

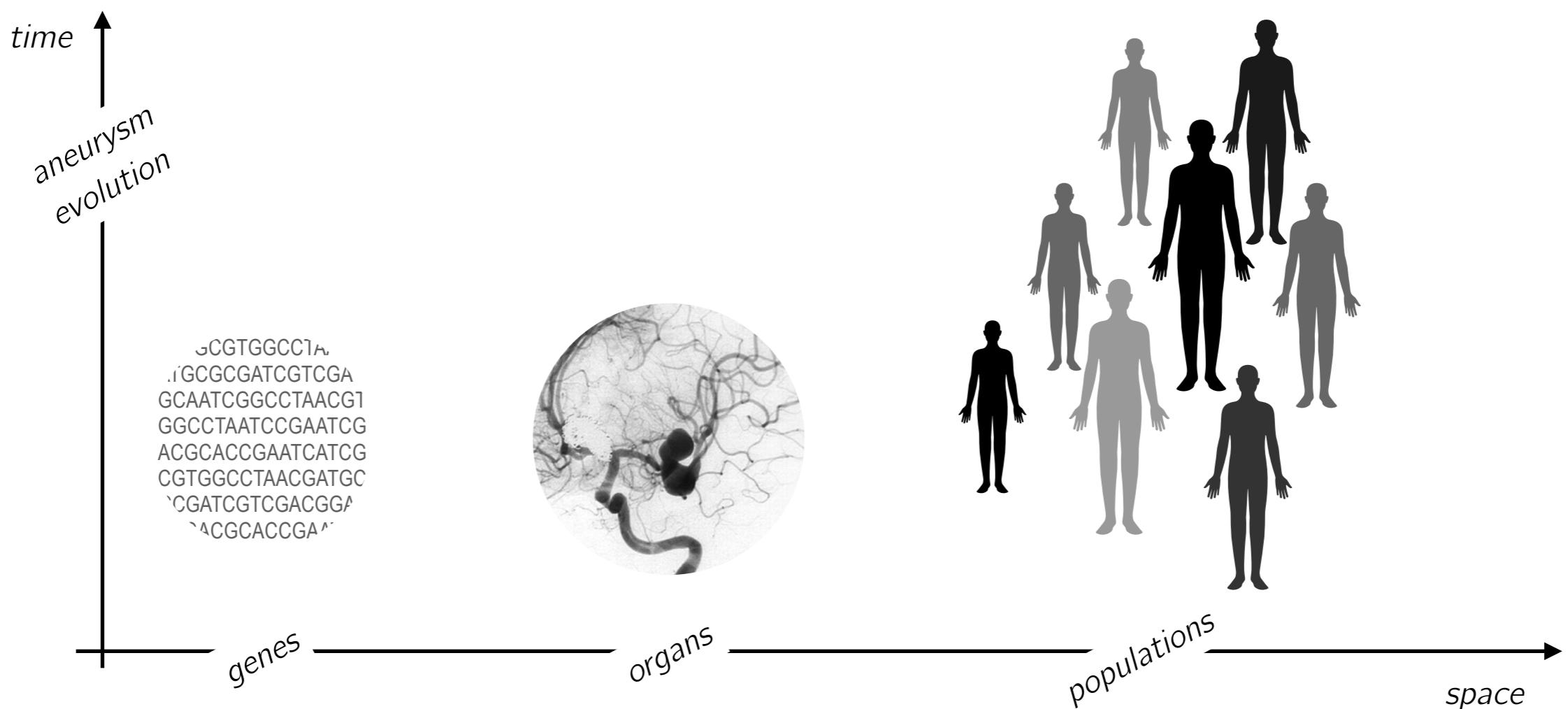
Linking and sharing multi-modal data with Knowledge Graphs, application to intracranial aneurysms

Alban Gaignard, PhD, CNRS

19th Interdisciplinary Cerebrovascular Symposium
17-19 August 2023, Geneva

multi-factorial disease → multi-scale data

- ▶ Intracranial aneurysms: a complex & multifactorial disease
- ▶ Inter-disciplinary efforts needed for a better understanding of the pathology
- ▶ Specific data produced at very specific scales



How to **jointly analyse**
these
diverse datasets ?

① what is observed with data ?

Computational ontology

« a **formal specification** of a **shared conceptualization** » (Borst, 1997)
→ 1,049 life science ontologies registered in BioPortal (2023)

