✅ Capturing health data
- *Storing digital health data*
✅ Integrating digital heath data
✅ Sharing digital health data in science
- *Exploiting digital health data in care*

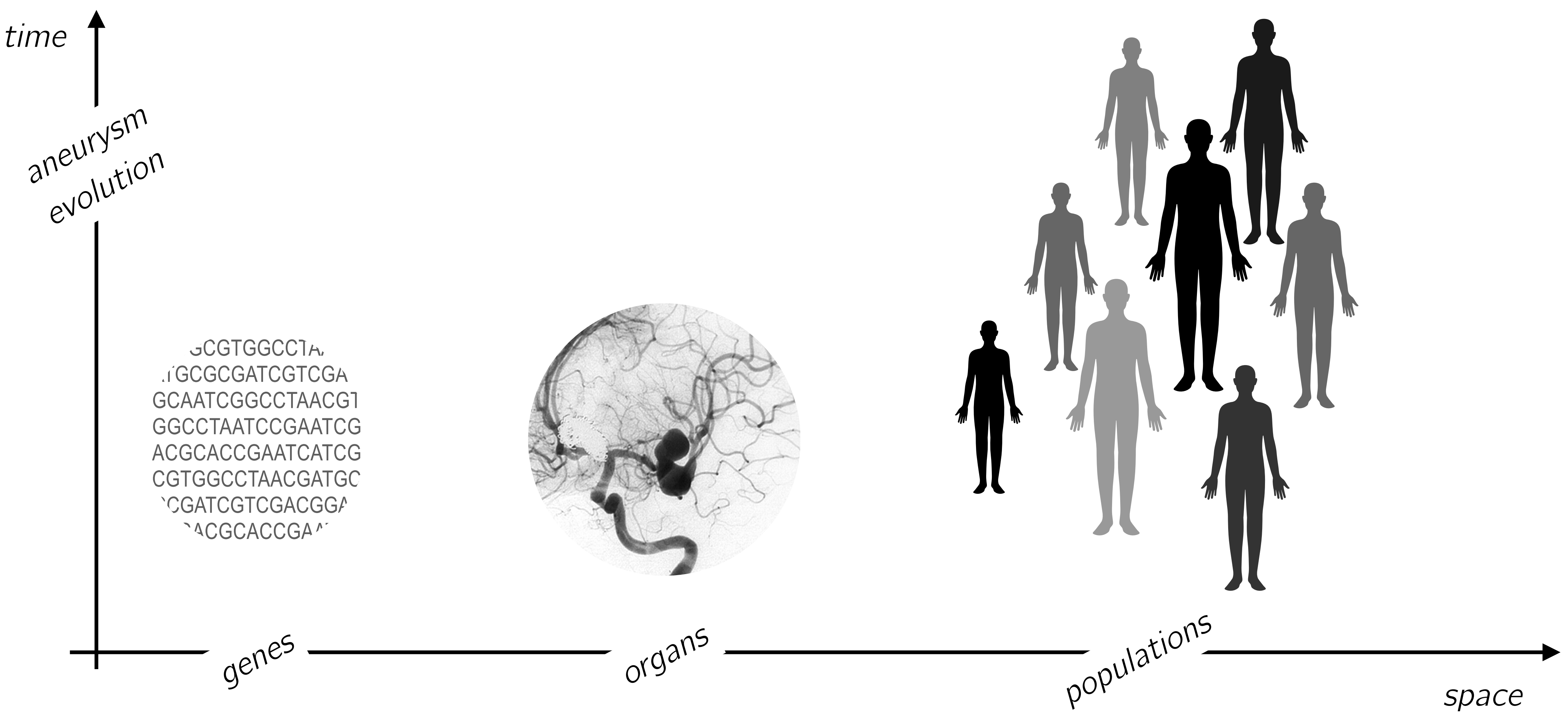
# Data integration and sharing in the context of intracranial aneurysms

Alban Gaignard

ISGC IA group premeeting session Seoul
24th of September 2025

# Multi-factorial disease → multi-scale data

- Inter-disciplinary efforts needed for a better understanding of the pathology
- Specific data produced at very specific scales



# Neurovasc project

- Neurovasc: a national programme funded for 4 years by the french research agency to build a digital infrastructure to manage and exploit intracranial aneurysm data
  - 3 Research Institutes (Inria, Inserm, IMT Atlantique)
  - 2 Clinical Research Teams (Brest & Nantes academic hospitals)
  - 3 Universities (Bordeaux, Paris-Saclay, Nantes)

**WP 1: A model of interoperable infrastructure between research and healthcare**

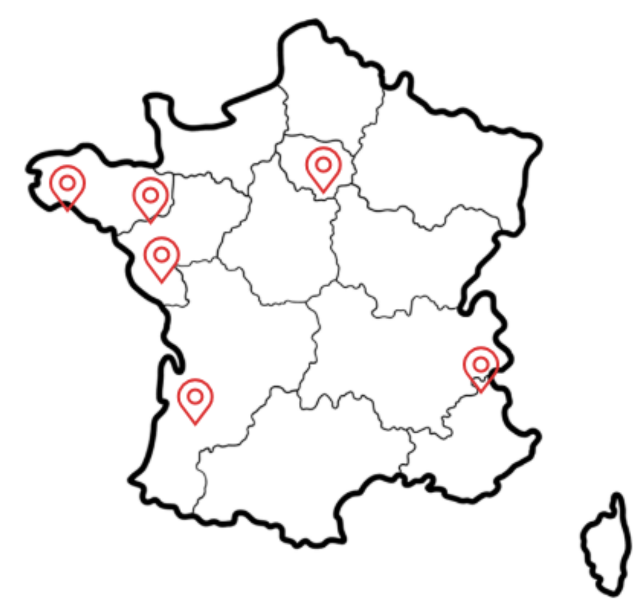
- Task 1.1. Managing clinical data
- Task 1.2. Managing imaging data
- Task 1.3. Managing genetic data

