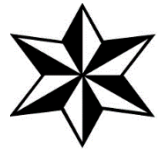


Managing Workflows in a Big Data World

Universidad Complutense, Madrid

Facultad de Informática

April 21, 2017

Dr Stelios Kapetanakis

University of Brighton

Speakers Intro

Dr Stelios Kapetanakis

- Associate Professor in Business Intelligence
- Head of the Knowledge Engineering Group, CEM, Brighton
- Consultant for Banking, UK Railways, US Healthcare
- Consultant for Clarksons, Airbus

- What if data could talk to us..

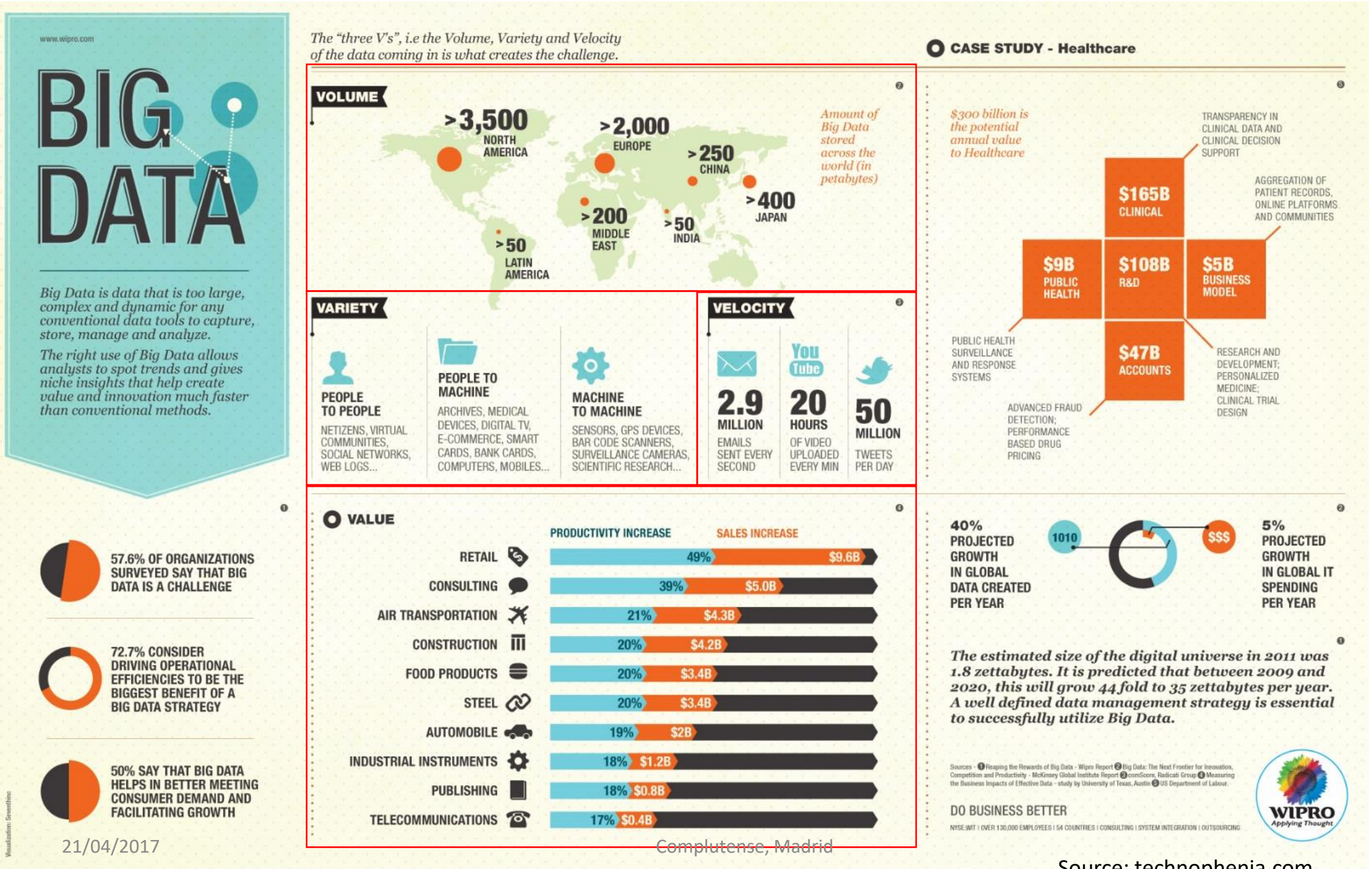
Contents

- Big Data
- Business Workflows
- Real-life Workflows
- Intelligent management of Workflows

Revision

byte	b	8 bits	1 byte
kilobyte	Kb	1024 bytes	1024 bytes
megabyte	MB	1024 KB	1 048 576 bytes
gigabyte	GB	1024 MB	1 073 741 824 bytes
terabyte	TB	1024 GB	1 099 511 627 778 bytes
petabyte	PB	1024 TB	1 125 899 906 842 624 bytes
exabyte	EB	1024 PB	1 152 921 504 606 846 976 bytes
zetabyte	ZB	1024 EB	1 180 591 620 717 411 303 424 bytes
yottabyte	YB	1024 ZB	1 280 925 819 614 629 174 706 176 bytes
brontobyte	BB	1024 YB	1 237 940 039 285 380 274 899 124 224 bytes
geopbyte	GB	1024 BB	1 267 650 600 228 229 401 496 703 205 376 bytes