Human Phenotype Ontology

Abnormality of the cardiovascular system

Abnormal vascular morphology

Abnormal vascular morphology

Abnormal systemic arterial morphology

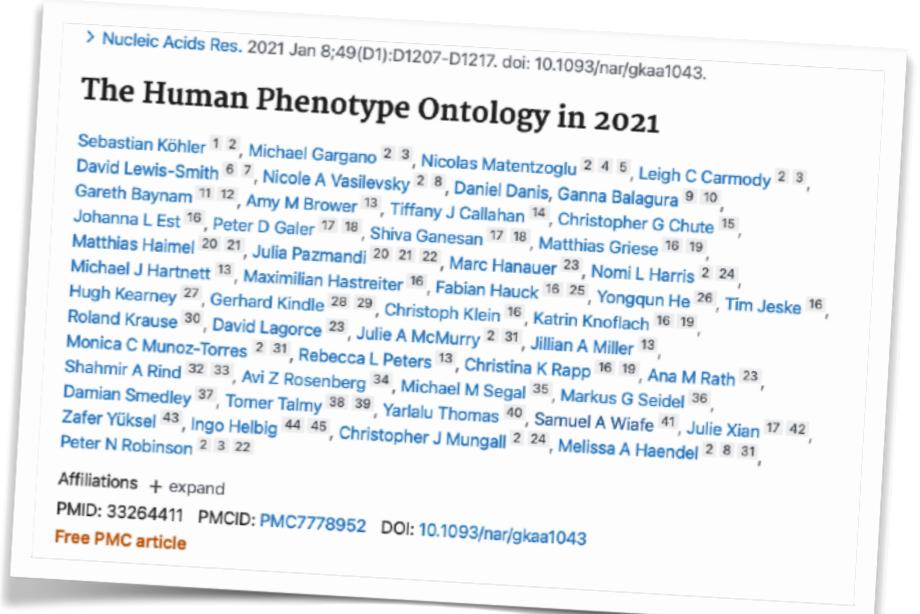
Abnormal cerebral artery morphology

Dilatation of the cerebral artery

is a

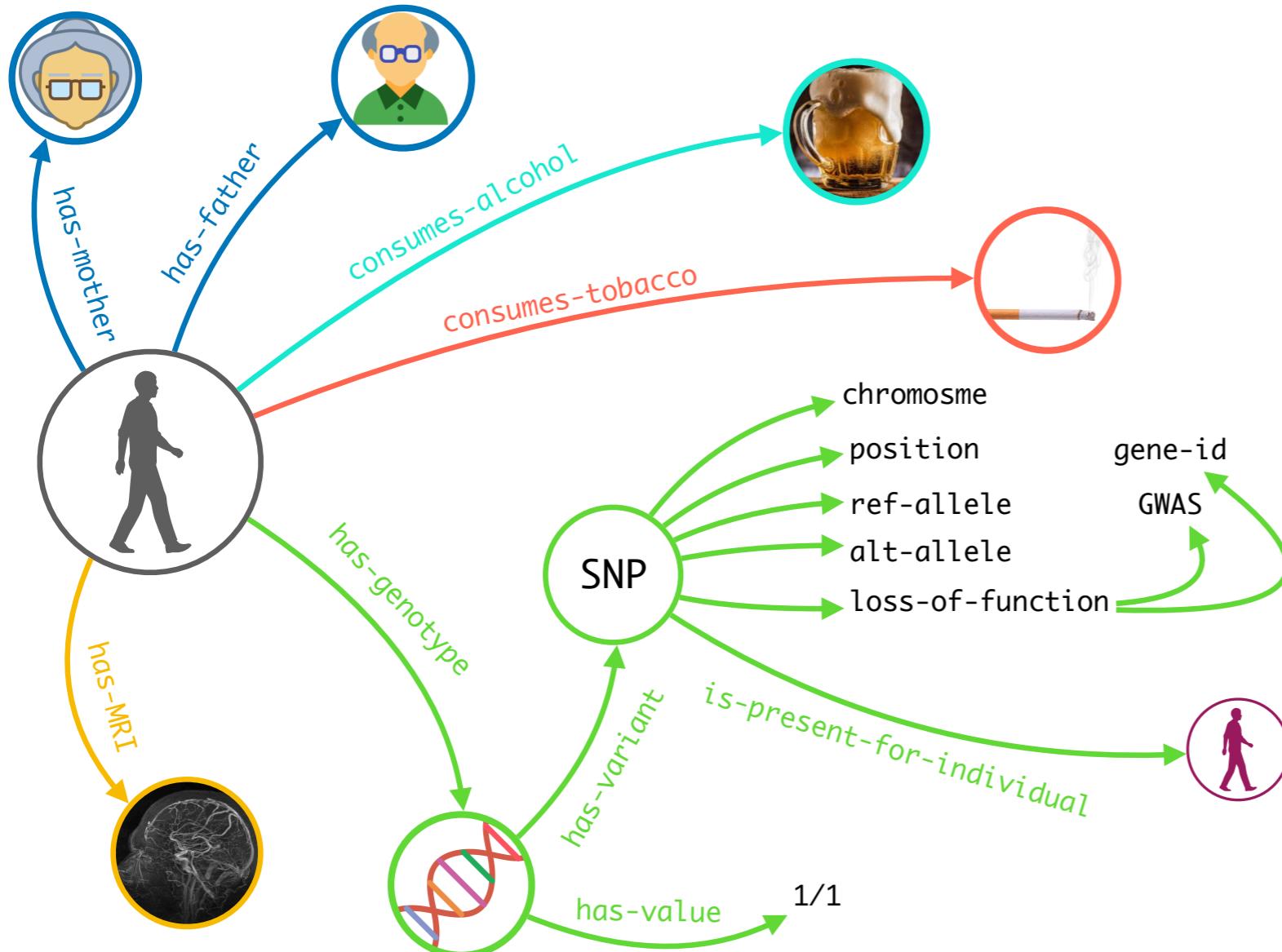
Cerebral berry aneurysm

" Concepts (nodes) are connected
with generalization relations (edges) "



② how to link data from multiple scales ?

Knowledge graphs: « a collection of interlinked descriptions of things (real-word objects, abstract concepts, events, etc.) »



