Dr Xosé M Fernández

Chief Data Officer, Institut Curie

AI and Big Data in Cancer Outcome Research

OECI2022 Oncology Days, València 16th June 2022

institut Curie

Disclosure

Roche speaker honoraria and research contract
Janssen speaker honoraria
Bristol-Myers-Squibb speaker honoraria
Novartis research contract
Leo Pharma research contract
Amgen research contract

institut Curie

Inventing Tomorrow's Cancerology Today

Chemotherapy
will be increasingly administered by mouth and will be done at home.

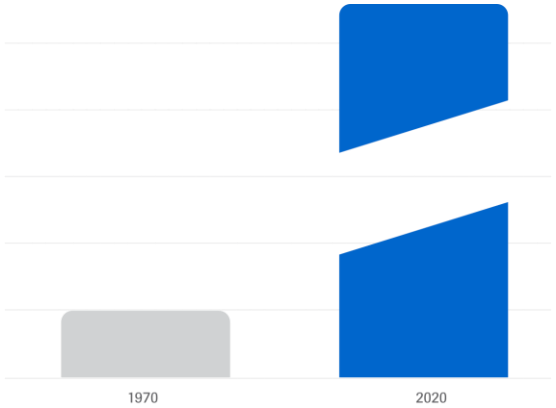
Radiotherapy
will be more targeted, less invasive, and more secure.

Interventional radiology
will replace certain complex and invasive surgeries.

Molecular biology
will allow for better tumour characterisation.

A Transformational Moment for Healthcare

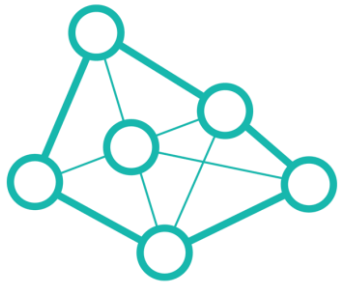
Knowledge acceleration



Medical knowledge doubled **every 50 years** in 1950.

Today, it doubles **every 72 days.**

Disease complexity

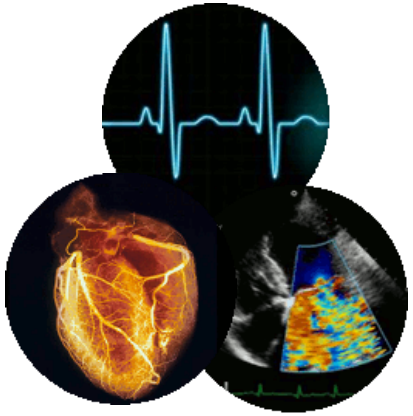


There are **+200** tumour types, which can have up to **1.2 million mutations**.

Diagnostic complexity

We are witnessing an overlap between disciples. Which leads to increased complexity when dealing with **multimodal data**.

Multimodal EHR



Diagnostic exams
(RIS, PACS)

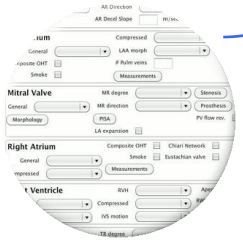
Genomic Data

Patient family
history

Lab Results

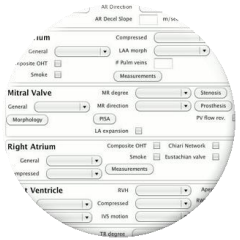
Pathology reports

Longitudinal patient record



Reimbursement

Prescriptions



Problem lists,
symptoms, diagnosis

Discharge reports

The Data Flood

60%

Images
- Radiographic imaging
- 2D/3D ultrasound
- Electron microscopy
- 4D microscopy
- Histopathology
- Single-cell imaging…