Source: es.tableworld



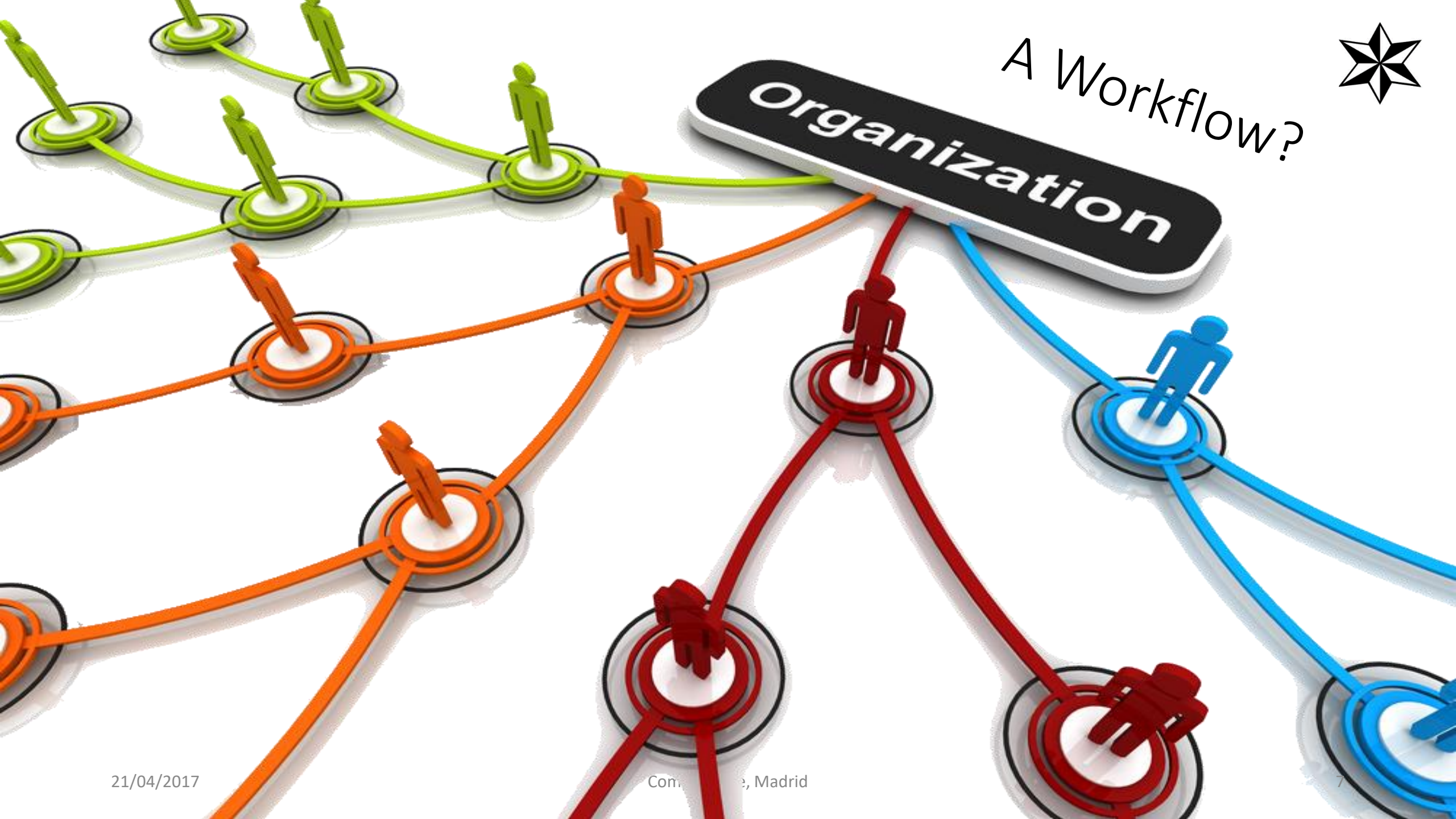
Source: technophenia.com

Big Data what is really..

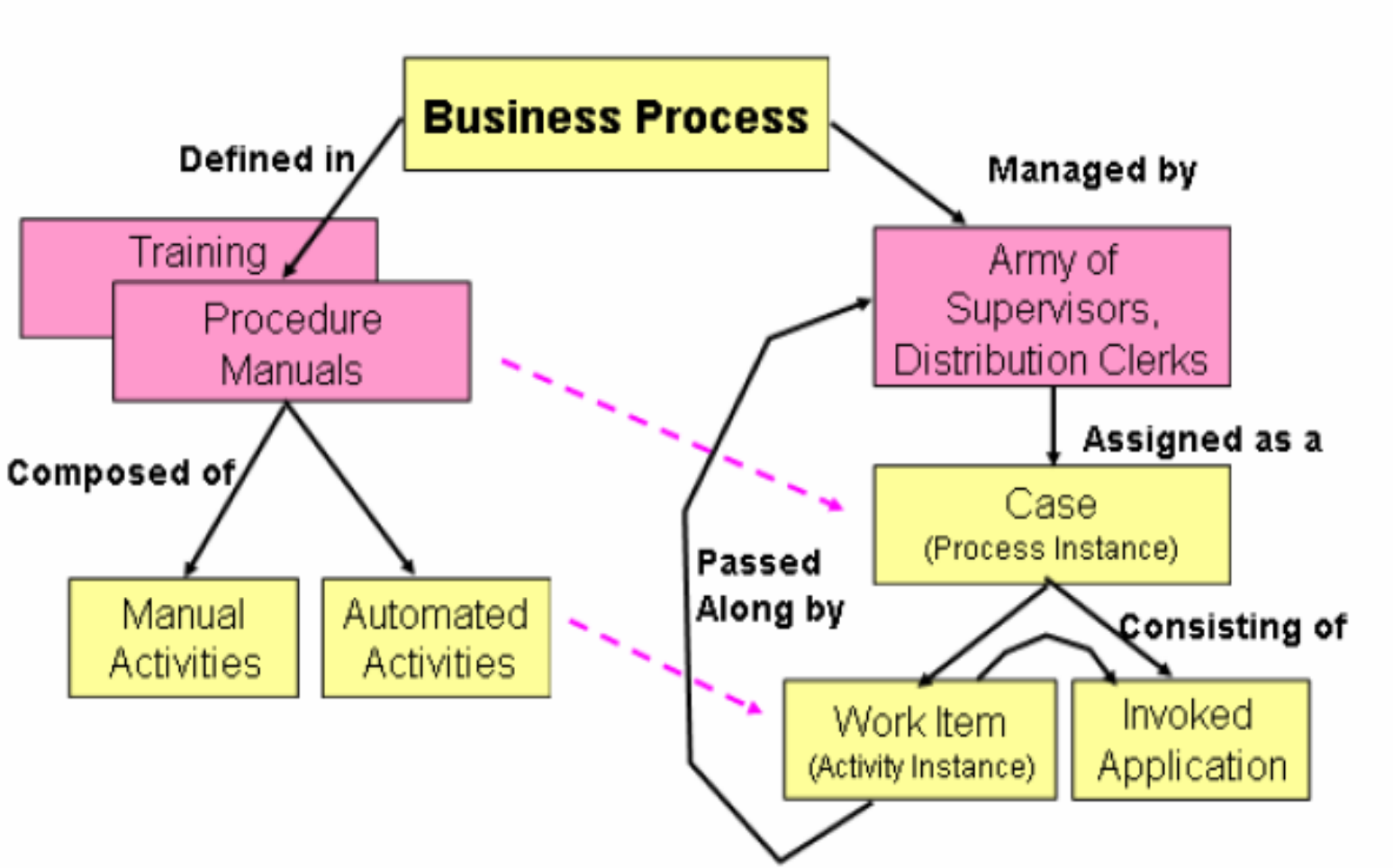
- Big Data is a marketing term
- Since 2012 everybody has.. some..
- Big Data assumes that bigger is better
- “Big data is high volume, high velocity, and/or high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization”

Gartner, 2012

A Workflow?

