Big Data and machine learning in service of antimicrobial stewardship

Guillaume Béraud

Université Laval, Canada

October 5th, 2017

Resistance to antibiotics is on the rise

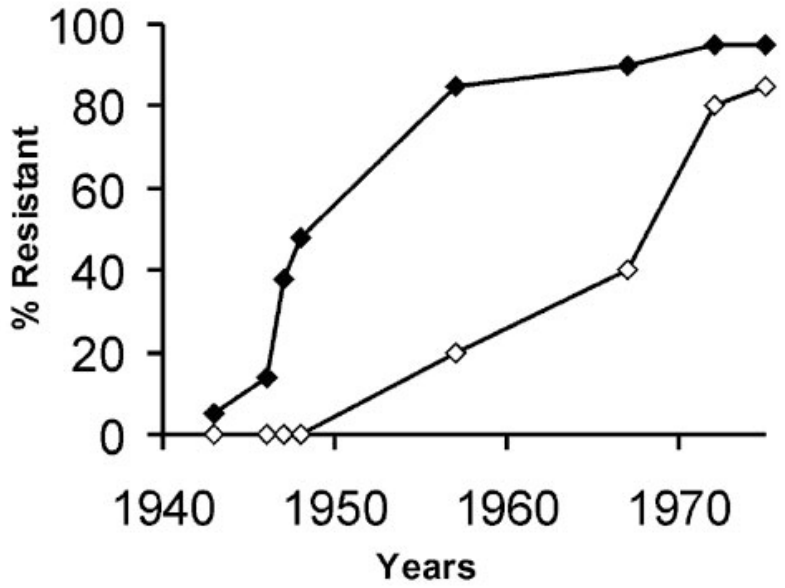


Figure – Secular trends of approximate prevalence rates for penicillinase-producing, methicillin-susceptible strains of Staphylococcus aureus in hospitals (closed symbols) and the community (open symbols). Chambers EID 2001

No new antibiotics

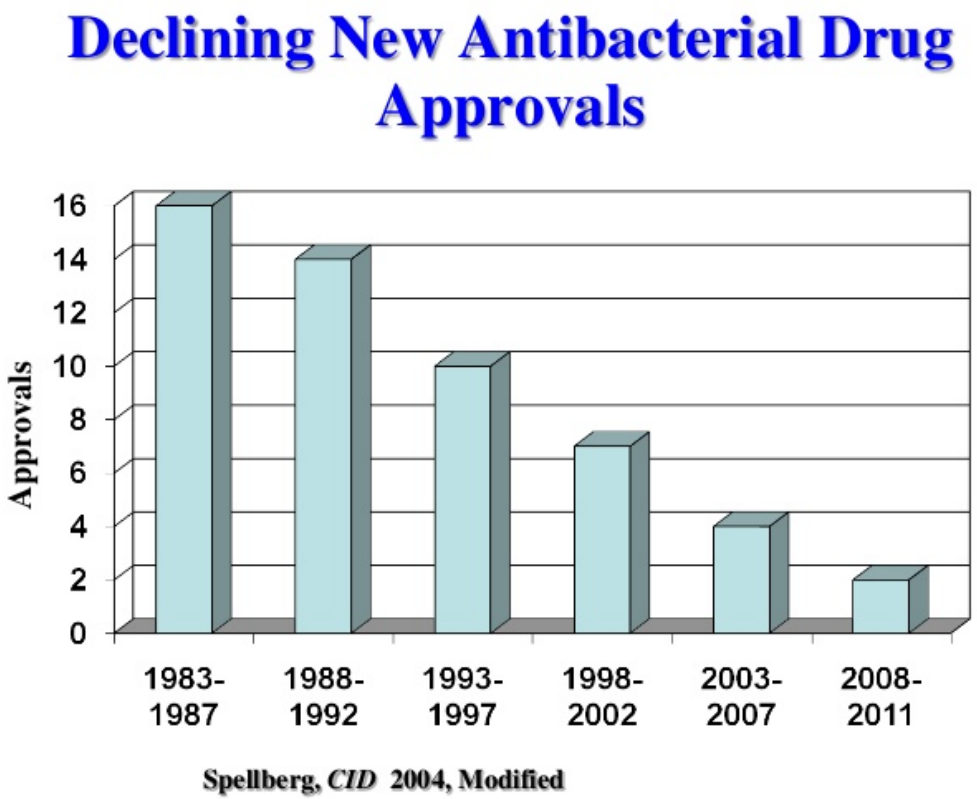


Figure – Adapted from Spellberg CID 2004

Besides

- Infections are often poorly diagnosed (e.g. viral vs. bacterial)
- Antibiotics are not prescribed adequately
- What should we do ?

