UNIVERSITY OF NEBRASKA AT OMAHA

NexTech 2019: Panel on Advances in Data Processing

Hesham H. Ali
UNO Bioinformatics Research Group
College of Information Science & Technology
University of Nebraska at Omaha

Panel on Advances in Data Processing

Moderator: Hesham Ali, University of Nebraska Omaha, USA

Panelists:

- Dimitris Kardaras, Athens University of Economics and Business, Greece
- Wiam Ramadan, Lebanese International University, Lebanon
- István Vassányi, University of Pannonia, Hungary
- Hesham Ali, University of Nebraska Omaha, USA

Advances in Data Processing – The Debate

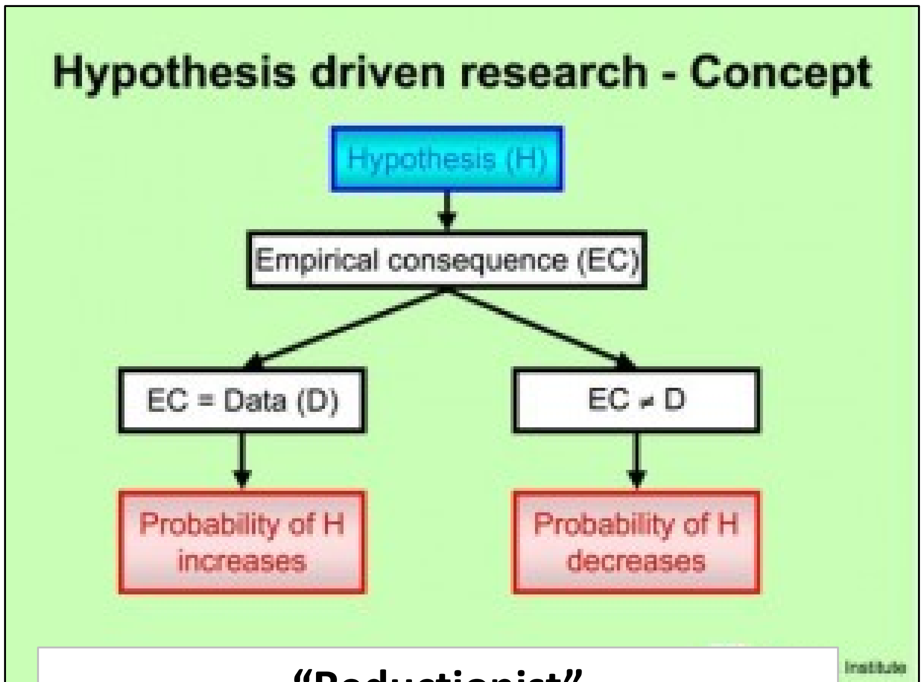
- On the one hand, recent advances in data availability and use have touched many aspects of our lives and influenced various scientific areas. Advances in Data Processing has been a big Success.
- On the other hand, it can be argued that significant impact on critical domains remains limited. Many scientific areas enjoy now more available data but such data don't always transfer into impactful knowledge. Healthcare still struggles with chronic diseases and advances in this area remain below expectations. Advances in Data Processing has some chanlleges to overcome.

Scientific Research will never be the same

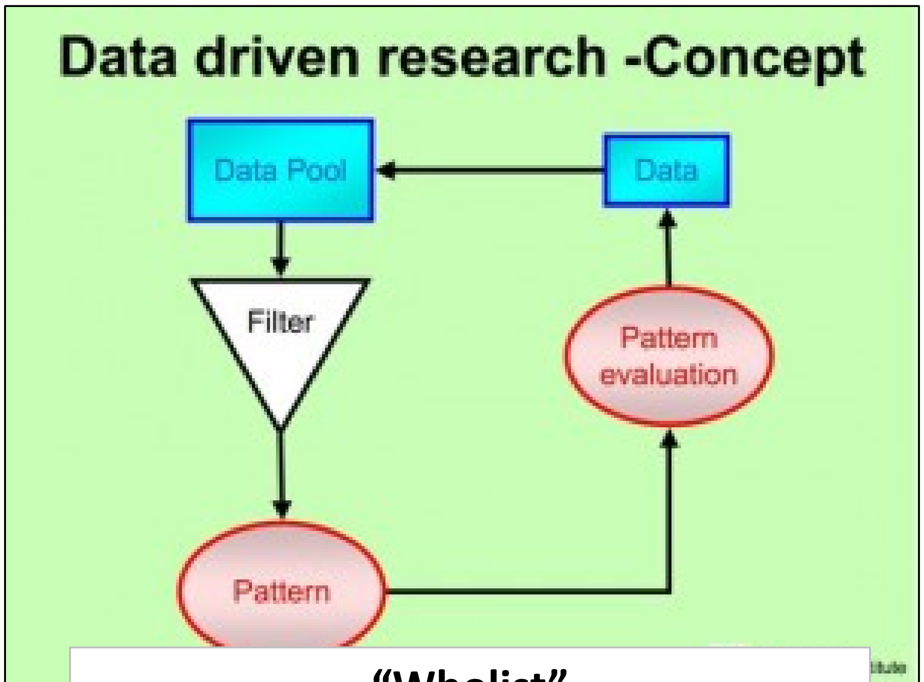
- IT is changing many scientific disciplines
- So much relevant data is currently available
- The availability of data shifted many branches in sciences from pure experimental disciplines to knowledge based disciplines
- Incorporating Computational Sciences and other branches of sciences is not easy
- Interdisciplinary Research? Translational Research? Big Data Analytics?