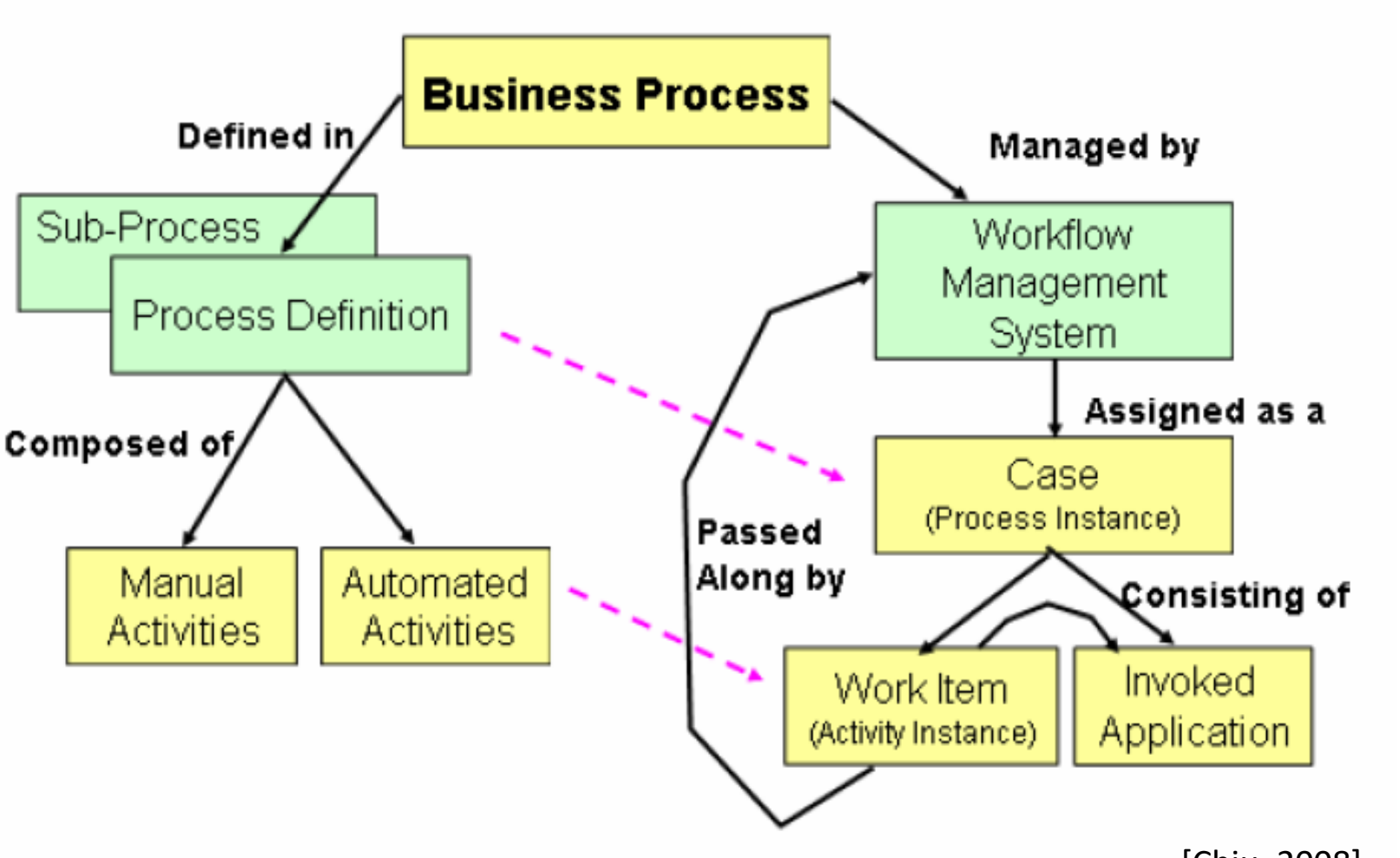


Traditional Workflows



[Chiu, 2008]

Automated Workflows



[Chiu, 2008]

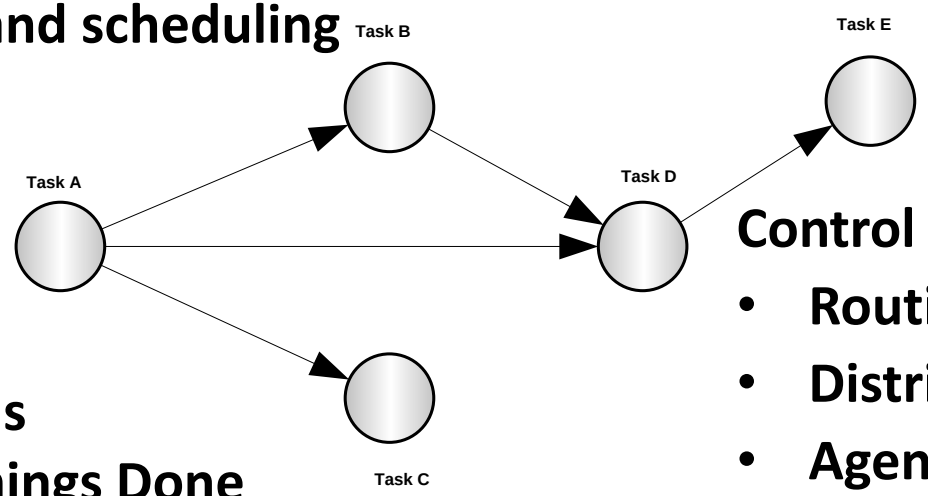
Managing workflows

Categories

- **Flow control**
- **In-transit visibility**
- **Processes**
- **Planning and scheduling**

