

Big Data & Health

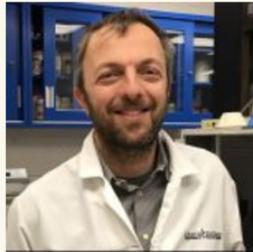
Enrico M. Bucci

Bologna, 22-04-2017



Why I am entitled to talk

Cancer Systems Biology & eHealth Programs



Enrico Bucci, PhD

Professor in Systems Biology

[Publications](#)

Cancer Systems Biology Program

This program relies on the use of (a) collective technologies used to explore the roles, relationships and actions of the various types of molecules that make up the cells of an organism and (b) informatics approaches with the goal to rationally construct an integrated picture to better understand the dynamic behavior and regulation of integrated intra- and intercellular biomolecular networks deregulated in cancer. Special interest is given to new computational approaches, able to get insight on how measurements of global network



[Follow Sbarro Health Research Organization](#)

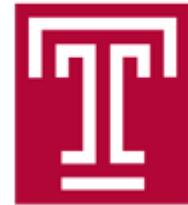
FACEBOOK SHRO



You and 30 other friends like this

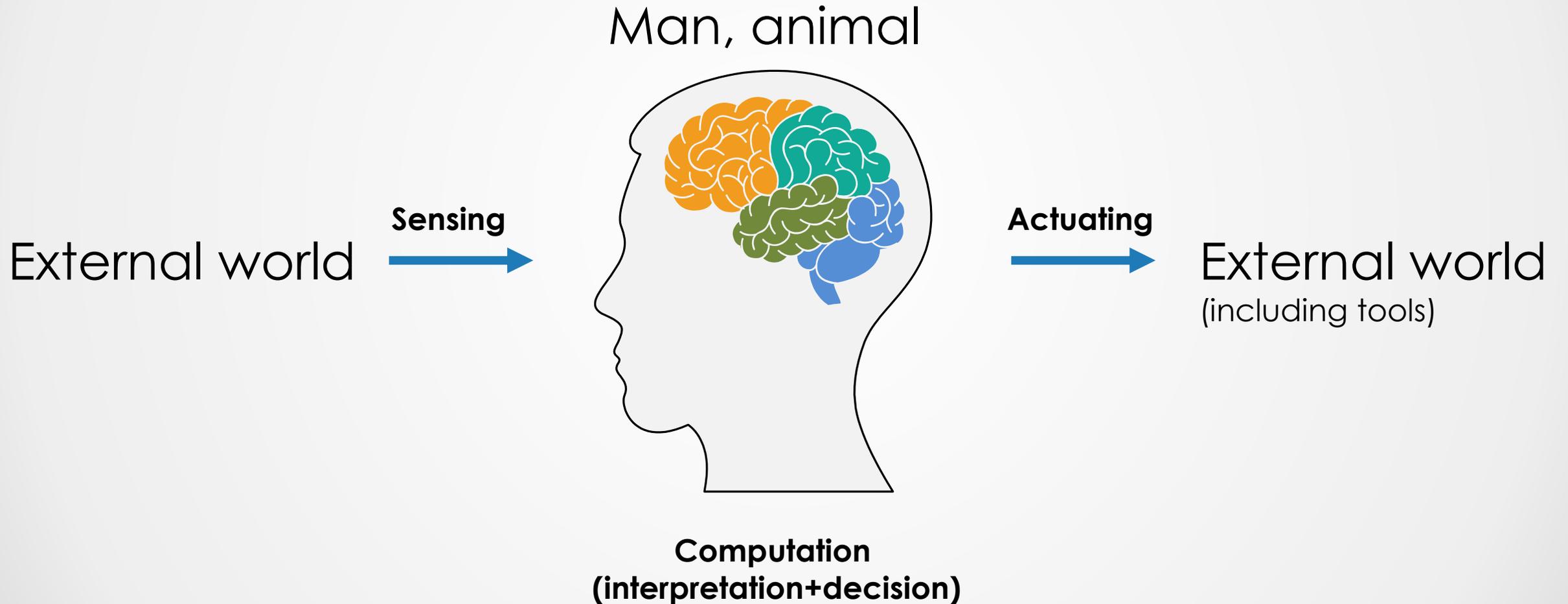


<https://shrodotorg.wordpress.com/blog/cancerbio/>

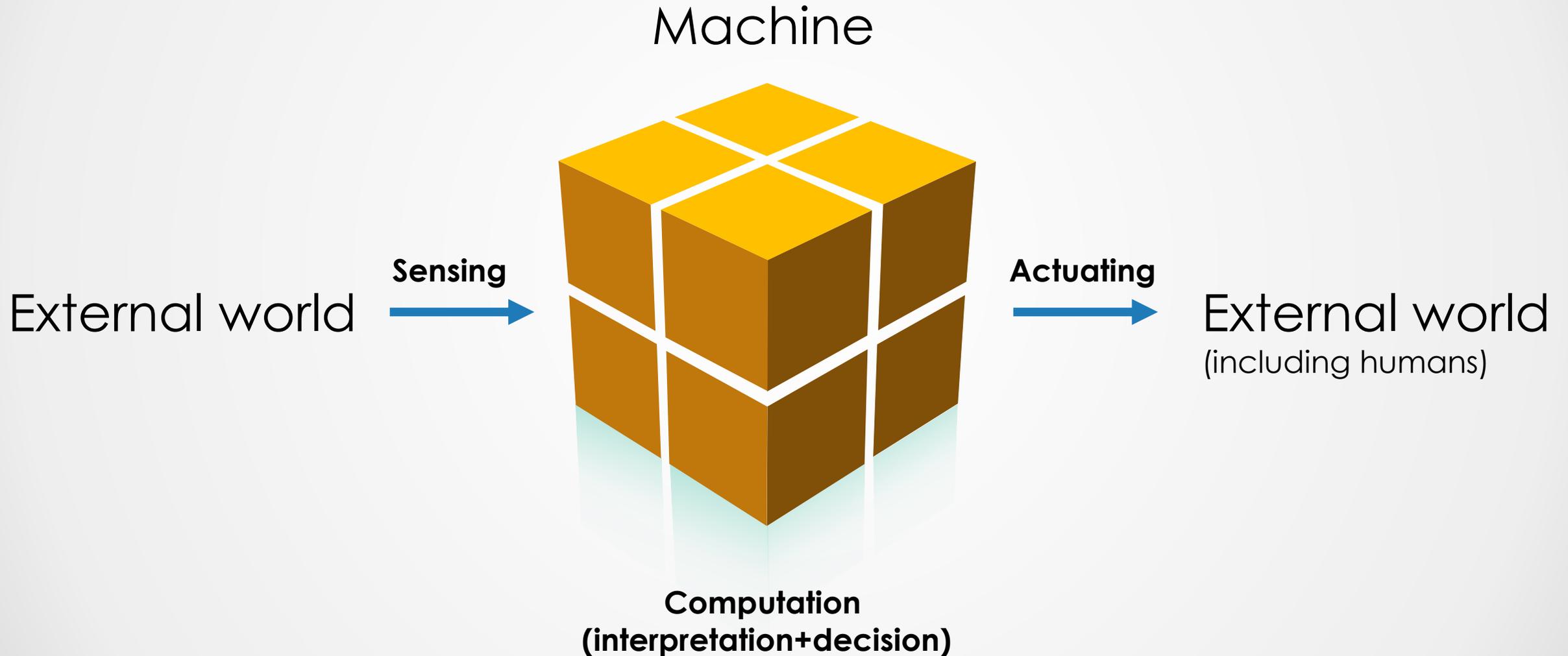


TEMPLE
UNIVERSITY®

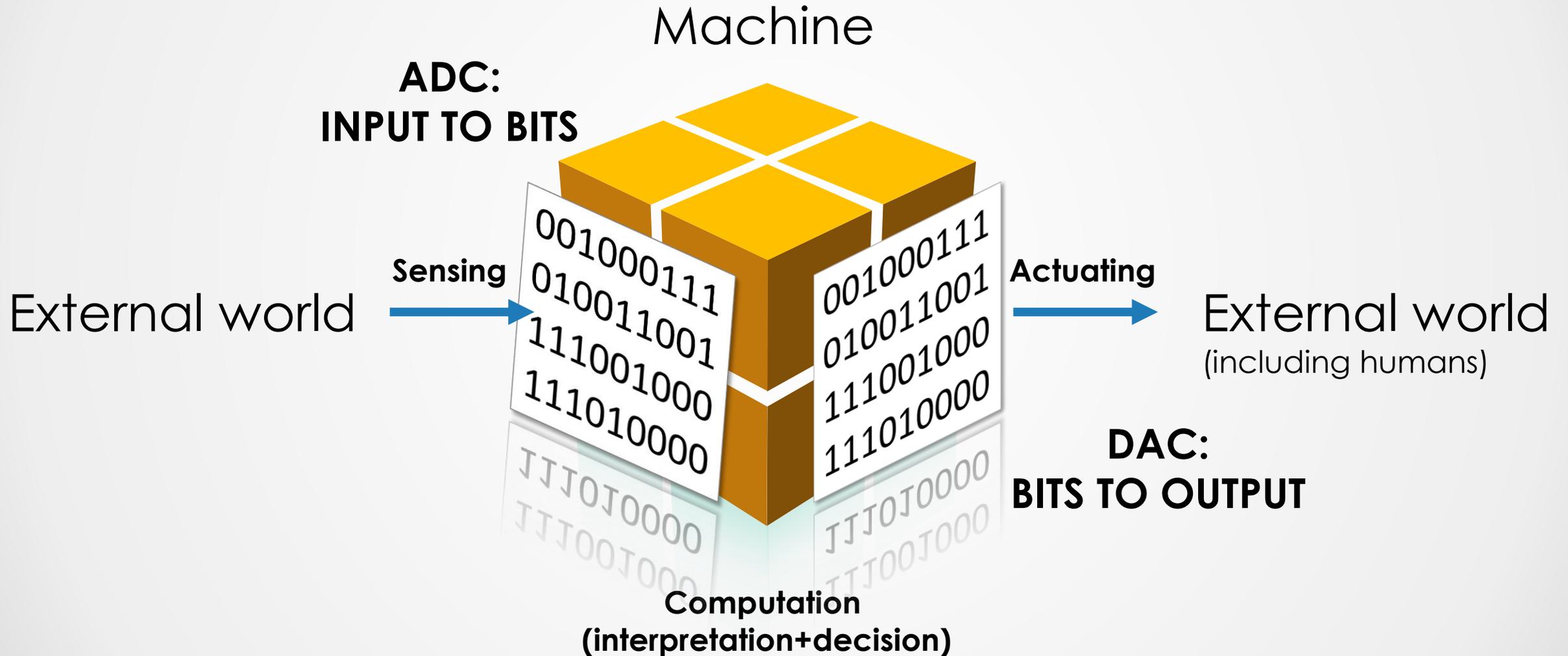
Before us: analog World



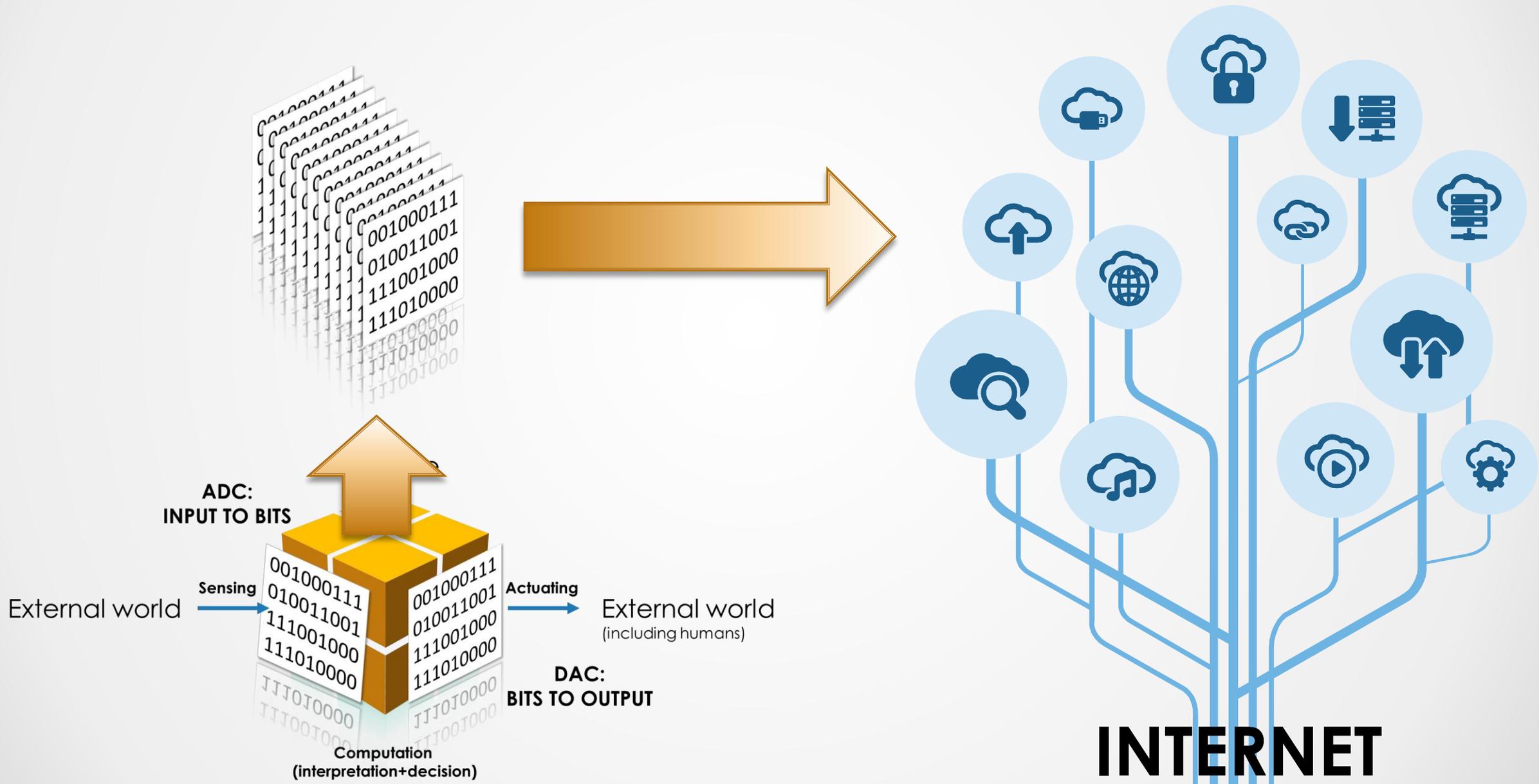
Slightly before us: a digital World



Slightly before us: a digital World



Our time: a Big Data World

