

Najväčšie riziko Big Data — nejasný účel projektu?

Ivo Kovačič

<https://sk.linkedin.com/in/ivokovacic>

ABIT 2015

Ivo Kovačič

ivo.kovacic@atos.net

Atos

Atos IT Services and Solutions

Head of Solutions
Slovakia and Czech Republic

- Aplikačná integrácia, ESB, BPM
- Vývoj aplikácií
- Document management, document workflow, podpisovanie dokumentov
- Data and information analysis
- IT bezpečnosť - biometria
- SAP

Obsah

- Kto je vlastníkom Big Data príležitosti a problému
- Kto má predpoklady na úspešný Big Data projekt
- Biznis problém – rozhodovanie – informácie
- Hypotéza – model – predikcia a ich rola v rozhodovaní
- Interpretácia dát
- Aký je recept na úspešný projekt

Image je nanič,

počúvaj smäd !

Image je nanič,

počúvaj smäd !

Intuícia je nanič,

počúvaj Big Data!

**Prečo pri takom množstve marketingu,
osvety a predávania nie je naše okolie plné
šťastných používateľov big data riešení?**

**Prečo napriek zrejmemu obchodnému
úsiliu nevidíme množstvo úspešných
projektov?**

Image je nanič,

počúvaj smäd !

Big Data sú nanič,

počúvaj ... ???

„Business has no Big Data problem“

Treating big data as a problem which must be addressed is misguided. “It is only a problem when viewed from an IT-centric perspective,” says Lounsbury, and he advises discovering the opportunities business analytics presents. (Dave Lounsbury, chief technology officer (CTO) at standards body the open Group.)

Business2020 - A special report on the future of business and IT from Computer Weekly, in association with ORACLE • issue 2 of 4, 2013

**“BUSINESS LEADERS
KNOW THAT THE ABILITY
TO GET AND UNDERSTAND
COMPETTIVE DATA IS
GOLD DUST”**

DAVE LOUNSBURY,
THE OPEN GROUP

**“IT’S LIKE SPONSORING A PEARL
DIVING EXPEDITION – THERE
HAS TO BE SOME RISK AS WITH
ANYTHING SPECULATIVE IN R&D”**

BOB HARRIS, CHANNEL 4

**“THE TRUE VALUE OF DATA
COMES FROM BEING ABLE
TO CONTEXTUALISE AND
UNDERSTAND IT, IN ORDER
TO DELIVER INSIGHTS”**

SURANJAN SOM, IMGROUP

**“THE BEST TOOLS OF
TOMORROW WILL NOT JUST
REPORT WHAT HAPPENED,
OR WHAT IS HAPPENING IN
REAL TIME, BUT WILL OFFER
ALTERNATIVE SCENARIOS”**