Components

- **Input(s)**
- **Transformations**
- **Outputs**

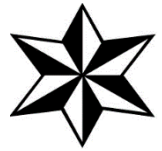


Examples

- **Shops**
- **Bank Loans**
- **Getting Things Done**

Control Systems

- **Routing**
- **Distribution systems**
- **Agent systems**
- **Expert systems**



Workflow Management Systems

- Over the last years standards have been developed in enterprise (W3C standards, BPEL, BPMN, XML, XPDL)

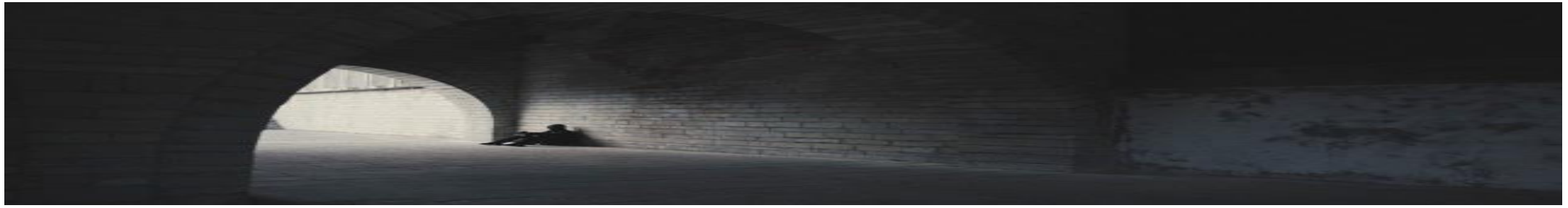
- Recipe of success:
  - Broadly used.. Manually used..
  - Different actors, systems, aims..
  - Coordinate their actions towards a set target
  - Large volumes of data

- Need for intelligent systems
  - Assist with the management
  - Make desirable, feasible DSS
  - Identify volumes and repeated patterns
  - Give ground to AI technologies and in particular CBR

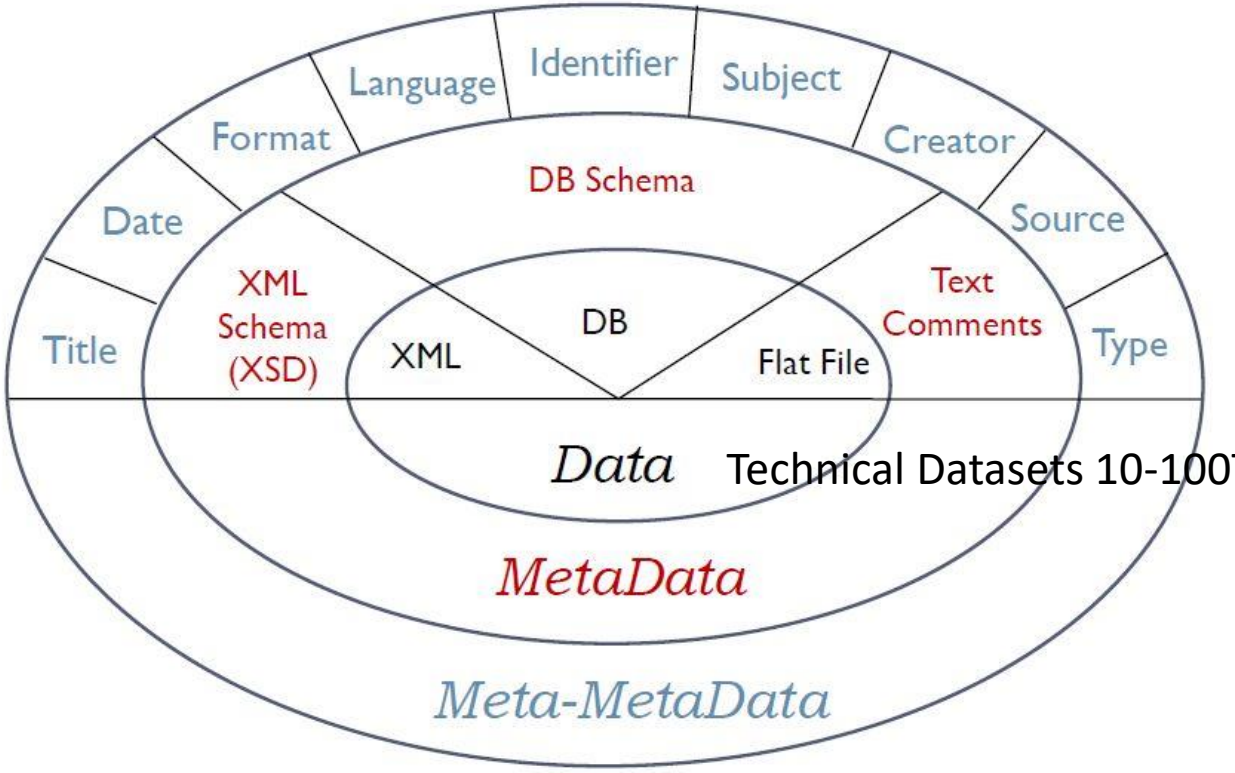
Challenges

Human – related workflows

- Uncertainty
- Inconsistency
- Incompleteness
- Large volumes of workflow instances