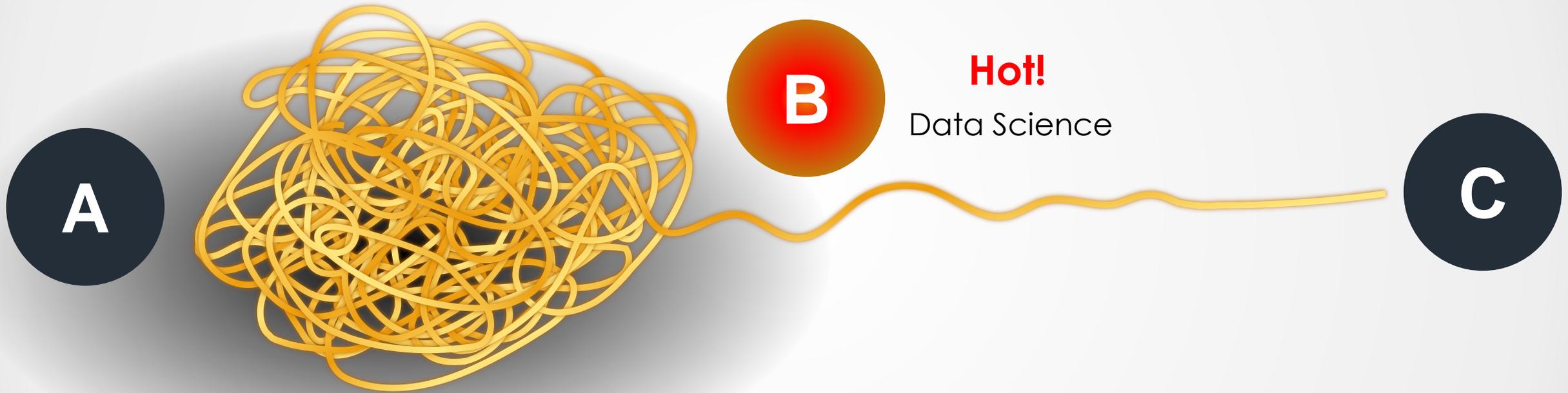


Our time: a Big Data World



We all shop, bank, chat, post, meet online, use phone, travel, read, feed, get diagnosed or cured ... every day we all create a digital sea of data which encode our entire daily life.

Data Science: from Big Noise to information



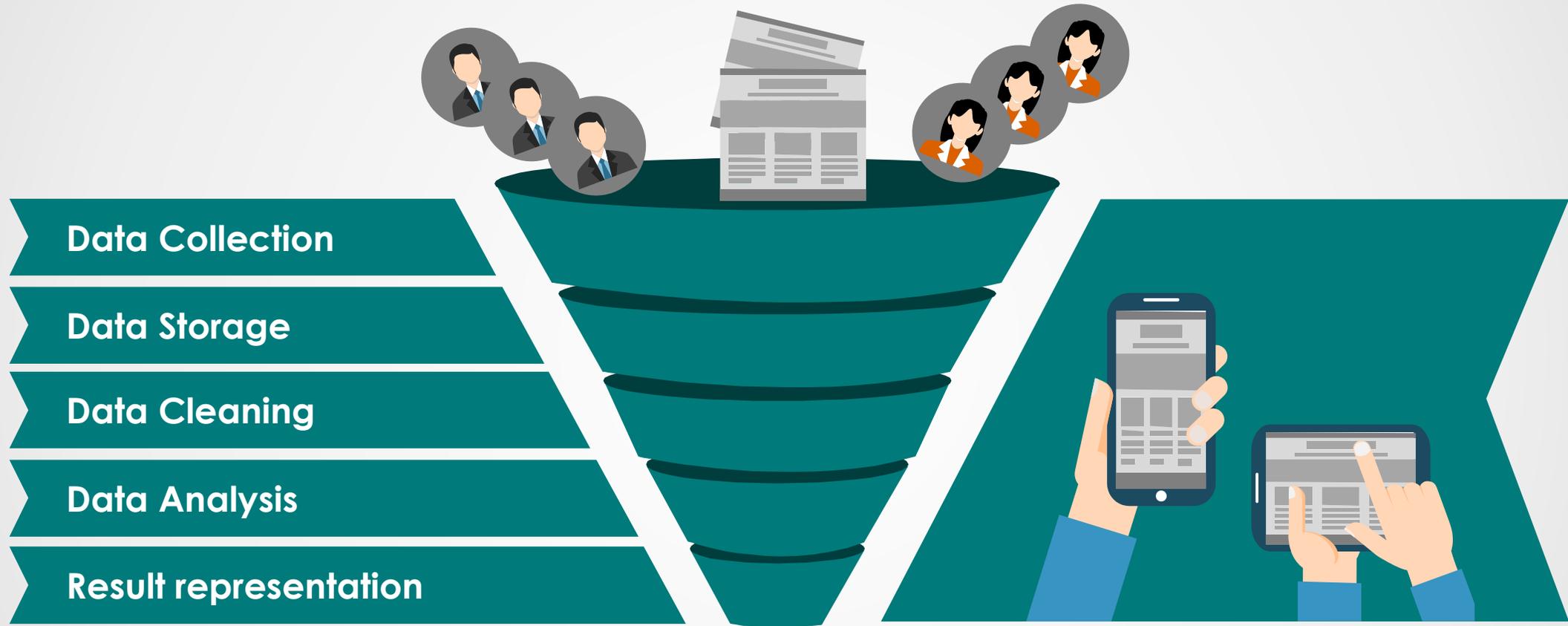
Unstructured, noisy data

Data flows are noisy, rich and unpredictable exactly like the world generating them

