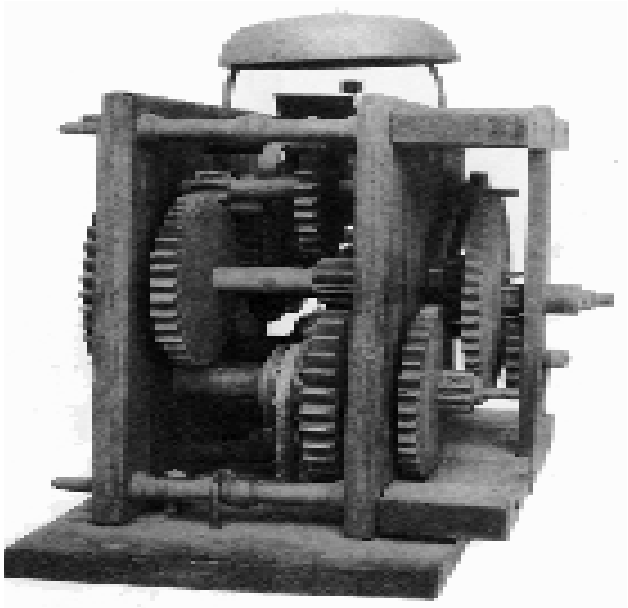
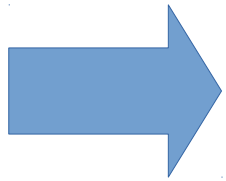
Aalto University
School of Science

Probabilistic Modelling of Multiresolution Biological Data

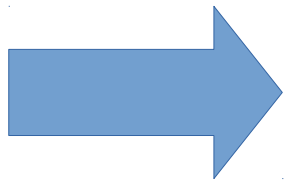
Prem Raj Adhikari

Lectio Praecursoria
21.11.2014






SONY
SONY
SONY
12
9
23
6
Emma Brown
Amanda Jones
Jacob Moore



If someone from the 1950s suddenly appeared today, what would be the most difficult thing to explain to them about life today? (self.AskReddit)

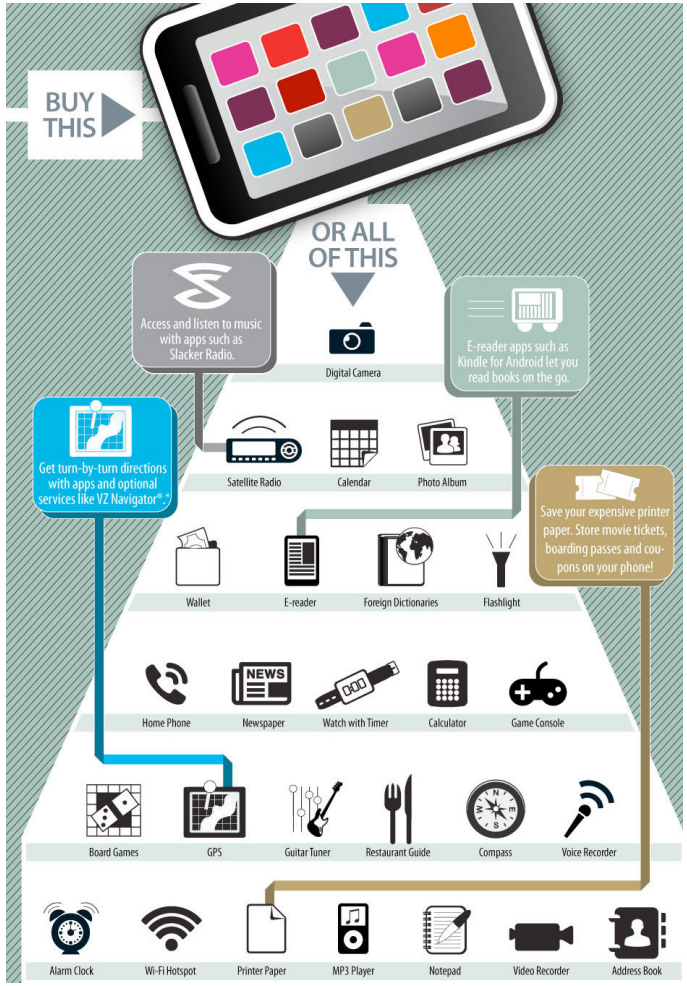
submitted 1 year ago * by [deleted]

[–] **nuseramed** 3833 points 1 year ago

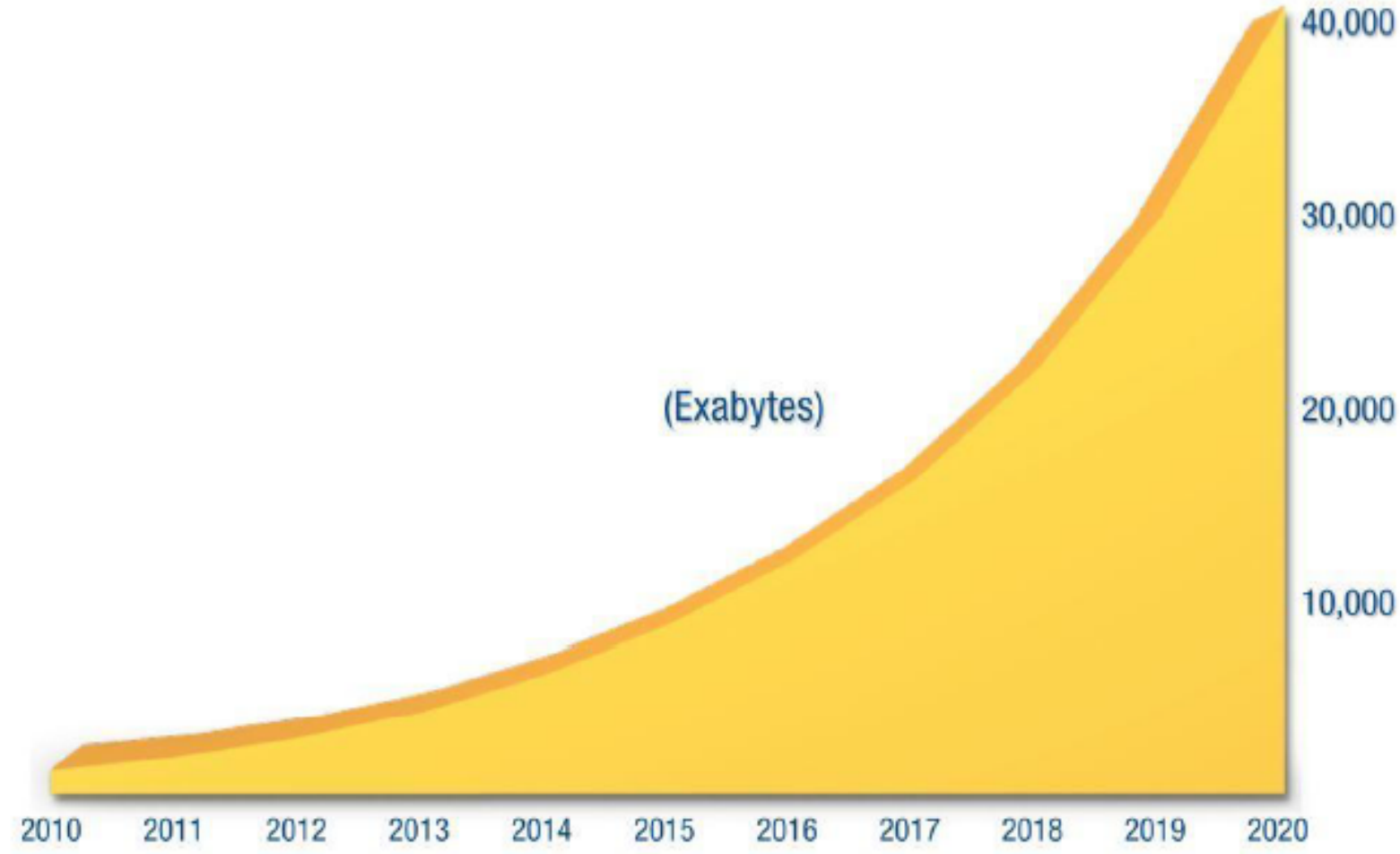
I possess a device, in my pocket, that is capable of accessing the entirety of information known to man.