Big Data -> Big Problems



Technical Datasets 10-100Tbytes in scale

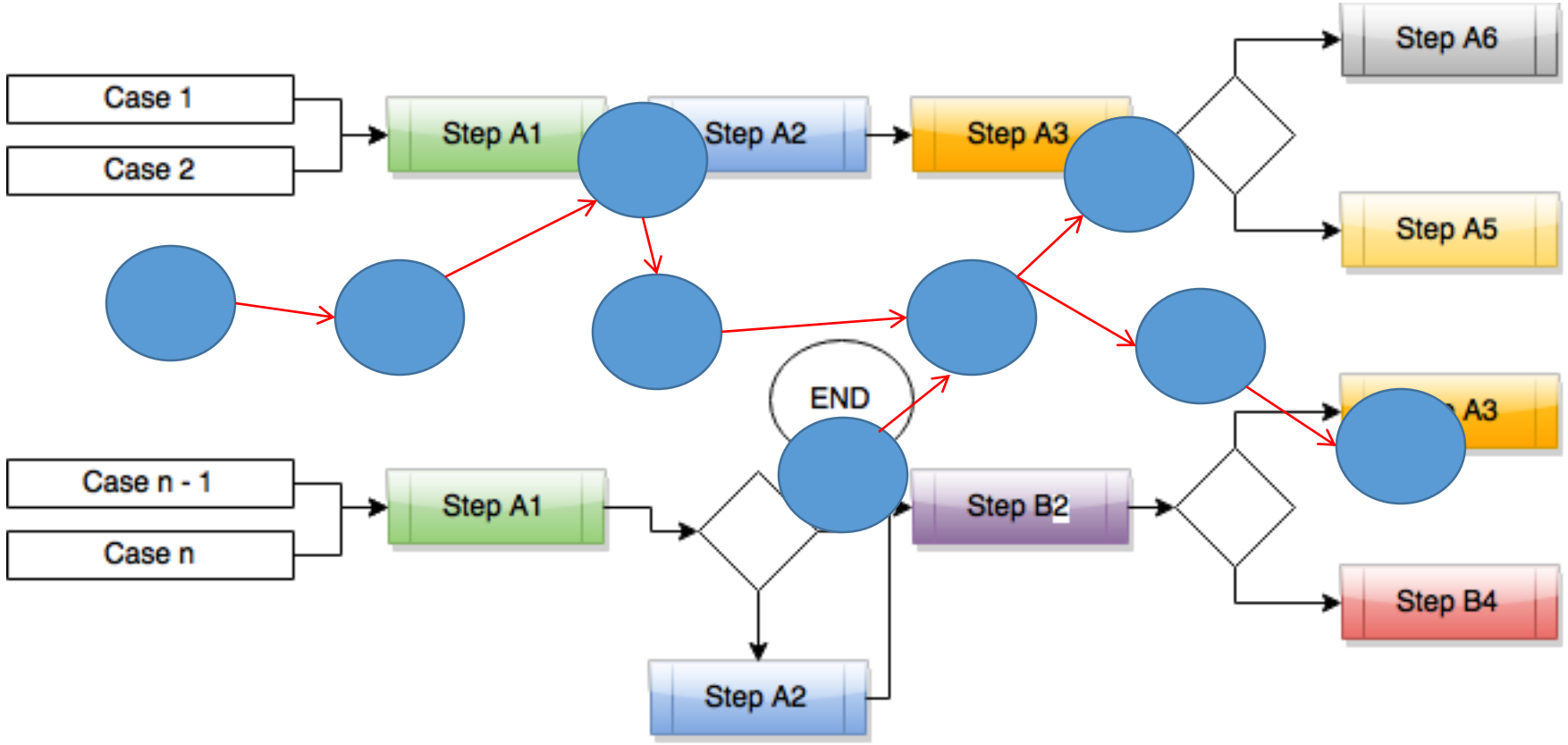
- The problems really start when you **reach back into the tree and change something**; for example, a new business strategy is defined which invalidates the interpretation and everything downstream of that