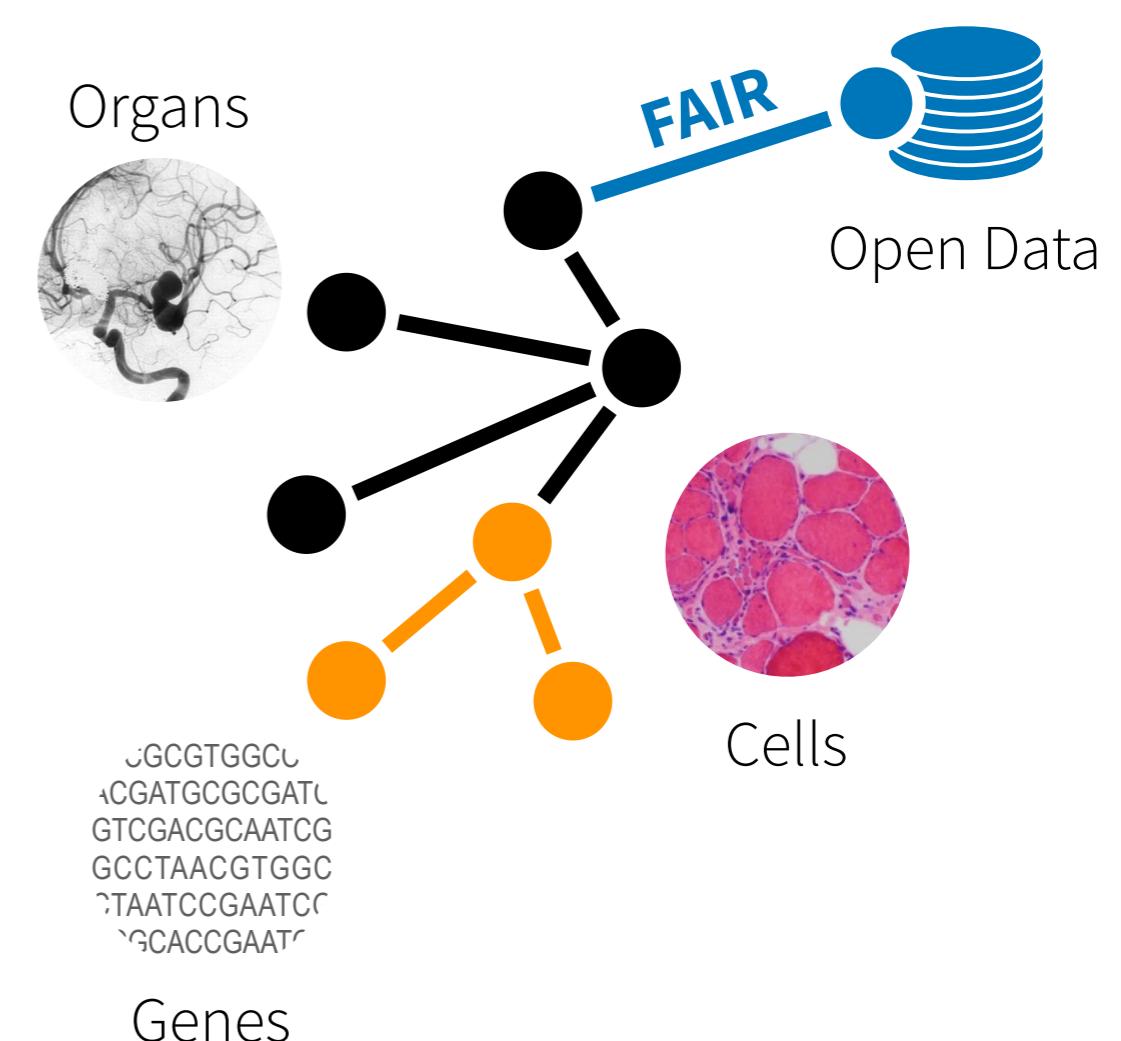
= **database:** information storage / extraction

= **graph:** network analysis

= **knowledge base** with formal semantics :

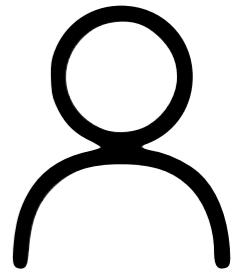
- + logical facts
- + logical inferences

