

Maximizing Insights from Clinical Data: Exploiting the Potential of Data Farming in Health Care

H.U.B

HÔPITAL UNIVERSITAIRE
DE BRUXELLES
ACADEMISCH ZIEKENHUIS
BRUSSEL



INSTITUT
JULES BORDET
INSTITUUT

Hôpital
Erasmé



Hôpital Universitaire
des Enfants
Université Kinderziekenhuis
Sintogen Polikliniek

Thomas GUIOT

BELNUC – 22/04/2023



OpenAI ChatGPT



Seminar title about data farming, exploiting data from clinical routine





OpenAI ChatGPT



Seminar title about data farming, exploiting data from clinical routine



"Maximizing Insights from Clinical Data: Exploiting the Potential of Data Farming in Healthcare"



nuclear medicine

AND

machine learning

cancer

OR

artificial
intelligence

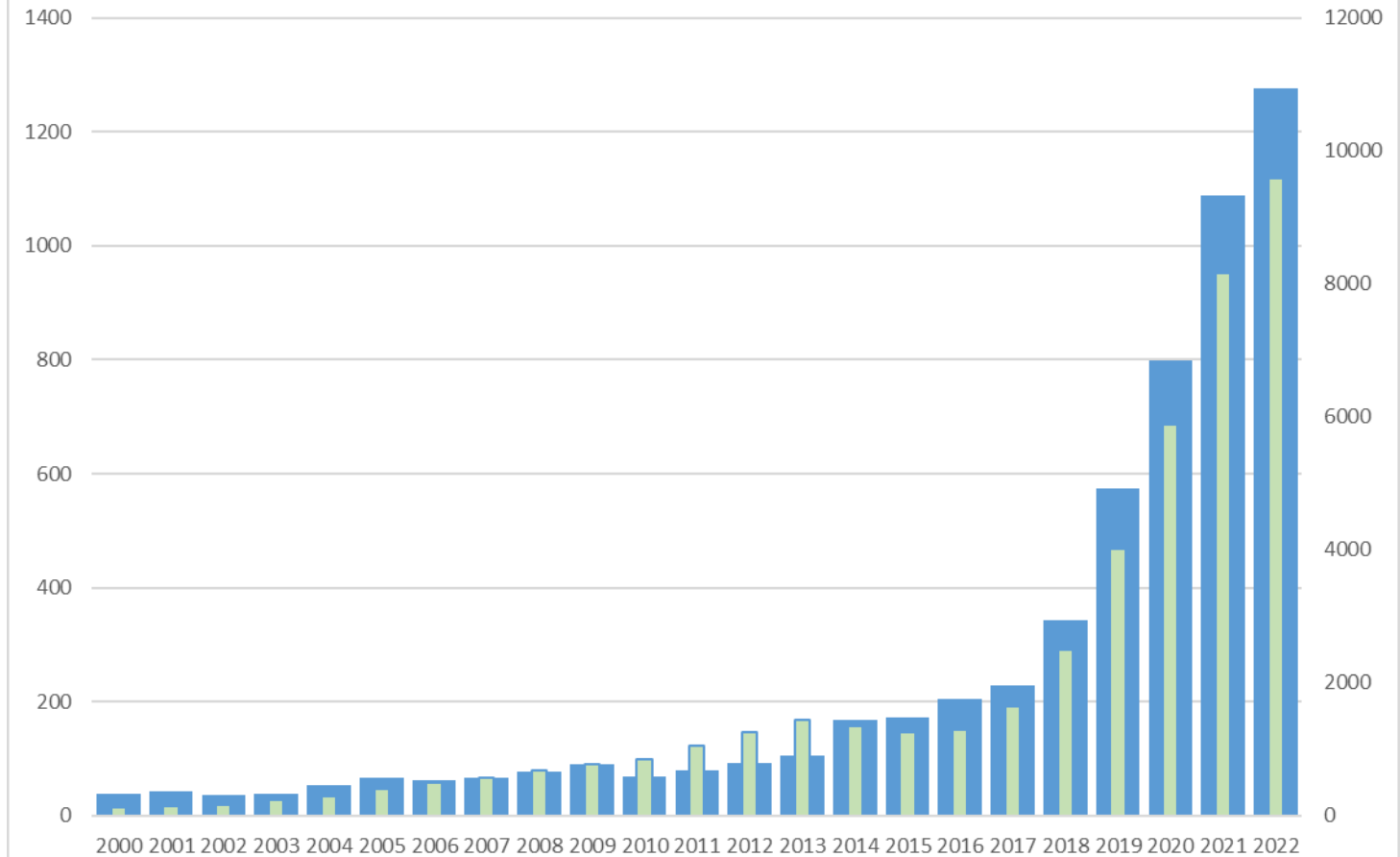
OR

deep learning

OR

neural network

Number of publications per year





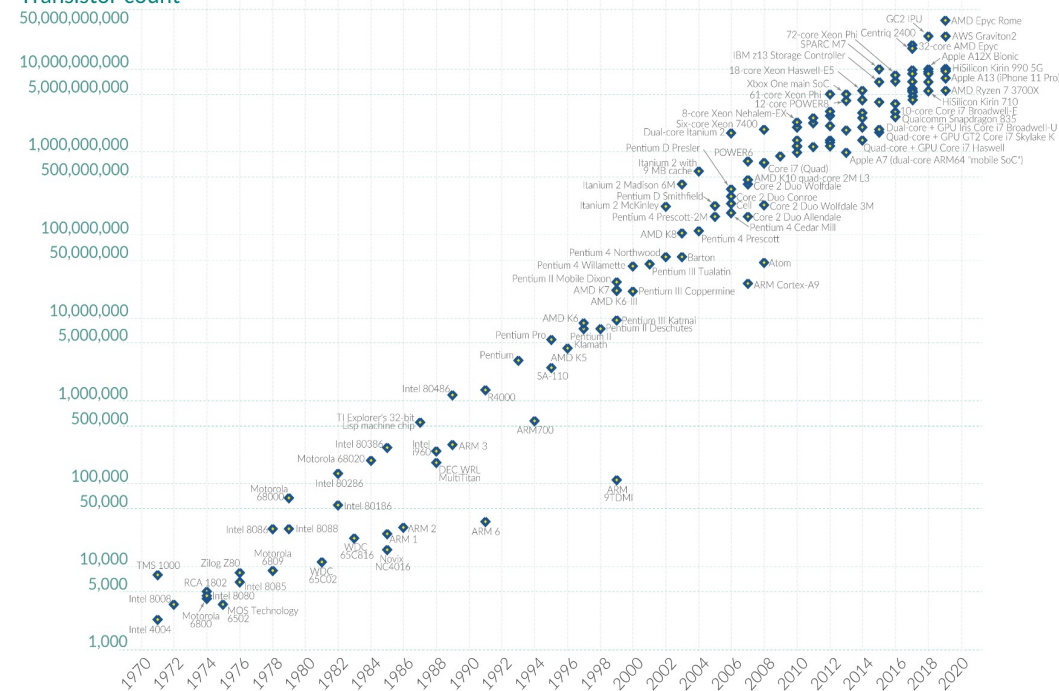
Why now?

Computing power

Moore's Law: The number of transistors on microchips doubles every two years Our World in Data

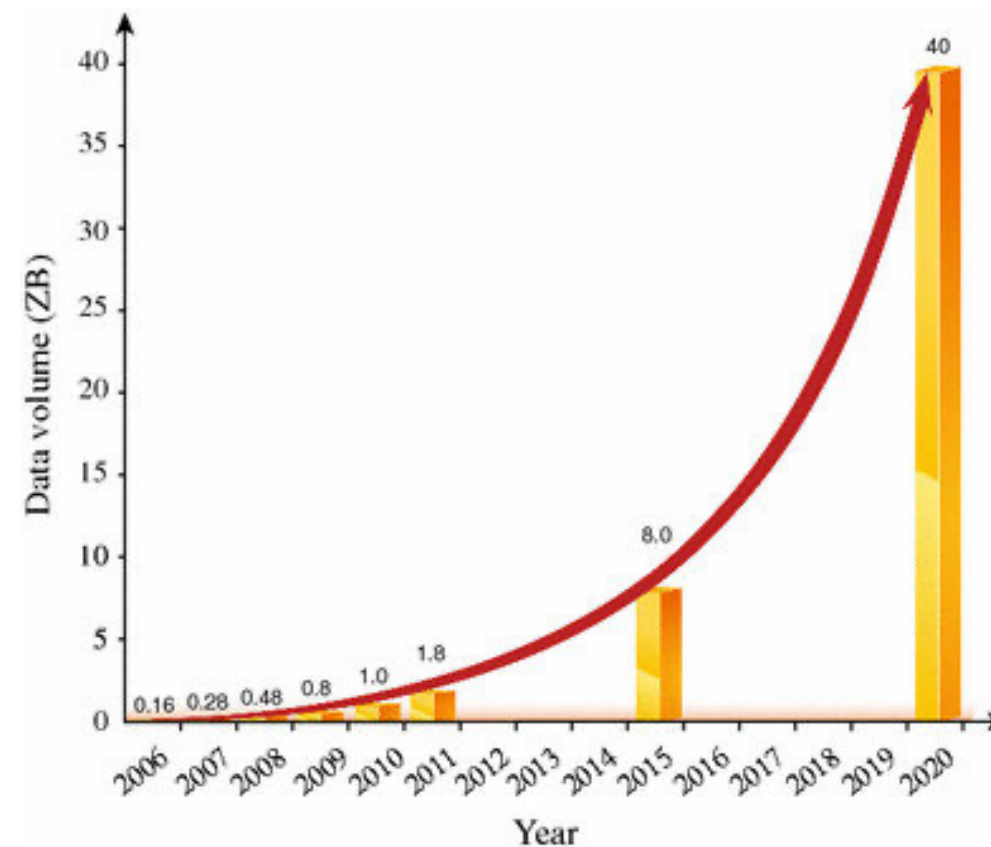
Moore's law describes the empirical regularity that the number of transistors on integrated circuits doubles approximately every two years. This advancement is important for other aspects of technological progress in computing – such as processing speed or the price of computers.