Pinterest.com

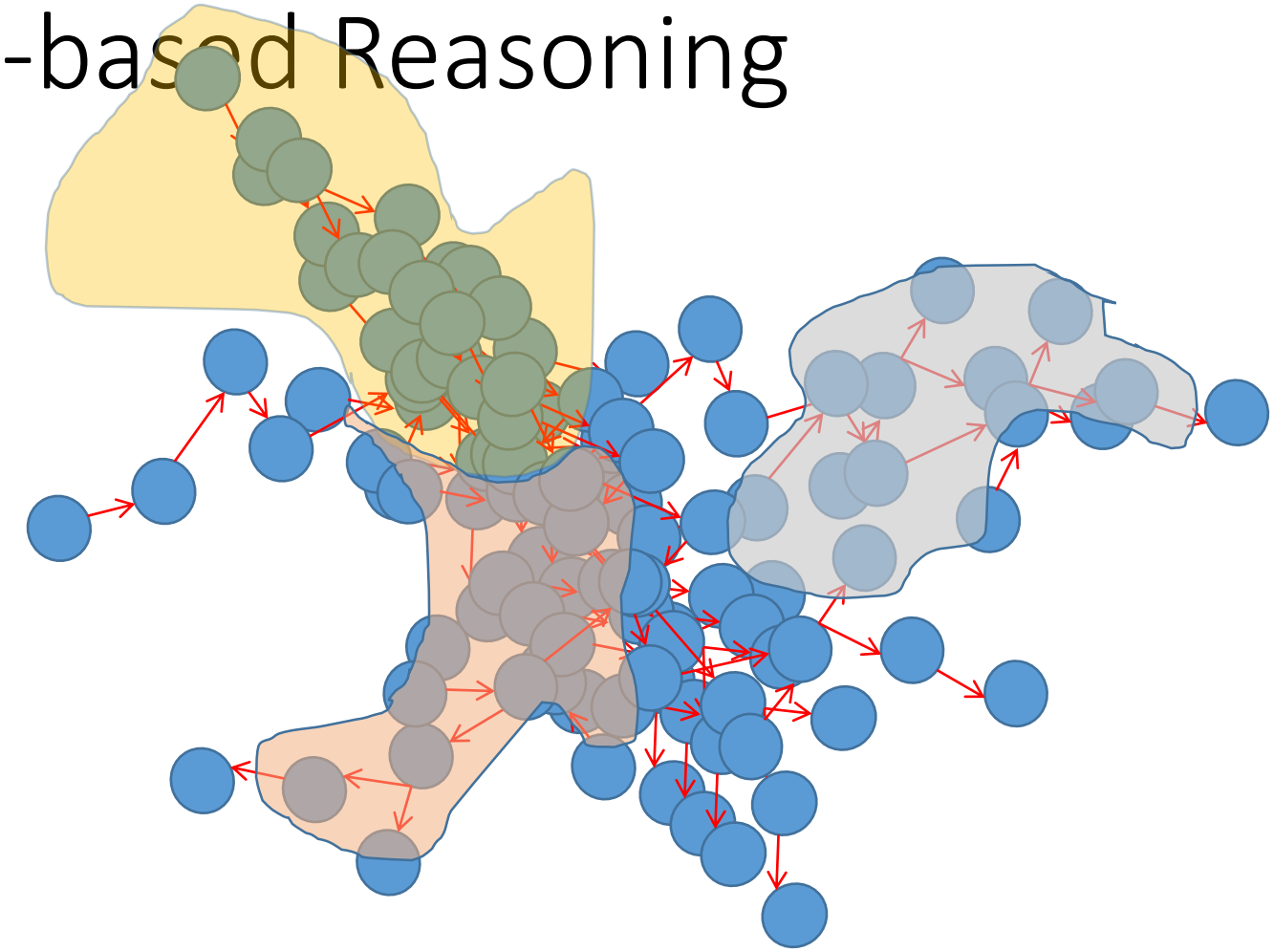
Intelligent Workflow Monitoring

- Data Mining techniques
- Machine Learning
- Artificial Intelligence
- Case-based Reasoning