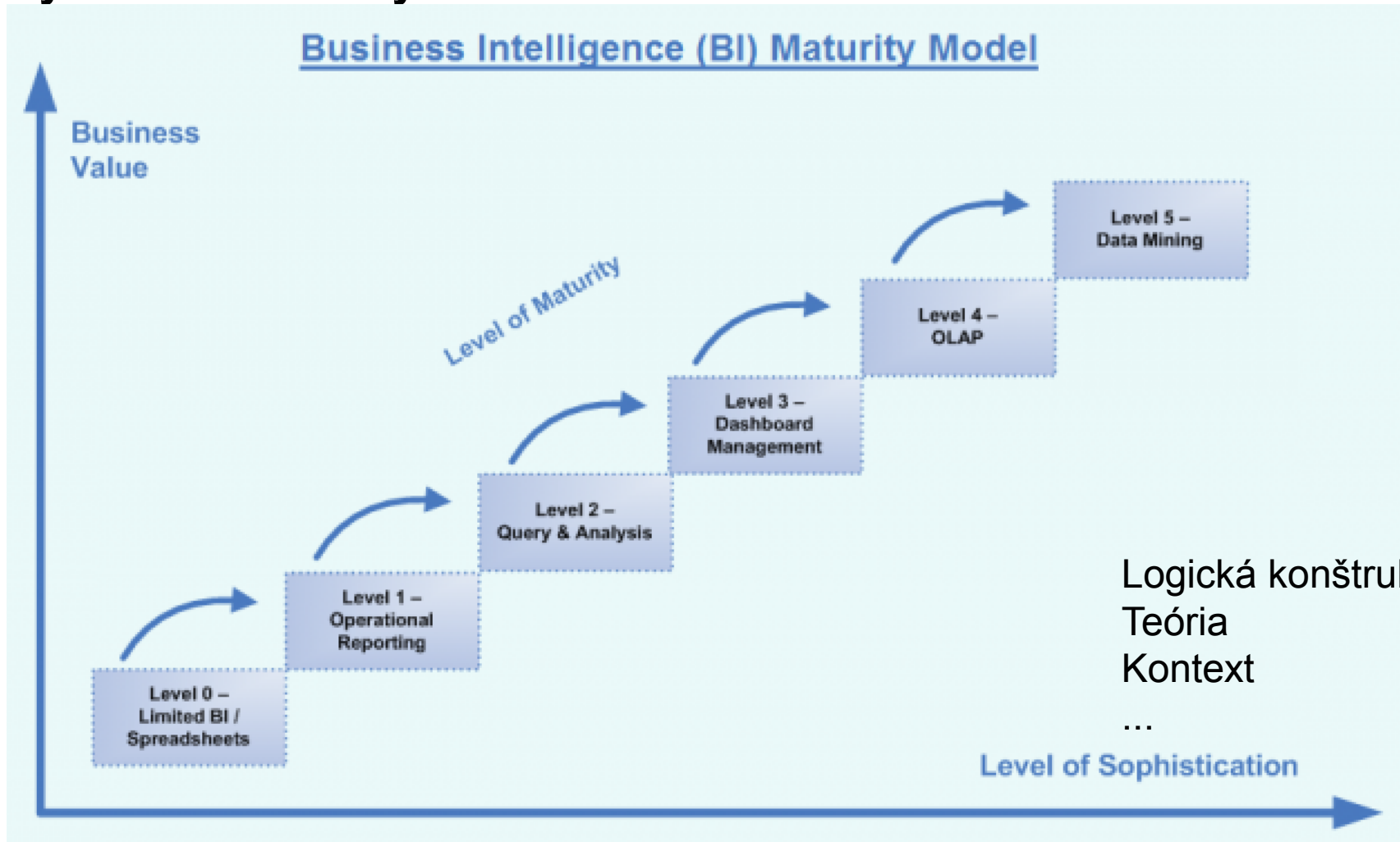
DARREN VENGROFF,
RICHRELEVANCE



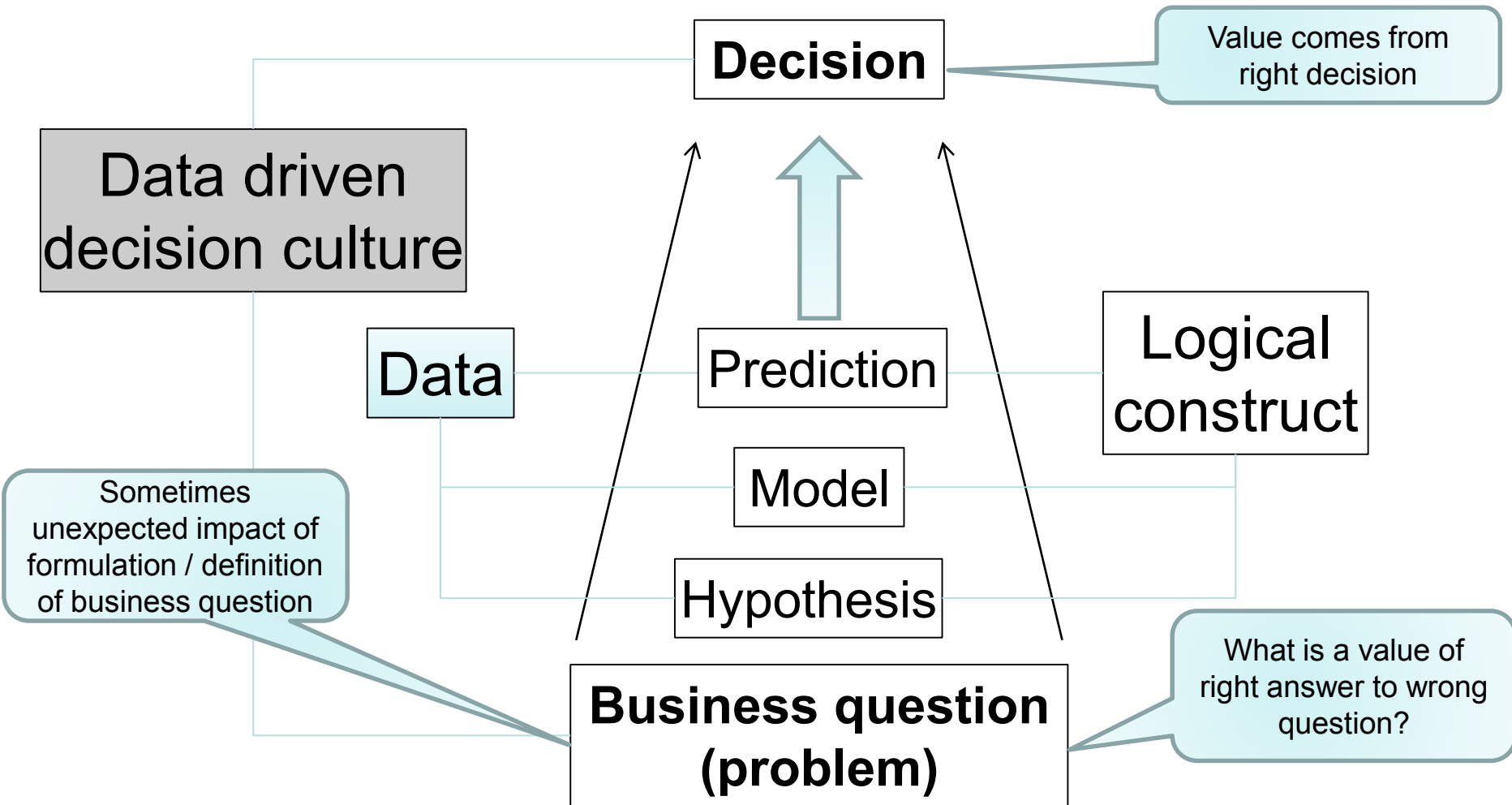
ComputerWeekly.com

- Understanding big data adoption and motivation in the UK - study
 - Gaining better customer understanding is expressed as the main motivation for big data programmes (50%), and organisations see big data as way to gain competitive advantage (48%).
 - Some 33% are looking to track sentiment towards their organisations on social networks. But more important still is machine data, with 38% seeking to make better use of sensor and log data to improve operations.
 - Although the survey shows that big data is for real, advanced technologies that pre-date the big data hype of recent years, such as data mining and predictive analytics, are still ahead in adoption. Currently 63% of respondents carry out data mining, 62% use predictive analytics, and 48% use big data analytics.
 - **However, in a press statement, IBM and the business school said 75% of organisations doing big data start with a strong core of analytics capabilities, such as query and reporting and data mining to address structured data, and 67% use predictive modelling.**