Transistor count



Data source: Wikipedia (wikipedia.org/wiki/Transistor_count)
OurWorldinData.org – Research and data to make progress against the world's largest problems. Licensed under CC-BY by the authors Hannah Ritchie and Max Roser.

Data availability



Guo, Huadong & Wang, Lizhe & Chen, Fang & Liang, Dong. (2014). Scientific big data and Digital Earth. Chinese Science Bulletin. 59. 1047. 10.1360/972013-1054.

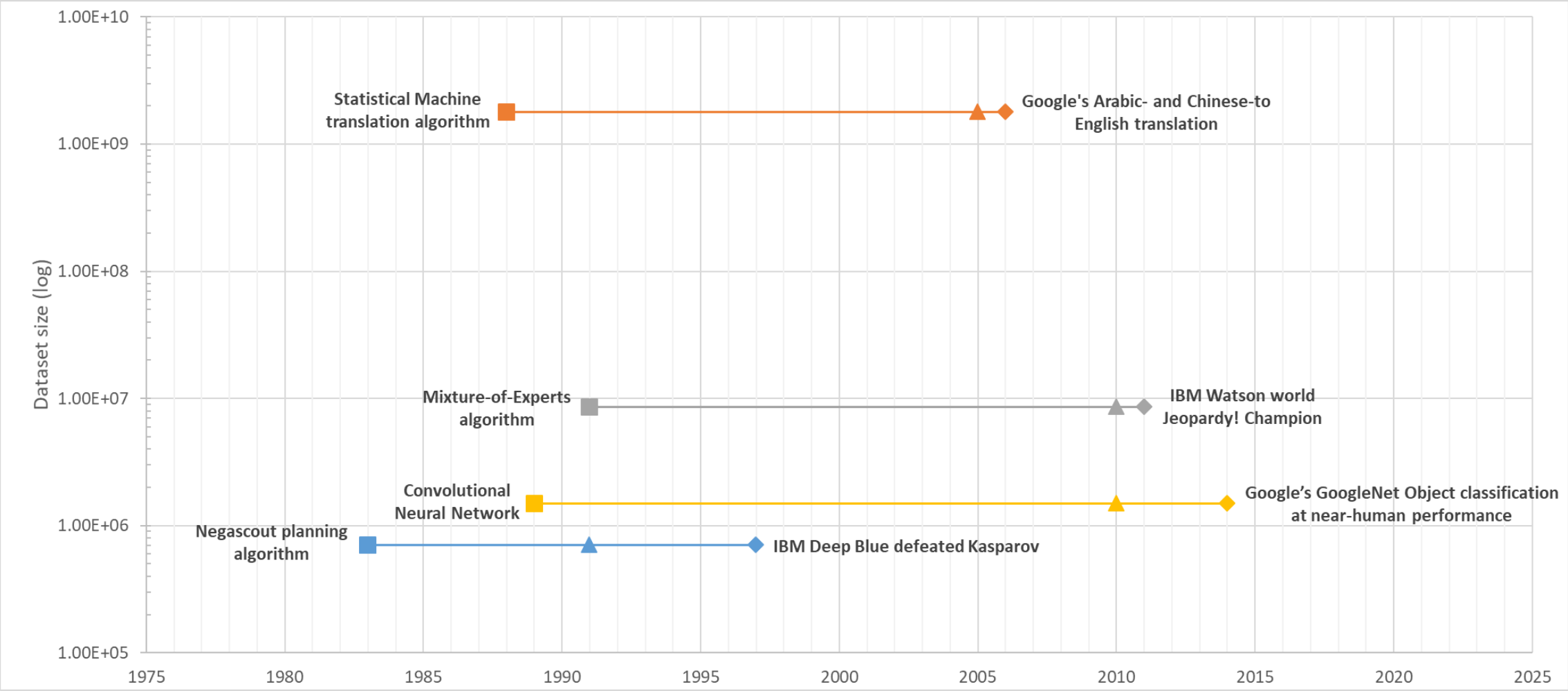


Why now?