I use it to look at pictures of cats and get in arguments with strangers.

permalink

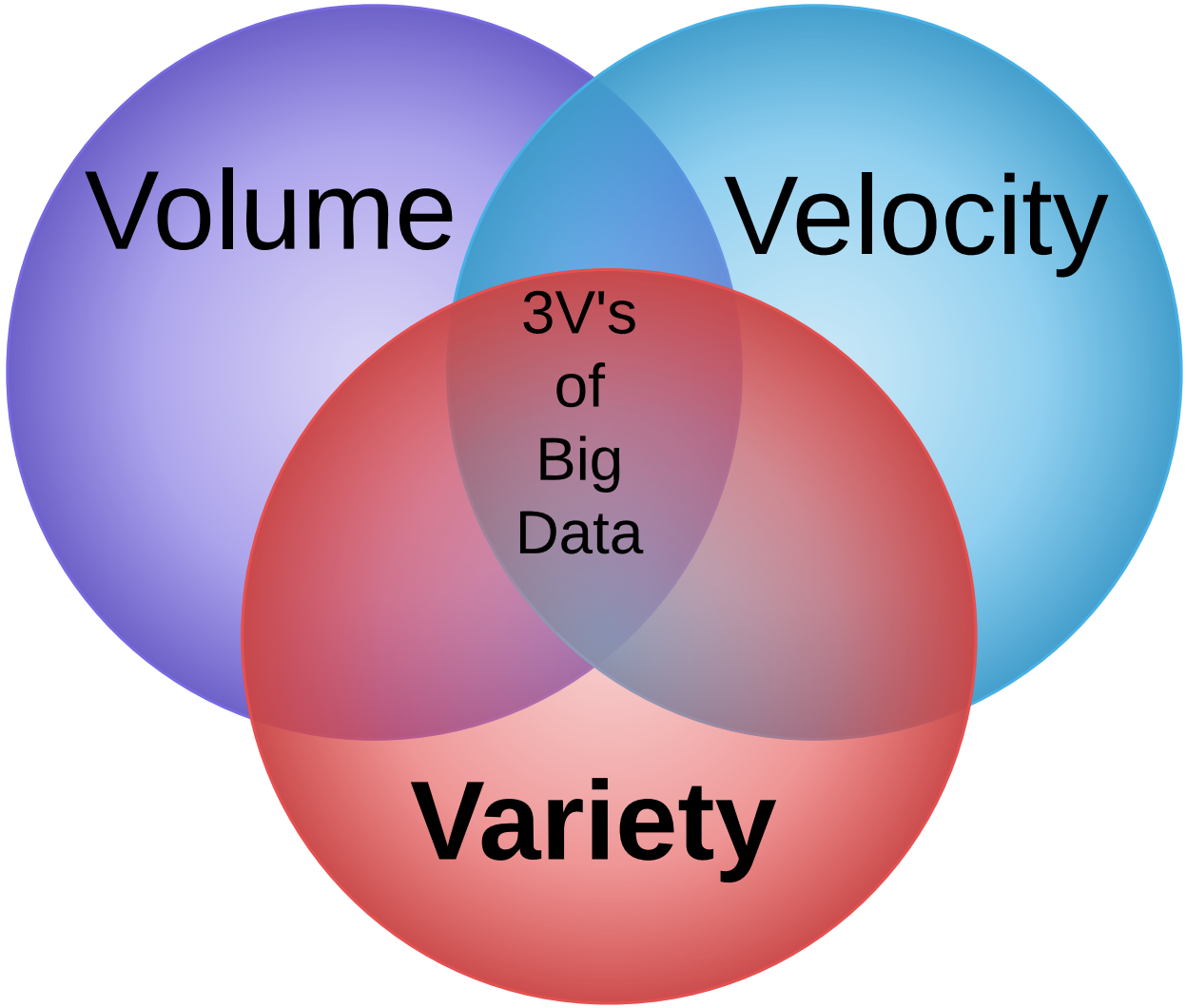


50-Fold Growth from the Beginning of 2010 to the end of 2020

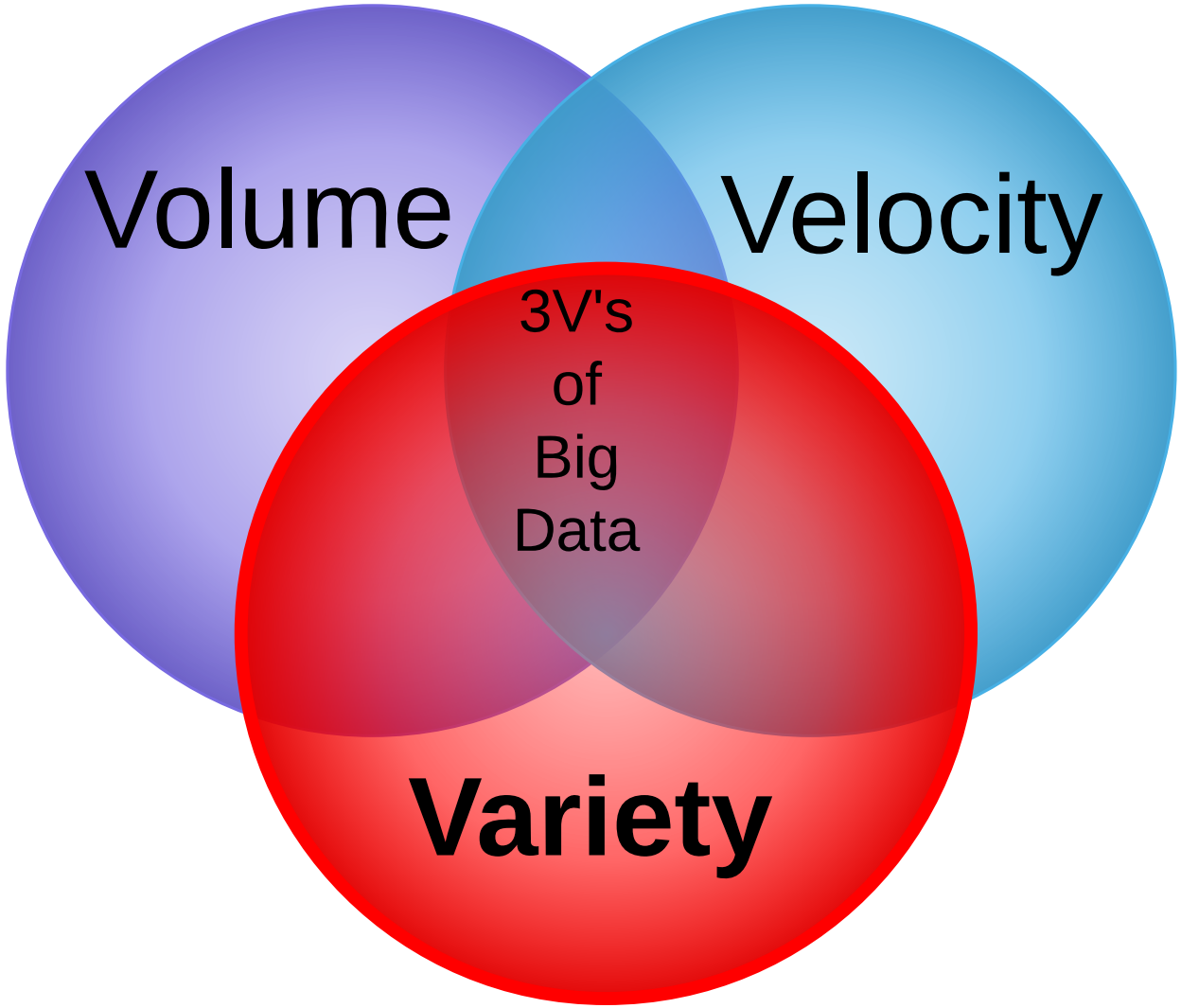


Source: IDC's Digital Universe Study, sponsored by EMC, December 2012

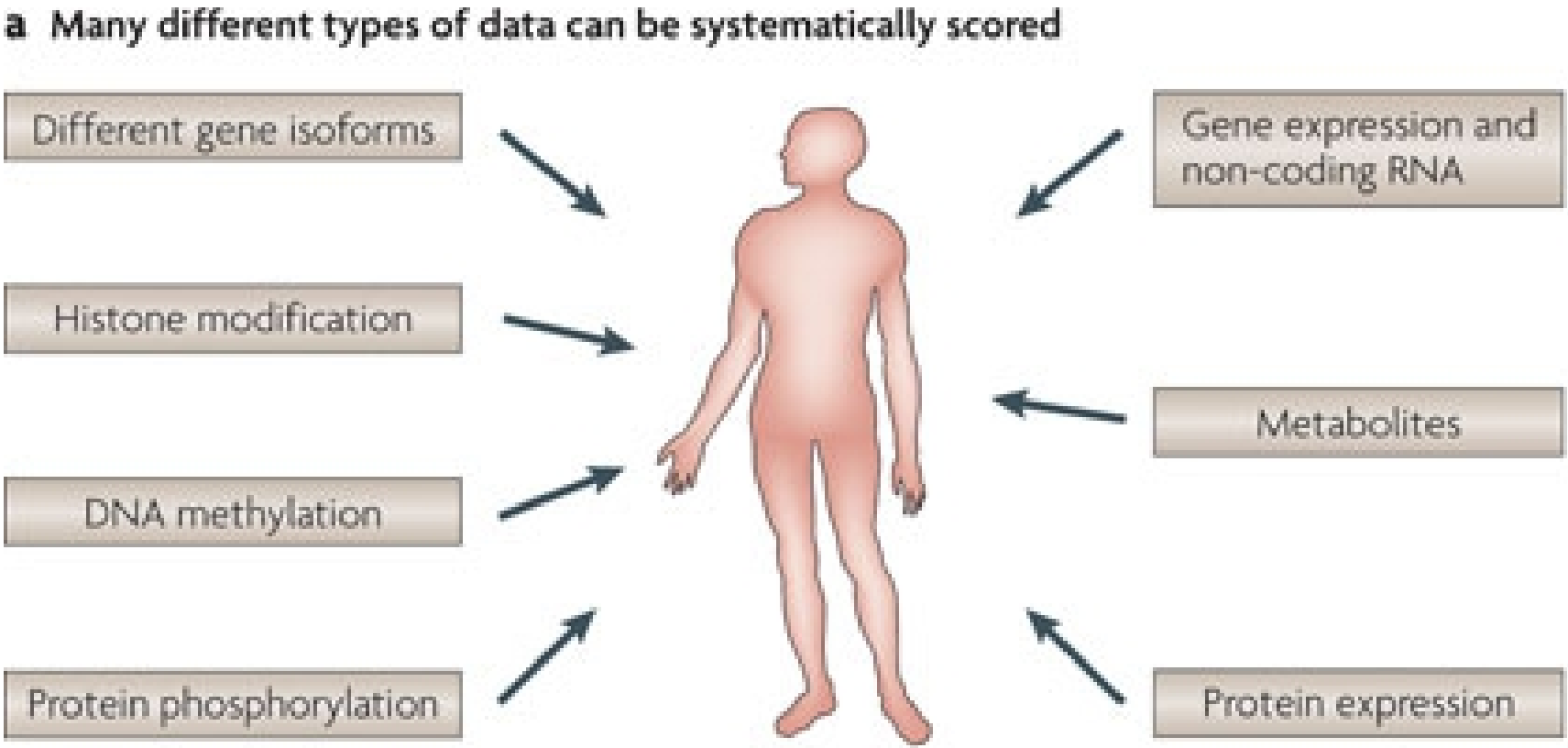