5%

EHRs
- Dossier médicale personnel
- Dossier Communiquant en Cancérologie

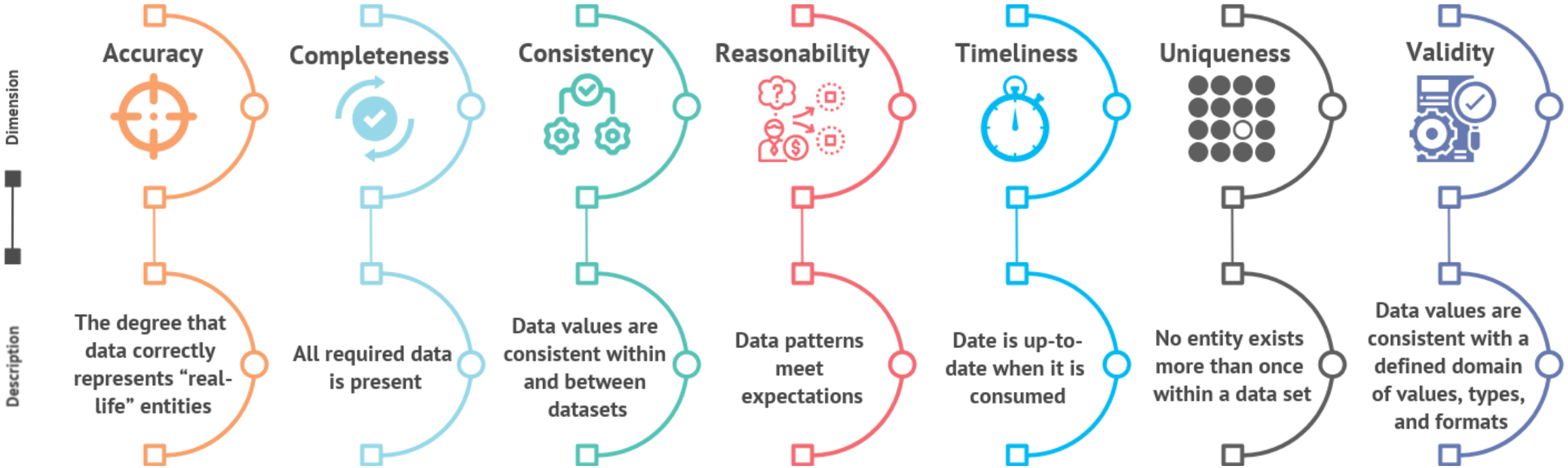
20%

Genomic data
- France Médecine Génomique 2025

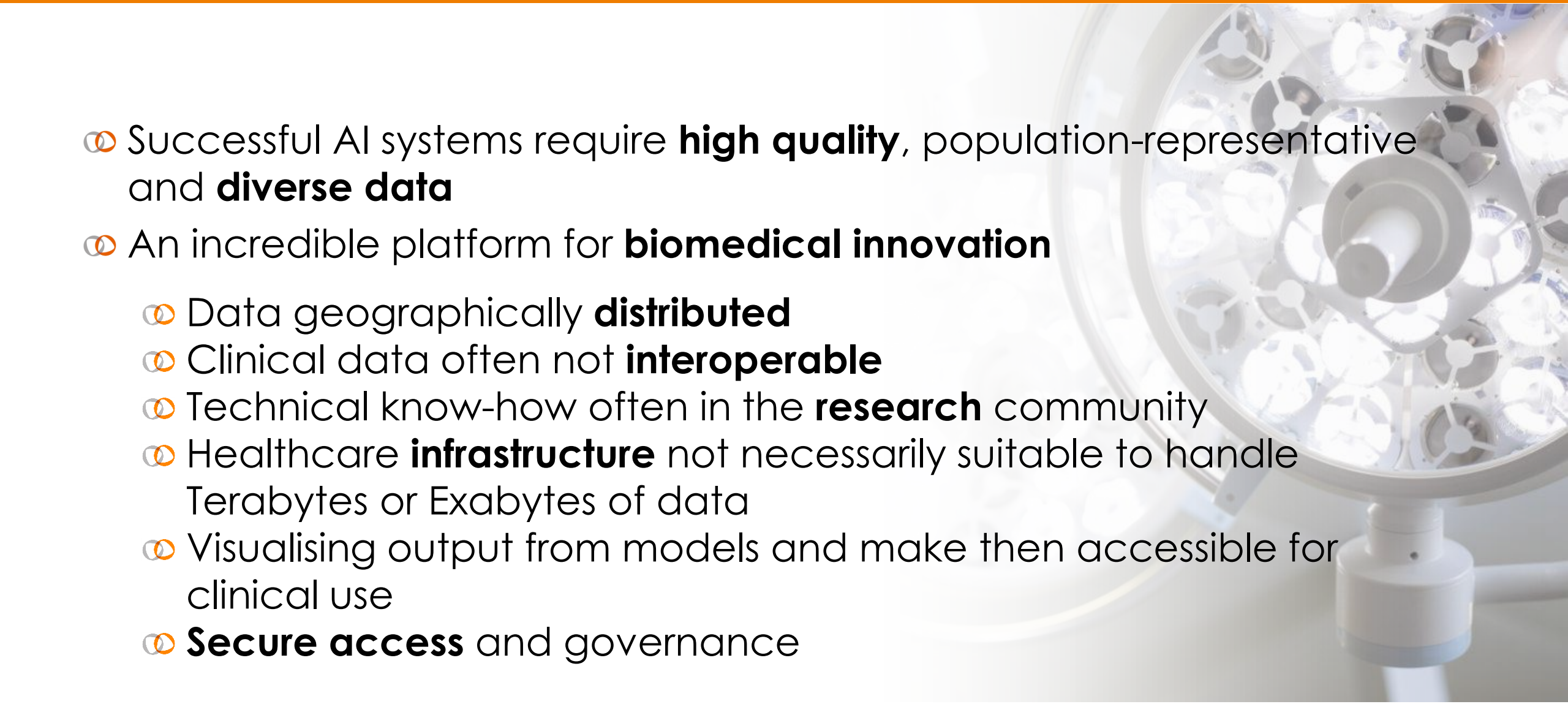
15%

Research data
- Clinical trials
- Phenotypes
- Various research databases and registries…