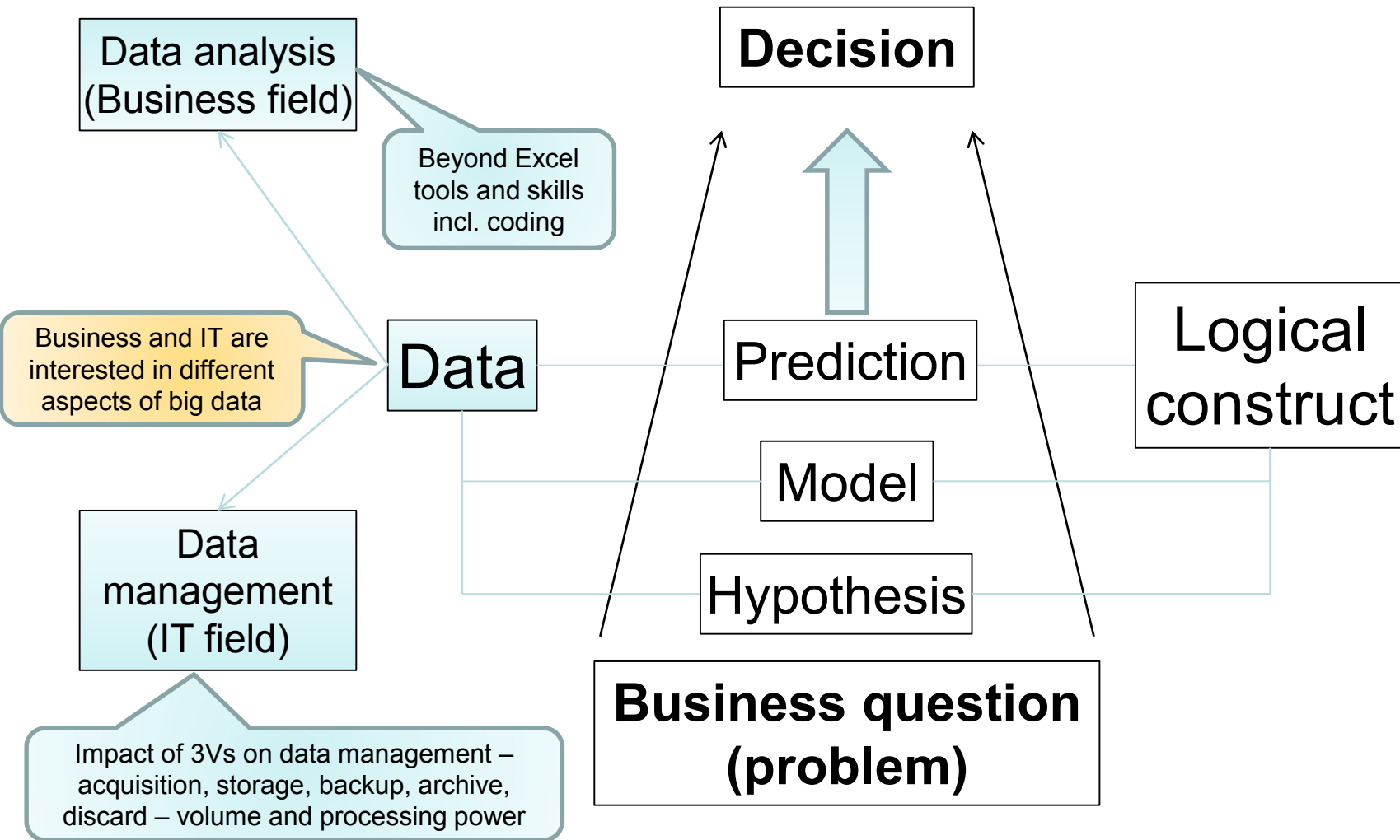
Význam BI maturity level



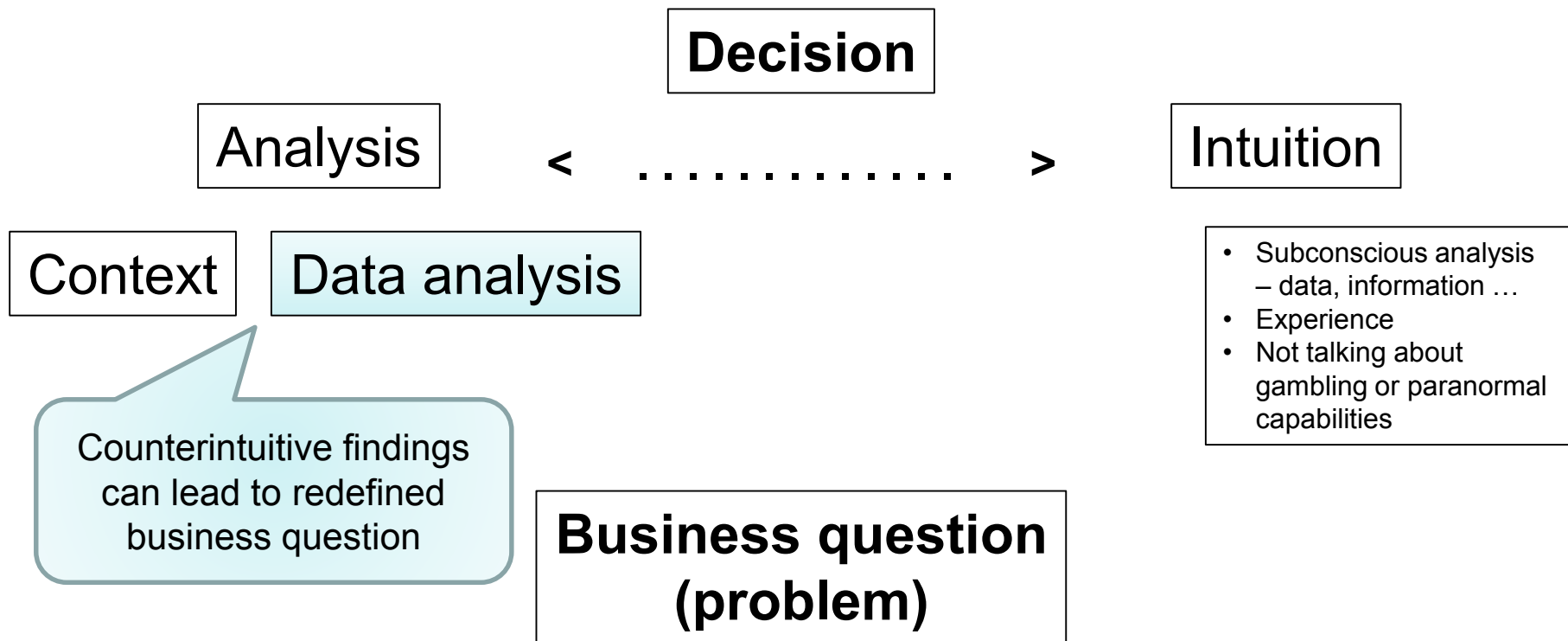
Business management – decisions and data



Business management – decisions and data



Decisions and data - big data specifics?



Basic principles of data analysis are independent from big data phenomenon. No business question, no decision needed.

Oblasti uplatnenia Big Data – prípadové štúdie

- Široká oblasť
 - Vedecké problémy – CERN, genóm ...
 - Technicko-prevádzkové – výroba, riadenie zariadení, doprava ...
 - Obchod a marketing – trendy v správaní sa zákazníkov, ovplyvňovanie rozhodovania ...
 - Spoločenské – voľby, referendá, prieskumy verejnej mienky – predpovede výsledkov a vedenie kampane – ovplyvňovanie ...
 - Športová výkonnosť – úžasné štatistiky ukazovateľov, ale čo s nimi ?

- Ciele
 - Predpovede
 - Rozhodnutia

Čo vyžadujú predpovede a rozhodnutia?

- Vzťah minulosť – budúcnosť
- Vzťah špecifické (konkrétne) – všeobecné
- Vzťah kauzalita – korelácia
- Vzťah determinizmus – náhodnosť
- Význam interpretácie - pochopenie zmyslu je kľúčové
- Analógia nie je vedecká metóda 😊

Je viac informácií viac?