Bridging imaging-omics- clinical data: **INEX-MED**

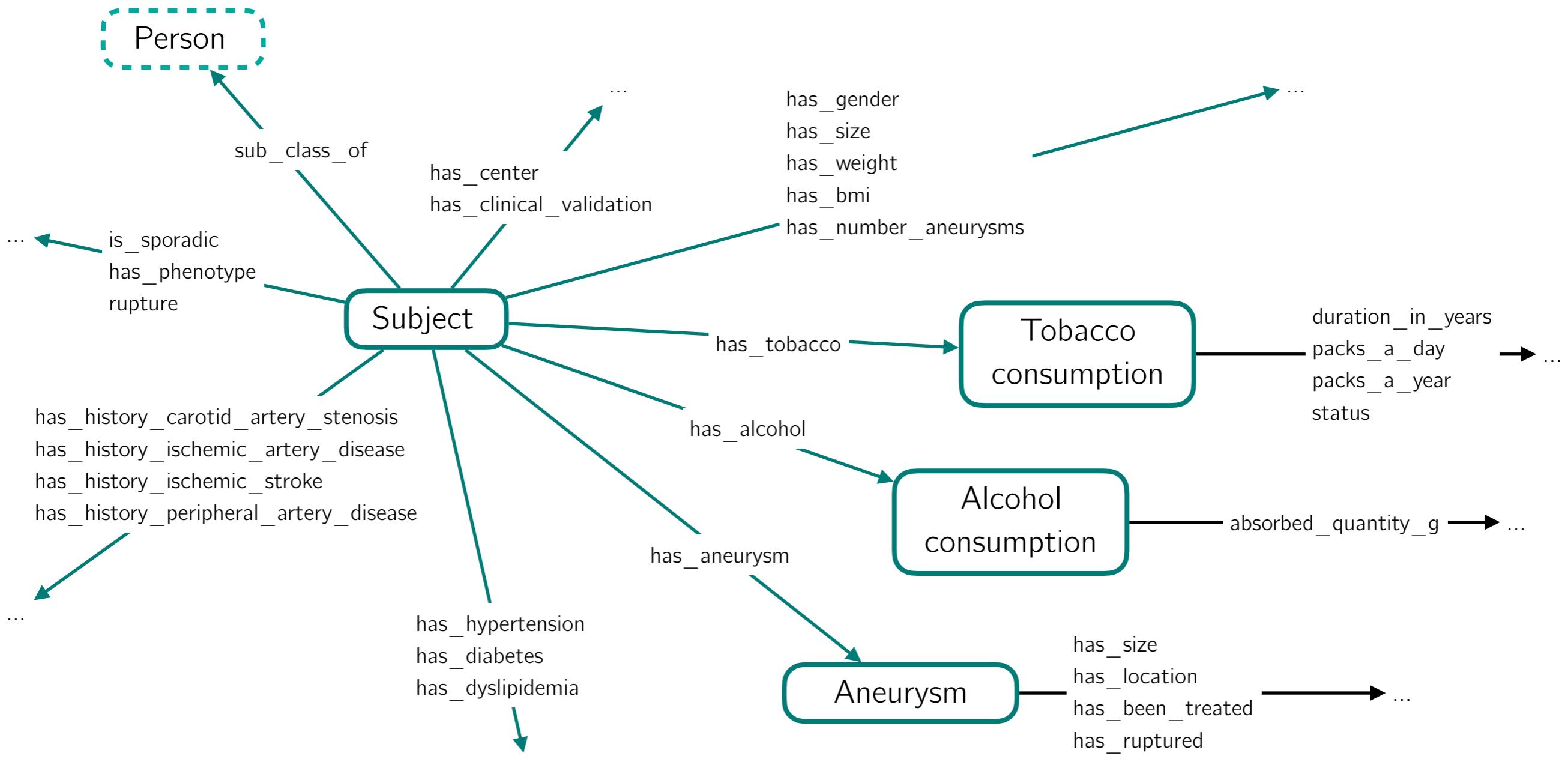


Multi-scale « Knowledge Graph »
145 Million facts

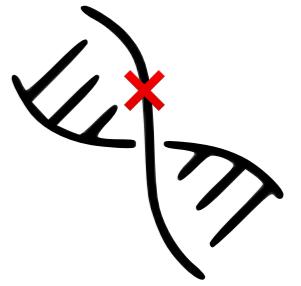
Clinical observations



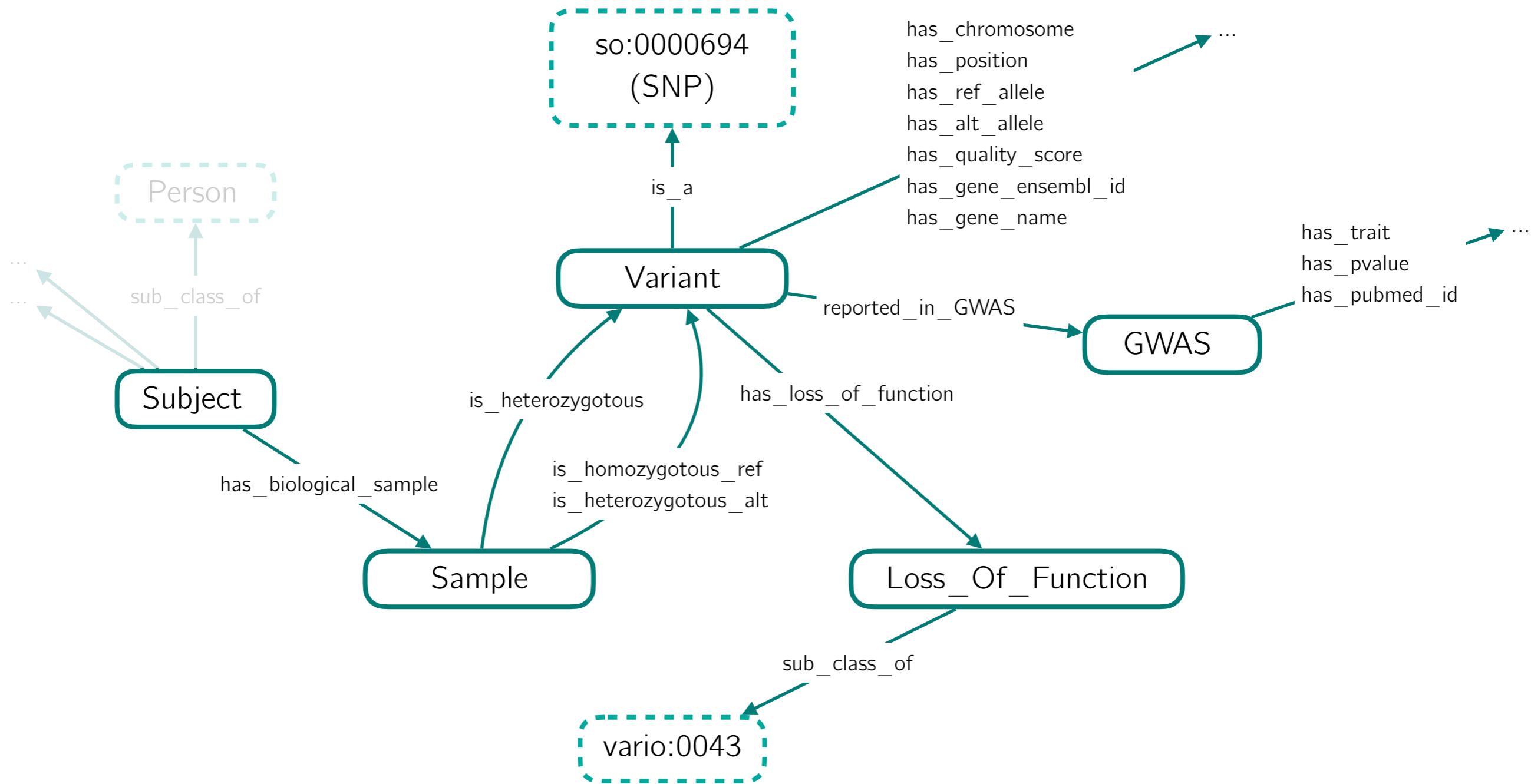
- ▶ Aim : representing key characteristics of ICA patients



Genomic markers



- ▶ Aim : representing associated genetic variations by **extending reference ontologies** (Sequence Ontology, VARIO)