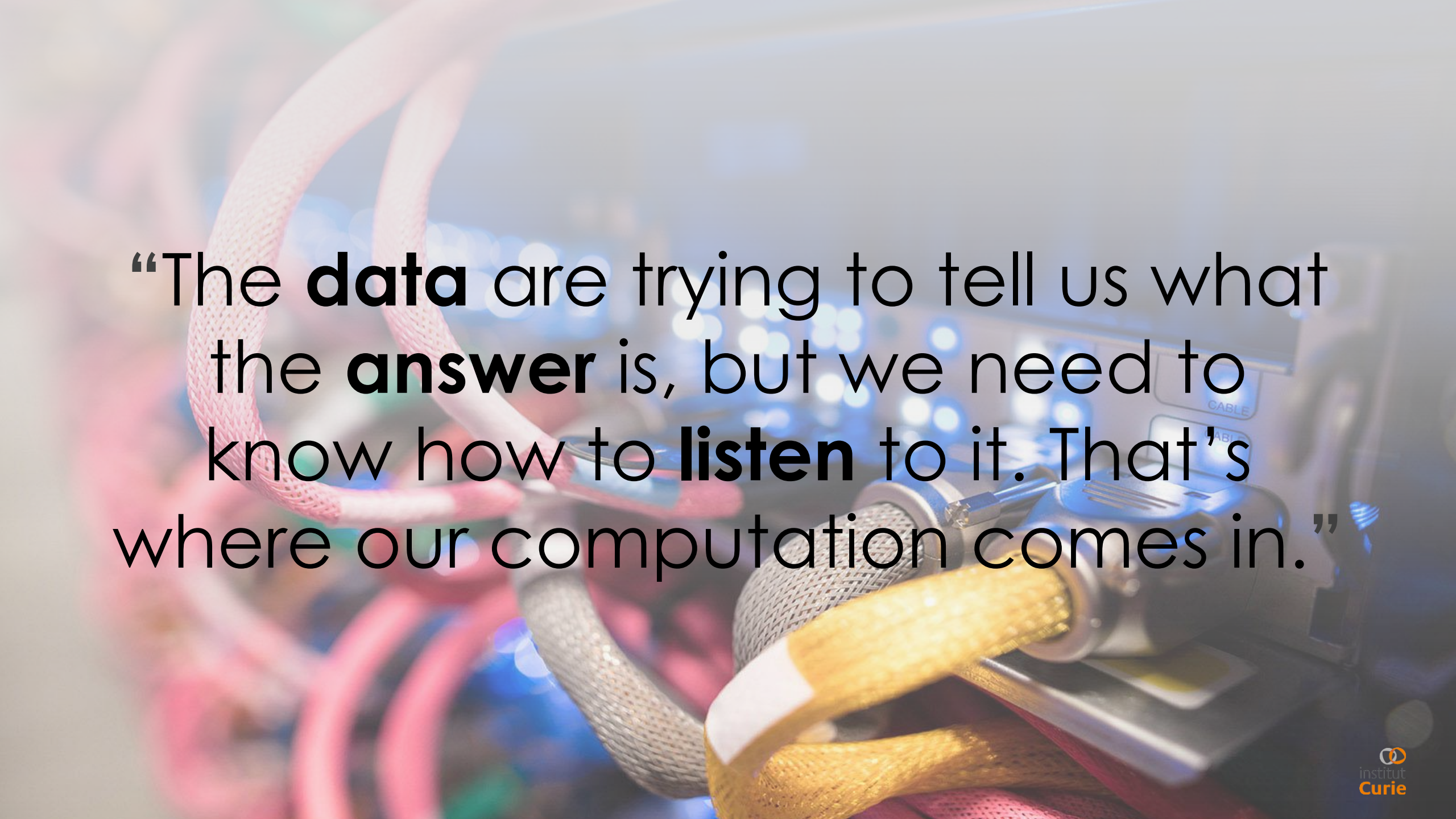
Dimensions of Data Quality



AI in Healthcare



- Successful AI systems require **high quality**, population-representative and **diverse data**
- An incredible platform for **biomedical innovation**
  - Data geographically **distributed**
  - Clinical data often not **interoperable**
  - Technical know-how often in the **research** community
  - Healthcare **infrastructure** not necessarily suitable to handle Terabytes or Exabytes of data
  - Visualising output from models and make then accessible for clinical use
  - **Secure access** and governance


“The **data** are trying to tell us what the **answer** is, but we need to know how to **listen** to it. That’s where our computation comes in.”
institut
Curie

Three Questions that every Cancer Patient Asks

A core ethical principle of medicine since **Hippocrates**, the imperative to learn from current patients

- Who's like me?
  - **DIAGNOSIS**
- How long have I got?
  - **PROGNOSIS**
- What are my options?
  - **PREDICTION**

Consore – Natural Language Processing

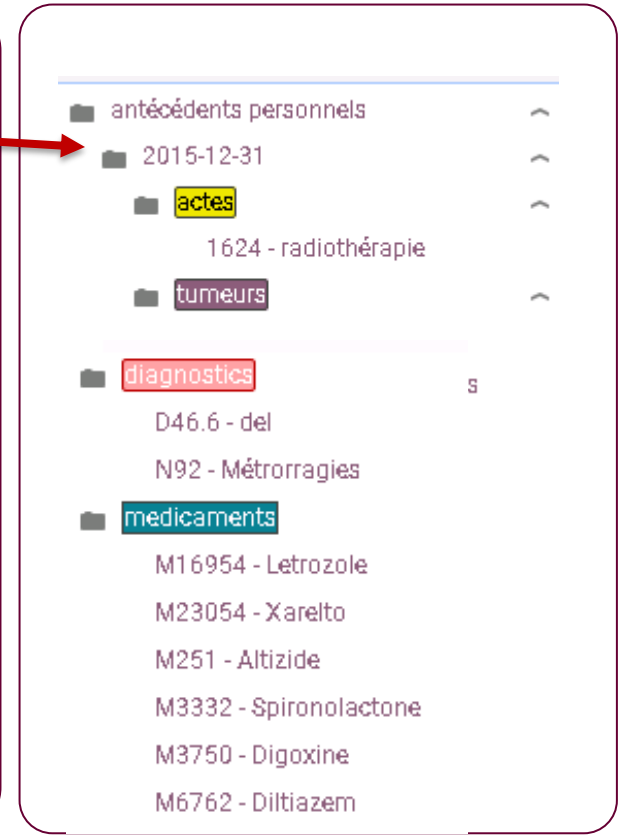
- Identifying medical concepts
- Timeline

CCI mammaire gauche : mastectoie + curage axillaire + radiothérapie + hormonothérapie en 2015.
cancer anal traité par radiothérapie

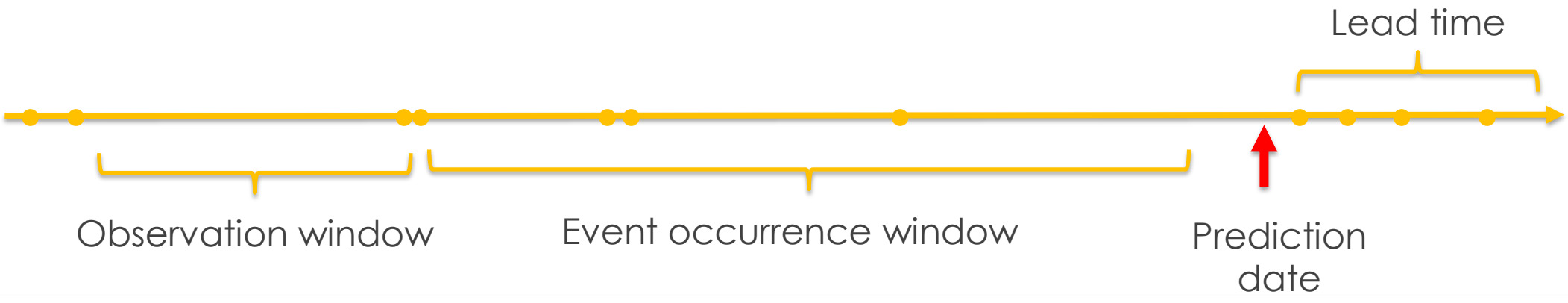
Antécédents familiaux :
soeur cancer du sein
mère et tante cancer utérin

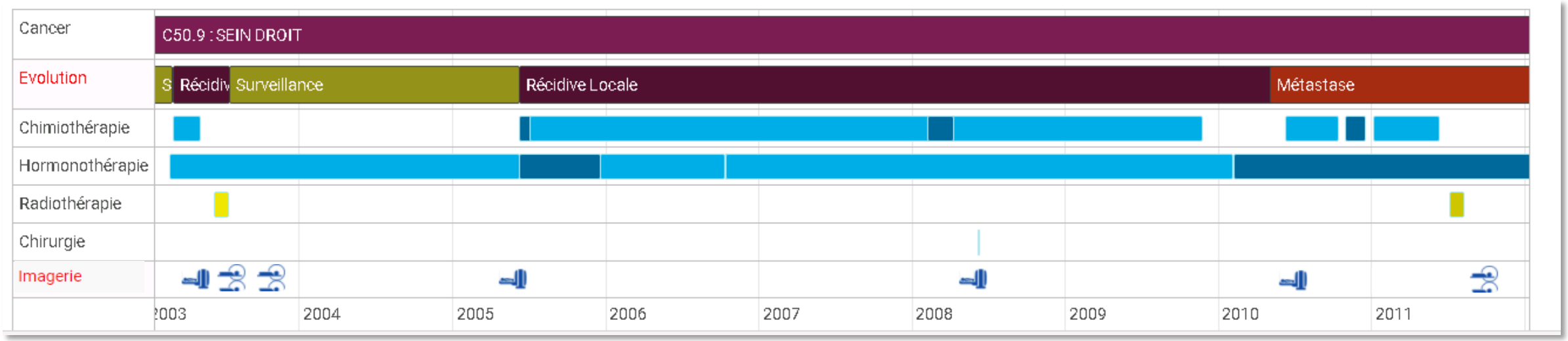
Traitement actuel Xarelto - Altizide - Spironolactone - Diltiazem - Digoxine - Rivarosaban - Letrozole.