Information

Useful information is based on the interpretation of structured and well-ordered data

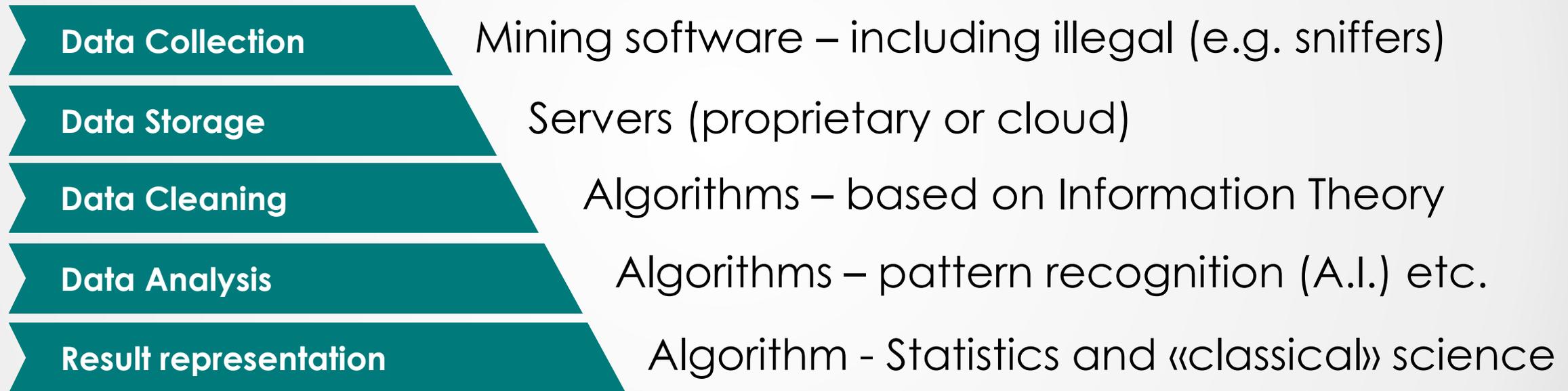
Data Science: from Big Noise to information



Information

Useful information is based on the interpretation of structured and well-ordered data

Data Science: from Big Noise to information



The Big Divide (individual vs tech-enabled entities)

Data Storm

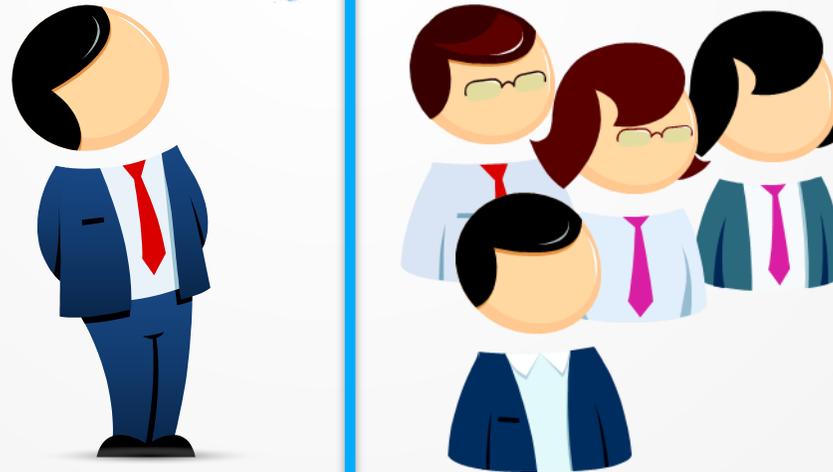
Individual

One way to reduce the information overload:

Picking the «liked» content (relying on cognitive bias)