■ Algorithm, first proposed

▲ Datasets, first available

◆ First Breakthrough

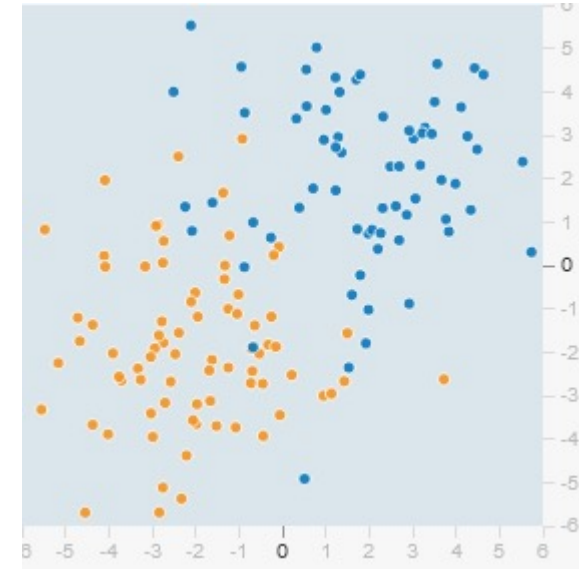


Why is (big) data useful ?

AI

- Better accuracy
- More robustness
- Generalization

Avoid
overfitting



Why is (big) data useful ?

AI

- Better accuracy
- More robustness
- Generalization

Avoid
overfitting



Why is (big) data useful ?

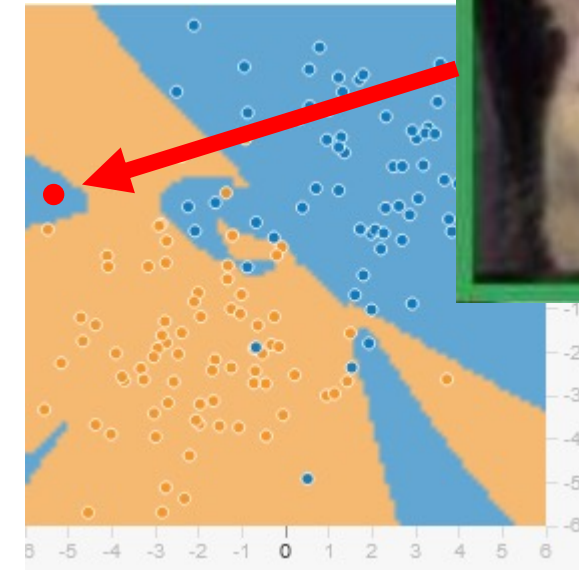
AI

- Better accuracy
- More robustness
- Generalization

Avoid
overfitting



Min 10-20 data points per feature



AI

- Better accuracy
- More robustness
- Generalization

Other uses

- "Classical" statistics
- Practice quality improvements
- Institutional analytics

Why is (big) data useful ?



1977

National registry
creation

99%

Consent rate

150K

Patients registered
up to 2016

~1.5M

Variables collected
annually

81%

Patients enrolled in
trials

631

Publications up
to 2020



How do we get data?

Clinical
trials

Data
mining

Data
farming

Volume



Quality

Time/effort



How do we get data?