A Potential Major Change

Data driven research vs. Hypothesis driven research



"Reductionist"
"Traditional"

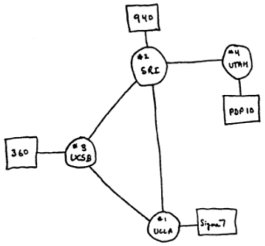


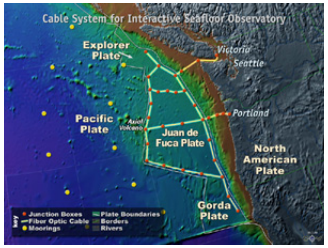
"Wholist"
"Systems"

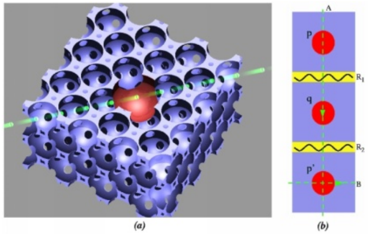
http://www.genome-engineering.com/from-hypothesis-to-data-driven-research.html

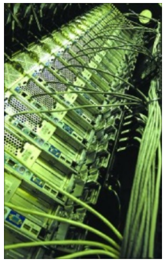
The Future is Full of Opportunity

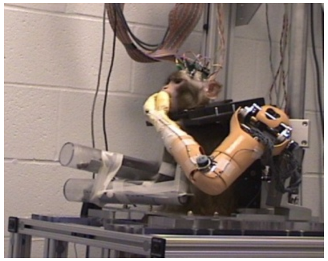
- Creating the future of networking
- Driving advances in all fields of science and engineering
- Revolutionizing transportation
- Personalized education
- The smart grid
- Predictive, preventive, personalized medicine
- Quantum computing
- Empowerment for the developing world
- Personalized health monitoring => quality of life
- Harnessing parallelism
- Neurobotics
- Synthetic biology