**WP 2: Interoperable datasets and predictive models for ICA diagnosis and outcomes**

- Task 2.1: FAIR genomic data demonstrator
- Task 2.2: Mining healthcare circuits following ICA diagnosis
- Task 2.3: A non-additive model for global genetic-risk prediction

**WP 3: Proof-of-Concept studies with digital companions**

- Task 3.1: Accompanying patients with diagnosed unruptured ICA
- Task 3.2: Post-stroke prevention through mobile applications and digital monitoring

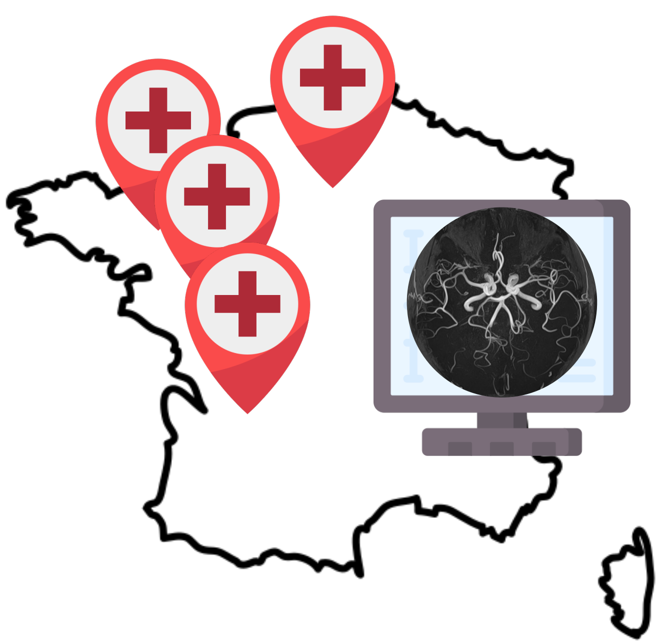


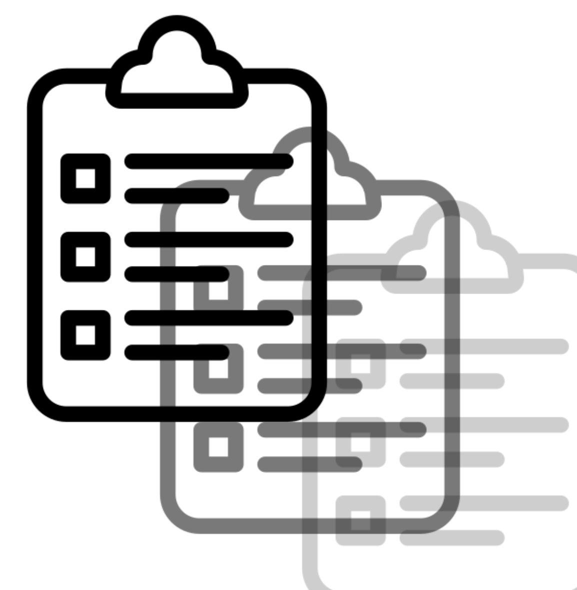
# Multiple data integration sharing issues

➊ How to **collect high-quality medical images** from multiple hospitals/MRIs ?

➋ How to **interlink and query multi-modal and multi-scale data** while preserving privacy constraints ?

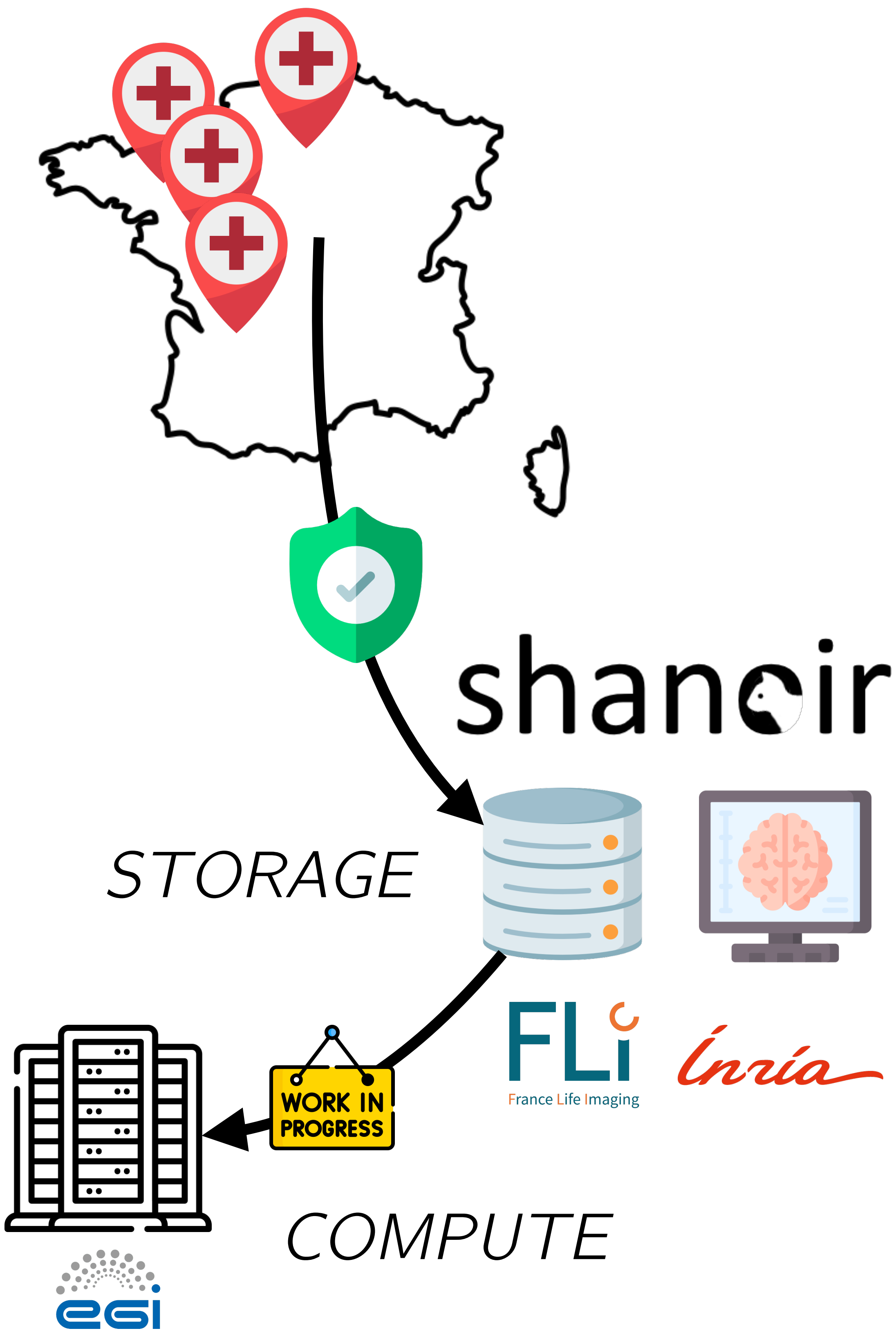
➌ How to mine and **model patient trajectories** from eHR data ? can we predict clinical outcomes ?