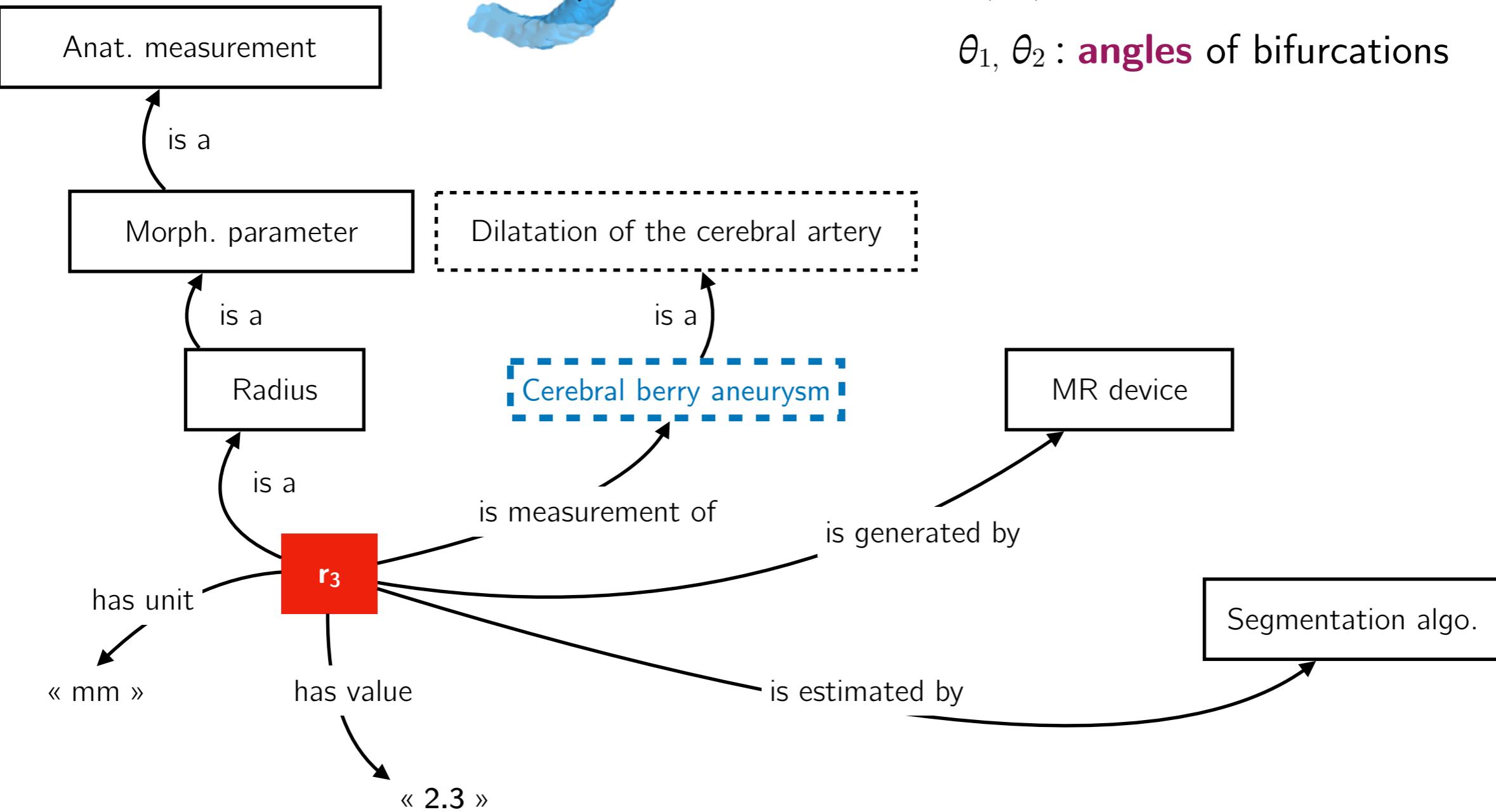
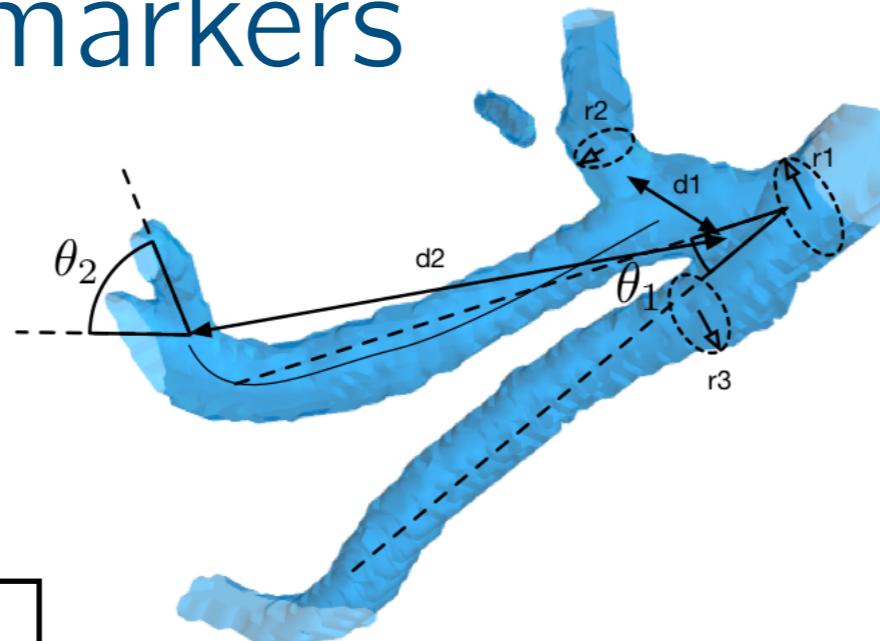