Big Data: Big Challenges



Big Data: Big Challenges



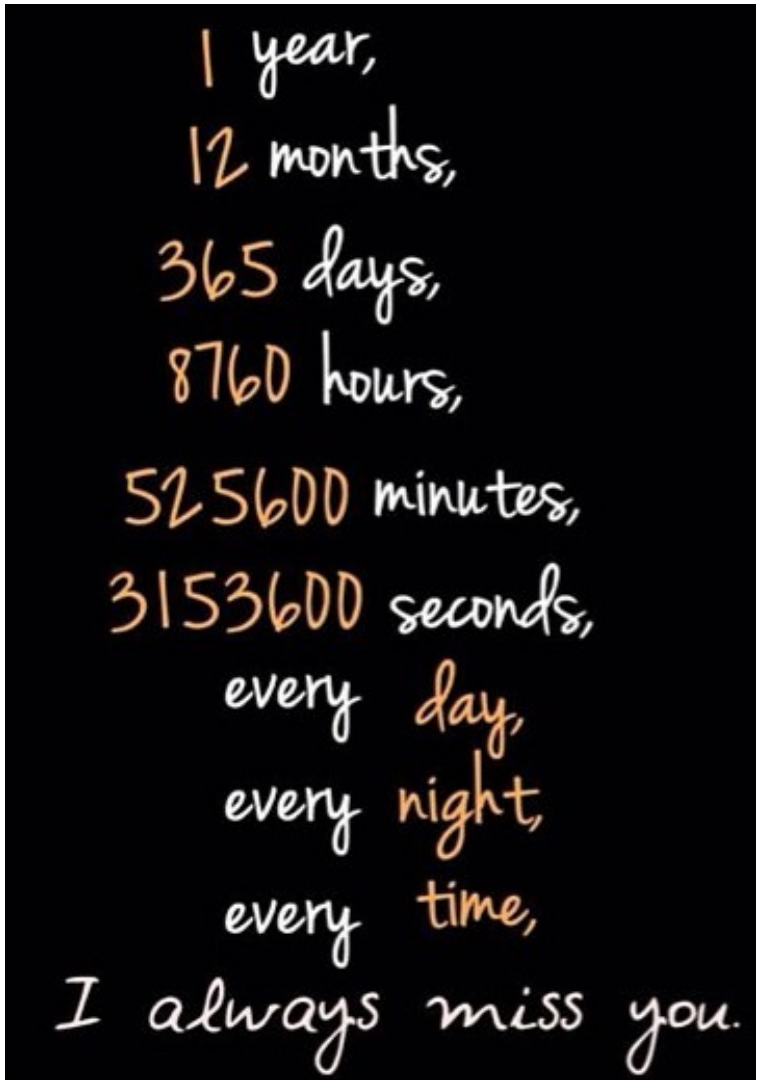
Big Data: Big Challenges



Adapted from E. E. Schadt, et. al., Nature Reviews Genetics, 2010

Multiresolution Data

Multiresolution data results when a phenomena is measured in different levels of detail.