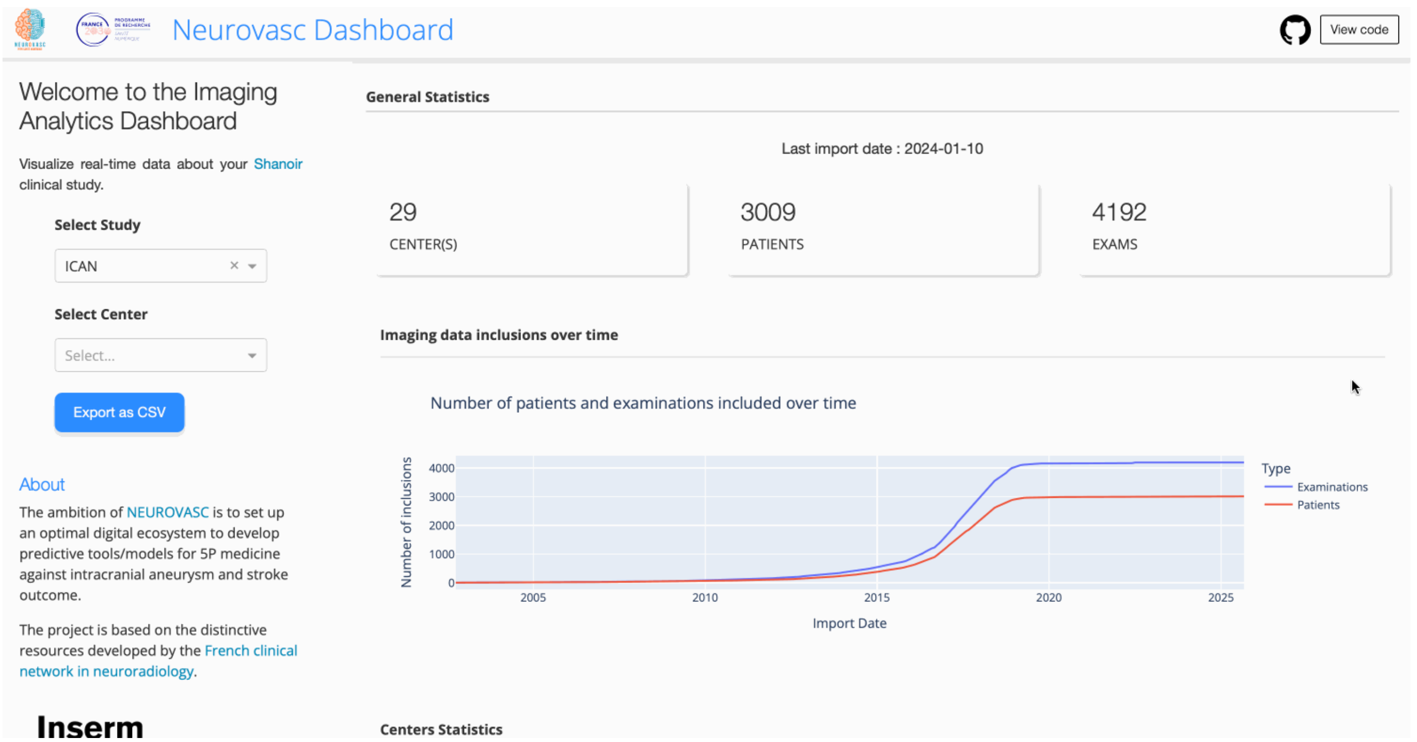


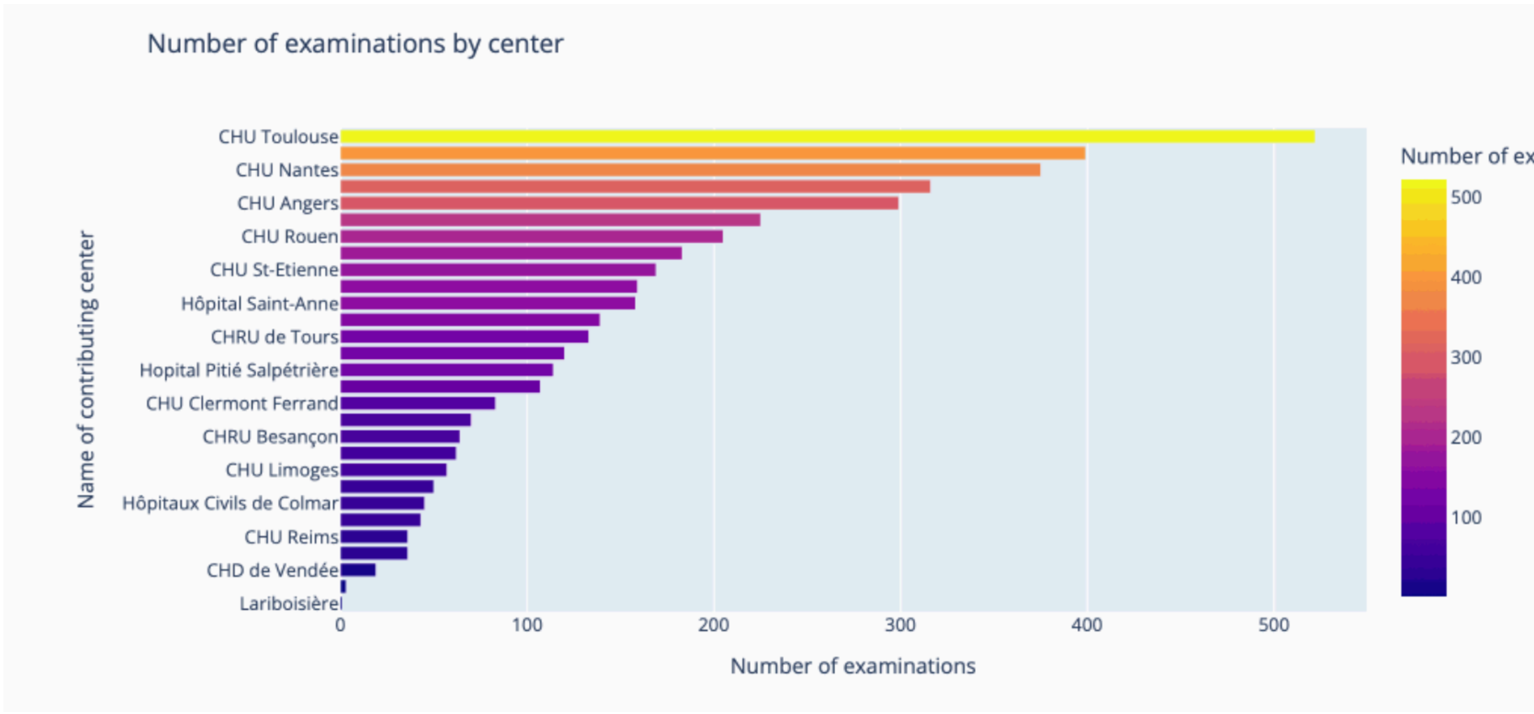


❶ How to collect high-quality medical images from multiple hospitals/MRIs ?