Unreliable information



Organized entity

Access to more data

Science-empowered

Tech-empowered



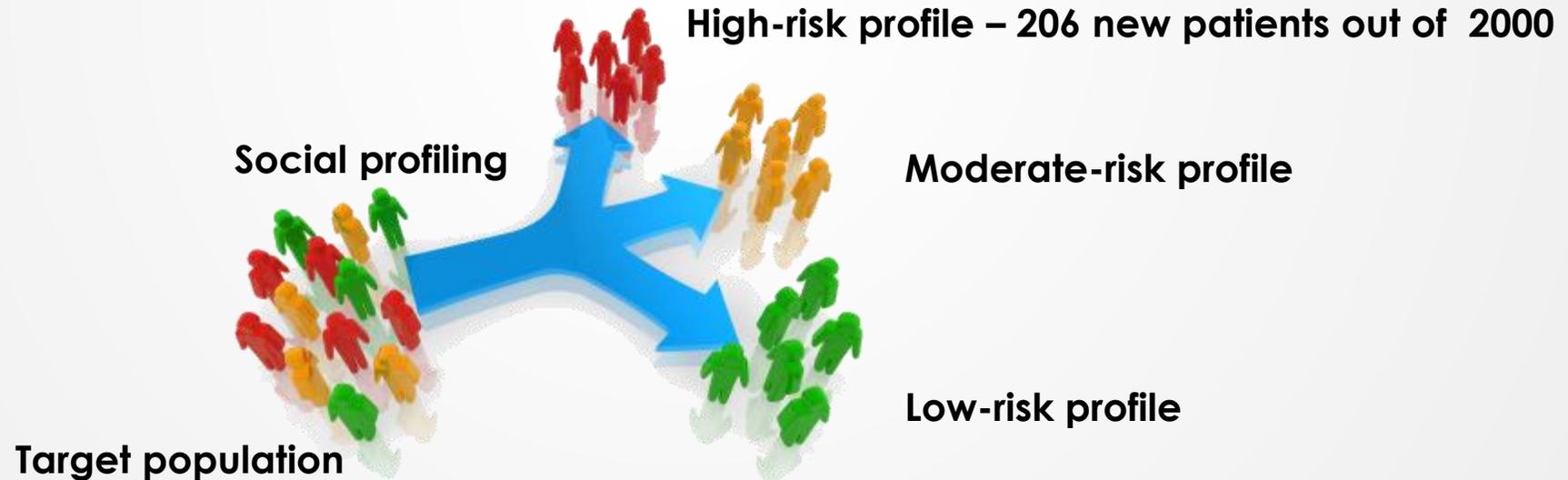
Useful information

The Big Divide: public organizations

People in UK were profiled according to several demographic, social and lifestyle habits. The corresponding data formed a commercially available database.

Key idea: to match this huge consumer database (more than 12.000.000 individuals) versus hospitalization census, so to have a demo-social description of people hospitalized for type II diabetes.

This social descriptor, validated in the clinical population, was subsequently used for targeting “at risk” people in Slough (UK) by a marketing campaign, aimed to encourage screening for diabetes. At least 206 cases of unreported type II diabetics were identified.



The Big Divide: private organizations

In 2012, BMS wanted to monitor social forums focused on colorectal cancer (CRC) to answer the following questions:

1. What is the impact of standard chemotherapy on the emotional well-being of patients and on their daily life (i.e. what BMS can do to improve the patient experience of chemotherapy).
2. What are the most prevalent reported side effects (i.e. what the patients perceive as relevant side effects)

They started monitoring 2 highly populated social forums. The CRC web forums were identified based on four criteria: active for \geq five years, >12,000 total posts, >20 individuals currently browsing, and \geq 10 new posts/day.



The Big Divide: private organizations

Physical and emotional well-being impacts	<i>N</i>	<i>N</i> = 330 (%)
Activities of daily life (eating, dressing, walking, etc.)	41	12.4
Activity disruption due to chemotherapy administration	45	13.6
Activity disruption due to treatment schedule	19	5.8
Work	41	12.4
Interpersonal relationships	23	7.0
Physical activities	18	5.5
Social activities	17	5.2
Sexual functioning	1	0.3
<i>Emotional impacts</i>		
Hope/appreciation	42	12.7
Anxiety	29	8.8
Depression	22	6.7
Emotional	15	4.5
Concern with physical appearance	7	2.1
Anxiety from carcinoembryonic antigen (CEA) testing	6	1.8
Guilt	2	0.6
Trying to act normal	1	0.3
Embarrassment	1	0.3

Side effect codes	<i>N</i>	<i>N</i> = 588 (%)
GI problems (diarrhoea, stomach pain, nausea, etc.)	123	20.9%
Skin problems (rash, itchiness, acne, etc.)	85	14.5%
Neuropathy	66	11.2%
Mouth problems (jaw pain, sores, dental problems)	57	9.7%
Bodily pain/headache	48	8.2%
Fatigue/tiredness	35	6.0%
Temperature-related issues (sensitivity, etc.)	34	5.8%
Chemo brain and dizziness	29	4.9%
Side effects unspecified	25	4.3%
Blood issues (bleeding, bruising, blood clot, etc.)	13	2.2%

This are very relevant information to establish what do people expect from a treatment for CRC. This information **could** be used to create advertisements luring the patients and the doctors into asking for the company treatment.

The Big Divide: individual bias to reduce complexity

Autism Andy



Andy is convinced vaccines cause autism. No amount of peer review will dissuade him from that one article he read in 1998. To him Dr. Wakefield is a martyr.

Poisonous Pete



Pete is terrified by all the words he can't pronounce on the vaccine ingredient list. He doesn't know much about chemistry, but he does know formaldehyde is used in embalming and mercury is bad.

Loony Lucy



All antivaxers are a bit conspiratorial, but Lucy takes the cake. Vaccines are not just a ploy to make money they are tools of the illuminati. Vaccines were invented by the rich to sterilize poor children and depopulate the earth.

Hoaxy Holly



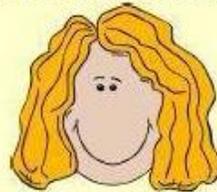
Holly knows the real reason for vaccines; PROFIT. Everyone from the local doctor to the World Health Organization is in on it. Vaccines don't work and never did. Parents who vaccinate their children are just lining the pockets of big pharma.