Métrorragies sous Xarelto.
Examen : lésion bourgeonnante occupant les 2/3 inférieurs del a paroi latérale droite du vagin jusqu'à la fourchette vulvaire ulcérée, pigmentée avec atteinte de la paroi postérieure du vagin jusqu'au 2/3 moyen et latérale gauche sur 2-3 cm.
Biopsie vaginale : localisation d'un mélanome (Breslow 1,4 mm)



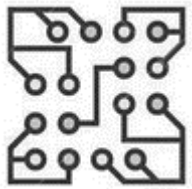
Patient Timeline and Associated Data Gathering





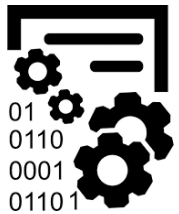
institut Curie

Advanced ML beyond Standard Statistical Approaches



More variables

Patients can be described with thousands to millions of variables, from diverse data types



Capturing interactions

Non-linear machine learning models can reveal complex interactions between variables

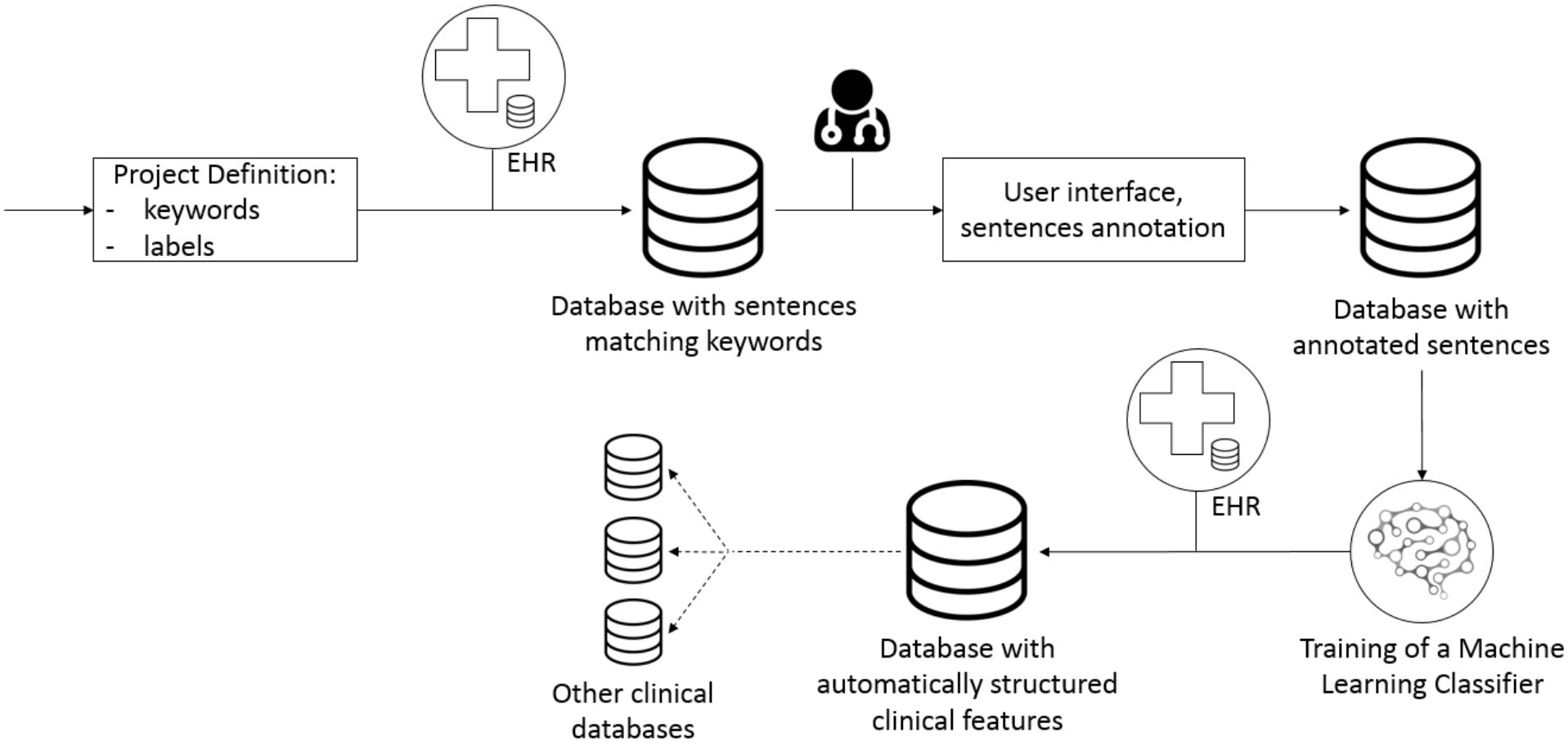
Clinical history

We can identify patterns in the clinical history of each patient, not just instantaneous snapshots

Unstructured data

Working with structured and unstructured data (capable of integrating EHR, free text and medical imaging)