# ❶ Collecting multi-source medical images

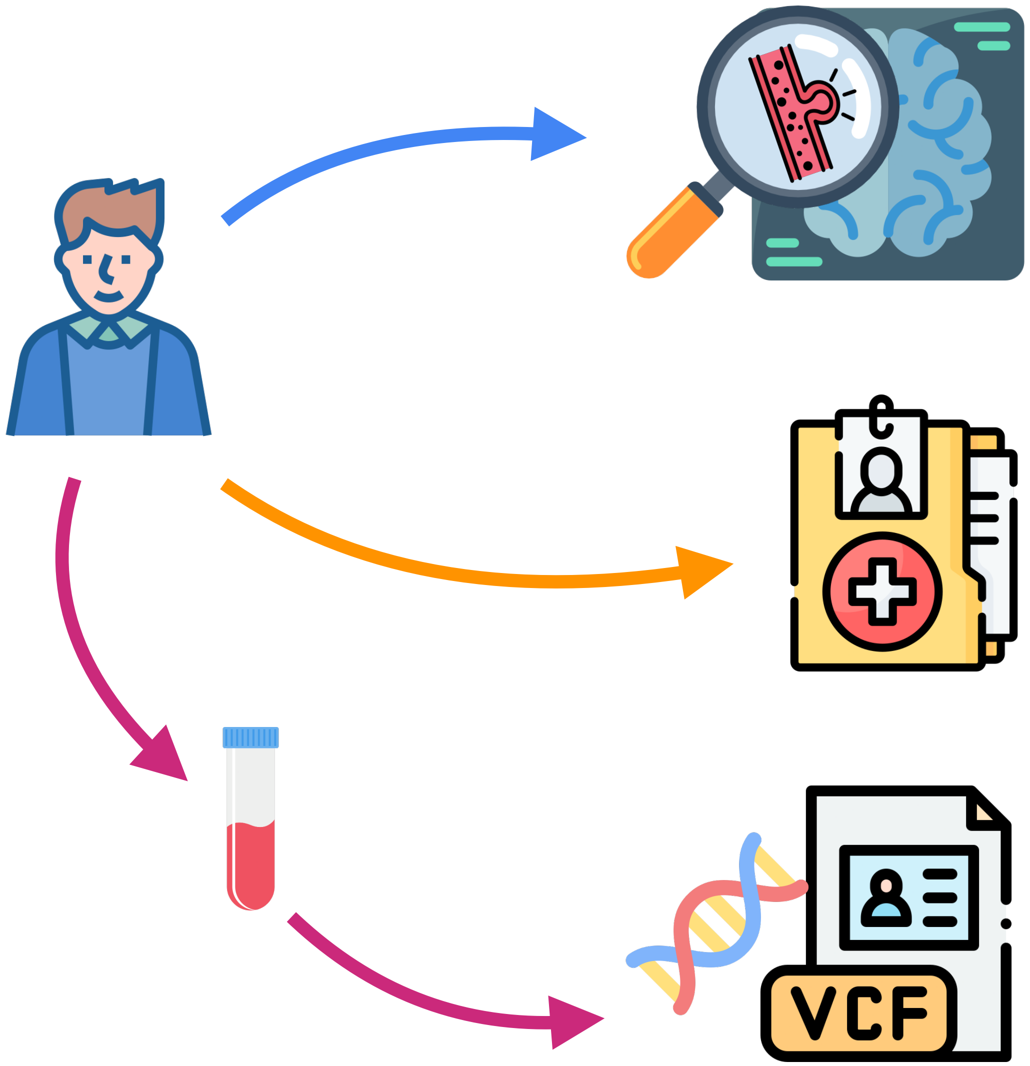


❷ How to interlink and make query-able multi-modal and multi-scale data while preserving privacy constraints ?