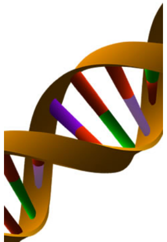


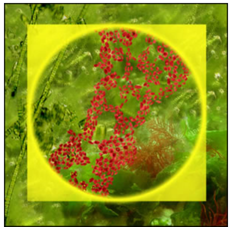








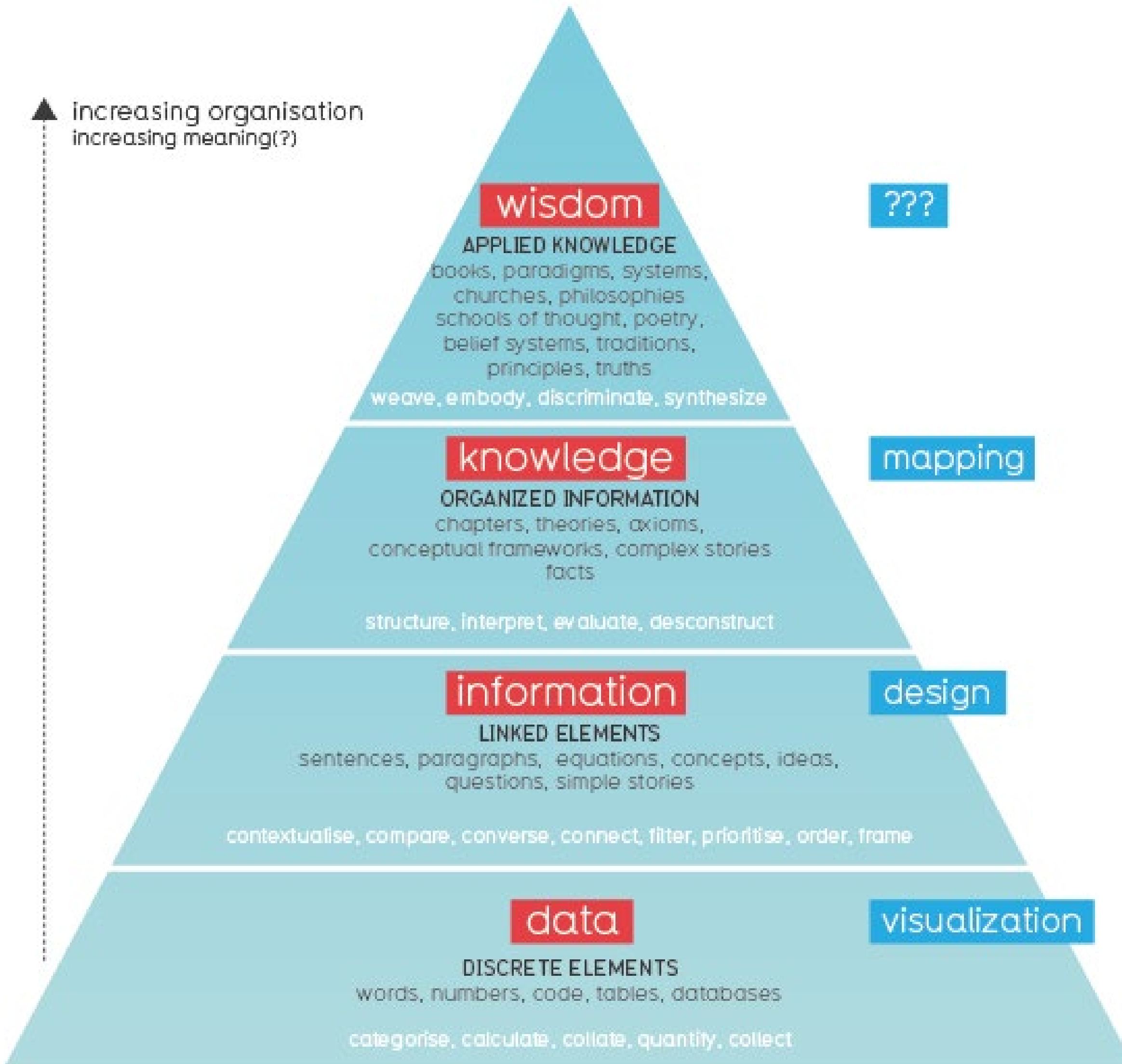




It’s all about the Data!

- How it all began:
  - Advances in medical instruments and computational technologies led to new new research directions
  - Massive accumulation of Biomedical data led to investigating new potential discoveries
  - The availability of enormous various types of public/private Biomedical data
  - How to take advantage of the available data
- Bioinformatics - Health Informatics - Biomedical Imaging - Public Health Informatics Biomedical Devices
- A new direction is now possible

Data-Information-Knowledge-Wisdom



http://www.ritholtz.com/blog/wp-content/uploads/2010/11/data_info_knowledge_wisdom.png

Biomedical Informatics in 2019

- Each generation, a scientific discipline emerges with a bang and promises to change the way we do things – a game changer.
- Back at 1995, one would have expected Bioinformatics to be further along after over 20 years.

The Debate

- On the one hand, IT has delivered the goods: recent advances in data availability and use have touched many aspects of our lives and influenced various scientific areas.
  - The telecom revolution and the associated connectivity, the impact of social networks on political movements, smart devices and environments, many advances in medical research, just to name a few.
- On the other hand, IT has not delivered at the big stage: significant impact on critical domains remains limited.
  - Many scientific areas enjoy now more available data but such data don't always transfer into impactful knowledge.
  - Healthcare still struggles with chronic diseases and advances in this area remain below expectations. Similar arguments can be made related to environmental studies, mental illness, and care for elder population.
  - Many scientists in sensitive or critical domains have resisted full adaptation of IT and remain suspicious, in part due to lack of robustness and reproducibility in IT studies.