In absence of resources :

Start to be smart

Changing the paradigm

We need data

Changing the paradigm

We need **BIG DATA**

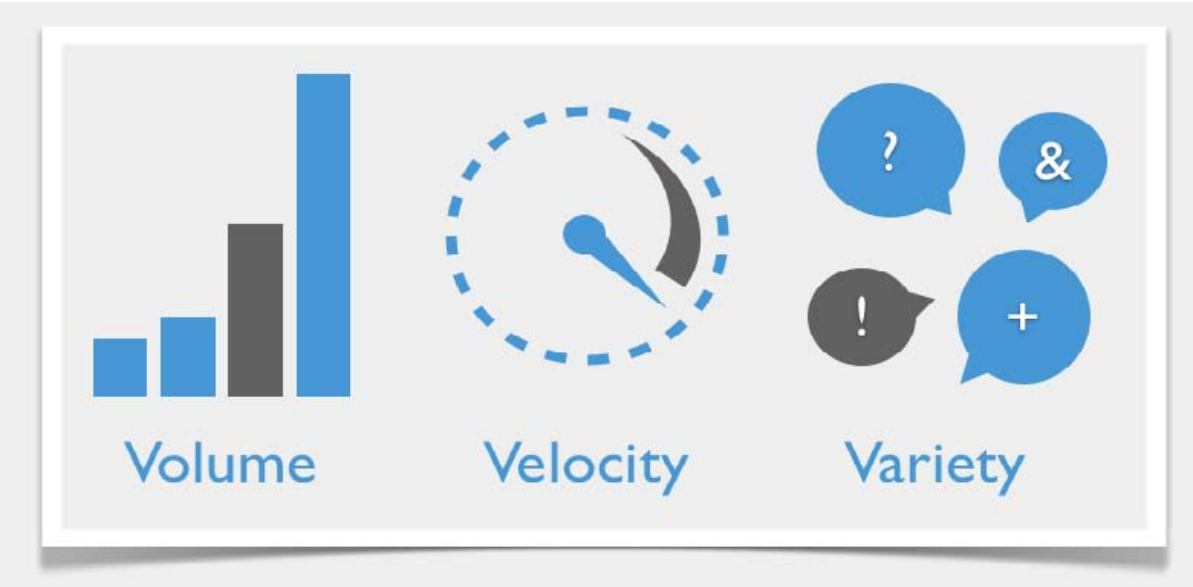
What is *Big data*

According to Wikipedia

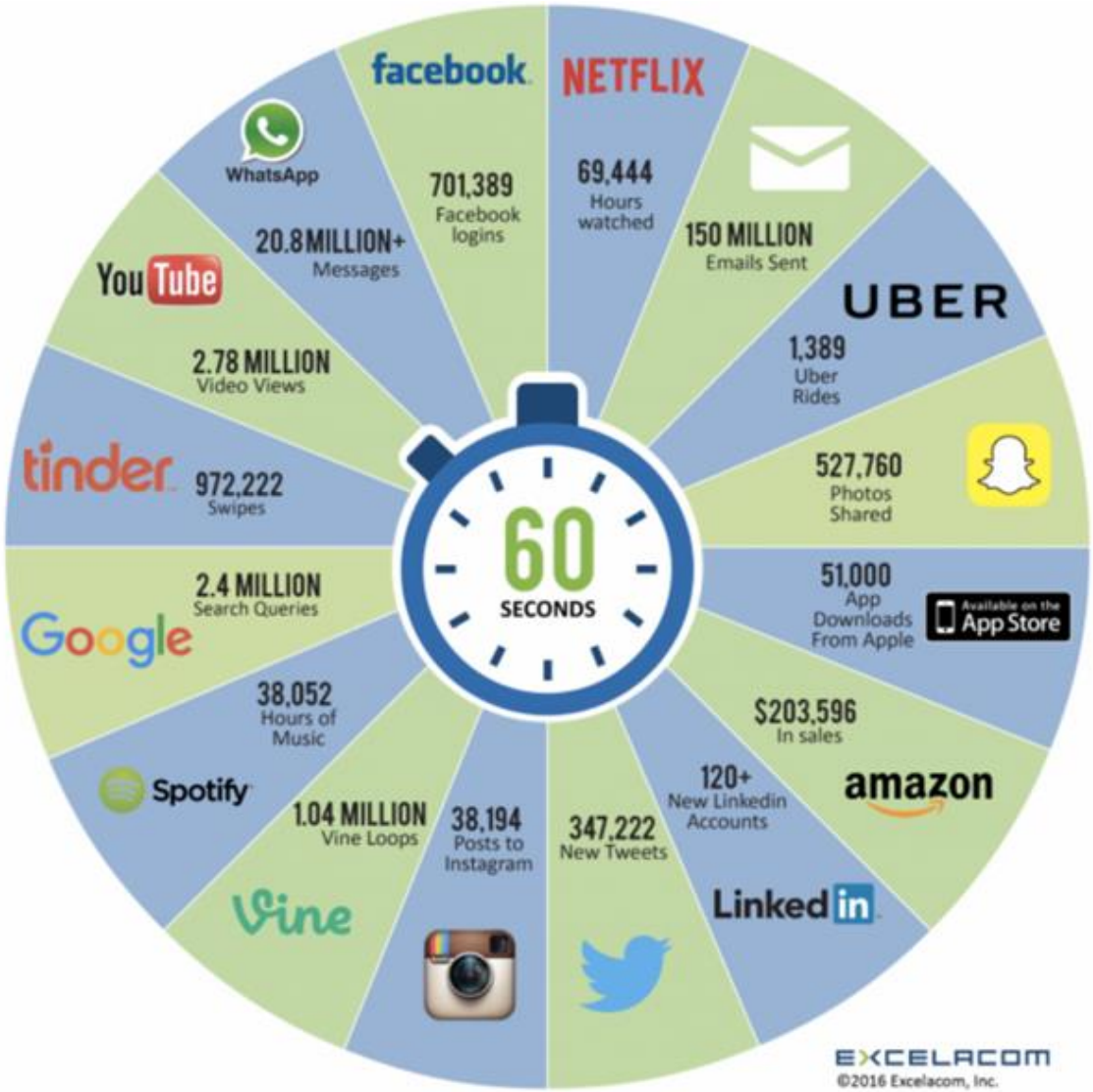
Big data is a term for data sets that are so large or complex that traditional data processing application software is inadequate to deal with them