Hygienic Helen



Unlike most denialists Helen actually believes vaccines are safe and effective, she just doesn't think they are needed. Disease was eradicated by soap and clean water. Vaccines had nothing to do with it, they just happened to be invented around the same time.

Naturalist Nancy



Nancy is sure all this modern medicine is killing us. If we all went back to nature we would have eternal life. Natural Immunity is the way to go (even if 2 in 1000 cases of measles are fatal). Just don't tell her life expectancy before modern medicine was around 40....

When a flood of information reaches people having no access to statistics and scientific tools for pre-treating information, data meet biases.

Every individual (including scientists and medical doctors) select a tiny bit of the available data, according to its pre-conception.

Social forums further reinforce this mechanism by creating communities where all the exchanged information is already biased (and presenting selectively those communities which users are expected to like)

HAVING ACCESS TO THE (WRONGLY SELECTED) DATA, EVERYONE THINKS TO KNOW BEST THAN DOCTORS.

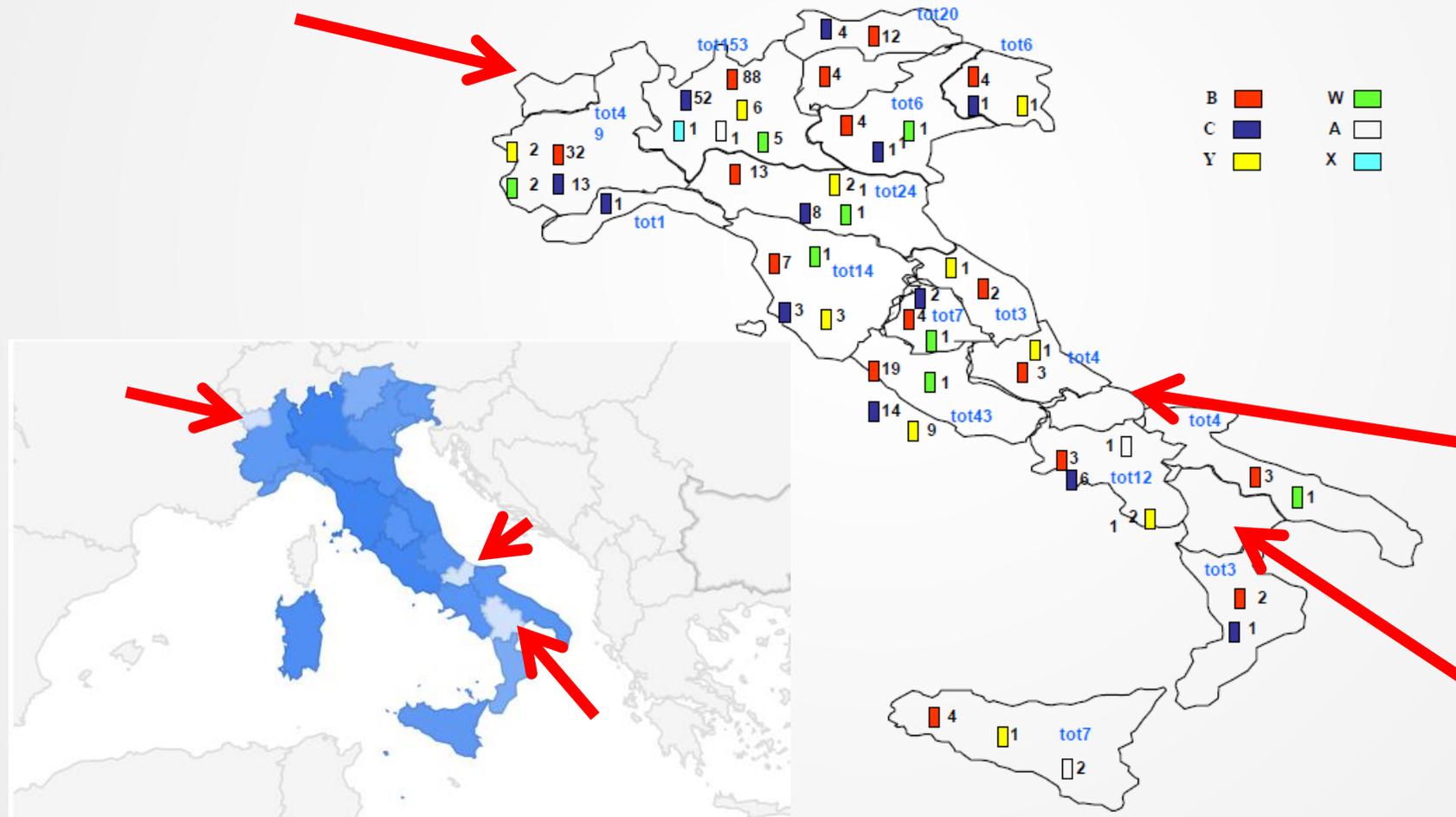
Empowering digital democracy: (data) science literacy

1. Everyone needs to know what information and data really are and how Big Data (i.e. Internet information) is biased, noisy, flawed if taken as it is
2. Everyone needs to know that before assuming data, they have to be processed, and what this actually means (in very simple terms)
3. Everyone needs to know how to extract a larger data set of internet data than those accessible by direct reading and how to use simple tools for there analysis (which is perfectly feasible at school)

Big Data @ Home: an example for Meningitis

ISS epidemiology data (on the right) are mirrored in google volume queries (on the left), when the query examined is properly set.