Clinical
trials

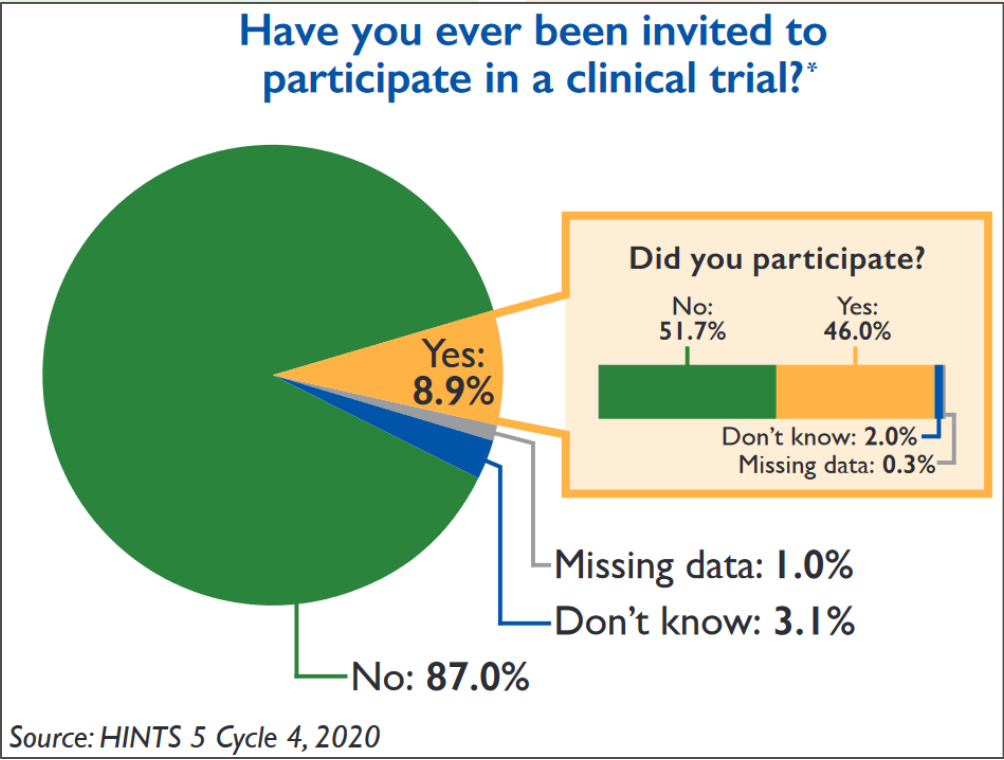
Data

Data

Volume

Quality

Time/effort





How do we get data?

Clinical
trials

Data
mining

Data
farming

Volume



Quality



Time/effort





How do we get data?