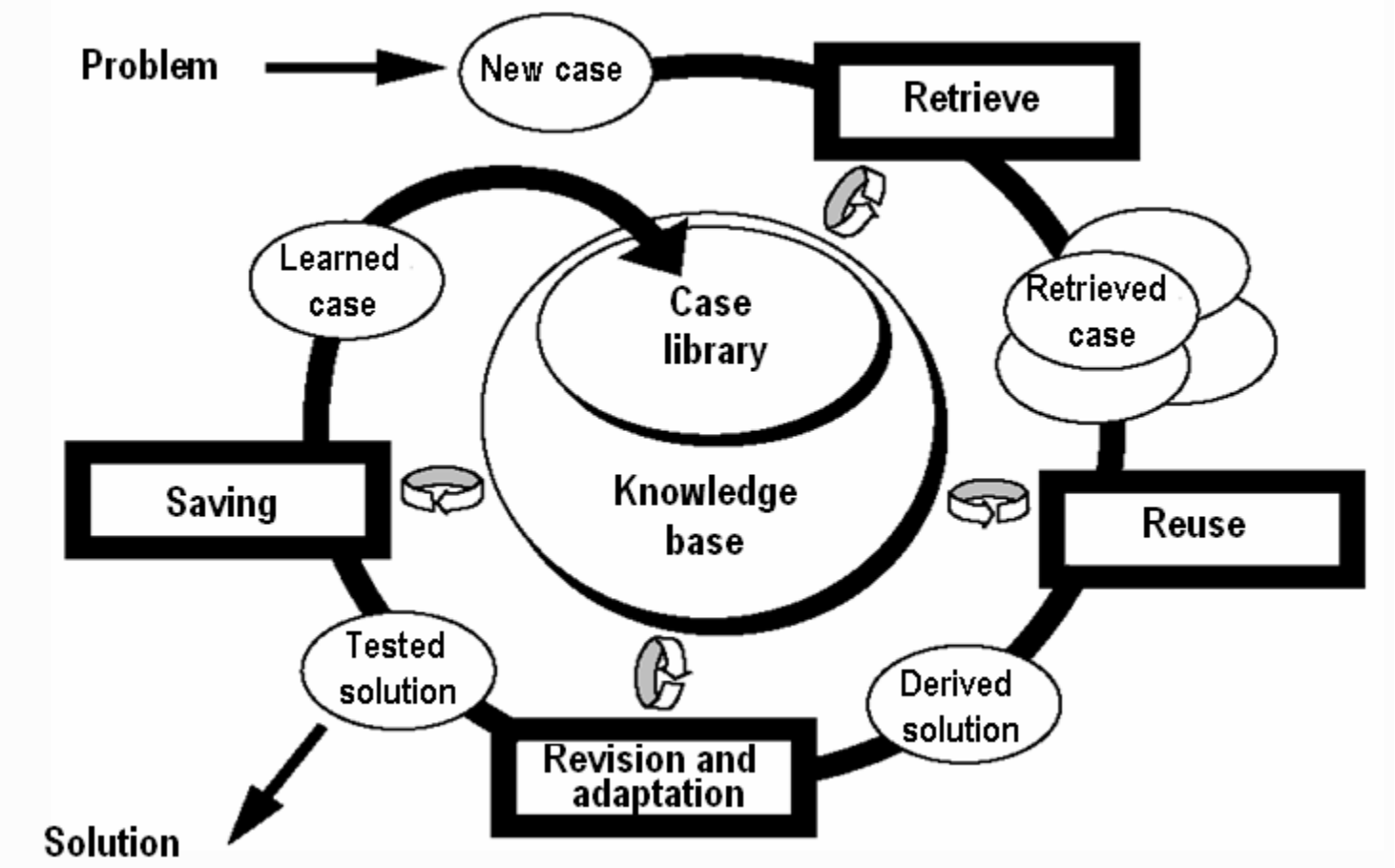
Workflows as cases



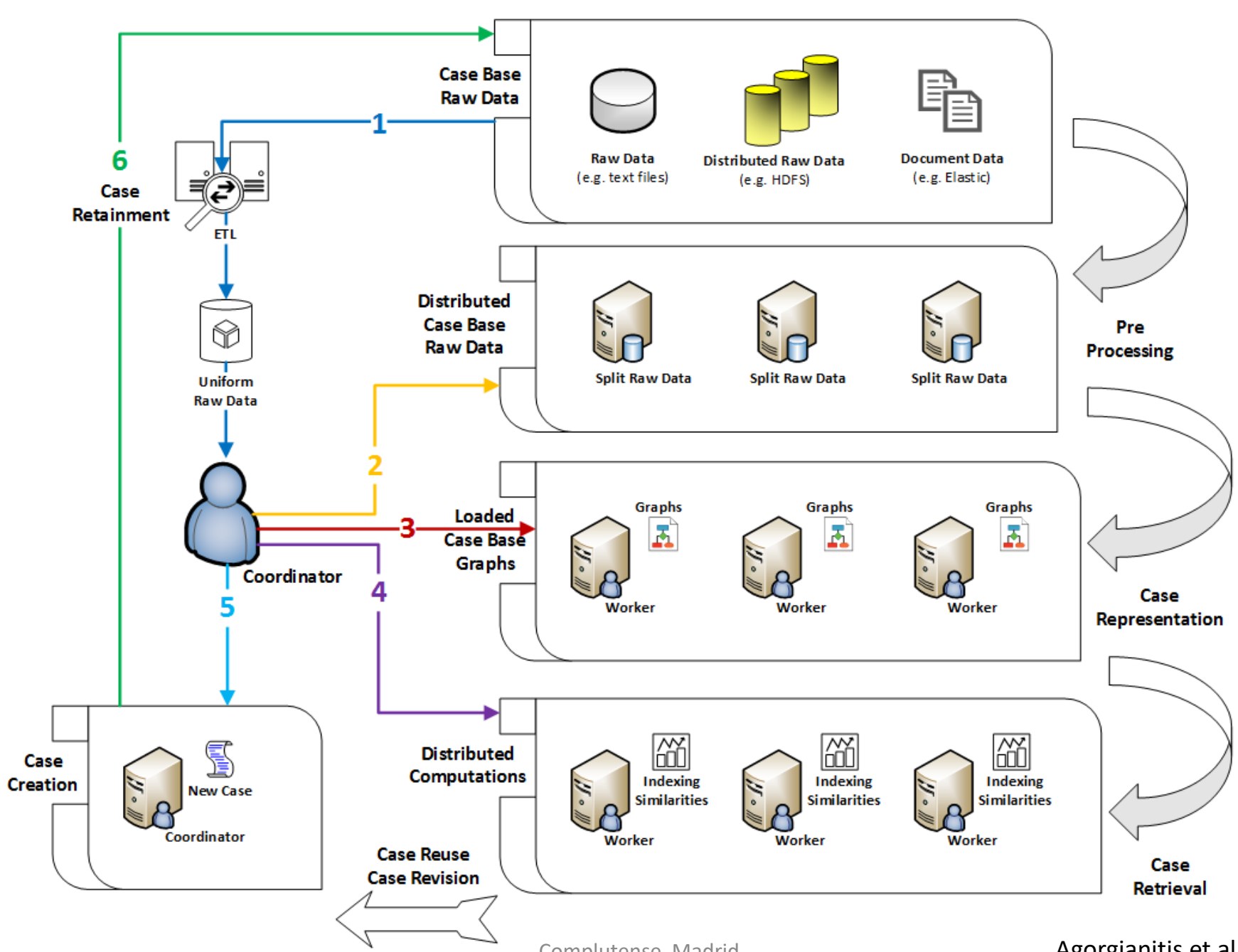
Case-based Reasoning



CBR Mechanics

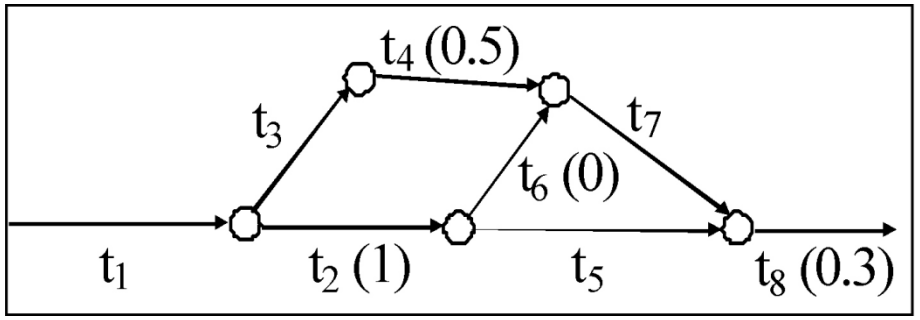


Source: Eremeev &. Vagin, 2011


Case Base
Raw Data
Raw Data
(e.g. text files)
Distributed Raw Data
(e.g. HDFS)
Document Data
(e.g. Elastic)
1
6
Case
Retainment
ETL
Uniform
Raw Data
Distributed
Case Base
Raw Data
Split Raw Data
Split Raw Data
Split Raw Data
Pre
Processing
2
3
4
5
Loaded
Case Base
Graphs
Graphs
Graphs
Graphs
Worker
Worker
Worker
Coordinator
Case
Representation
Case
Creation
New Case
Coordinator
Distributed
Computations
Indexing
Similarities
Indexing
Similarities
Indexing
Similarities
Worker
Worker
Worker
Case Reuse
Case Revision
Case
Retrieval

Time Mechanics

- GTT: General Time Theory [Ma & Knight, 1994]: Improved General Theory of Time



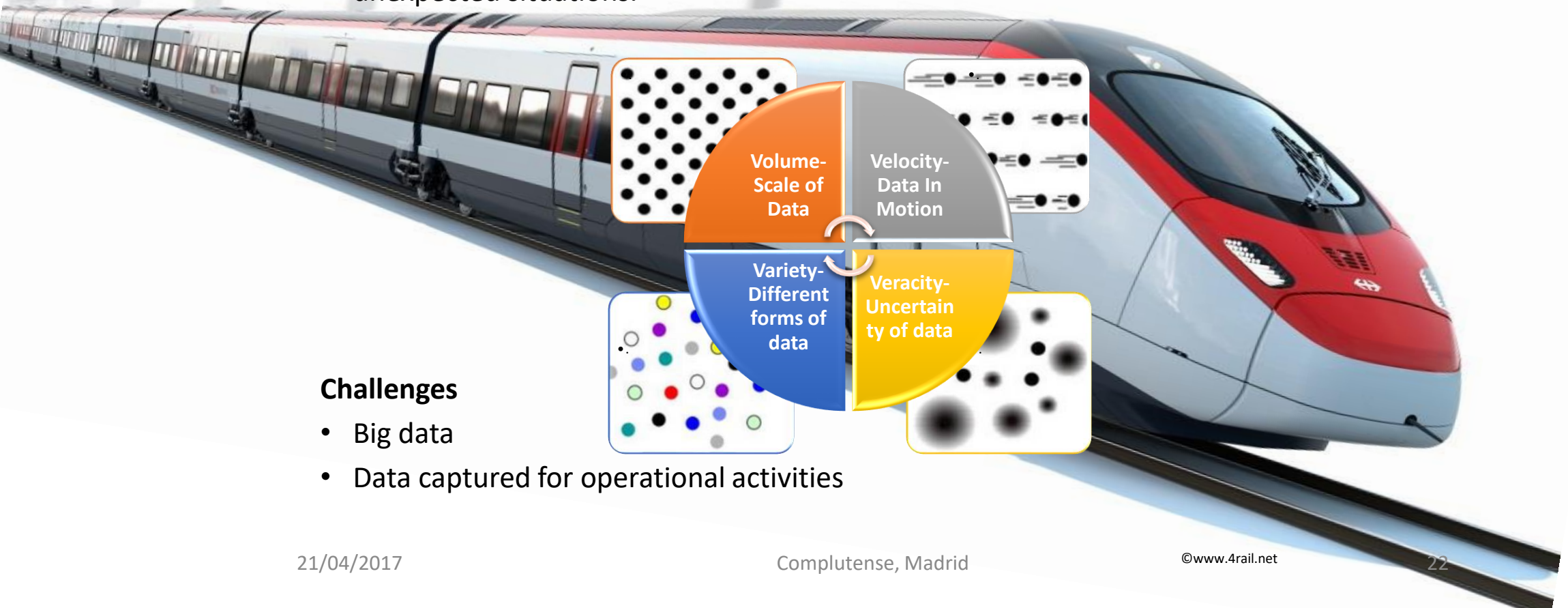
- Maximum Common Sub-graph [D. Conte et al., 2007]