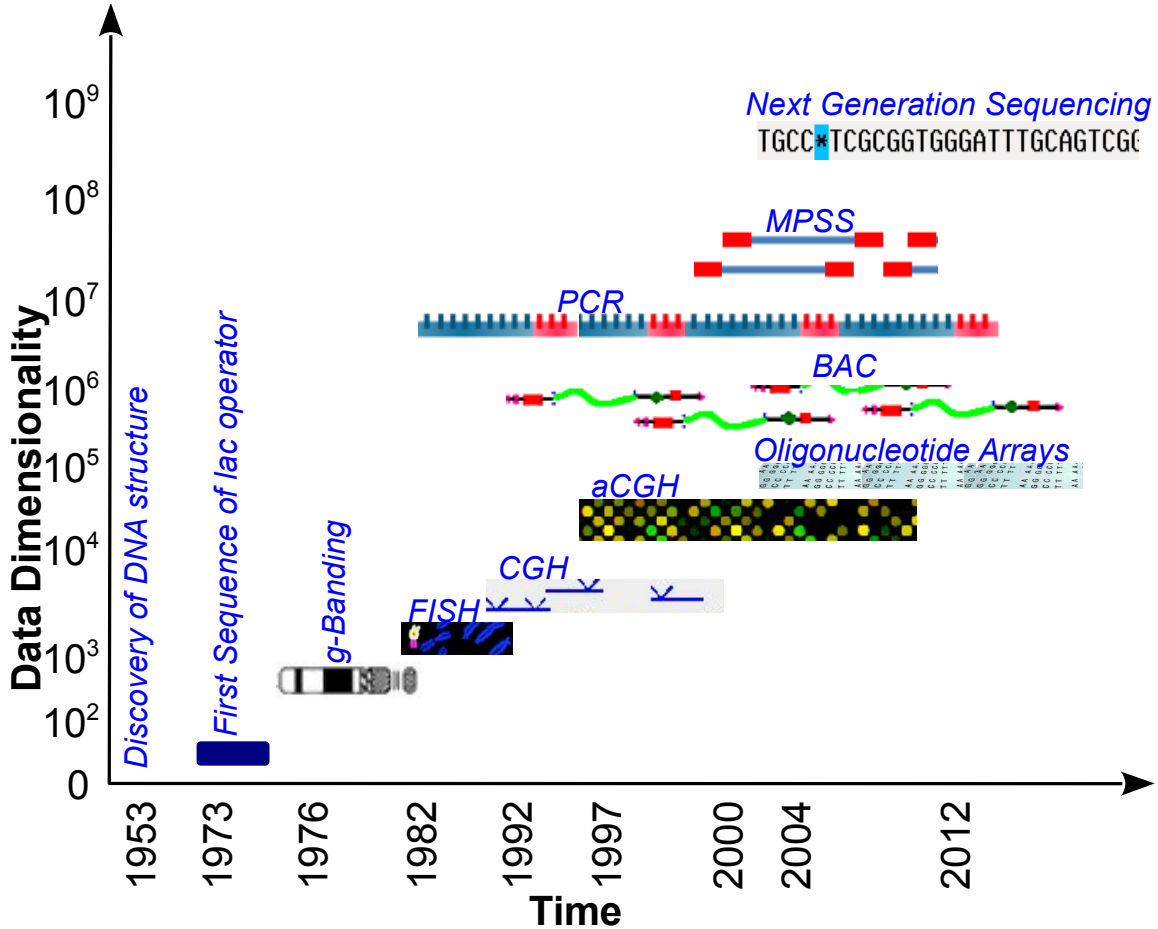
Chromosomal Aberrations Patterns in Cancer

- Abnormality in the normal chromosomal content of a cell
- Different cases of DNA copy number aberrations
  - Deletion: When the copy number $<$ **2**
  - Duplication: When the copy number $>$ **2**
  - Amplification: When the copy number $\gg$ **5**
- Why detect copy number aberrations?

Chromosomal Aberrations Patterns in Cancer

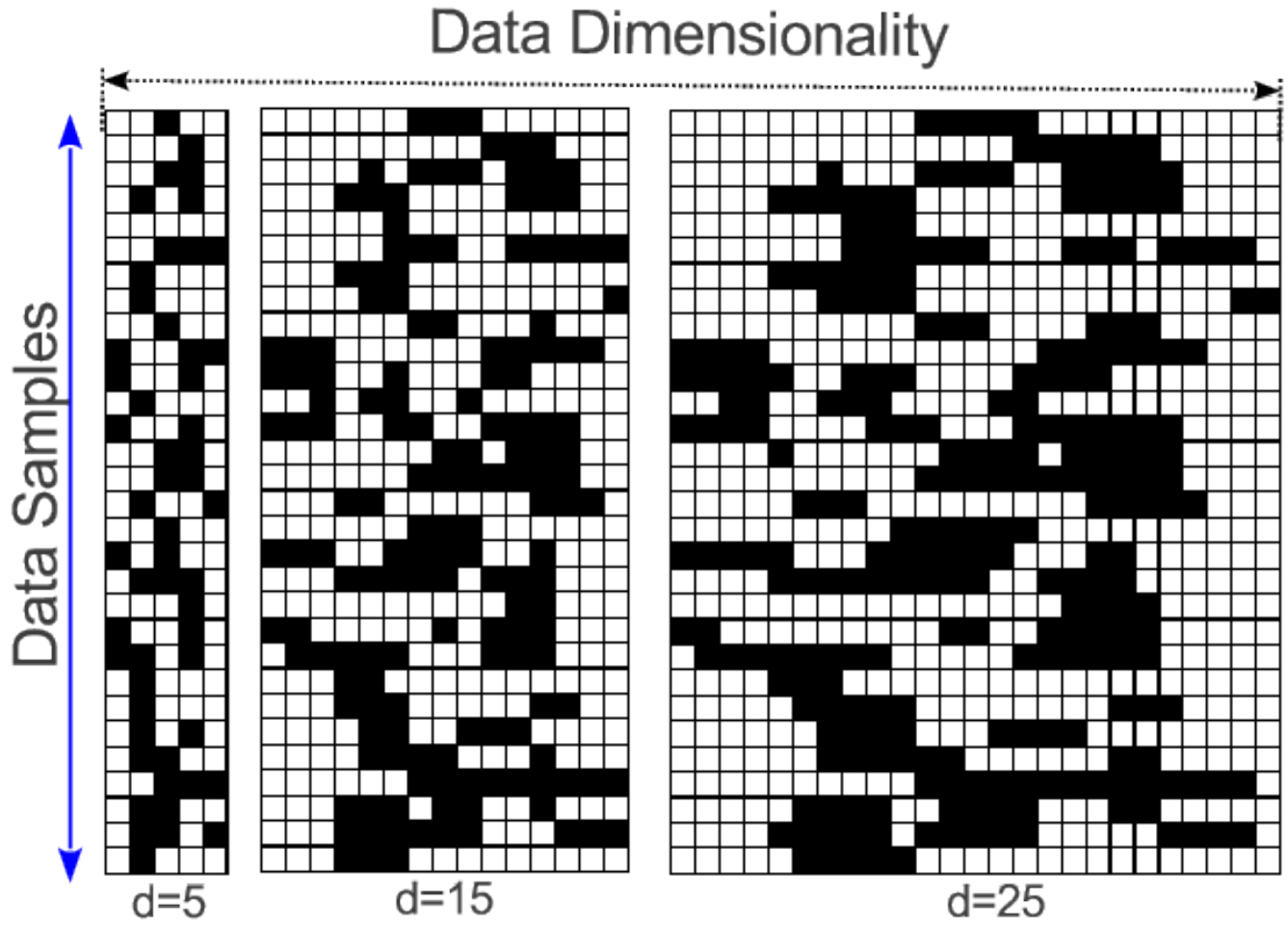
- Abnormality in the normal chromosomal content of a cell
- Different cases of DNA copy number aberrations
  - Deletion: When the copy number $<$ **2**
  - Duplication: When the copy number $>$ **2**
  - Amplification: When the copy number $\gg$ **5**
- Why detect copy number aberrations?
- **DNA copy number aberrations are hallmarks of cancer**