The 3 Vs

- **Volume** : No sampling or aggregation. All the data are potentially useful, therefore the amount is sizable.
- **Variety** : It may combine text, photo and numbers ...
- **Velocity** : data are often available in real-time.



One minute on Internet



From Big data to Machine Learning

- In traditional data analysis :
  - You have an hypothesis
  - You test your hypothesis
  - Then you draw conclusions and try to generalize
- In machine learning, you don't start with an hypothesis
- You search for patterns without "a priori".

From Big data to Machine Learning

- Instead of giving an algorithm or "rules" to the computer, we'll give :
  - a task,
  - a lot of training data sets (Big Data !) to play with,
  - maybe one or two example,
- Then, the computer will be able to distinct patterns in the next set of data and to solve a problem (the task)
- The key is to search (and find) patterns without "a priori".
- Discovering a pattern won't give you the explanation for the mechanism !

Practical example

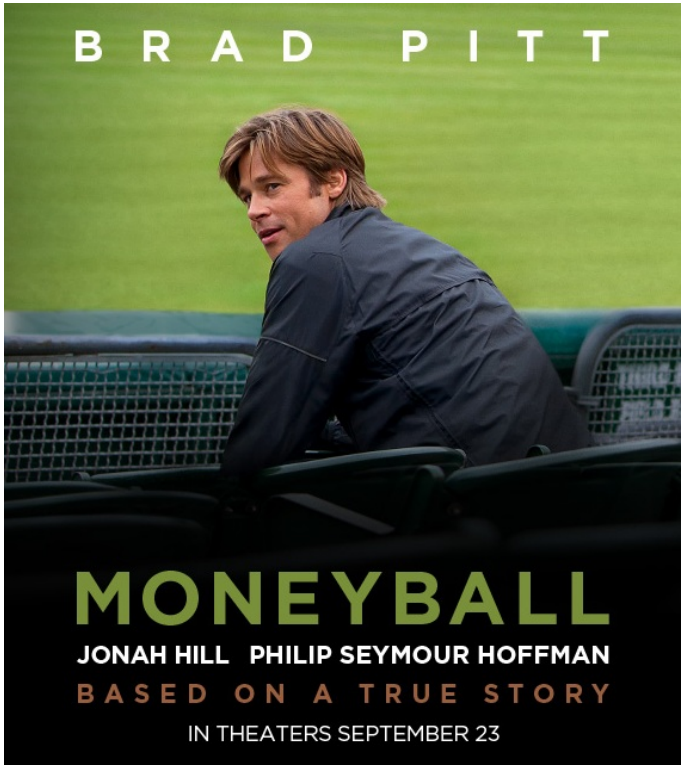
- With data mining, an American supermarket company discovered that placing diapers close to the beers resulted in an increase of beers sales . . . but only on the Friday evening and on Saturday.
-
-

