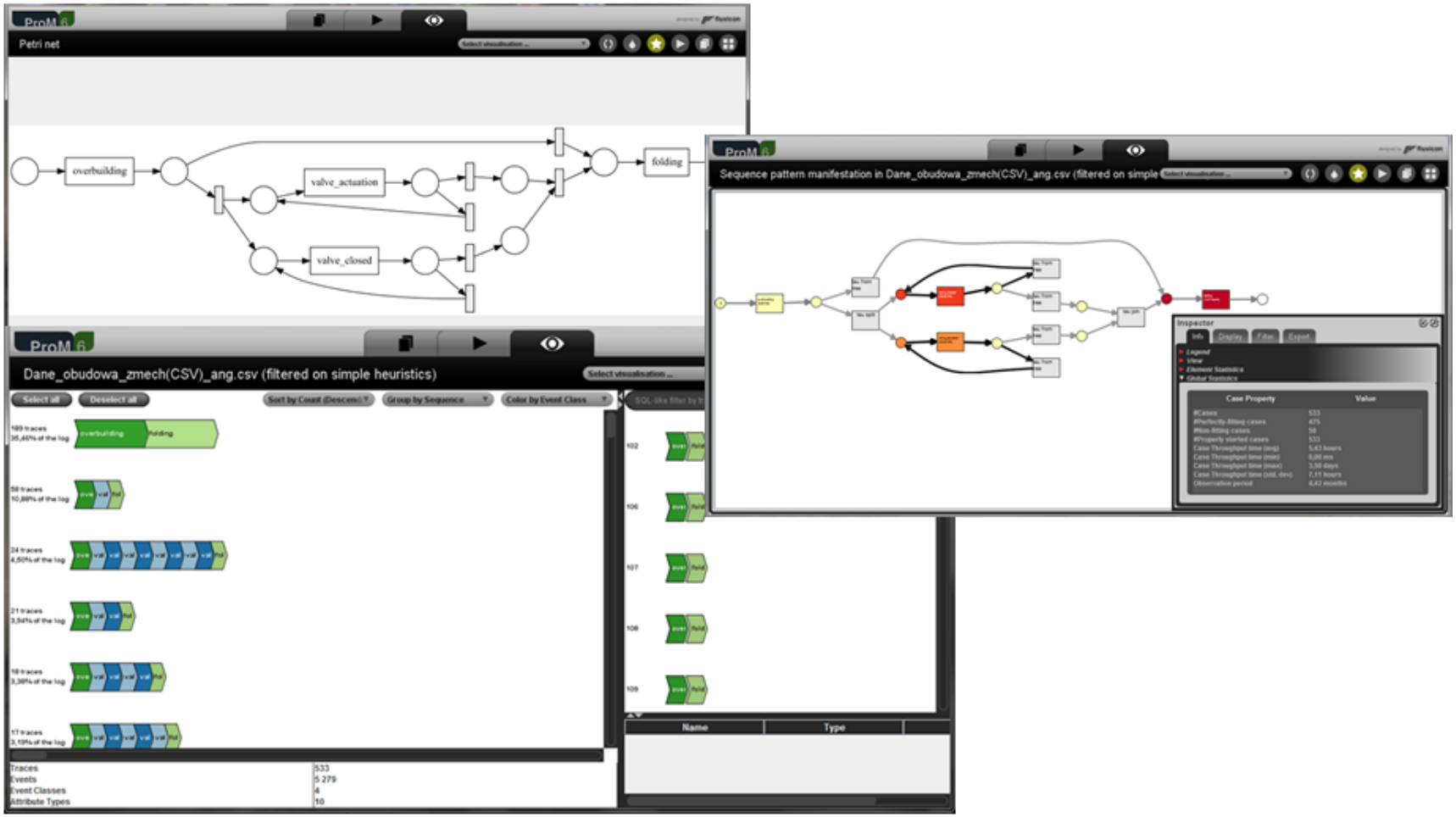




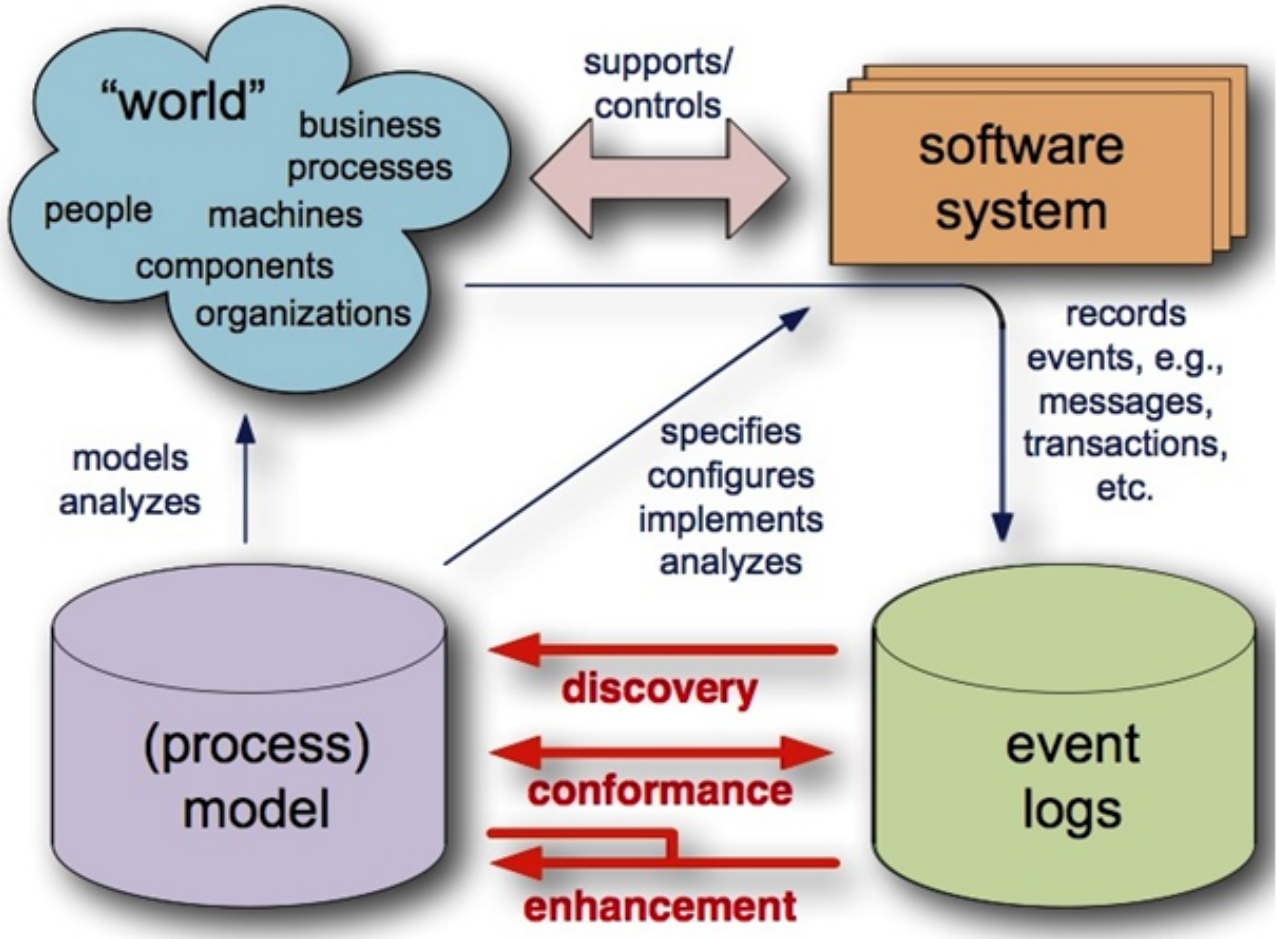
Time off work
Work time with mining
Work time with moving
Work time without moving
Time without data transmition

Source: Korski. J, FAMUR S.A.

How about real process mining?