$$S(G,G') = \frac{\left(\sum_{\substack{matches\\ C,C'\\ in\\ MCSG}} \sigma(C,C')\right)^2}{count(G).count(G')}$$

Case study: UK Rail Industry

Q: **How train operational data could be used in order to provide information/insights of trains' behaviour and performance?**

- Visualise RCM data to understand the granularity of delays
- Understand how trains reacted over signals, system failures and other various unexpected situations.

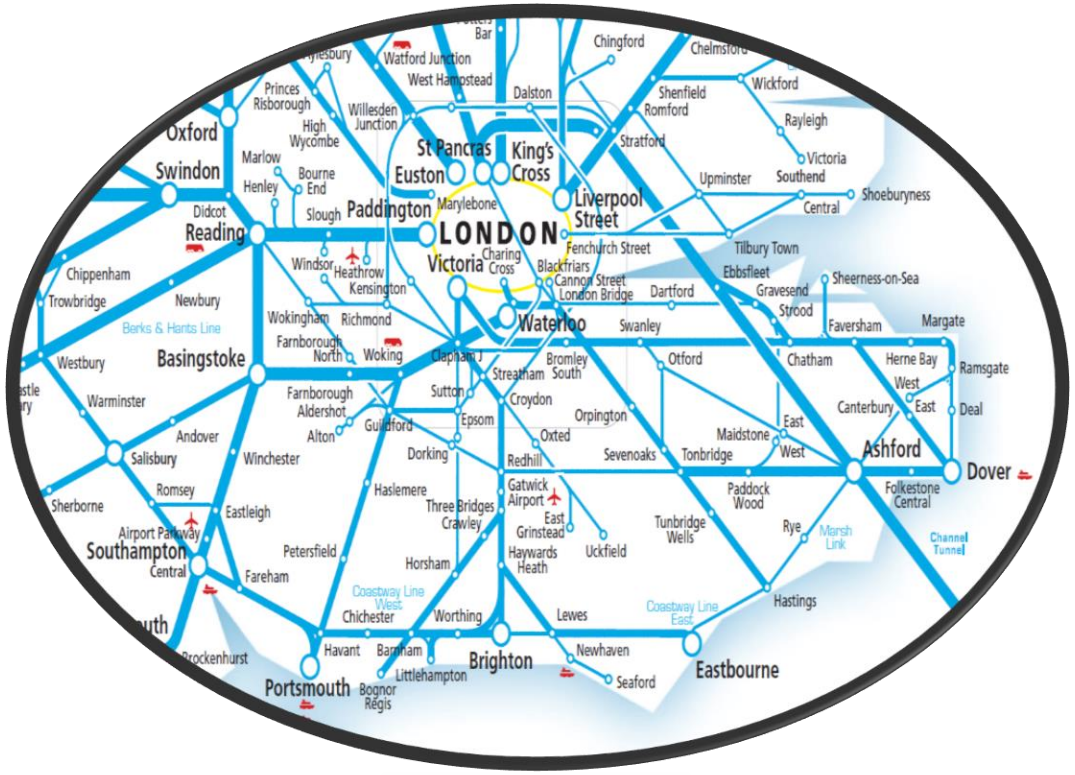


Challenges

- Big data
- Data captured for operational activities

British National Rail

- Case & Solution Architecture
- Process
- Implementation time
  - 4 years (4 phases)
- Future perspectives
  - Now on phase 3
  - 4 phases (Visualisation, ..Monitoring, Data Mining, Simulation)



Real Challenges

- **Data silos (dlis, las, lis)**
- **Application silos (apis)**
- **Library style data management**
- **Project, corporate, or master?**
- **Never fixing the data (gps)**
- **Big data vs “lots of data”**
- **Decide by PowerPoint**
- **We do what everybody else does**

The End

Thank you!

Contact Details

Dr Stelios Kapetanakis

email: s.kapetanakis@brighton.ac.uk

linkedin: stelios-kapetanakis-ba816440

References

- [Agorigianitis, 2017]: Agorgianitis, I., Kapetanakis, S., Petridis, M., Fish, A. (2017) Business Process Workflow Monitoring using Distributed CBR with GPU Computing, In the 30th International FLAIRS Conference, Florida, Marco Island, pp.48-57, May 2017

- [Chiu, 2008]: Chiu, D. (2008), Business Process and Workflow Management

- [D. Conte et al., 2007]: Conte, D., Foggia, P., Vento, M., (2007) Challenging Complexity of Maximum Common Subgraph Detection Algorithms: A Performance Analysis of Three Algorithms on a Wide Database of Graphs, Journal of Graph Algorithms and Applications , 11(1) pp. 99–143

- [Eremeev &Vagin, 2011]: Eremeev, A.P., Vagin, V. N. (2011) Common Sense Reasoning in Diagnostic Systems, Efficient Decision Support Systems - Practice and Challenges From Current to Future, Prof. Chiang Jao (Ed.), ISBN: 978-953-307-326-2

- [Ma & Knight, 1994]: Ma, J., Knight, B. (1994) A General Temporal Theory, The Computer Journal, Vol.37(2), 114-123, 1994.