Champollion – Automatic Structuration of Clinical Features from EHR



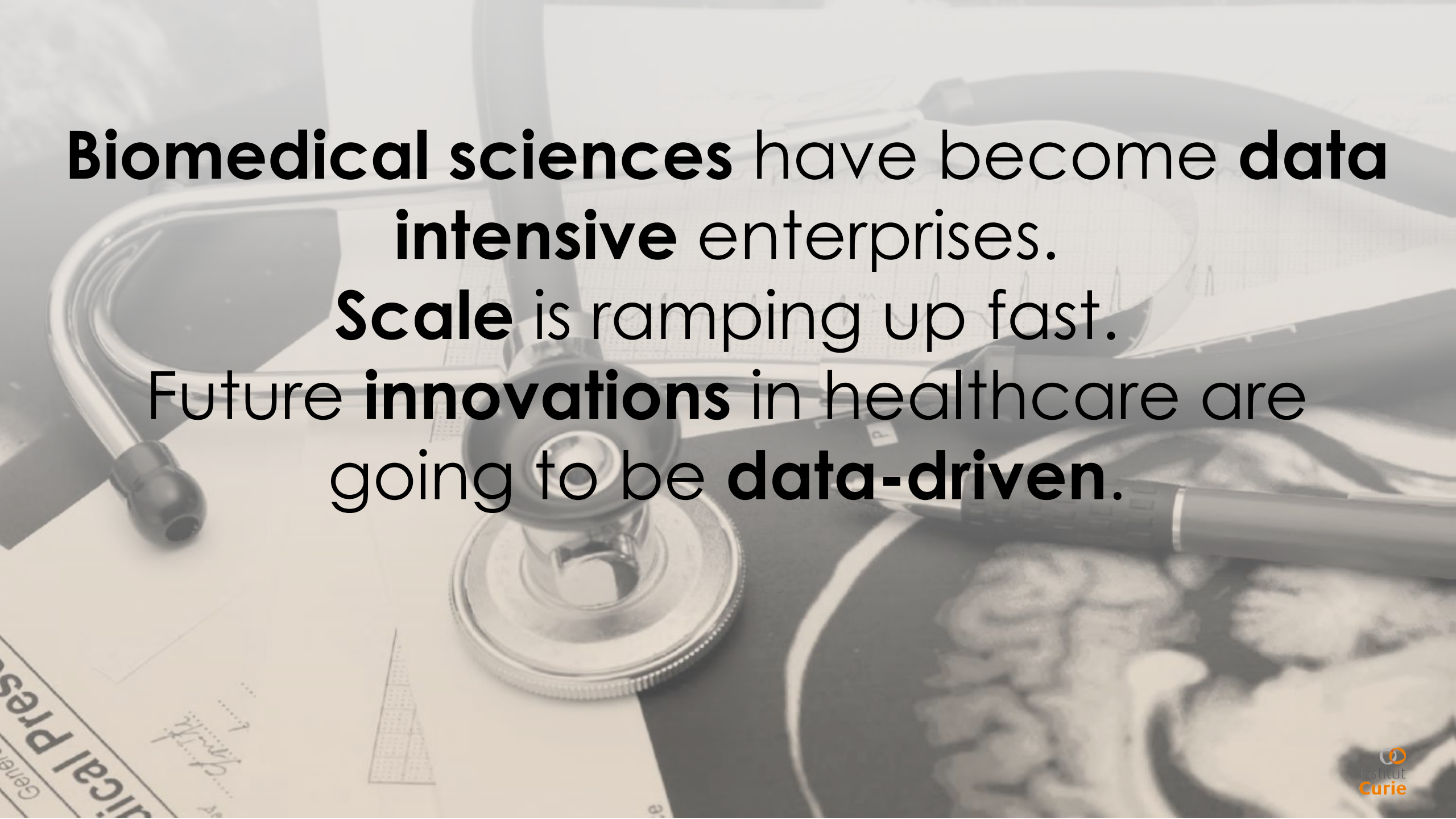
Hijano *et al*. (2021) doi: 10.1007/978-3-030-77211-6_54

institut Curie

Champollion – Evaluation of performance of ML Classfiers

Project	Labels	# annotated sentences	Score
Inflammatory Cancer Assessing whether a tumour is inflammatory or not	**Yes, no, n/a**	1846	**95%**
Genetic Mutations Mutation status of a given a set of genes whose mutations (or absence thereof) are known to impact the evolution of the tumour	**Mutated, not mutated, n/a**	1031	**93%**
PD-L1 Immunohistochemistry Percentage of PD-L1 immunohistochemistry	**Percentage in the text, n/a**	508	**97%**
Performance Status Described using the Zubrod scale	**Score 0-4, n/a**	480	**98%**

Hijano *et al*. (2021) doi: 10.1007/978-3-030-77211-6_54



Biomedical sciences have become **data intensive** enterprises.
Scale is ramping up fast.
Future **innovations** in healthcare are going to be **data-driven**.

Institut Curie

AI & Medical Imaging

Abnormality detection

Lung
Early detection of lung cancer is associate with improved outcomes

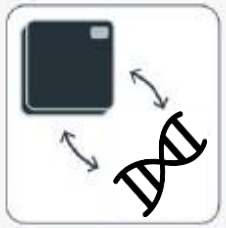
Breast
More robust screening mammography interpretation and analysis

Characterisation of suspected lesions

Segmentation
Defining the extent of an abnormality in terms of 2D or full 3D assessments

Staging
Categorising tumours into predefined groups based on expected course & treatment strategies