# ❷ FAIRifying clinical & genomic data




Neuro-vascular imaging / tissues ?

- UBERON
- NCIT

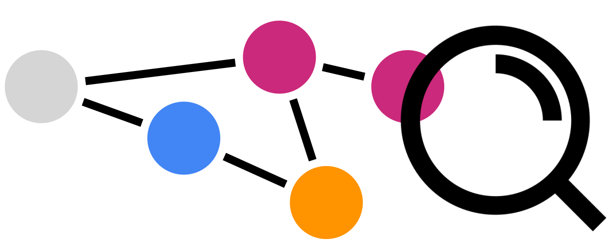
Clinical data / phenotypes ?

- SPHN
- HPO
- DUO

Genomic data ?

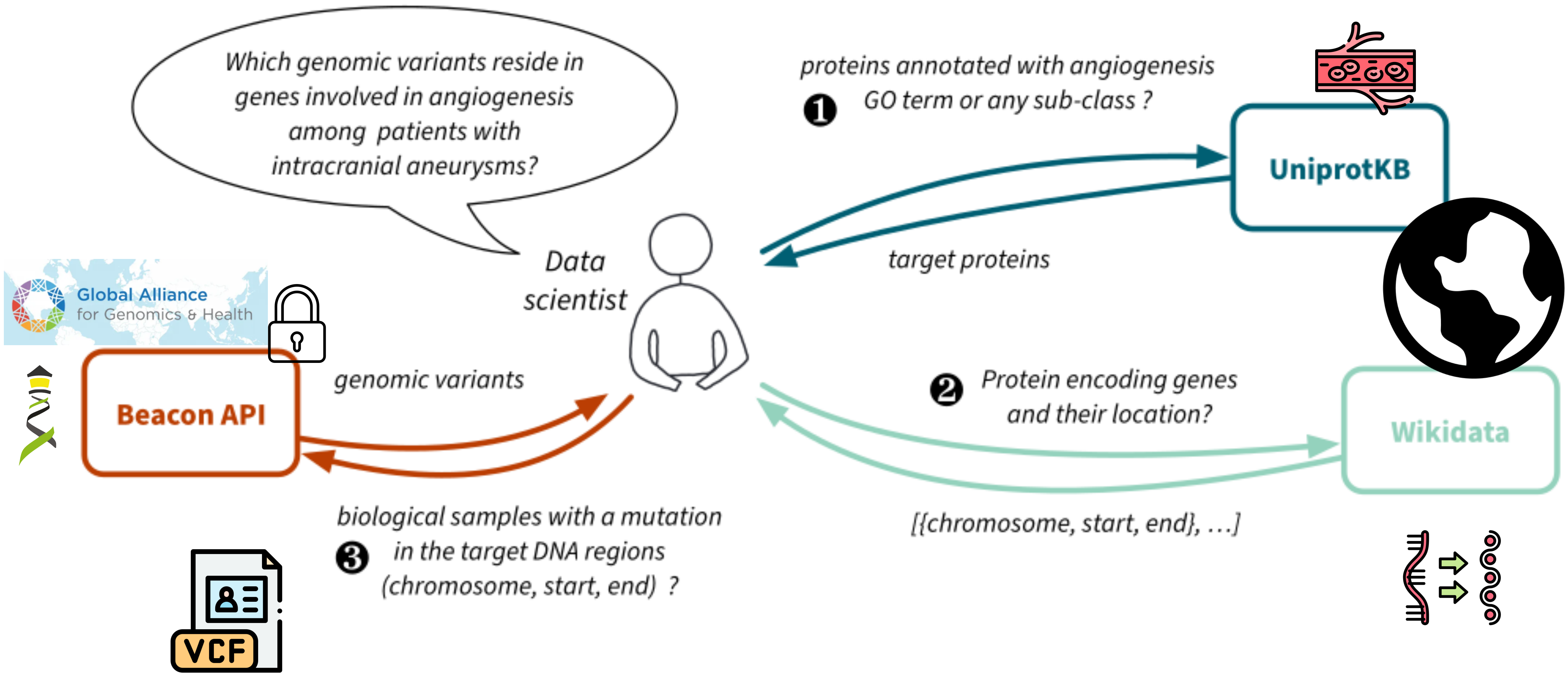
- FALDO
- SO / GENO
- SIO

**A clinical and genomic intracranial aneurysm knowledge graph**



Find/Exchange
phenoytypes - variants
with refererence terminologies !

# ❷ Making genomic data interoperable with public knowledge bases




→ **A single fedetared SPARQL query over 3 data sources (1 non-RDF)**

Alexandrina Bodrug-Schepers, Hugo Chabane, Gabriela Montoya, Richard Redon, Alban Gaignard, et al.. **Semantic Beacons: a framework to support federated querying over genomic variants and public Knowledge Graphs**. SWAT4HCLS 2025, Feb 2025, Barcelona, Catalonia, Spain. ⟨hal-04908530v2⟩

❸ How to mine and model patient trajectories from eHR data ? can we predict clinical outcomes ?

## ❸ Evaluating clinical data models for patient outcome prediction

- Predict outcome of intracranial aneurysm patients after certain medical procedures or treatments.