Je viac informácií viac?



Čo vieme o veličinách, metrikách a ich vzťahoch?

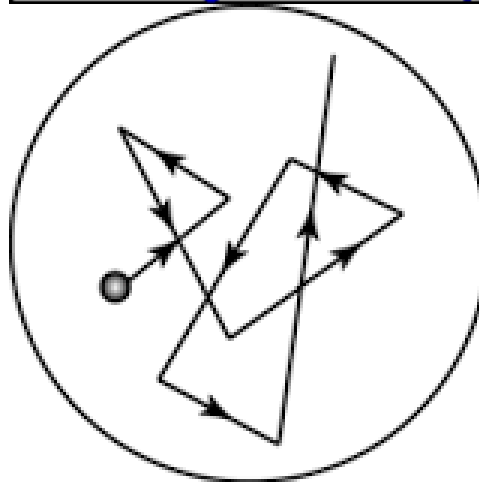
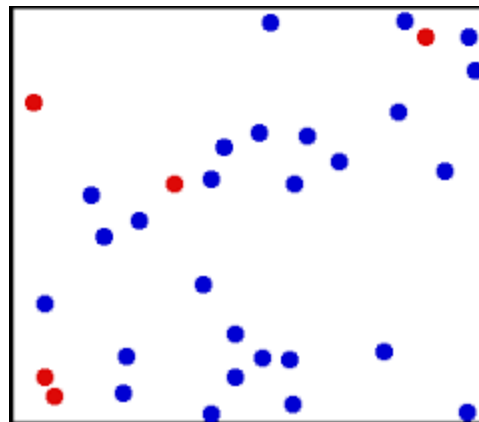
Mikro