Clinical
trials

Data
mining

Data
farming

Volume



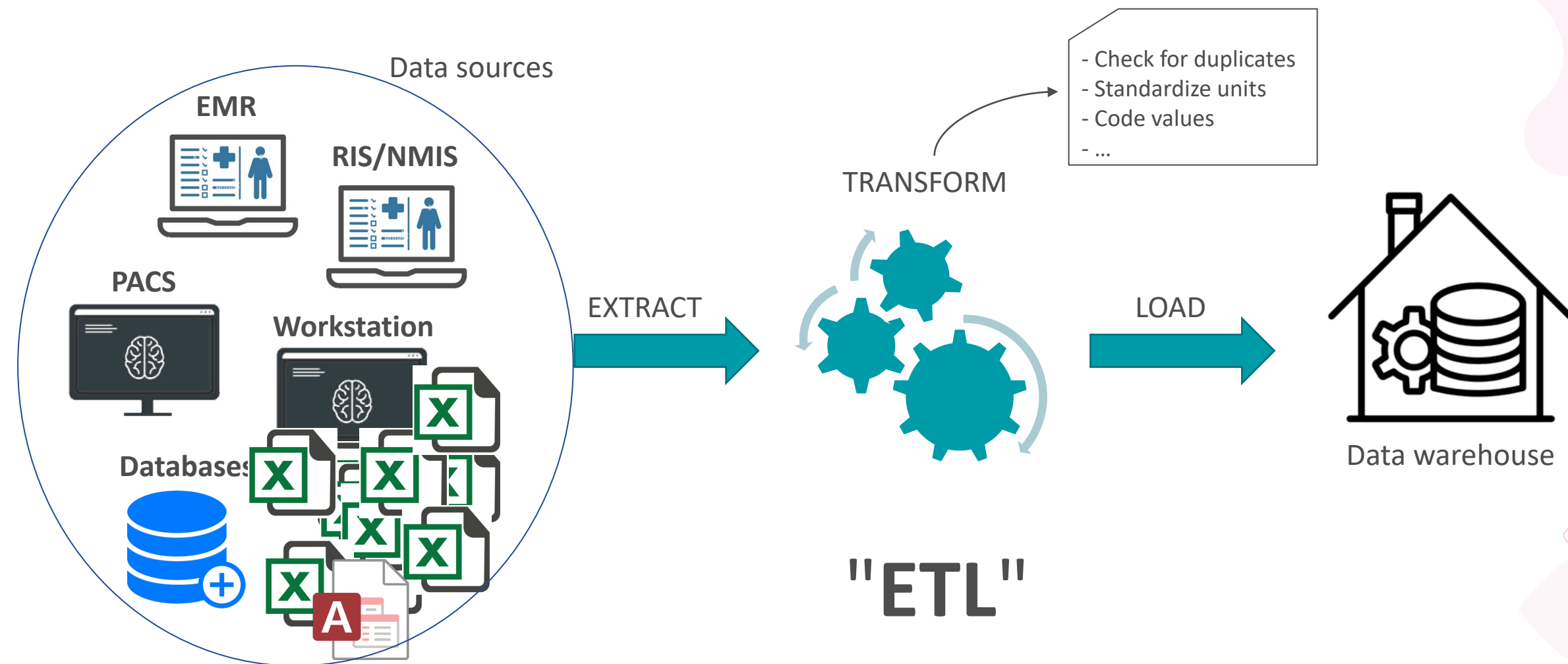
Quality



Time/effort



Process of extracting and analyzing large datasets to generate new information





Main problem: Assumes the data are there, ready to be mined

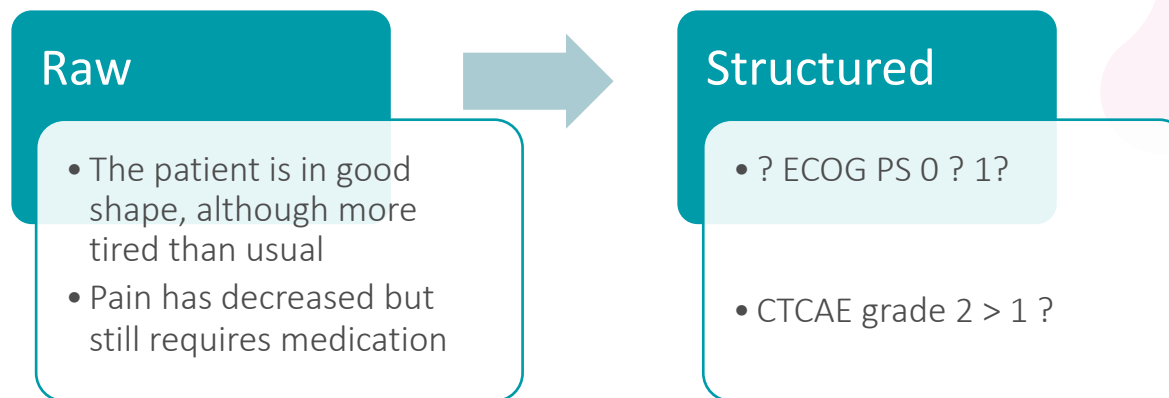
! Missing data

! Unstructured data

Free-text variables

! No metadata

Who, when, how



MATV = 153 mL	
Made by	Dr Doe, J
Date	20230422
Contouring method	PERCIST-based thresholding
Software	LifeY
Software version	7.1.6534
...	...



How do we get data?

Clinical
trials

Data
mining

Data
farming

Volume



Quality



Time/effort





Data
farming

"Best of both worlds" approach

High volume
from clinical care

High quality
by cultivating your data sources



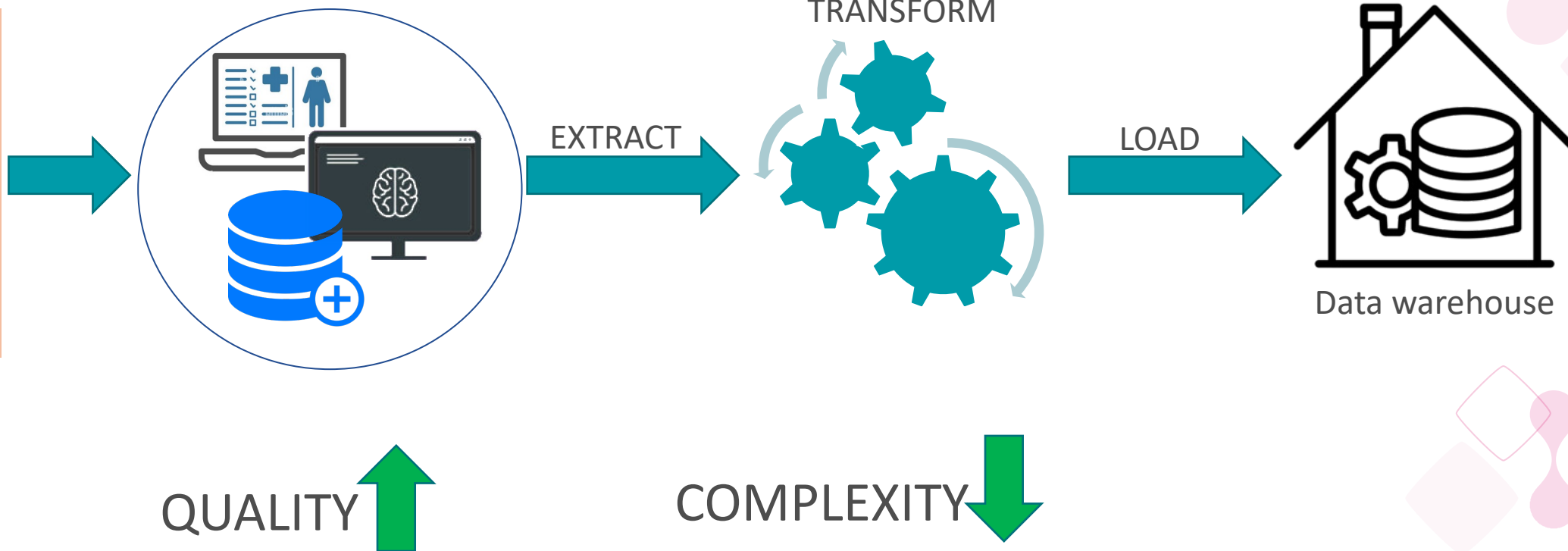
HÔPITAL UNIVERSITAIRE
DE BRUXELLES
ACADEMISCH ZIEKENHUIS
BRUSSEL



« Good
Planting
Practice »