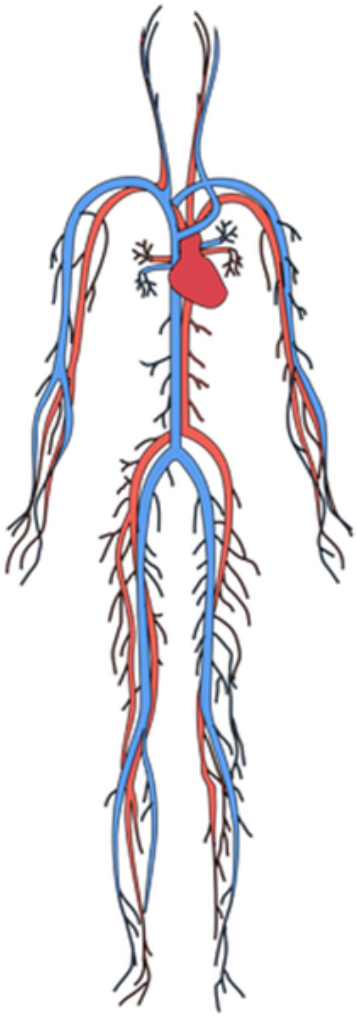


Process mining



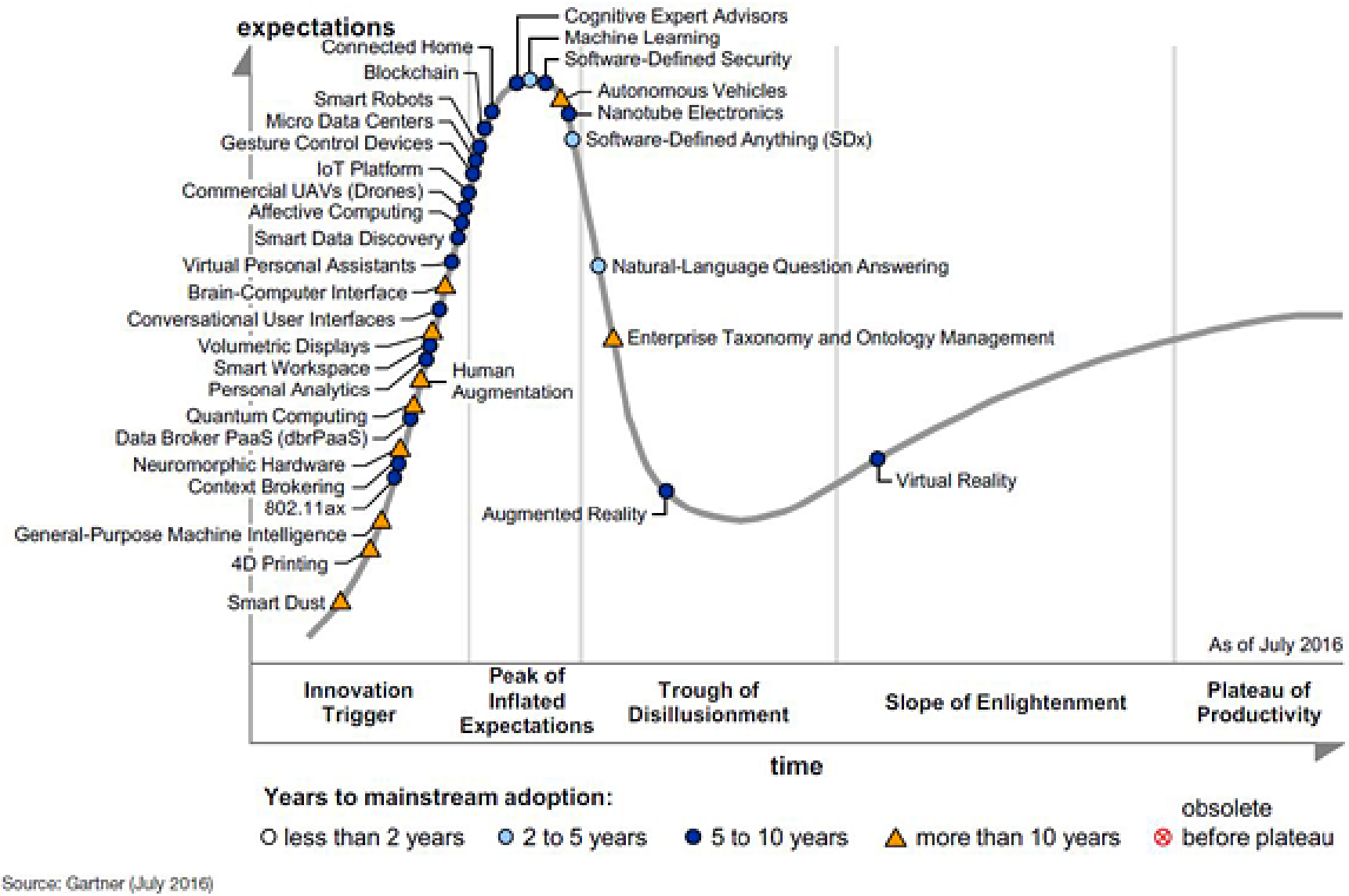
http://prompt.processmining.it/static-images/pmini.jpeg

Smart factories need deep process analysis in process-oriented way



https://www.vexels.com/png-svg/preview/141827/cardiovascular-system-blood-human-body

Gartner technology hype cycle (2016)



Source: Gartner (July 2016)

Take away messages

1. **Machine learning (incremental) needed for BD**
2. **Semantics needed for interoperability in IoT**
3. **Context and uncertainty handling for reasoning**
4. **Process mining and management for analytics**

Thank you for your attention!
Do you have any questions?


`home.agh.edu.pl/gjn`, `geist.re`, `bigdata.agh.edu.pl`