The Multiresolution Data in Biology

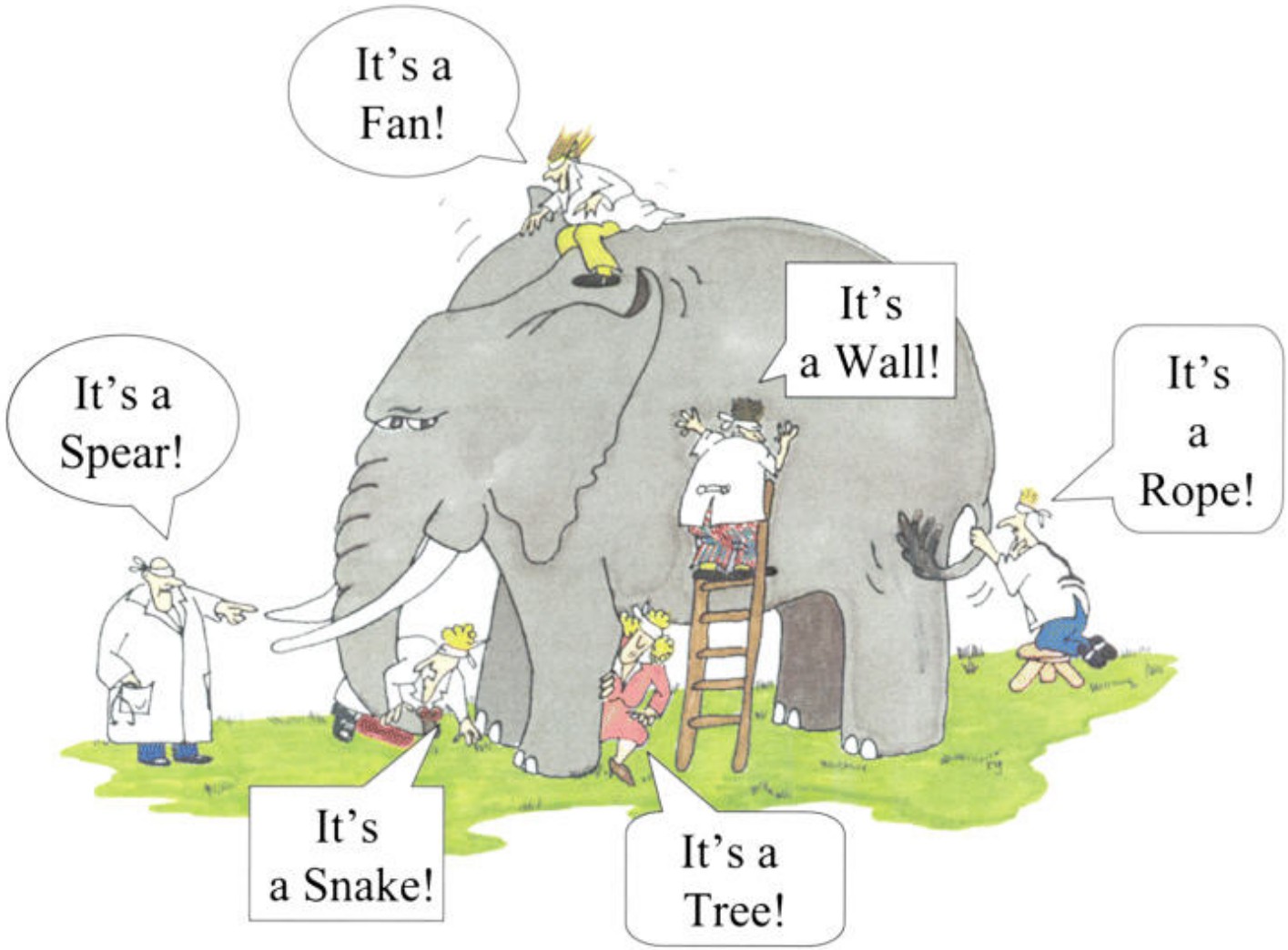


- Multiresolution data is everywhere: biology, computer vision, telecoms ...
- Older Generation Technology $\Rightarrow$ Data in Coarse Resolution
- Newer Generation Technology $\Rightarrow$ Data in Fine Resolution