Ceppi e campioni di meningococco inviati all'ISS 2008-2011 (tot.386)

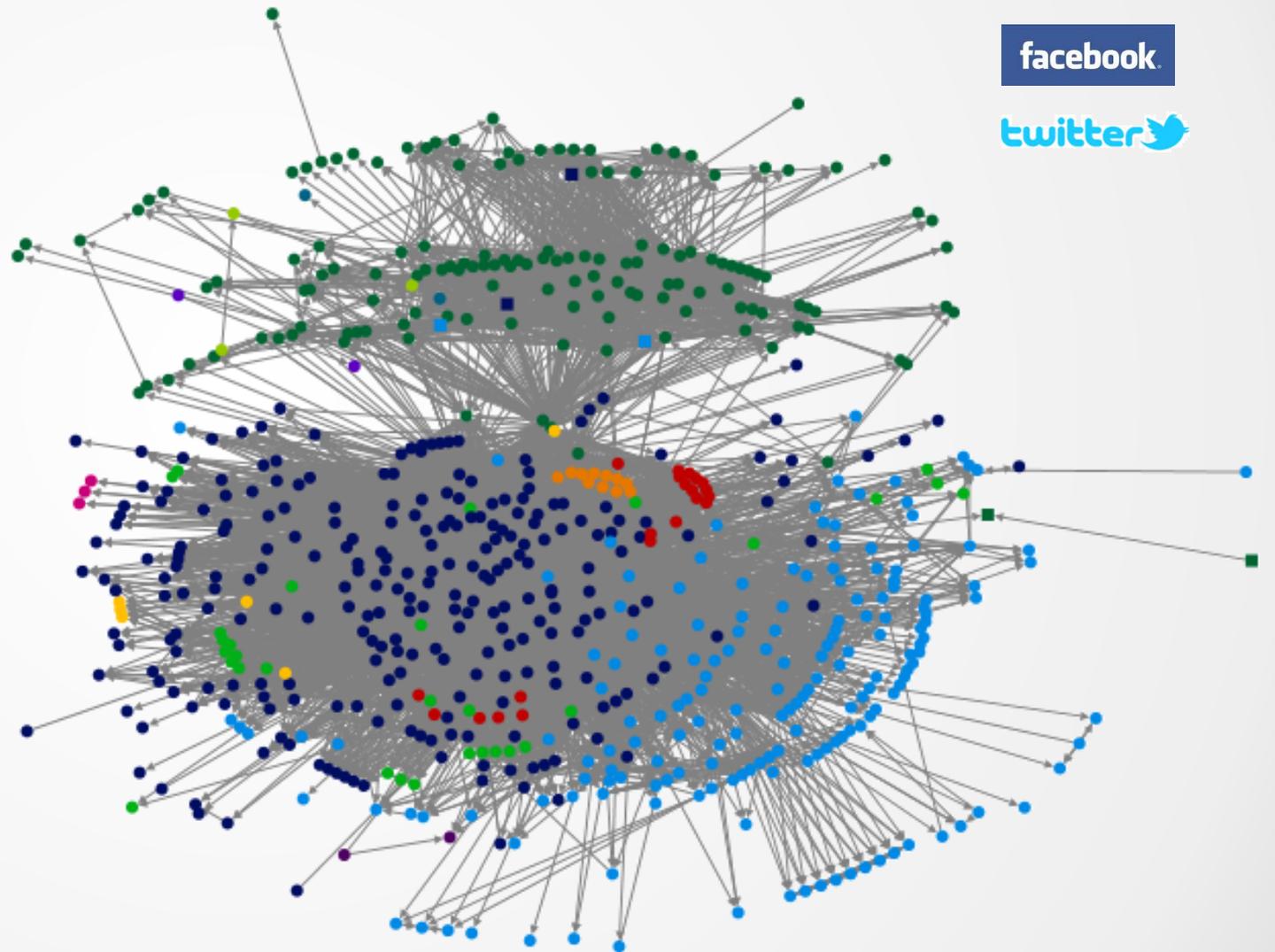


Big Data @ Home: vaccine discussions in Italy

To understand bias in vaccine communication, one can identify groups discussing vaccine over Facebook and Twitter (Italian groups).

Open-source tools can be used to group people according to their community, and then examine the discussions in each community to measure how polarized it is.

As a results, one can show that groups over Facebook and Twitter are always polarized: they are not a source of unbiased information on vaccines, and they can be labeled according to the bias degree (i.e. internal homogeneity).



Conclusions

1. Big Data are a consequence of the industrial revolution: when equipments went digital and Internet connected, data flowed
2. Big Data applications for health care are already emerging and promising results were already presented in several different domains
3. Big Data are not information: processing is required
4. A Big Divide due to the asymmetric availability of minimal information processing abilities and to the biased selection used by individuals to reduce complexity is rendering the promising world of Big Data biased and also dangerous, including
5. This Big Divide can and must be reduced to preserve society