Practical example

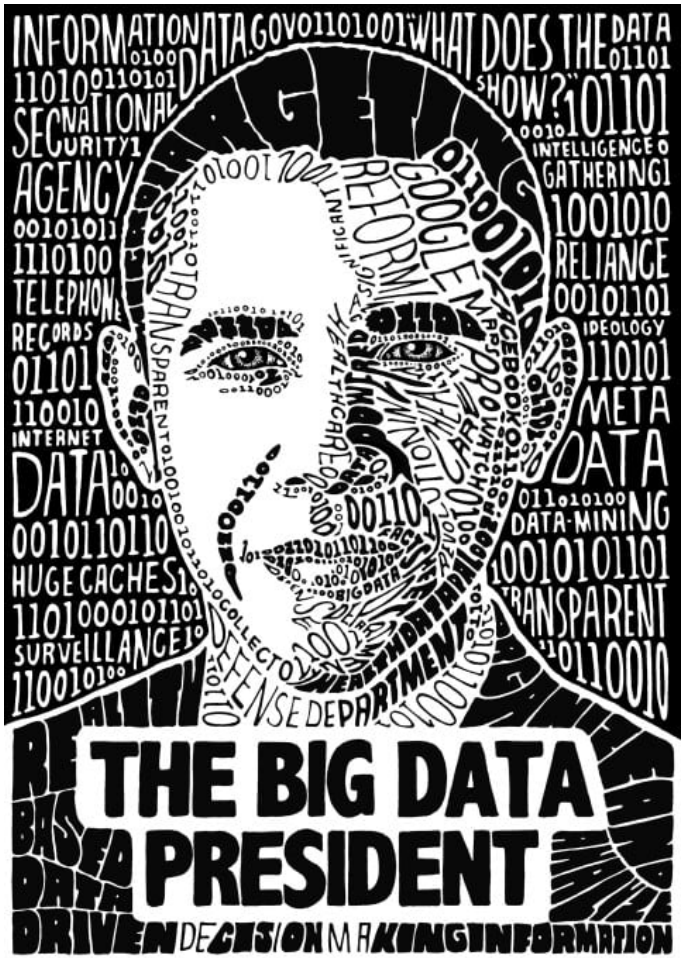
- With data mining, an American supermarket company discovered that placing diapers close to the beers resulted in an increase of beers sales . . . but only on the Friday evening and on Saturday.
- Because men sometimes go to the supermarket to buy diapers, but usually during Friday afternoon and Saturday...
-

Practical example

- With data mining, an American supermarket company discovered that placing diapers close to the beers resulted in an increase of beers sales . . . but only on the Friday evening and on Saturday.
- Because men sometimes go to the supermarket to buy diapers, but usually during to Friday afternoon and Saturday...
- *Therefore, if it works for marketing, why not antimicrobial stewardship ?*



- Having to deal with one of the poorest base-ball team in USA
- Billy Beane, and a young economist Peter Brand were able to build a competent team in Oakland, using data mining
- The rest is now legendary ...



- Obama used Big Data with the company HP Vertica to build his relection.

Usually, how do we think in infectious diseases ?

- Treating an infectious diseases consists in distinguishing colonization from infection.
- Infection is determined according to reaching a certain threshold for the most abundant bacteria.
- Only the most abundant bacteria is usually considered.
- But what about the other bacterias ? *Opportunistic criminal association* ?

Villains and sidekicks ?

- Does every villain need a sidekick ?
- Is a villain as bad as he usually is without his sidekick ?

Wolves hidden among sheeps

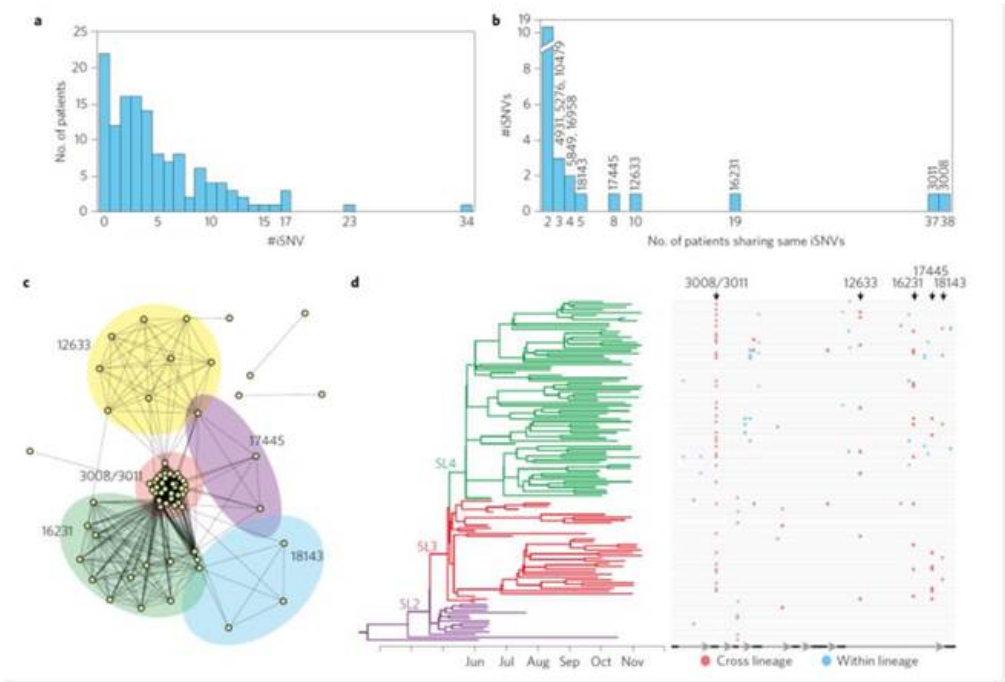
- Small colony variants of S. aureus : a bacterial survival strategy
- Responsible of chronic, relapsing infections
- A practical application : hVISA

Aggregated data

- A coarse way to analyze infectious diseases,
- but acceptable as long as it was the only feasible way.