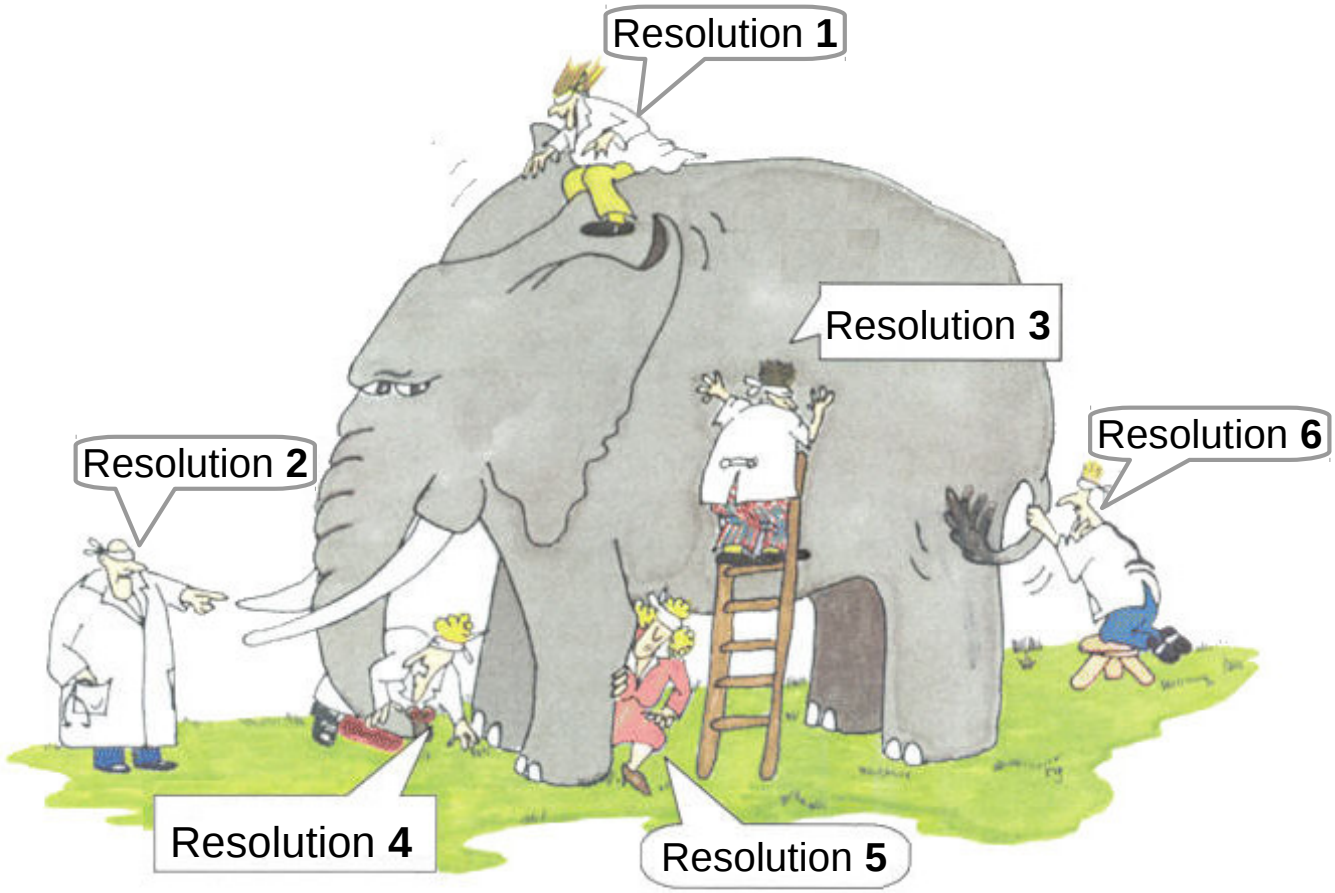
DNA Copy Number Amplification Dataset



How to analyze data in multiple resolutions in a single analysis?


It's a Fan!
It's a Spear!
It's a Wall!
It's a Rope!
It's a Snake!
It's a Tree!

Multiresolution Data in Biology



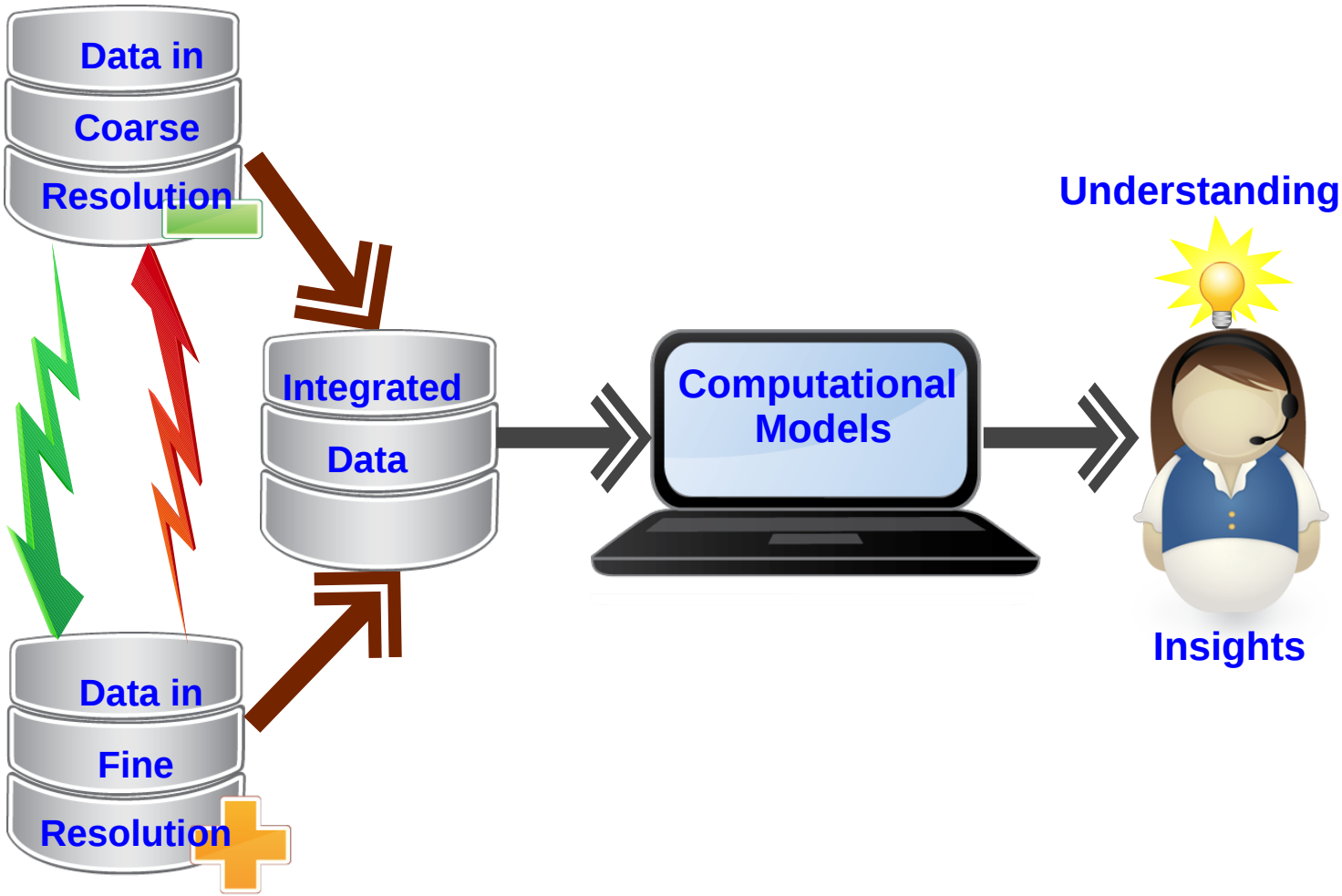
Adapted from Y. Moreau, University of Leuven, Belgium

Mixture Modelling of Multiresolution 0–1 Data

> *Medicine is a science of uncertainty and an art of probability.*
>
> — Sir William Osler
>
> *Father of modern medicine, 1849–1919*