*Figures from Wikipedia*

# Synthetic Clinical Data

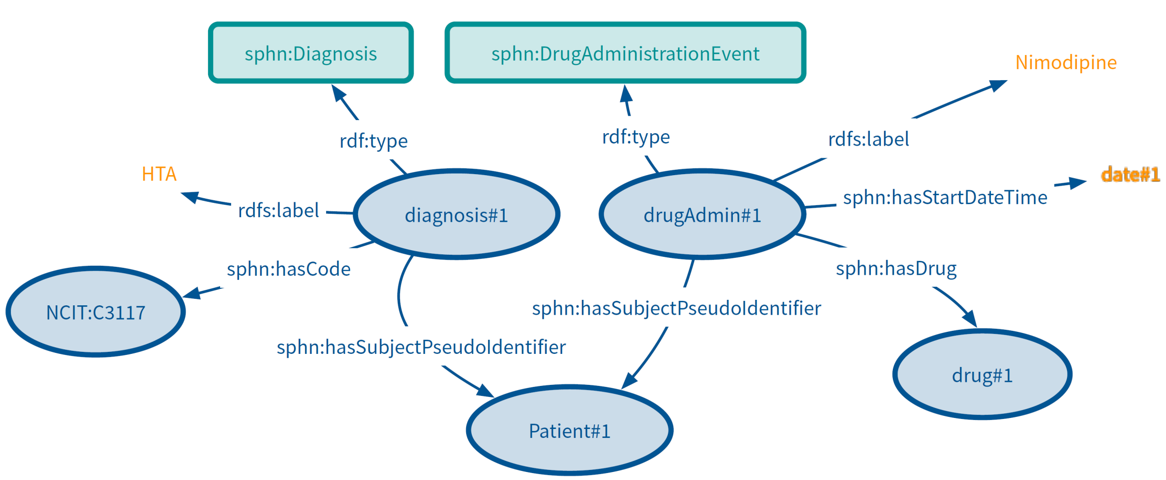
- We generated **synthetic dataset** that follows the distribution and the correlations of variables from **real-world clinical data** provided by the Nantes University Hospital.
  - Synthetic tabular data of intracranial aneurysm (10,000 individuals)
  - 22 non-temporal features (4 numerical, 6 categorical, 12 binary)
  - 8 temporal features (e.g., drug administration)

| ID | Hospital stay length | Age | Gender | Entry unit | ICA | Diabetes | O2 clinic | Nimodipine | Paracetamol | Outcome |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 41.09 | 38 | 0 | 0 | 0 | 0 | 0 | 2023-04-08 | 0 | 0 |
| 2 | 21.70 | 58 | 0 | 1 | 2 | 0 | 0 | 2022-12-09 | 2022-12-09 | 1 |
| 3 | 4.63 | 76 | 0 | 2 | 2 | 0 | 0 | 0 | 2021-05-23 | 0 |
| 4 | 12.83 | 87 | 1 | 1 | 4 | 0 | 1 | 0 | 0 | 2 |
| 5 | 75.68 | 75 | 0 | 3 | 5 | 0 | 1 | 2020-11-23 | 0 | 1 |
| 6 | 41.95 | 59 | 0 | 2 | 5 | 0 | 0 | 2019-07-19 | 2019-07-26 | 0 |
| 7 | 27.54 | 42 | 0 | 4 | 5 | 0 | 0 | 0 | 0 | 0 |

# Evaluating clinical data models for patient outcome prediction

- The structure of SPHN is more **patient-centric** compared to CARE-SM, which is more **diagnosis-centric**

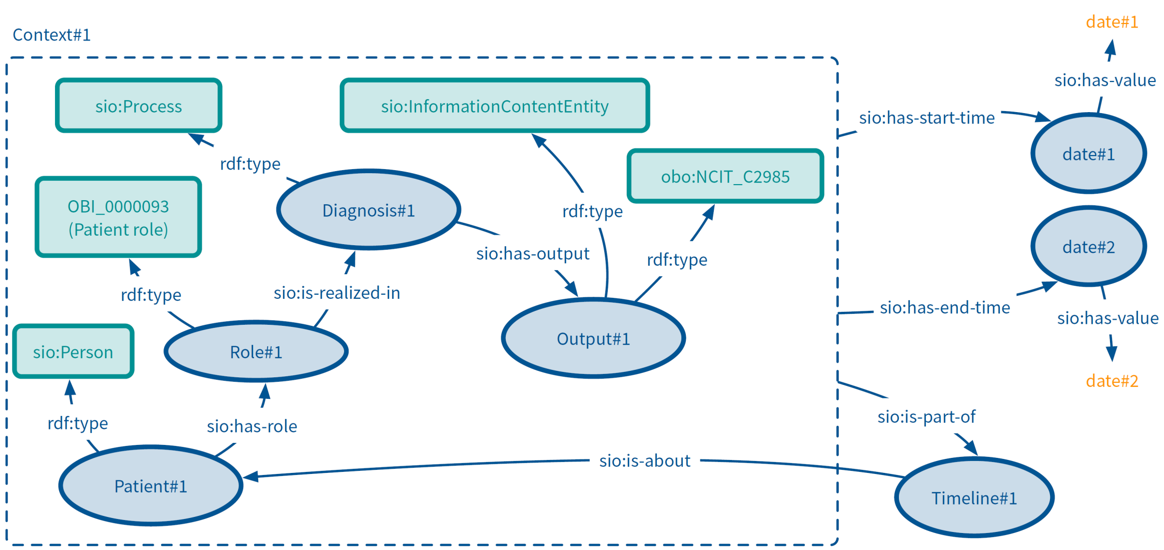
**SPHN**
**(Swiss Personalized Health Network)**




*Adopted by the five Swiss academic hospitals for better data sharing and integration*