Imaging markers



Anass Nouri

Florent Autrusseau



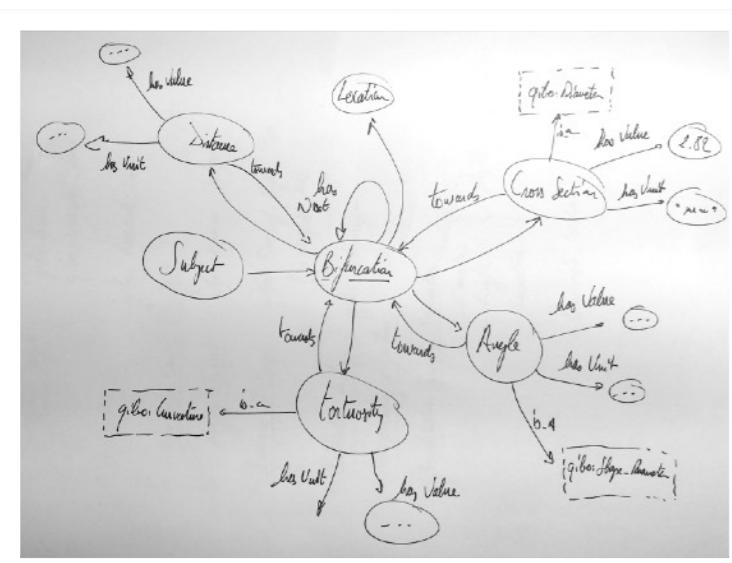
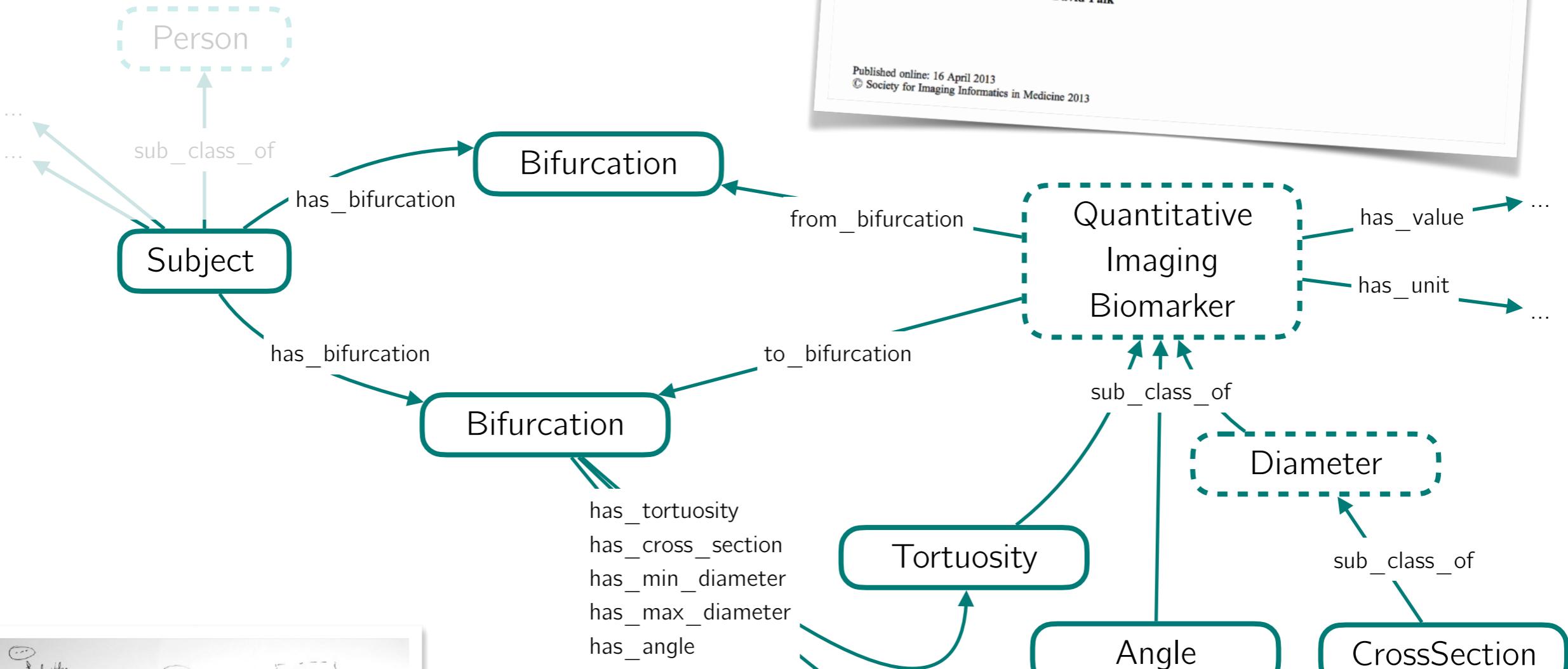
Imaging markers

J Digit Imaging (2013) 26:630–641
 DOI 10.1007/s10278-013-9599-2

Quantitative Imaging Biomarker Ontology (QIBO) for Knowledge Representation of Biomedical Imaging Biomarkers

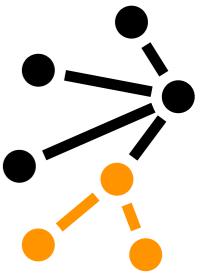
Andrew J. Buckler · M. Ouellette · J. Danagoulian ·
 G. Wernsing · Tiffany Ting Liu · Erica Savig ·
 Baris E. Suzek · Daniel L. Rubin · David Paik

Published online: 16 April 2013
 © Society for Imaging Informatics in Medicine 2013



Searching
in multi-scale data ?

Queries with graph patterns



? "Give all patients with a mutation in ANGPTL6, in the case of a rupture, give me also the age and the location of the ruptured aneurysm."

```
SELECT * WHERE {
  # Search variants
  ?subject ican:has_dna_sample ?sample .
  ?sample ican:is_heterozygous ?variant .
  ?variant ican:has_chromosome "19" ;
            ican:has_position ?position ;
            ican:has_alt_allele ?alt_allele ;
            ican:has_ref_allele ?ref_allele .
  FILTER (?position > 10092337 && ?position < 10106407)
  # dbSNP identifier
  OPTIONAL { ?variant ican:has_dbsnp_id ?dbSNPid . }
  # First aneurysm rupture: age and localisation
  OPTIONAL { ?subject ican:age_first_rupture ?ageRupt . }
  OPTIONAL { ?subject ican:loc_first_rupture ?locRupt . }
}
```

Select "extreme phenotypes" for genomic studies

< 30 years old ?