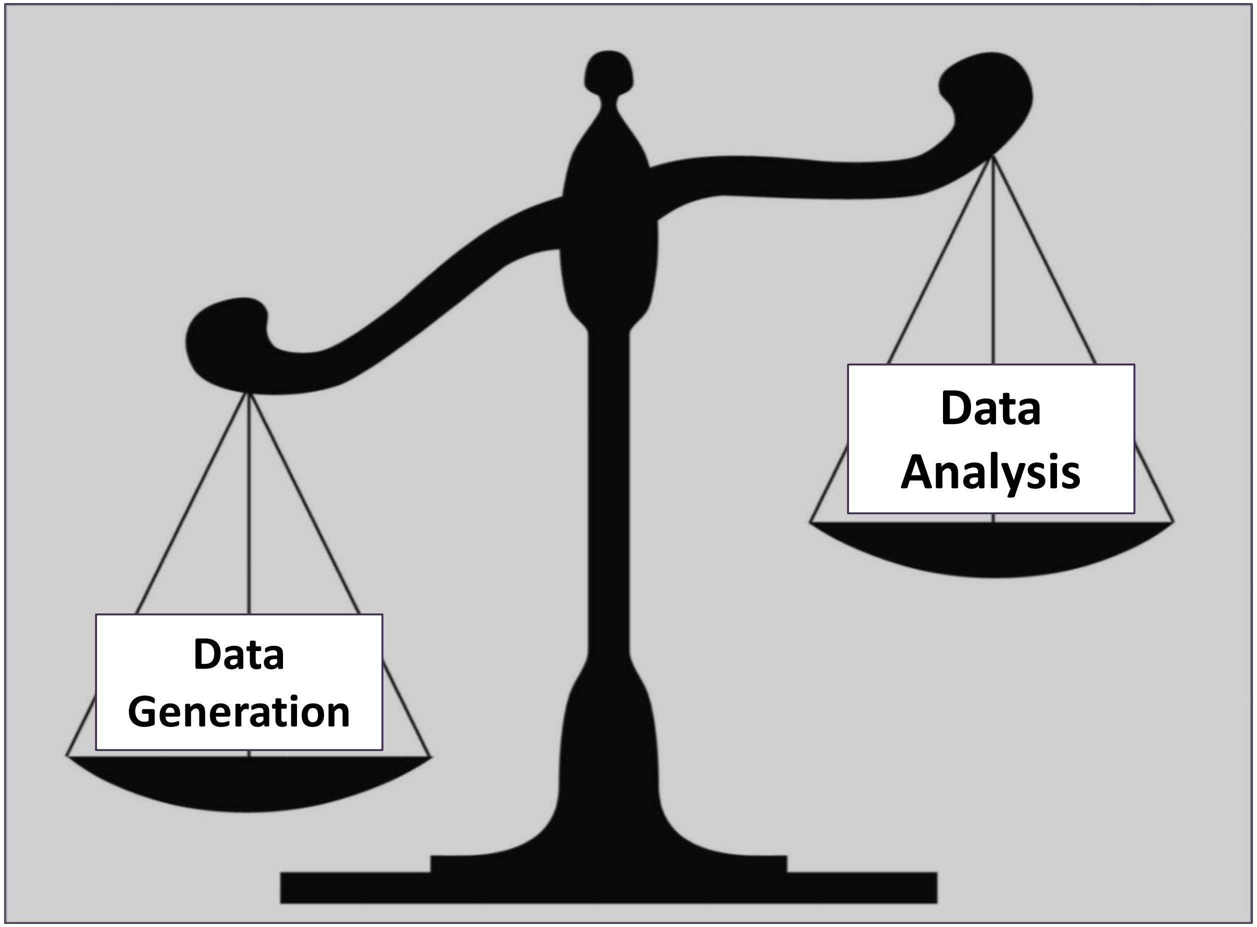
What are the barriers?

- On the biomedical (domain) side:
  - Too much focus on data collection
  - Competition to own the latest technology
  - Excitement associated with New technologies – however that leads to new data
  - The black box syndrome
- On the computational/informatics side:
  - Certain level of casualness remain a major concern – just another application domain
  - Inconsistent results – lack of robustness and reproducibility
  - Heuristics and thresholds
  - Lack of Biomedical-rich integration

Data Generation vs. Data Analysis/Integration

- New technologies lead to new data:
  - Competition to have the latest technology
  - Focus on storage needs to store yet more data
- Biomedical community needs to move from a total focus on data generation to a blended focus of measured data generation and data analysis/interpretation/visualization
- How do we leverage data? Integratable? Scalable?
- From Data to Information to Knowledge to Advanced Decision Making

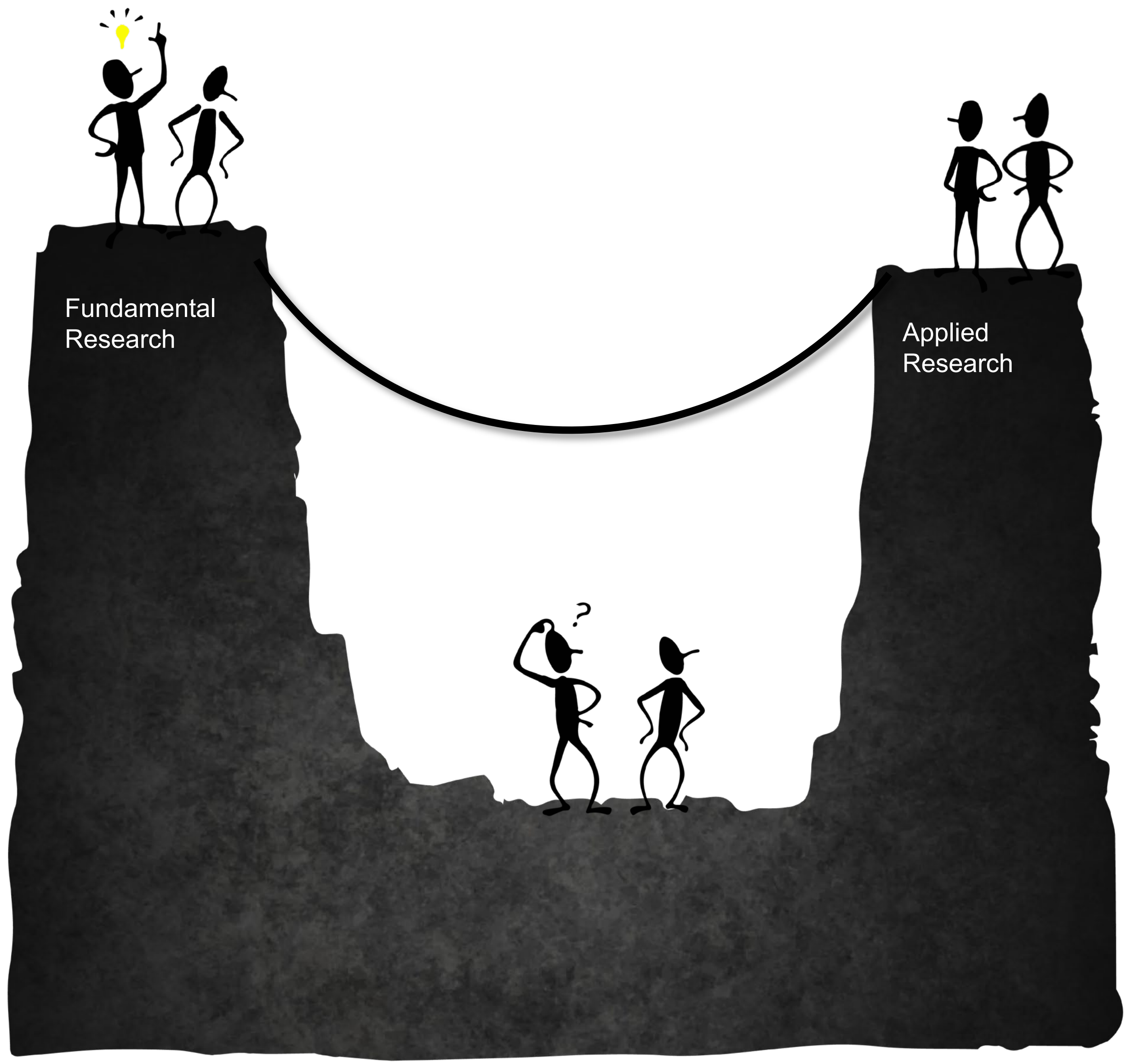
Time for a Better Balance?