Diagnosis
Classifying abnormalities as benign or malignant



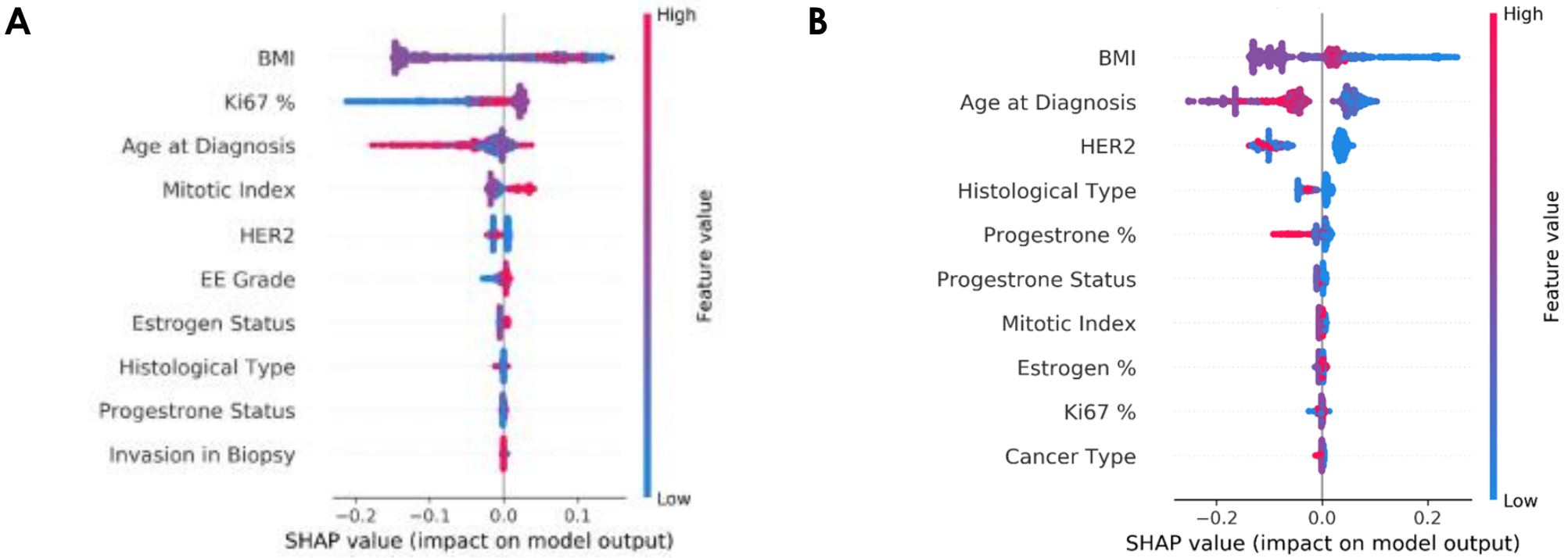
Imaging Genomics
Associating imaging features with genomics data for comprehensive tumour characterisation

Monitoring over time

Change analysis
Temporal monitoring of tumour changes either in natural history or in response to treatment

- Highlighting suspicious regions in images
- Detecting indeterminate nodules
- Addressing high- false-positive rates and over diagnosis
- Providing robust tumour descriptions to capture intra-tumour heterogeneity and variability
- Capturing a large number of discriminative features that go beyond those measures by traditional evaluation criteria

Multimodal AI Models for Predicting Outcomes of a NAC cohort

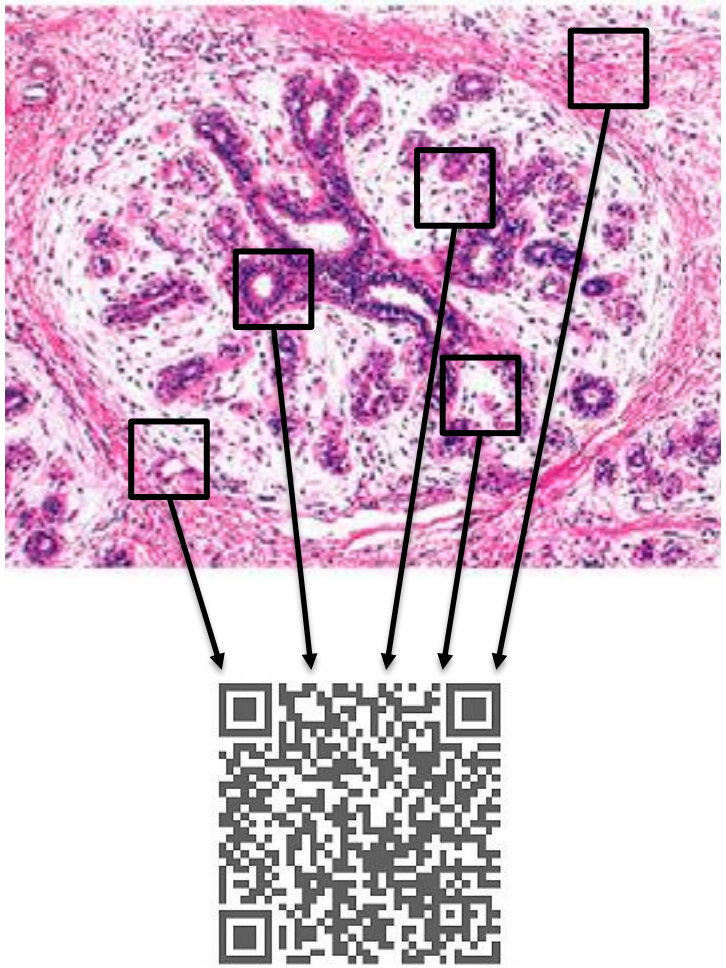


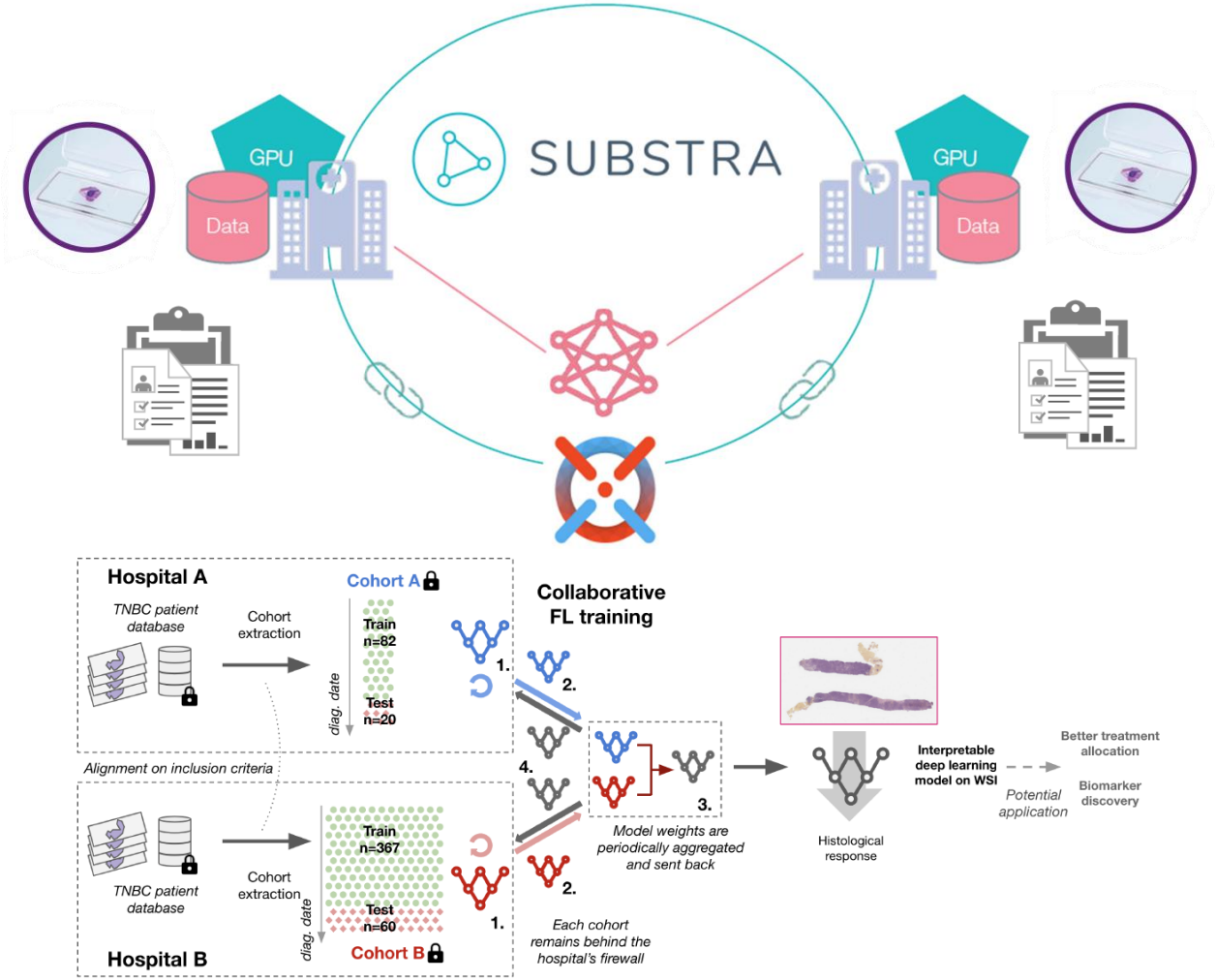
A. Prediction of **metastasis** in women with locally advanced breast cancer (0.747 AUC)
B. Multimodal prediction of **five-year** breast cancer **recurrence** in NAC treated women (AUC>0.89)