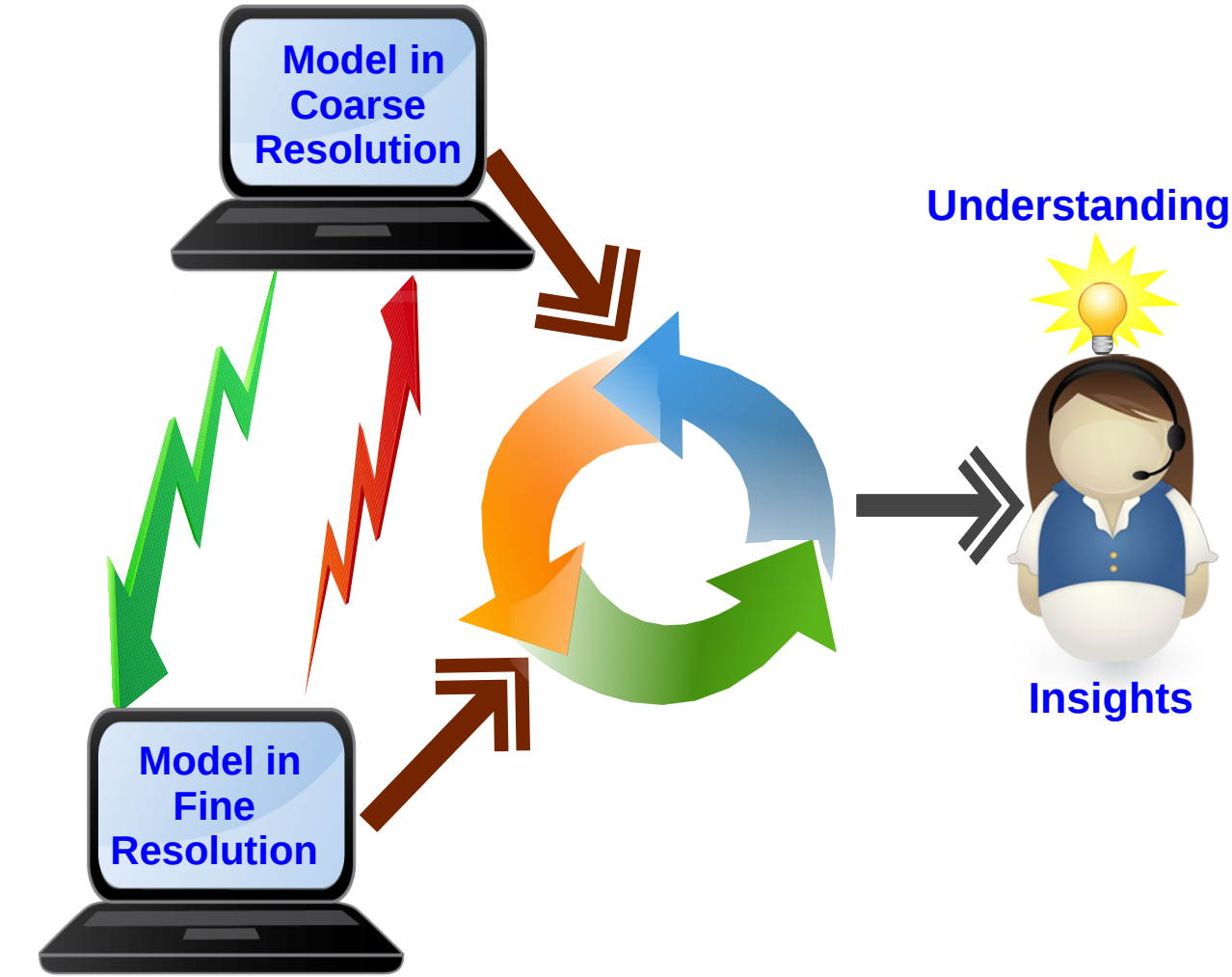
- Cancer is a heterogeneous collection of several diseases and mixture models are well known for their ability to model heterogeneity
- Mixture models generally cannot model multiresolution data
- Only mixture modelling solution to multiresolution data is to model each resolution separately