Challenges

- Technical Challenges:
  - The need to analyze existing data not just collect data Are they all accurate? Complete?
  - Aggregation versus no aggregation
  - Ideal level of analysis - granularity
  - How results can be verified? Validated?
- Philosophical Challenges:
  - Collaboration
  - Work on the correct problem
  - Need for genuine translational work

Transitional Research



Translational Research

- The right problem
- Basic research: fundamental results
- Translational research: interdisciplinary groups
- Applied work: new diagnostics or cure
- Validation: Experimental and computational

Biomedical Informatics and Big Data

- All the features of Big Data are represented:
  - Volume: New levels of massive data
  - Variety: Only one type of data is not enough
  - Veracity: Not always fully compete or fully trusted
  - Velocity: Data is collected continuously

- Multiple levels of Big Data analysis:
  - Populations
  - Individuals
  - Many granularities in between

Informatics and NexTech

- The NexTech format is unique – it has a room for all critical components of Informatics
- Many invited talks and contributed papers cross boundaries between multiple aspects of Informatics
- It would be helpful to look for ways to connect the various conferences more – highlight the common thread

Panel on Advances in Data Processing
(Data Analytics 2019, Porto Portugal)

Dimitris K. Kardaras
Athens University of Economics and Business,
Athens, Greece

Data Growth I

- **A study by International Data Corporation (IDC)** estimates that the global amount of data will grow from 33 zettabytes in 2018 to 175 by 2025.

- **Cisco** foresees a massive growth of IP traffic – 4.8 zettabytes per year by 2022; three times the 2017 rate – lead by the increased use of IoT device traffic, video, and sheer number of new users.

Data Growth II

- 90ZB of data **will be created on IoT devices** by 2025.

- By 2025, 49 percent of data will be stored **in public cloud environments.**

- Nearly 30 percent of the data generated will be consumed **in real-time** by 2025.

Challenges in Data Analytics I

- **Need For Synchronization Across Disparate Data Sources;** i.e. there is a big challenge to incorporate them into an analytical platform.

- **Shortage Of Professionals Who Understand Big Data Analysis**; i.e. the job of a data scientist is multidisciplinary, needs domain knowledge also…

Challenges in Data Analytics II

- **Data Storage And Quality;** i.e. problem arises when a data warehouses **try to combine unstructured and inconsistent data** from diverse sources, it encounters errors. **Missing data, inconsistent data, logic conflicts, and data duplicates** all resulting in data quality challenges.

Challenges in Data Analytics III

- **Security And Privacy Of Data;** high risk of exposure of the data
- **Uncertainty Of Data Management Landscape;** new technologies and companies are being developed every day, the challenge for businesses is to find out which technology will be best suited to them

Challenges in Healthcare I

- Data is expected to grow 13% **faster than other industries**.

- **Hypercritical data** is expected to grow more than 47%.

- There is **little data management synergy** or consistency across the industry that makes learning from and collaborating with peers.

Challenges in Healthcare II

- IT investment in healthcare is among **the lowest of all industries**. As a result, IT departments have difficult catching up with data management challenges.