*Touré, V. et al. (2023)*

**CARE-SM**
**(Care and Registry Semantic Model)**




*Initially designed to represent clinical data in the context of rare diseases*

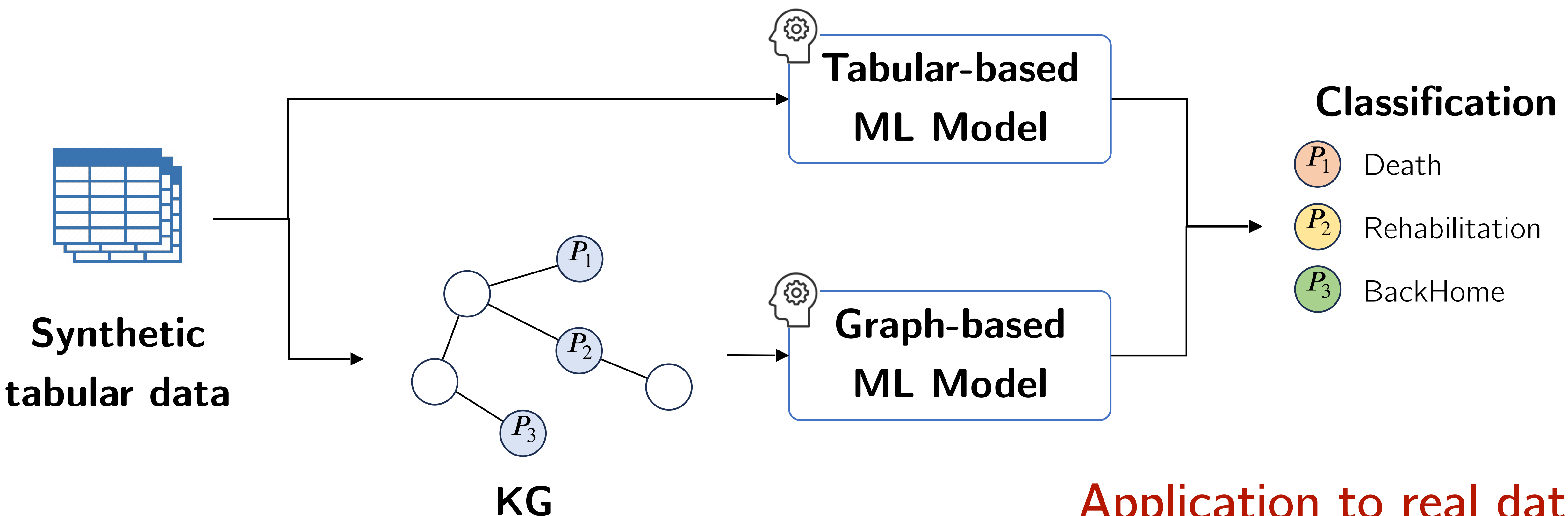
*Kaliyaperumal, R. et al. (2022)*

# Predict outcome of intracranial aneurysm patients after certain medical procedures or treatments.



## Key questions:

- Tabular data or Knowledge graph (KG)?
- What is the best KG structure for prediction tasks?
- How to represent temporal information in the KG?




**Application to real data ?**


Jong Ho Jhee, Alberto Megina, Pacôme Constant Dit Beaufils, and Matilde Karakachoff, Richard Redon, Alban Gaignard, Adrien Coulet, **Predicting Clinical Outcomes from Patient Care Pathways Represented with Temporal Knowledge Graphs**, ESWC 2025, https://doi.org/10.1007/978-3-031-94575-5_16

# Acknowledgments

Alexandrina
Bodrug

Laurent
Vallet

Richard
Redon

Adrien
Coulet

Matilde
Karakachof

Romain
Bourcier

Pacôme
Constant dit Beaufils

**Small Animal Shanoir (SAS) A Cloud-Based Solution for Managing Preclinical MR Brain Imaging Studies**

Michael Kain[1]*, Marjolaine Bodin[2], Simon Loury[2], Yao Chi[1], Julien Louis[1], Mathieu Simon[1], Julien Lamy[3], Christian Barillot[1] and Michel Dojat[2]*

[1] INRIA U1228, INSERM, Université de Rennes, Rennes, France, [2] INSERM U1216, Grenoble Institut des Neurosciences, Université Grenoble Alpes, CHU Grenoble Alpes, Grenoble, France, [3] ICube, University of Strasbourg-CNRS, Strasbourg, France

Clinical multicenter imaging studies are frequent and rely on a wide range of existing tools for sharing data and processing pipelines. This is not the case for preclinical

https://doi.org/10.3389/fninf.2020.00020

**Semantic Beacons: a framework to support federated querying over genomic variants and public Knowledge Graphs**

Alexandrina Bodrug-Schepers[1,†], Hugo Chabane[2,†], Gabriela Montoya[2], Patricia Serrano-Alvarado[2], Richard Redon[1] and Alban Gaignard[1,3]

[1] Nantes Université, CNRS, INSERM, l'institut du thorax, F-44000 Nantes, France

[2] Nantes Université, LS2N, Nantes, France

[3] IFB-core, Institut Français de Bioinformatique (IFB), CNRS, INSERM, INRAE, CEA, 91057 Evry, France

https://hal.science/hal-04908530v2

ARTICLE

**Predicting Clinical Outcomes from Patient Care Pathways Represented with Temporal Knowledge Graphs**

Authors: Jong Ho Jhee, Alberto Megina, Pacôme Constant Dit Beaufils, Matilde Karakachof Redon, Alban Gaignard, Adrien Coulet Authors Info & Claims

The Semantic Web: 22nd European Semantic Web Conference, ESWC 2025, Portoroz, Slovenia, June 1–5, 2025 Proceedings, Part I
Pages 282 - 300 • https://doi.org/10.1007/978-3-031-94575-5_16

Published: 01 June 2025 Publication History

https://doi.org/10.1007/978-3-031-94575-5_16

Inserm
cnrs
Nantes Université
CHU NANTES

contact: alban.gaignard@univ-nantes.fr

l'institut du thorax

# Backup slides

# Results 1: Patient Outcome Prediction

## Tabular vs. TKG (temporal knowledge graphs)

- Compared to tabular data, TKG showed better performance with RGCN3+lit.