Contribution

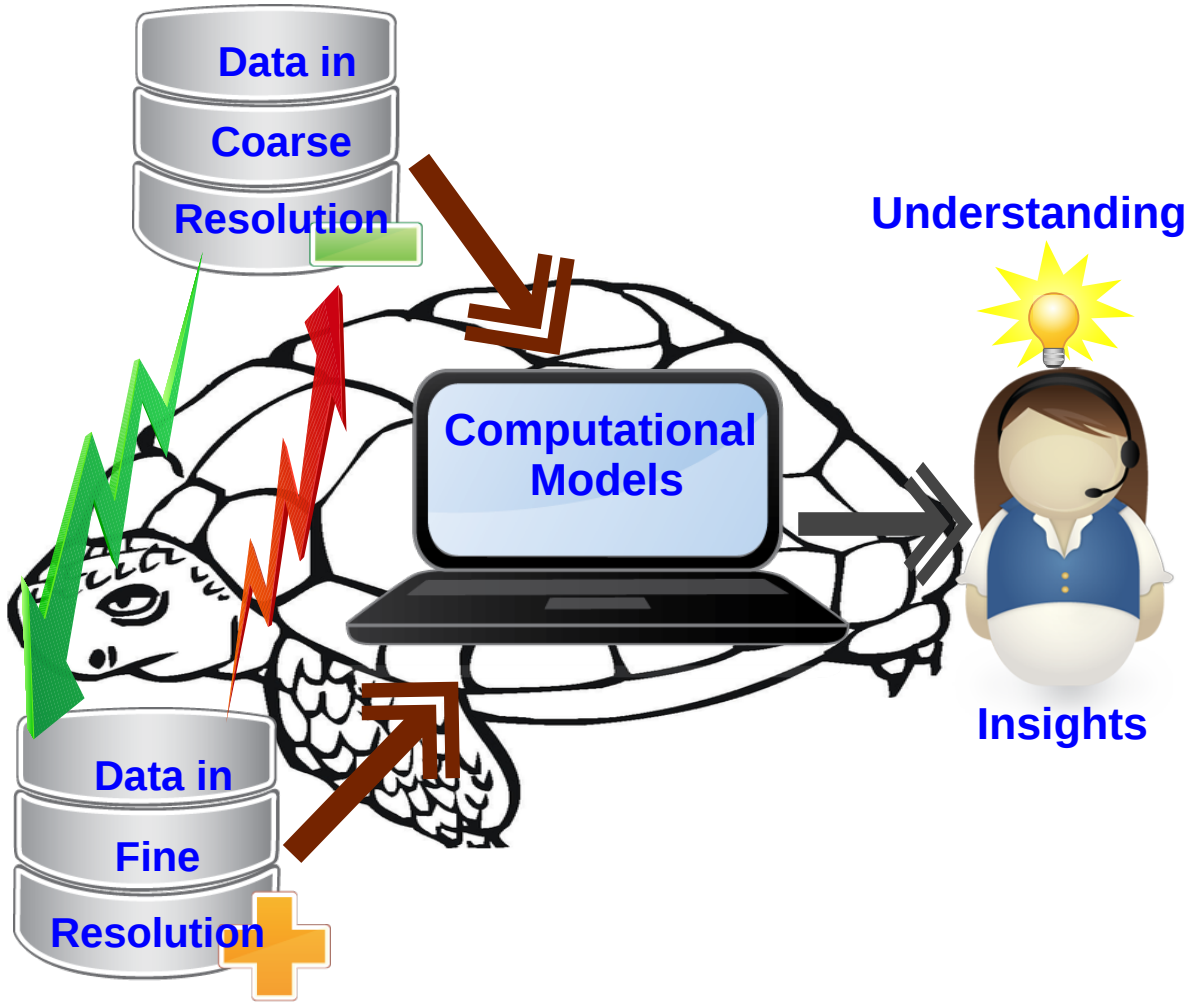


Publication I

Contribution

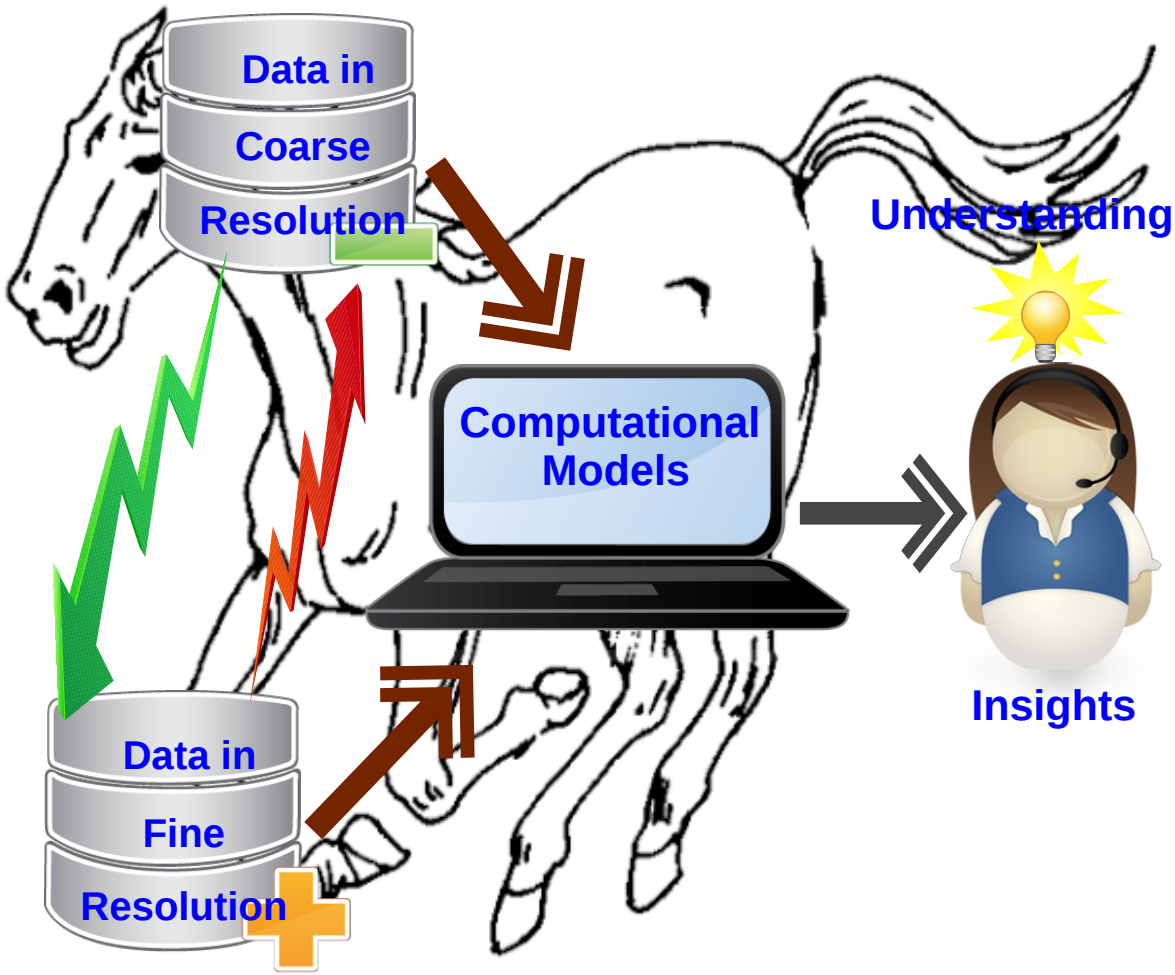


Contribution



Publication II

Contribution



Publication II

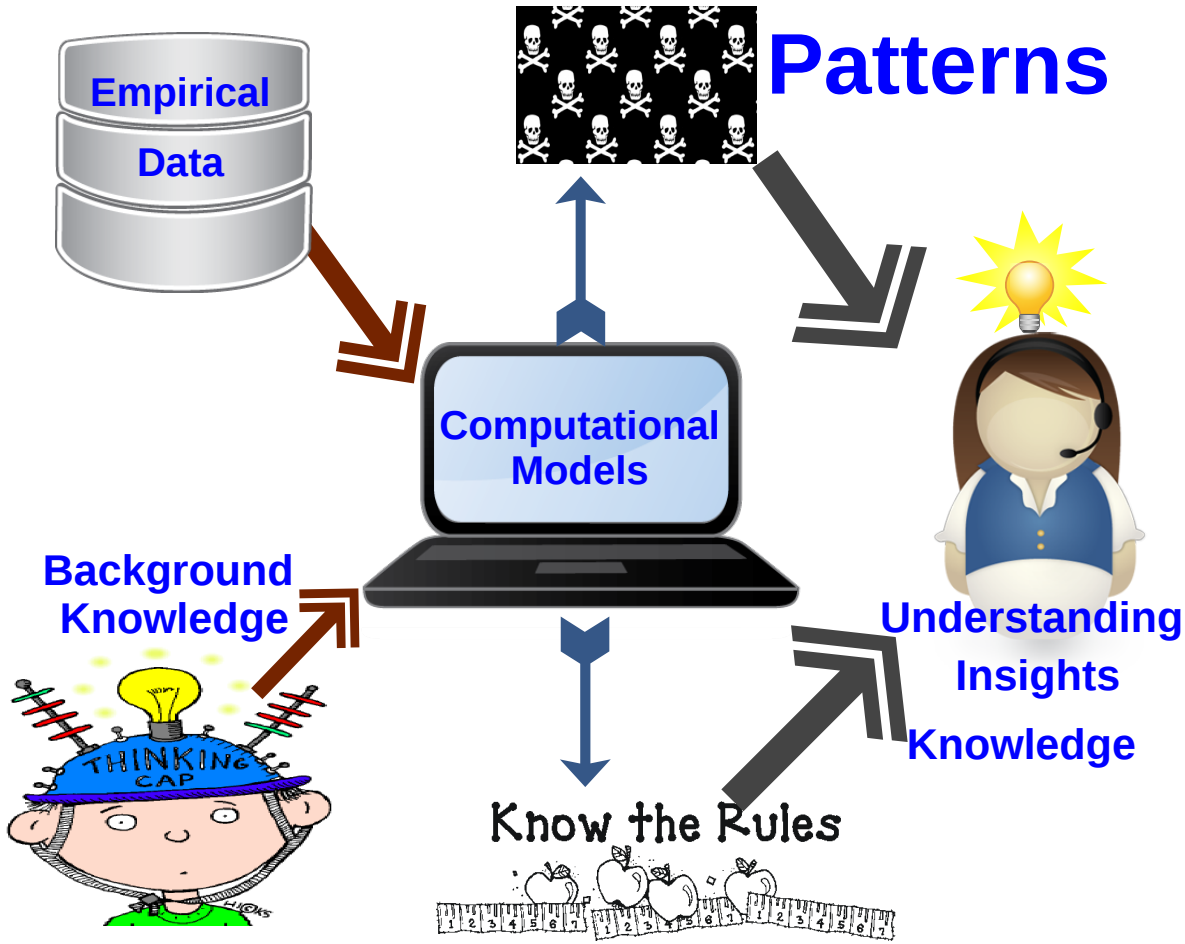
Semantic Data Mining

> *Data do not speak for themselves – they need context, and they need sceptical evaluation*
>
> — ALLEN WILCOX
> *Harvard Professor*

- Abundance of ontologies and semantically annotated data
- Biological systems are complex: interwoven subsystems
- Plenty of Semistructured, heterogeneous and distributed data

> *The purpose of computing is insight, not numbers.*
>
> — RICHARD HAMMING
> *1962*

Contribution



Publication IV & V

Summary and Conclusions

- Growth of data
- Variety in the data growth
- Copy number aberrations
- Analysis and Modelling of Multiresolution 0–1 Data
- Fast and efficient training of series of mixture models