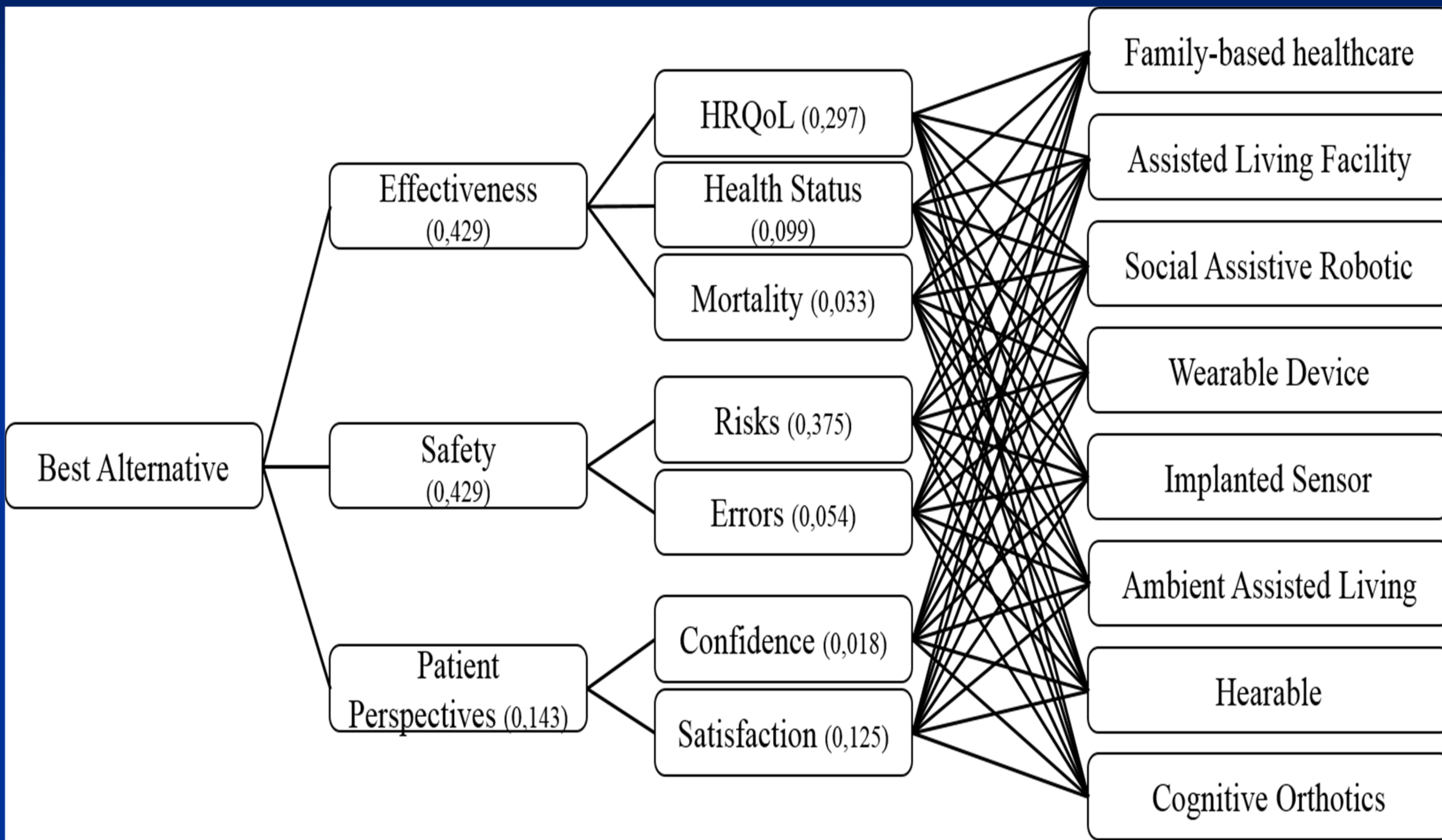
- Issues related to **data privacy and security** can hinder advancements in genomic research and personalized medication.

Our research in e-Healthcare…

- Also indicates that management and users in general are reluctant to adopt the new technologies.

- (Dimitrioglou N., Kardaras D., and Barbounaki S. (2017). **Multicriteria evaluation of the Internet of Things potential in health care: The case of dementia care, published in the 19th IEEE Conference on Business Informatics (CBI)**, Thessaloniki, Greece, 24-26 July, 2017, DOI 10.1109/CBI.2017.34, ISBN 978-1-5386-3034-1)

(IoT in Dementia) Relative weights of the criteria



Ranking per criterion and the final ranking of the alternatives

Alternative	Ranking per Criterion			Final Ranking
	Effectiveness	Safety	Patient Perspectives	
Family-based	1	1	1	1
Assisted Living Facility	2	2	2	2
Social Assistive Robotic	5	3	3	4
Wearable Device	6	5	7	5
Implanted Sensor	7	6	8	7
Ambient Assisted Living	3	4	4	3
Hearable	4	7	6	6
Cognitive orthotics	8	8	5	8

Conclusions I

- IoT technologies are not yet competitive, in accordance with expert's judgment

- AAL have the potential to overcome assisted living facilities, as it is very effective, but not yet…

- **However, it is not just the healthcare**… e.g. Data analytics and IoT in smart building and energy savings (proving the gains can be difficult…)

Conclusions II

- The use of technology adoption models and theories (TAM, Diffusion Theory, Theory of Reasoned Action) **can be used to address the challenges**.