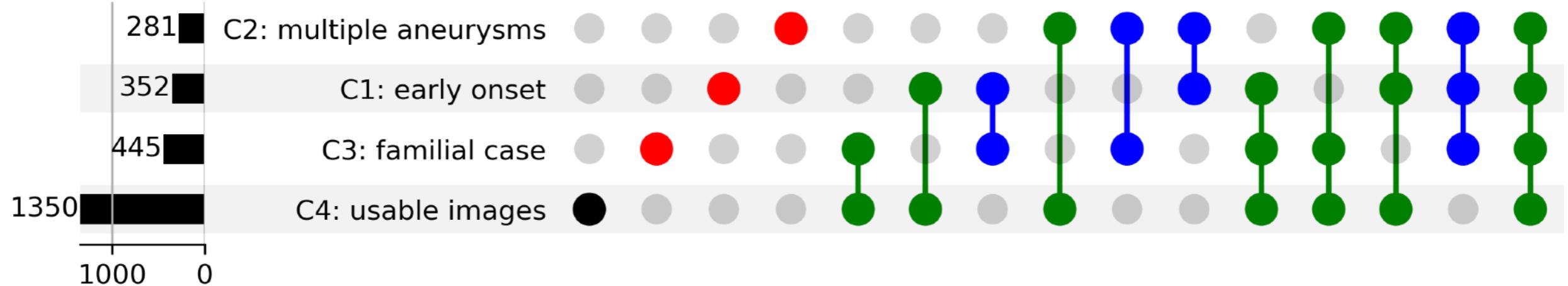
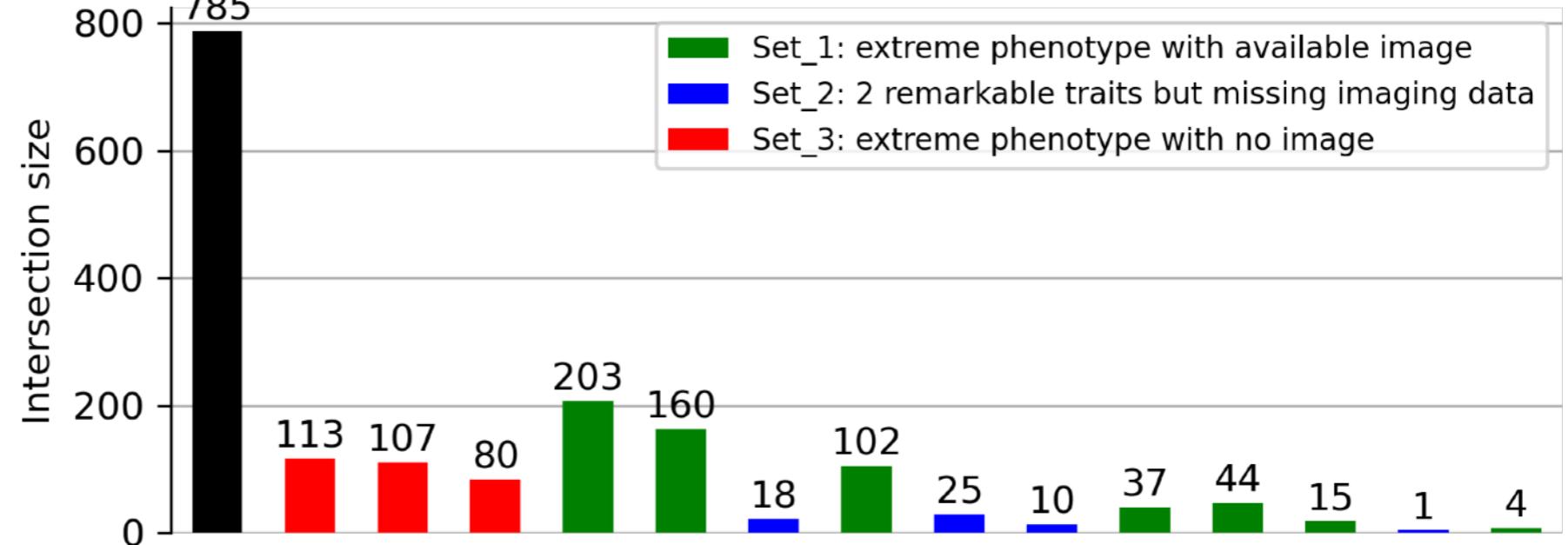
usable images ?

2+ aneurysms ?

familial forms ?

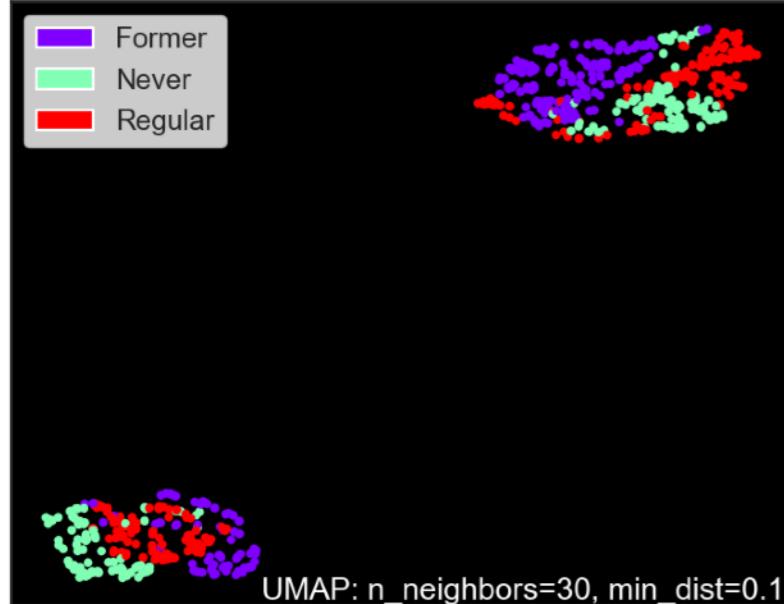
ICAN subjects

919 colored individuals, 565 in Set_1, 54 in Set_2 and 300 in Set_3.

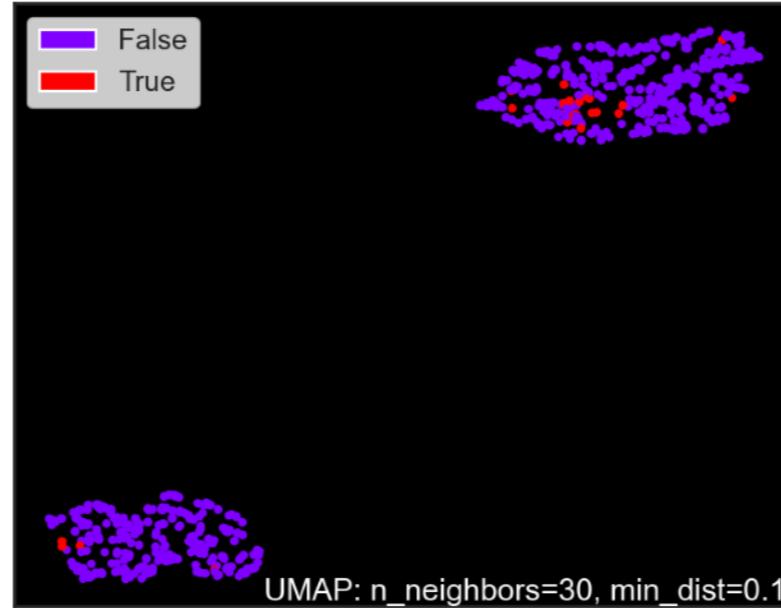


Train machine-learning models ...