A summary plot of SHapley Additive exPlanations (SHAP) values of top features in the clinical model. The figure depicts, in descending order, the top features that had the most influence on the multimodal prediction. Each point represents a single patient. The *X-axis* indicates the **effect** (either positive or negative) **of the feature** on the predicted score for the patient. The point's colour represents the **value** of the features (red = high value, blue = low value).

Rabinovici-Cohen *et al* (2022) doi: 10.1117/12.2613169

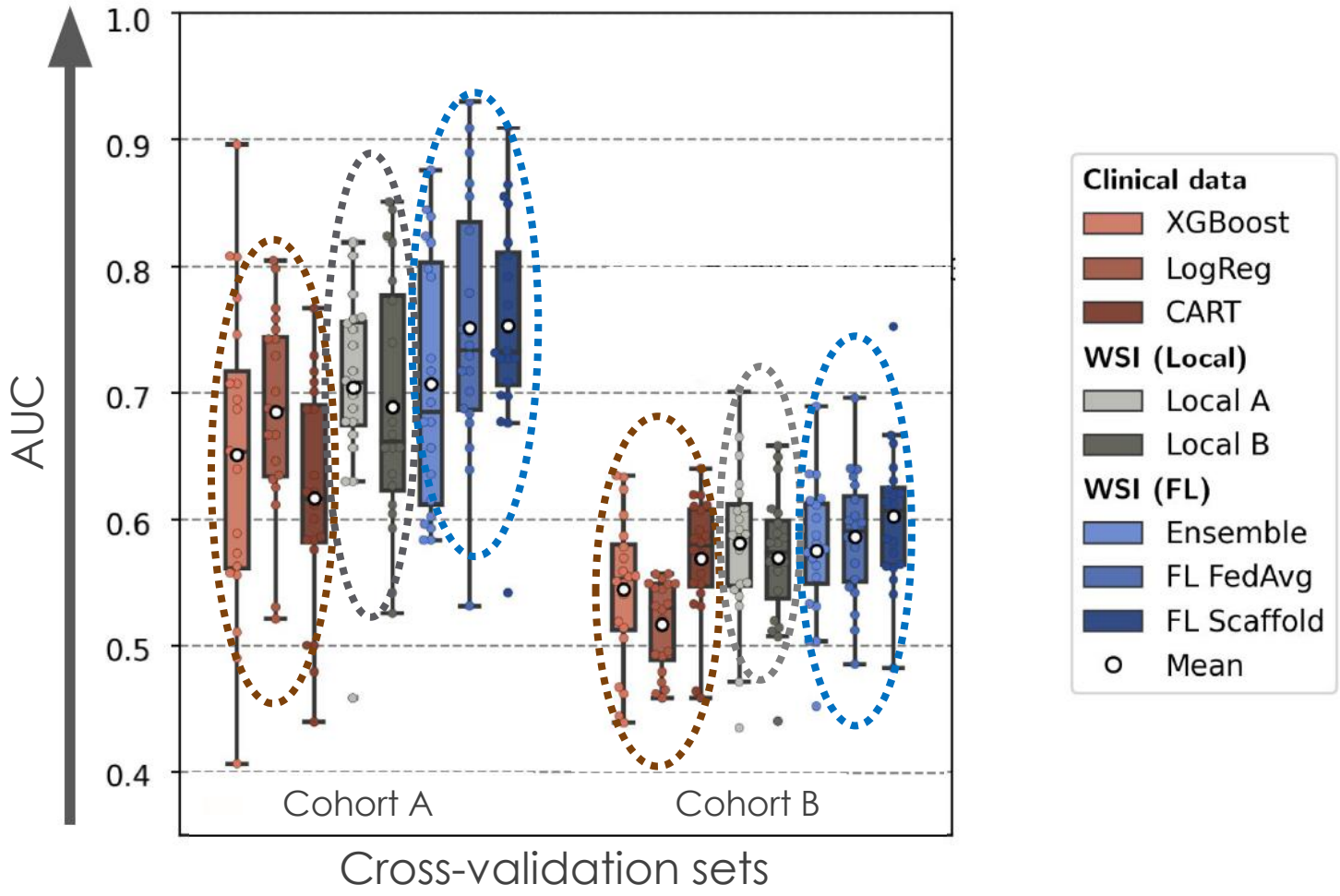
Digital Pathology – Collaborative Federated Learning