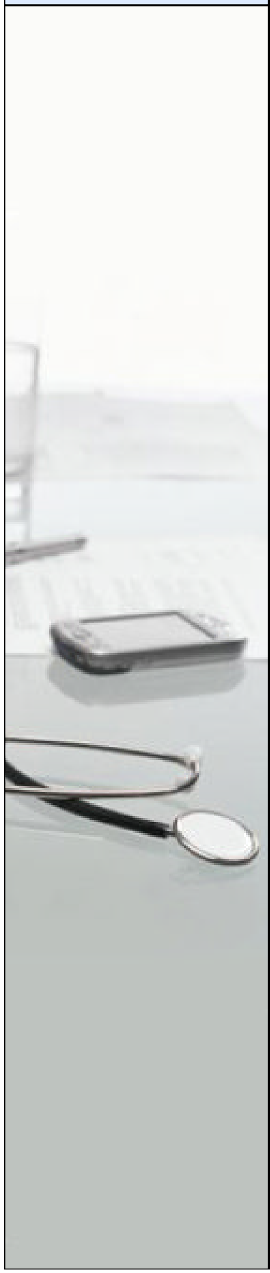
- Necessary are the…
  - ➢ **Usefulness,**
  - ➢ **Ease of Use,**
  - ➢ **Playfulness,**
  - ➢ **Cost effectiveness, etc.,**

Conclusions II

- Data analytics are used to produce services that are meant to be offered to users…

- Services need to be useful to users, to produce value and to be ease to understand…

Thank you!!!



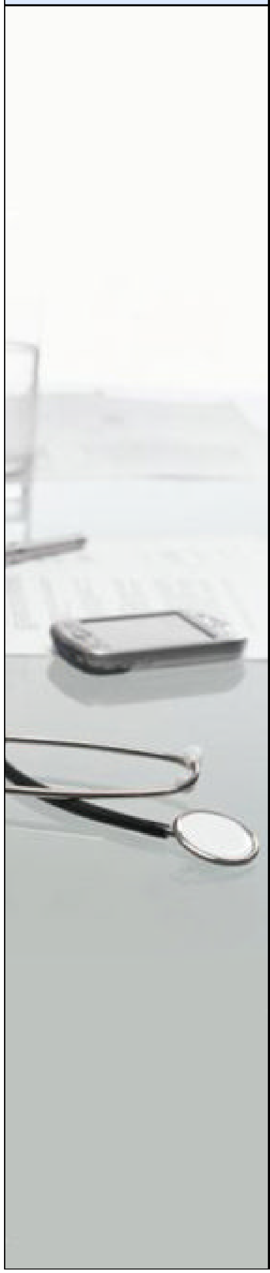
Specific challenges of BI in the health care domain

panel slides

István Vassányi
2019

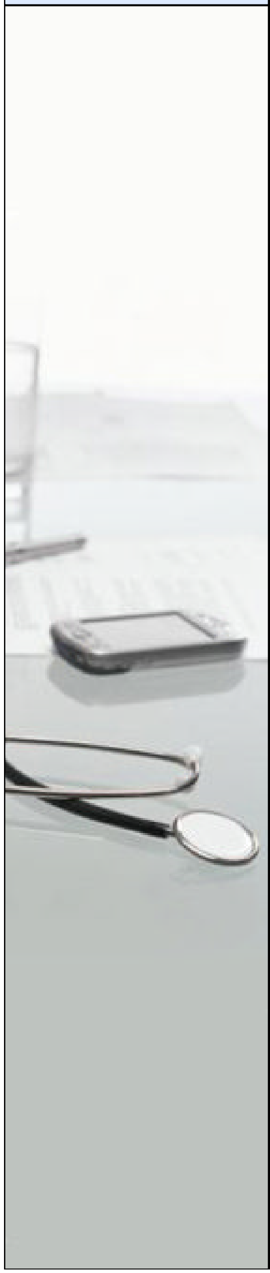
University of Pannonia
Medical Informatics Research and Development Centre