Availability of key data
elements

Professional standards



High volume

from clinical care

High quality

by cultivating your data sources

Data mining = Age of hunter-gatherers



Data farming = Age of agriculture





Ensure collaboration

- Team effort
- IT
- Head of department
- Colleagues



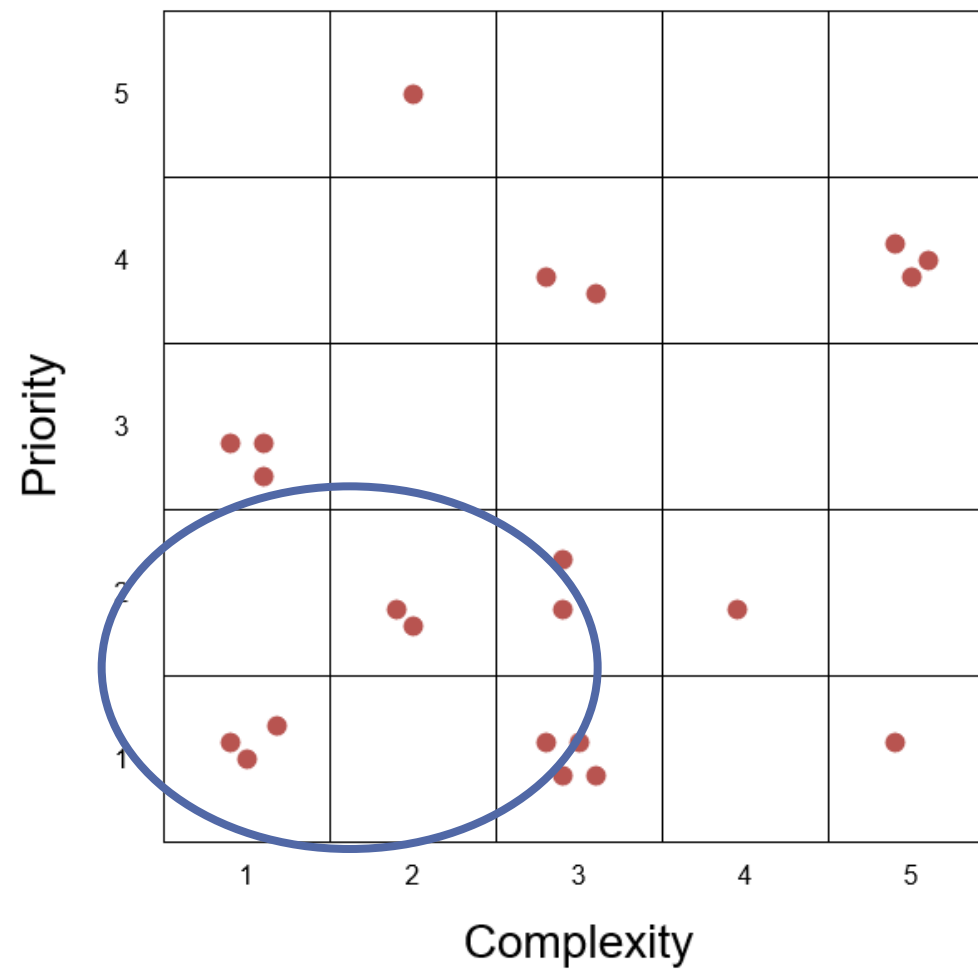
Define key data elements

- Goal
- Litterature
- Experience
- **Priority score**



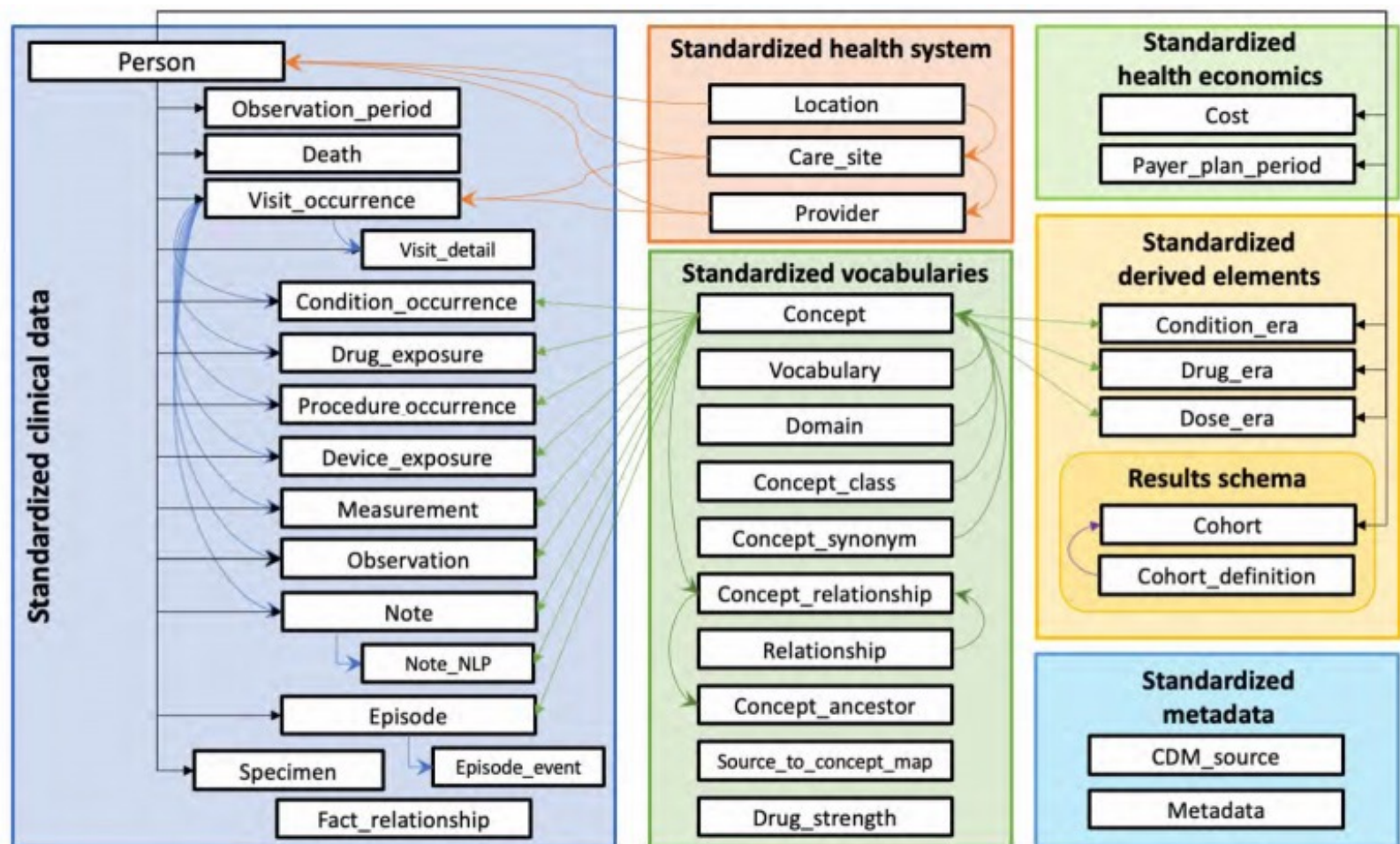
Assess ETL challenges

- Structured?
- Systematic?
- **Complexity score**



OHDSI OMOP CDM

OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS



- Large vocabulary
- Very flexible
- Big active user base



Data
farming

Clinical
trials

Data
mining

Data
farming

Volume



Quality



Time/effort



Scope: Treatment of metastatic prostate cancer patients with PSMA-RLT

Registry:

- Web-based
- Built with  REDCap®
- 260 variables
- Currently manual data entry
- Currently monocentric but ...

LuPSMA Registry

PID 28

Record Home Page

The grid below displays the form-by-form progress of data entered for the currently selected record. You may click on the colored status icons to access that form/event. If you wish, you may modify the events below by navigating to the [Define My Events](#) page.

Legend for status icons:

-  Incomplete
-  Incomplete (no data saved) ?
-  Unverified
-  Complete
-  Many statuses (all same)
-  Many statuses (mixed)

[Choose action for record](#)

Subject ID Test subject

Data Collection Instrument	History & General information	Pre-therapy	Therapy cycle	Post-therapy
	hist	pre_tx	tx_cycle	post_tx
Patient information				
Prostate Cancer History				
Other cancer history				
Treatment information				
Biology				
Clinical assessment				
Ga68-PSMA PET/CT				
FDG PET/CT				
End of treatment				
Follow-up				
Delete all data on event:				

MIM Workflows



mim[®]
SOFTWARE

- Pre-defined sequence of actions
- Standardized way of working
- Standardized reporting
- Centralized archive

DL PET segmentation

- MATV/TTB very cumbersome

Data farming/ETL implementations



PQI



Findable,
Accessible,
Interoperable,
Reusable

Facilitate international collaboration



*Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>

THANK YOU

H.U.B



Hôpital
Erasme



H.U.B.



Thomas GUIOT
Email : thomas.guiot@hubruxelles.be