| **Type** | **Model** | **F1-score** BackHome | Rehab | Death | Macro | Weighted | **Accuracy** | **AUC** |
|---|---|---|---|---|---|---|---|---|
| Tabular | LR | 0.63±0.02 | 0.55±0.02 | 0.25±0.05 | 0.47±0.03 | 0.55±0.02 | 0.56±0.02 | 0.70±0.02 |
| | RF | 0.63±0.01 | 0.55±0.02 | 0.28±0.04 | 0.49±0.02 | 0.55±0.01 | 0.56±0.01 | 0.71±0.01 |
| | NN | 0.58±0.03 | 0.48±0.03 | 0.26±0.04 | 0.44±0.02 | 0.50±0.02 | 0.50±0.02 | 0.63±0.02 |
| Graph (`SPHN-tr`) | TransE | 0.49±0.04 | 0.40±0.10 | 0.02±0.04 | 0.30±0.03 | 0.40±0.03 | 0.43±0.02 | 0.50±0.01 |
| | RDF2Vec | 0.50±0.05 | 0.39±0.14 | 0.01±0.02 | 0.30±0.03 | 0.39±0.04 | 0.44±0.02 | 0.49±0.01 |
| | RGCN3+lit | **0.84±0.01** | **0.76±0.02** | **0.64±0.08** | **0.75±0.03** | **0.75±0.02** | **0.78±0.01** | **0.91±0.01** |

# Results 2: Patient Outcome Prediction

## Patient-centric (SPHN) vs. Diagnosis-centric (CARE-SM)

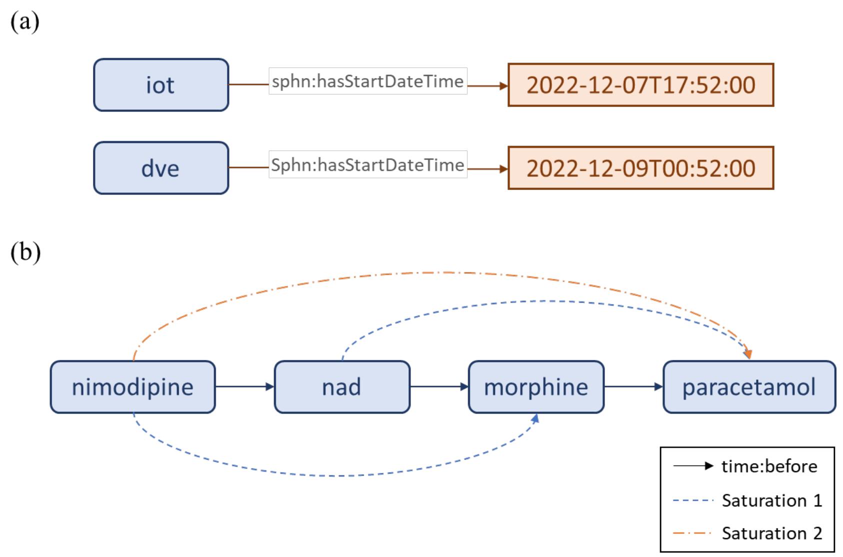
- This experiment shows the importance of the structure of a KG.

| KG | Model | F1-score | | | | | Acc. | AUC |
|---|---|---|---|---|---|---|---|---|
| | | Back2Home | Rehab. | Death | Macro | Weighted | | |
| SPHN | TransE | 0.510.07 | 0.330.16 | 0.020.04 | 0.290.04 | 0.370.05 | 0.430.02 | 0.500.01 |
| | RDF2Vec | 0.490.04 | 0.420.09 | 0.010.03 | 0.300.02 | 0.400.02 | 0.440.01 | 0.500.02 |
| | RGCN+lit | **0.830.01** | **0.760.02** | **0.660.04** | **0.750.02** | **0.780.01** | **0.780.01** | **0.910.01** |
| CARE-SM | TransE | 0.470.04 | 0.440.04 | 0.020.03 | 0.310.01 | 0.400.01 | 0.430.01 | 0.490.01 |
| | RDF2Vec | 0.510.07 | 0.380.11 | 0.000.00 | 0.290.02 | 0.390.03 | 0.440.02 | 0.500.01 |
| | RGCN+lit | 0.530.08 | 0.300.17 | 0.000.00 | 0.280.04 | 0.370.05 | 0.440.01 | 0.500.02 |

# Results 3: Patient Outcome Prediction

## Non-Temporal vs. Temporal

- Graphs with **temporal information showed improvements** in AUC compared to non-temporal graphs.




| KG | F1-score | | | | | Accuracy | AUC |
|---|---|---|---|---|---|---|---|
| | BackHome | Rehab | Death | Macro | Weighted | | |
| `SPHN-nl` | 0.64±0.03 | 0.46±0.11 | 0.05±0.07 | 0.38±0.06 | 0.49±0.06 | 0.53±0.04 | 0.64±0.06 |
| `SPHN-nt` | 0.75±0.02 | 0.65±0.02 | 0.55±0.06 | 0.65±0.02 | 0.68±0.01 | 0.68±0.01 | 0.85±0.01 |
| `SPHN-ts` | 0.83±0.02 | **0.76±0.02** | 0.66±0.08 | 0.75±0.03 | 0.78±0.02 | 0.78±0.02 | **0.91±0.01** |
| `SPHN-tr` | **0.84±0.01** | **0.76±0.02** | 0.64±0.08 | 0.75±0.03 | 0.75±0.02 | **0.78±0.01** | **0.91±0.01** |
| `SPHN-tsr` | 0.83±0.02 | **0.76±0.02** | 0.66±0.04 | 0.75±0.02 | **0.78±0.01** | **0.78±0.01** | **0.91±0.01** |
| `SPHN-sat1` | 0.83±0.01 | **0.76±0.02** | 0.64±0.06 | 0.75±0.02 | **0.78±0.01** | **0.78±0.01** | **0.91±0.01** |
| `SPHN-sat2` | 0.83±0.01 | **0.76±0.02** | **0.68±0.05** | **0.76±0.02** | 0.78±0.02 | 0.78±0.02 | **0.91±0.01** |