The situation



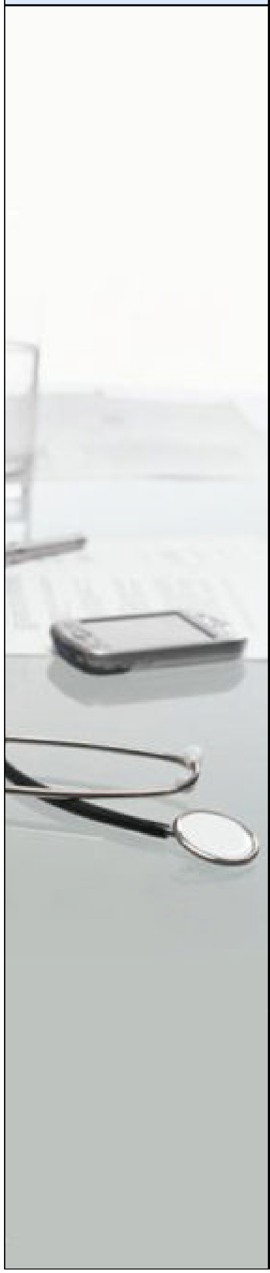
- **The public health care system is going to face severe difficulties across Europe due to**
  - **Ageing societies—ever less tax payers for more old age pensioners**
  - **The per-patient cost of the main chronic and age-related diseases (cancer, stroke, CVD) is rising due to enhanced imaging and treatment procedures**
- **The social security insurance databases hold per-case data for decades back, but this data is not analyzed**
- **Treatment protocols are available, but these are often not observed**
- **Yet classic BI methods and tools proven in other industries are not used to improve the efficiency—significant anomalies exist**

The causes



- **Compared to other industries, there is a more significant *semantic gap* between the IT providers and the medical professionals—due to the hyper-complex knowledge domain**
- **Also, this domain is changing faster than others**
  - **Special technologies like archetype based data models would be needed that the IT providers just don’t know**
  - **Current HIS solutions are not flexible and fail within years -> reluctant acceptance of IT in the medical community**
- **The HIS should support complex protocols but the users want support and not control over medical decisions -> intelligent HIS projects fail**
- **The costs are paid by social security all the same and there is weak professional control -> data analysis concluions have no consequences**

Conclusion



- The public health care domain lags behind other business domains wrt IT and BI utilization
- Causes are historical and structural
- The change will be enforced by worsening financial conditions within a decade

Data Processing In Toxicology From Animal Model To Human

Wiam Ramadan
Lebanese Institute for Biomedical Research and Application (**LIBRA**)
International University of Beirut (**BIU**) and Lebanese International University (**LIU**)

PhyToxE

Introduction

- Human health risk assessment is part of a broader risk assessment model, primarily concerned with scientific evaluation

- Toxicology research is most related to:
  - hazard identification:

    data are evaluated to determine if a material is likely to induce a hazard to human health

  - dose-response assessment

    data from studies demonstrating a hazard are translated into reference values for human exposure

Both require the evaluation of appropriate toxicological data

Hazard identification

- ❖ Which populations might be affected?

- ❖ What toxicity data are available?
  - ➢ Human Data
    - Epidemiology studies
    - Controlled human exposure studies

  - ➢ Animal Model Data (rats, mice…)

Advantages and Limitations of Using Animal Toxicity Data

Advantages	Limitations
Toxic effects expected to be similar in humans and animals	Results sometimes different from humans
Control of exposure and vehicle/negative controls	Anatomical differences between humans and animals
Less costly and timely than human exposure or epidemiology studies	Some serious chronic effects missed by standard toxicity studies
Enable proactive regulation and behavior	Variability in results

Dose-Response Assessment

Identify NOAEL or LOAEL or an LED10

NOAEL	LOAEL	LED10
No-Observed-Adverse-Effect Level. Highest dose at which no significant adverse effects are observed.	Lowest-Observed-Adverse-Effect Level. Lowest dose at which significant effects are observed.	Dose that produces an adverse effect in 10% of exposed, relative to control.

Conduct dose-response modeling

Identify critical effect(s) and level(s)

Extrapolating from animal study to humans

- When dose-response data from an animal study are used to assess risk in humans, methods are available for extrapolating between the doses used in the animal study and equivalent doses in humans:

  - Dosimetric conversions

  - Pharmacokinetic modeling

Exposure in toxicology studies

Exposure in toxicology studies depends on the:

- Route
- Magnitude
- Duration
- Frequency
- Timing

It's important that these parameters are carefully controlled in a study

Route of Exposure

- The route of chemical administration can affect toxicity because there are differences in
  - how much of an agent is absorbed into the body by different routes
  - how an agent is transported to major metabolic sites (e.g. the liver).
- For risk assessment purposes, the exposure route used in an animal study should mimic a potential route of human exposure:
  - Oral: food, drinking water, gavage, or capsule
  - Inhalation: via chamber exposure (whole body, head-only, or nose-only)
  - Dermal: applied to the skin of the animal

Parameters Examined

- In Vivo:
  - Respiratory function
  - Cardiac function
  - Sleep parameters
  - …
- In Vitro
  - Organ weights
  - Organ and tissue histopathology
  - Hematology
  - Clinical chemistry
  - Urinalysis
  - …

Summary

- Animal Toxicity Studies
  - Can provide data relevant to human exposure scenarios

- Elements Important for Study Use
  - Study design and reporting
  - Test substance characterization
  - Dose selection
  - Exposure characterization
  - Applicability to humans