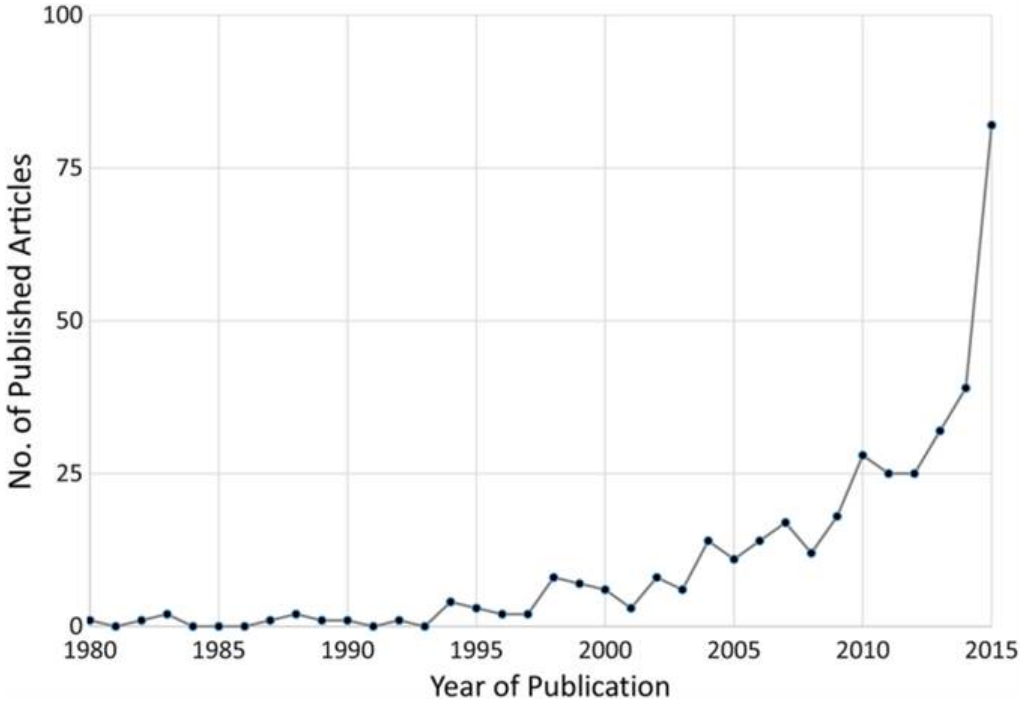
The times they are A-changin

- MALDI-TOF, RT-PCR, Point-of-care lab, crowdsourcing ...
- Exhaustive and precise data are available
- The time has come for big data in infectious disease