institut Curie

Quantitative Results in Cross-validation Sets

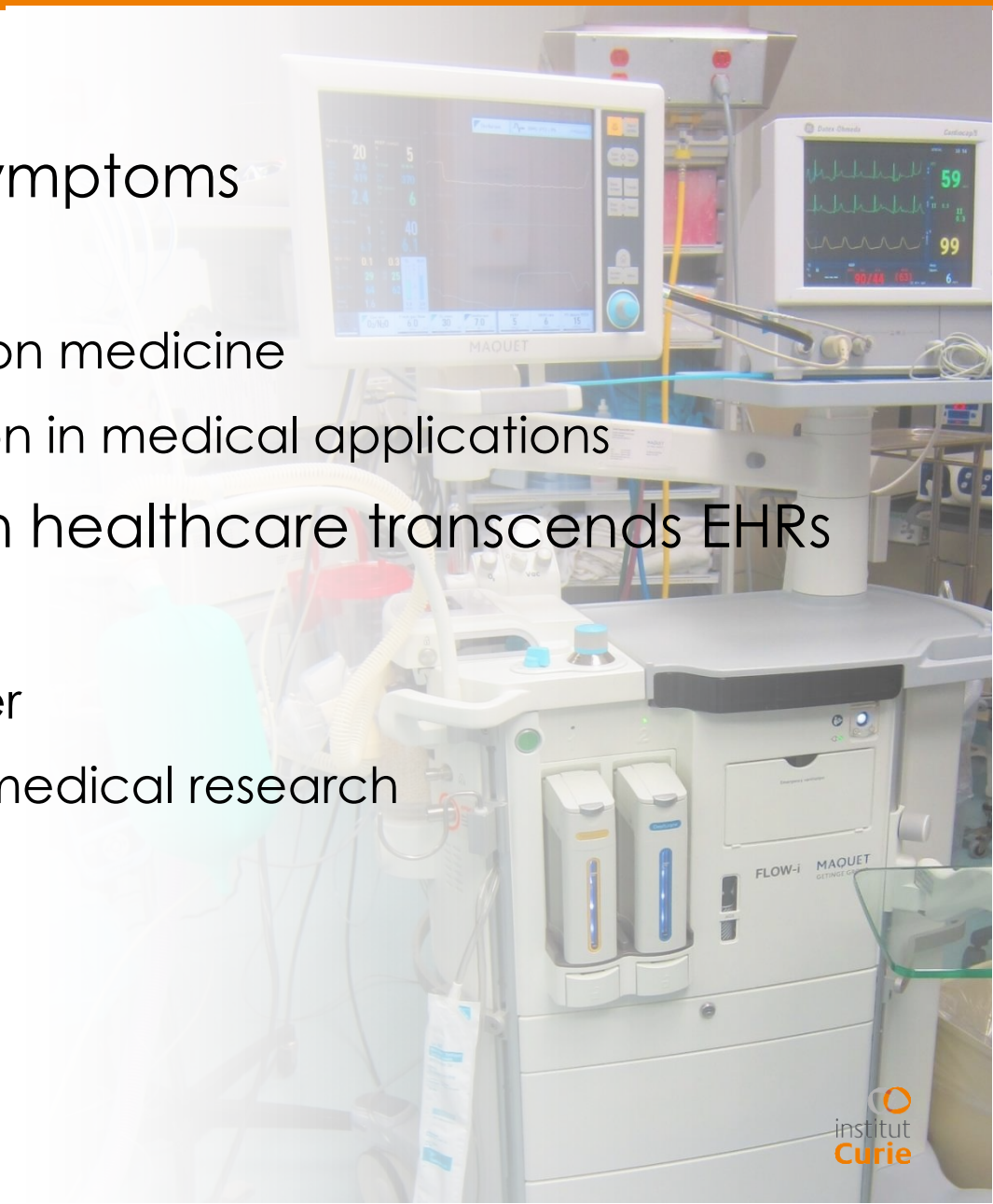


Ogier du Terrail *et al.* (2021) doi: 10.1101/2021.10.27.21264834

institut Curie

Big Data and Cancer Precision Medicine

- **Understanding illness** and **health** beyond symptoms
  - A detailed map of medicine at molecular level
  - Genomics is only one of *many* sources in precision medicine
    - However we are witnessing a rapid expansion in medical applications
- **Knowledge networks** (so-called *big data*) in healthcare transcends EHRs
  - Unifying medicine and data
  - Information-based approach to describe cancer
  - A framework for integrating and leveraging biomedical research

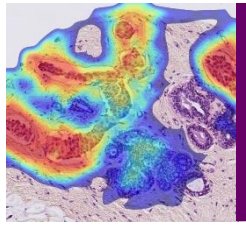


Toward Precision Medicine, US National Academy of Sciences (2011)
Personalised Medicine for the European Citizen, European Science Foundation (2012)

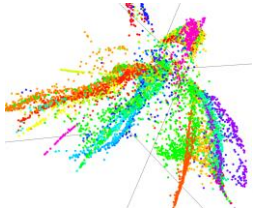
"**digital technology** is no longer in the cordoned-off domain of **IT**; it is being applied to almost **every part** of a company's value chain"

Digital Doesn't Have to Be Disruptive (Furr & Shipilov), *Harvard Business Review* 2019

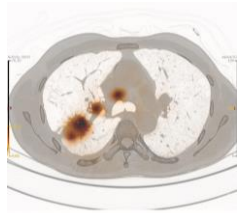
Enhancing Collaboration Opportunities



Digital Pathological Analysis

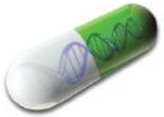


Data Visualisation Networks

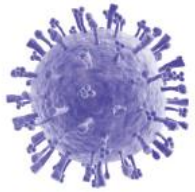


Digital Imaging Analysis

Biomarkers



Immunotherapy



Text mining (EHR)

Genomics

Data Science

Laetitia Chanas, Linda Ammoura-Haroun, Paulette Feghali-Salamoun, Raïna Ikonomova, Eve Lapouble, Maud Milder, Hurera Mohammed
Julien Guérin, Thomas Balezeau, Jessica Henao, Aurelien Legros, Armand Léopold, Victor Nguyen, Pier-Francesco Rocci, Christophe Sive

Catherine Manceau

Alain Livartowski

Alumni: Élise Dumas, Matthieu Bachelot, Xiaomeng Wang, Dido Carrero, Juliette Tran Lu Y

I am among those who think that science has great beauty

Marie Skłodowska-Curie

Gràcies

Thank you

xose.m.fernandez@gmail.com

@xosegb