- m

- v

- \vec{v}

- l



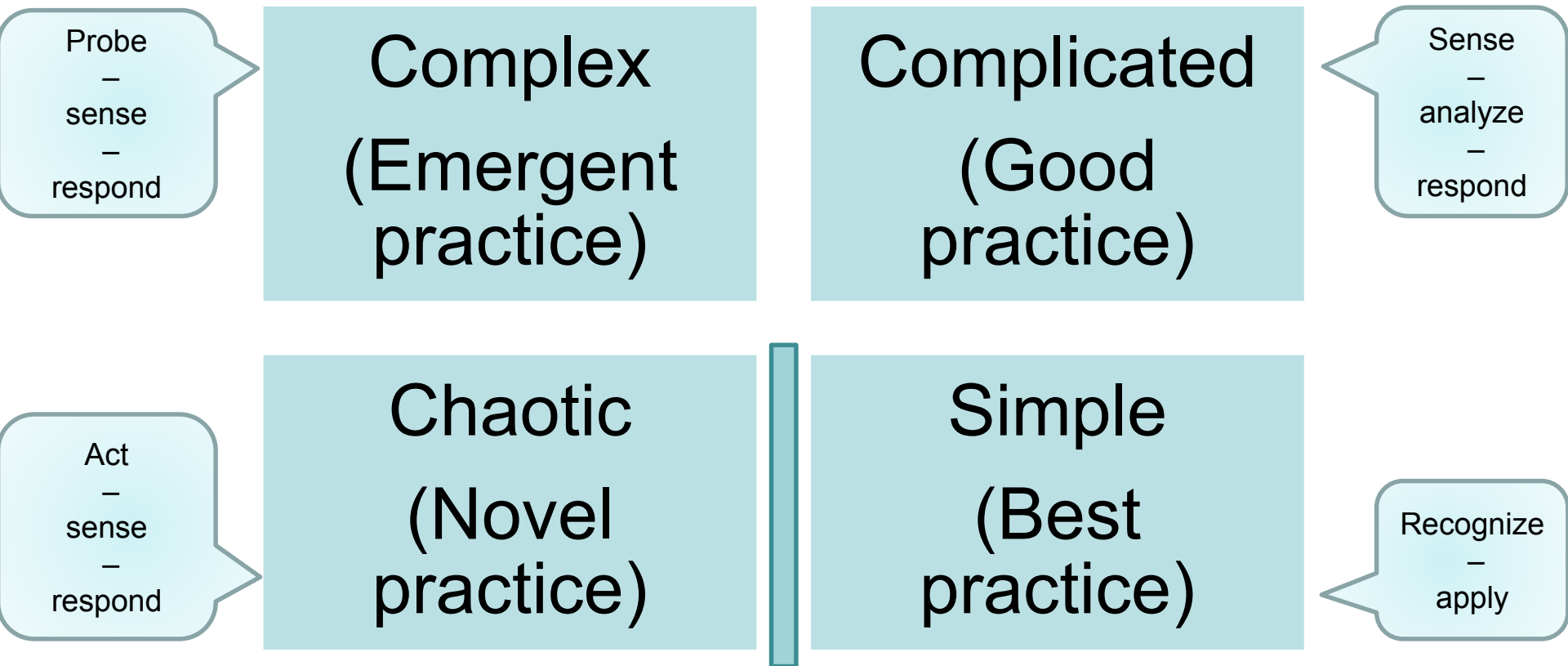
Brownian Movement

Makro

$$p \cdot V = n \cdot R \cdot T$$

Interpretácia – pochopenie zmyslu

Cynefin – sense making framework



Sense making framework

- Pomáha prechádzať elementárnym chybám
 - Chýbajúca formulácia problému / otázky
 - Charakter rozhodnutia – strategické vs operatívne
 - Záměna korelácie a kauzality
 - Výber metriky - vzťah dát a reality priamy / nepriamy –ukazovatele priame / proxy
 - Validácia počas interpretácie

- Žiaden nástroj a štatistické spracovanie nenahradí
 - Správne formulované otázky
 - Hypotézy podložené logickou konštrukciou
 - Teóriu ako základ predikcie

- Ale...
 - **Experiment a dáta sú kľúčom k potvrdeniu alebo vyvráteniu hypotézy**

HBR recommendation How to Benefit from Big Data

To improve performance with advanced analytics, companies need to develop strengths in three areas.

START WITH DEFINITION OF QUESTION

1

Multiple Data Sources

Creatively source internal and external data.

Upgrade IT architecture and infrastructure for easy merging of data.

2

Prediction and Optimization Models

Focus on the biggest drivers of performance.

Build models that balance complexity with ease of use.

3

Organizational Transformation

Create simple, understandable tools for people on the front lines.

Update processes and develop capabilities to enable tool use.

1. Chose right data
2. Build models that predict and optimize business outcomes
3. Transform your company's capabilities

Key elements of approach

- Problem definition
- Identification of knowledge or information gap
- Hypothesis (HBR - Big data's power does not erase the need for vision or human insight.)
- Verification (accept or reject)
- Prediction
- Continue in loop

- **Differentiate between correlation and causality**

Ako to súvisí s bezpečnosťou a ochranou osobnosti

- Ak je relevantne stanovená otázka (biznis problém) a teda cieľ analýzy ...
 - Obmedzenie zberu a skladovania dát
 - Výber techniky na anonymizáciu dát
 - Zachovanie vypovedacej schopnosti vzorky dát
 - Výnimky - trendy
 - Individuálne / konkrétne bude zachované alebo potlačené

Strategický element v príprave Big Data projektu

- Vylúčte technologické vplyvy / IT jednoduchým predpokladom: „Technické prostriedky sú k dispozícii“
- Čo sú rozhodnutia a informácie, ktoré vám prinesú efekt – profit, obrat, zákazníkov
- Aká je predpokladaná ekonomická hodnota prínosov?
- *Ak nepracujete v CERN, NASA alebo na výskume genómu, je dobrá šanca, že si tento prístup môžete dovoliť*

Image je nanič,

počúvaj smäd !

Big Data sú nanič,

***počúvaj rozum,
počúvaj zákazníka...***

Ďakujem za pozornosť