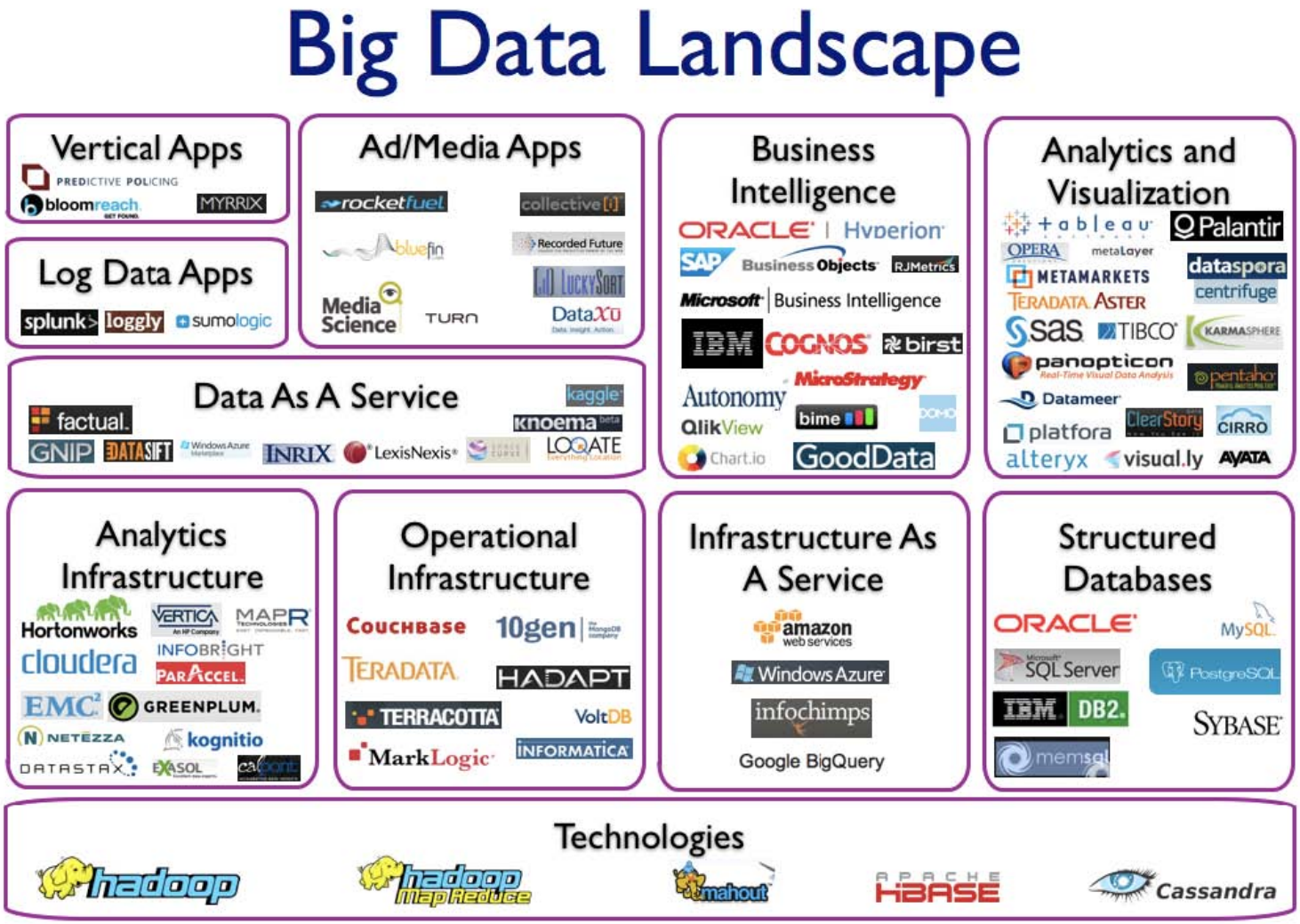
Big data and infectious diseases

- (big data AND infectious diseases) OR (big data AND epidemics) OR (digital epidemiology AND infectious diseases)

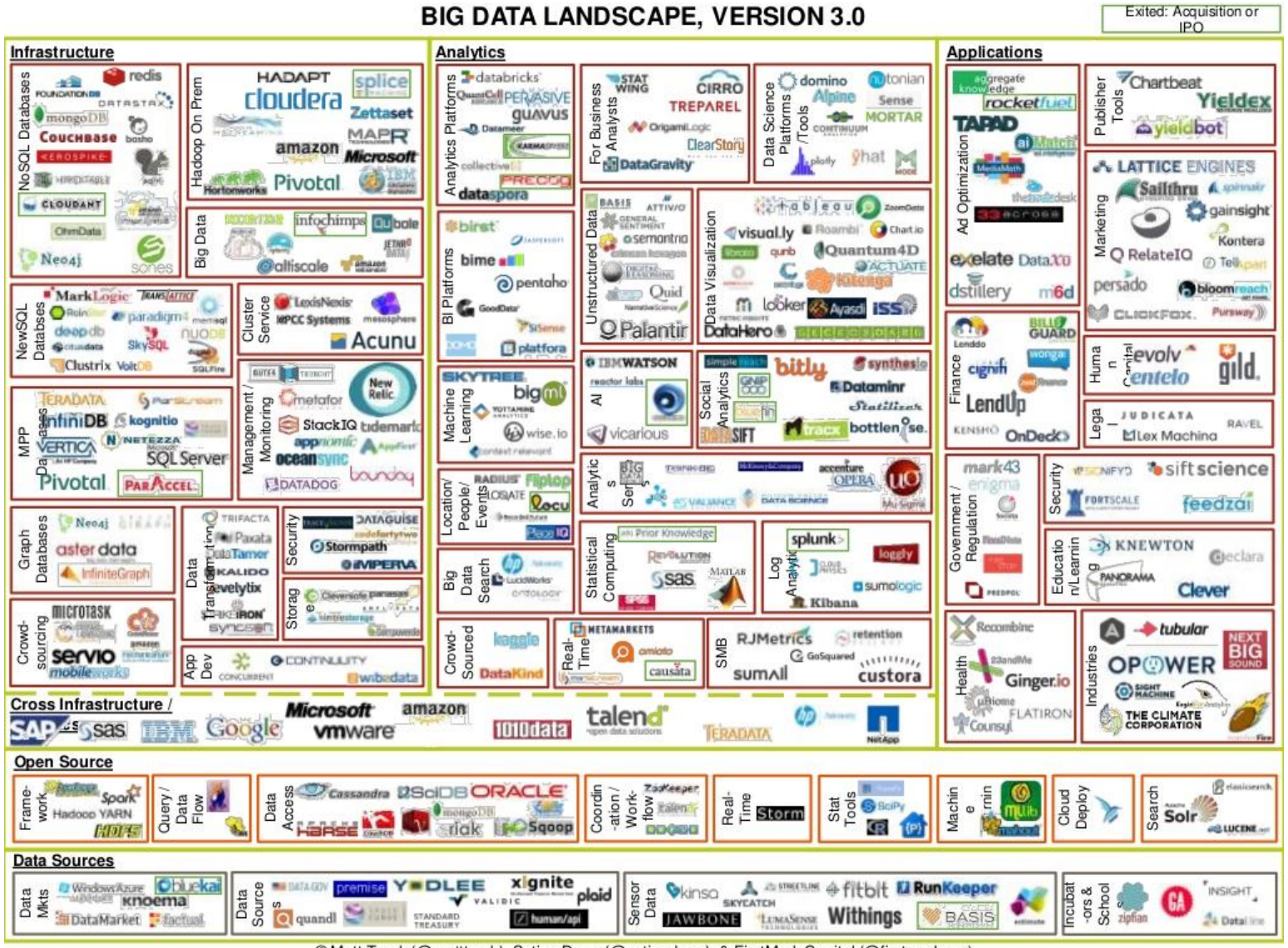


A rise also in providers and tools

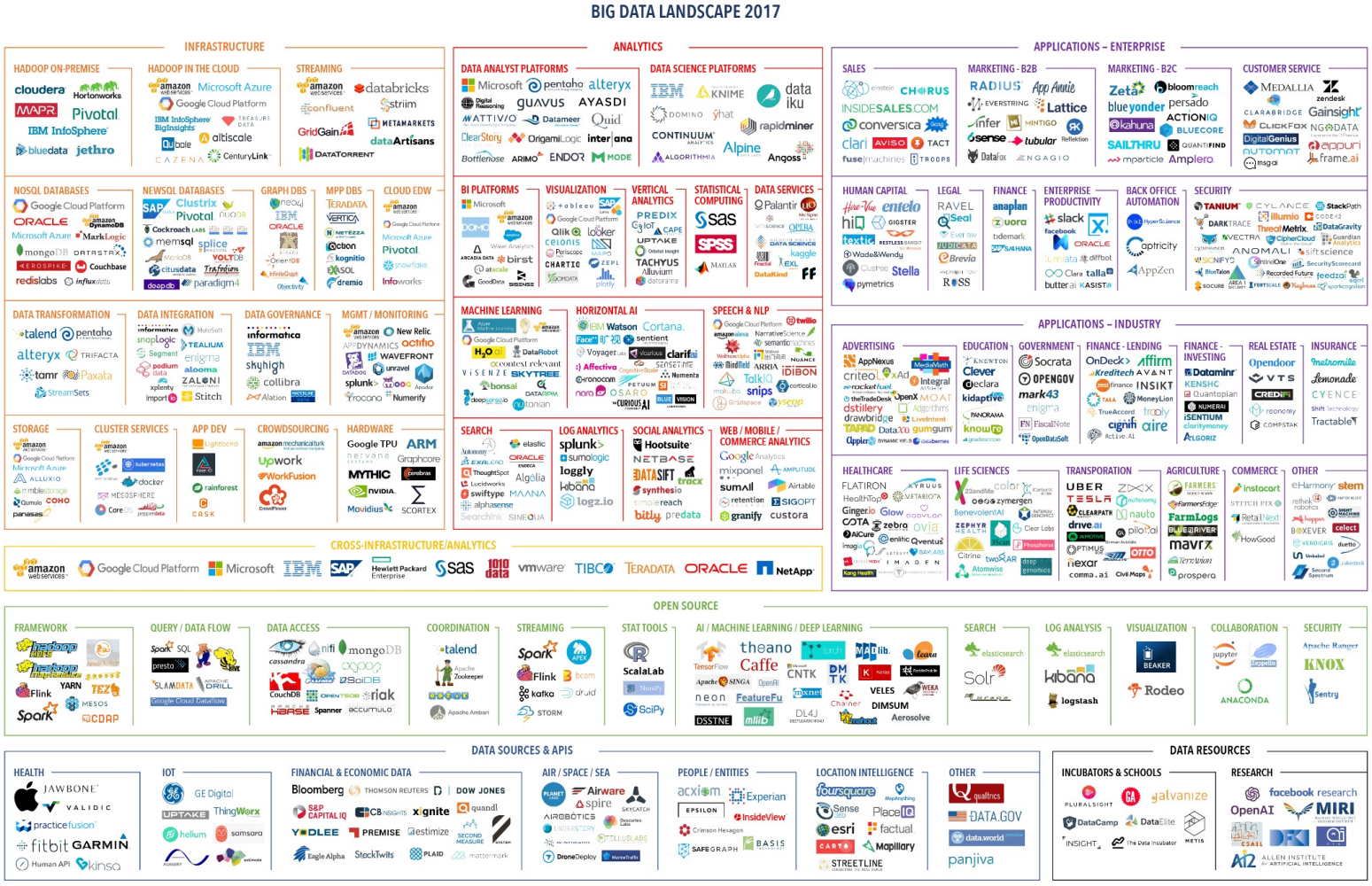
A rise also in providers and tools



A rise also in providers and tools

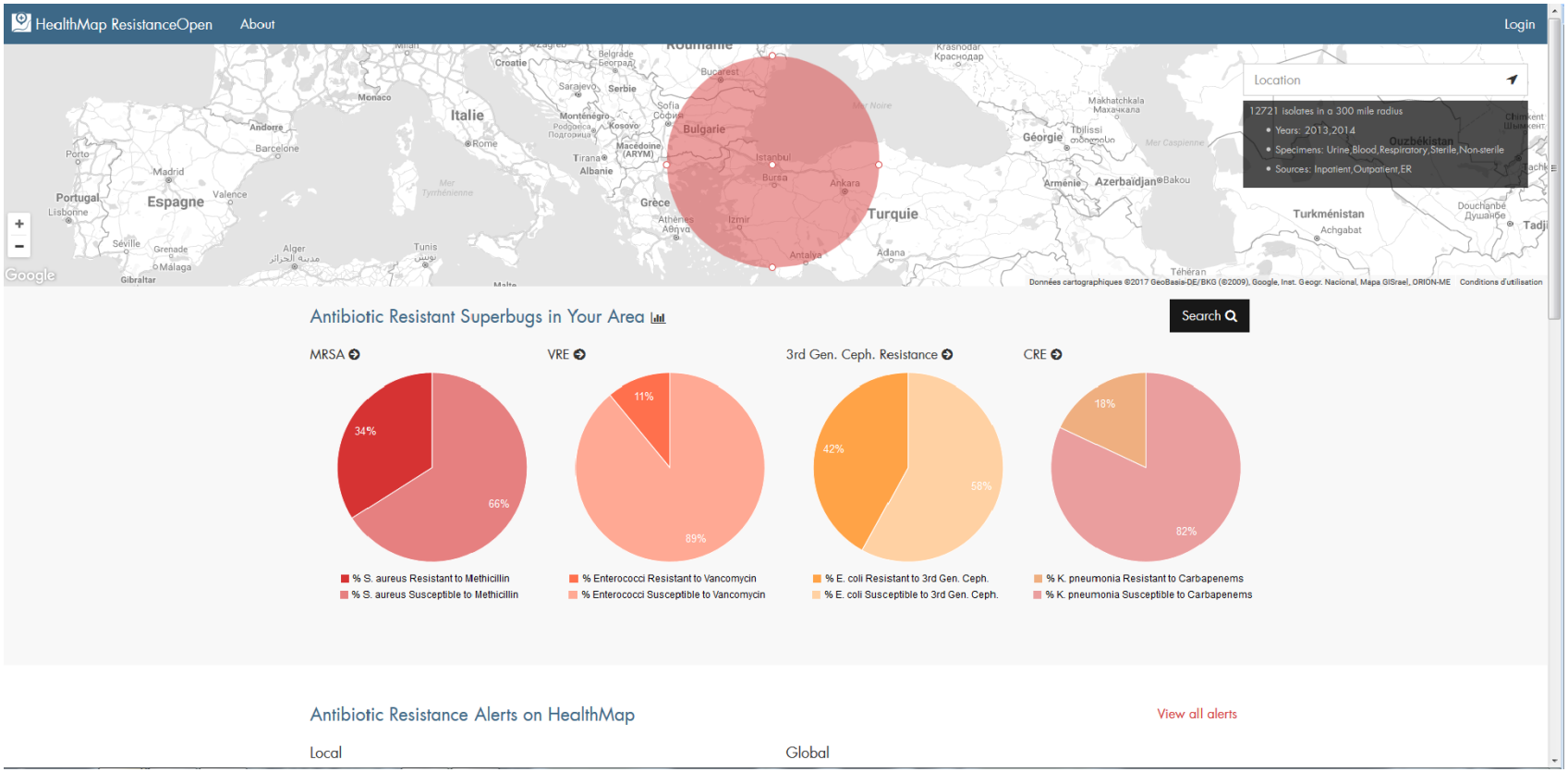


A rise also in providers and tools



A rise also in providers and tools

HealthMap : Resistance Open



Big data and infectious diseases

- Data available.
- Not so big compared to onco-genomics or theoretical physics
- Will allow understanding of infectious diseases dynamics without a priori
- Big data already successfully used in marketing, meteorology, baseball and US politics.

BUT

- Patterns, even new ones, shouldn't replace mechanistics studies
- It's a complement, and a source of hypothesis
- « Big data hubris » : An overload of data is not an excuse for a lack of rigorous analysis.
- Please welcome the next speaker to see Big Data & Infectious Diseases

Thank you for your attention