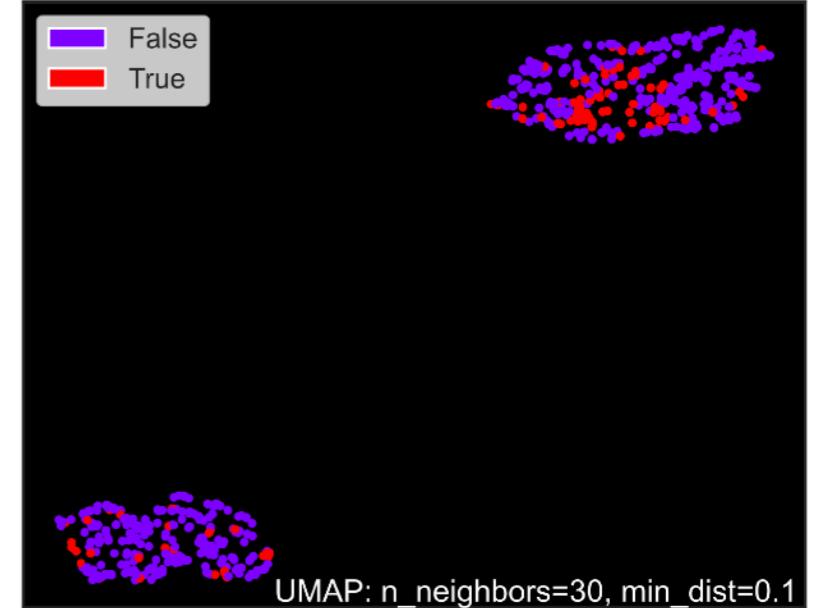
Tobacco consumption



Ischemic heart disease history



Platelet aggregation inhibiting treatment



Observational Study > J Neurol Neurosurg Psychiatry. 2021 Feb;92(2):122-128.

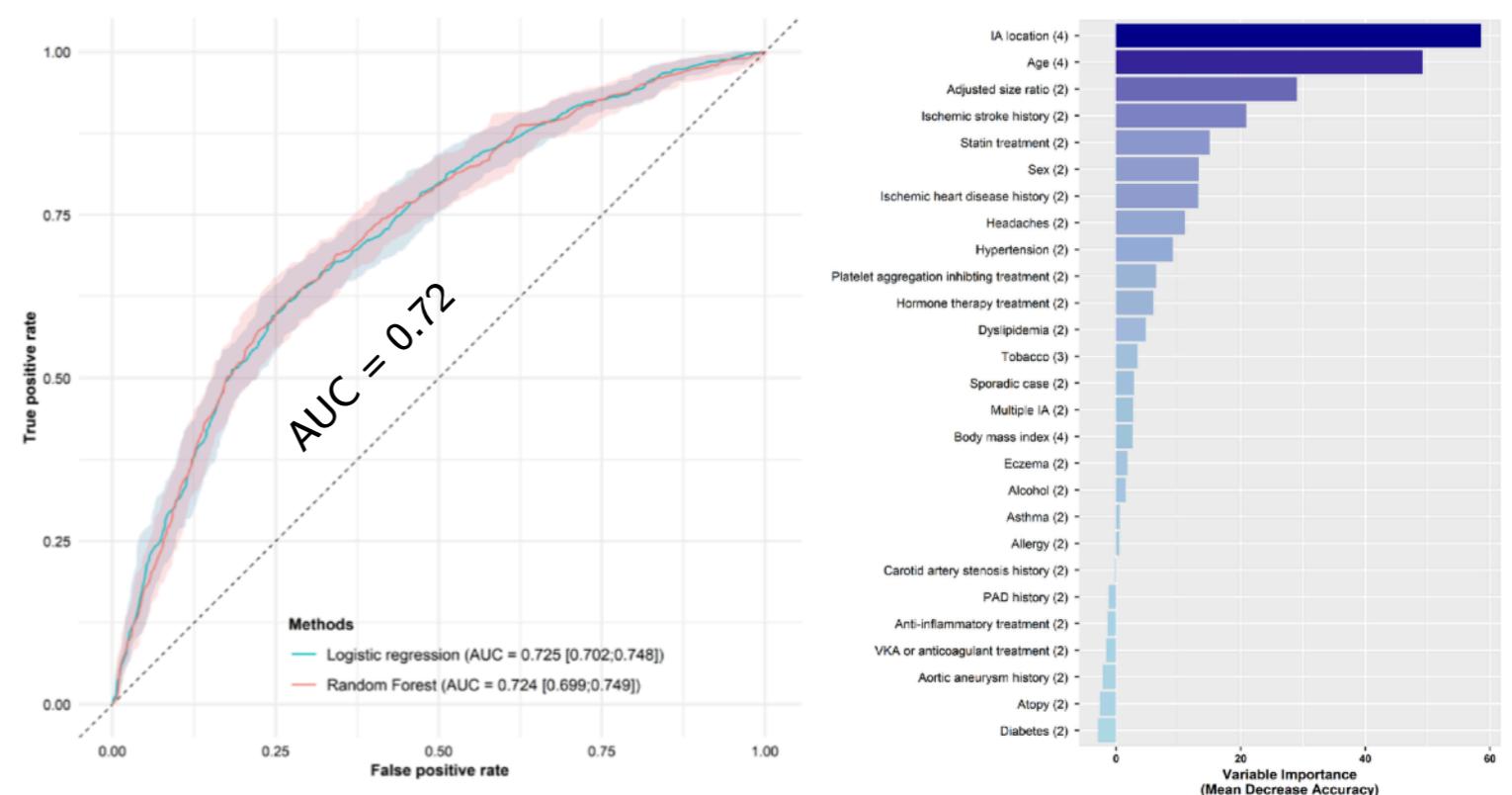
doi: 10.1136/jnnp-2020-324371. Epub 2020 Oct 23.

Location of intracranial aneurysms is the main factor associated with rupture in the ICAN population

Olivia Rousseau ¹, Matilde Karakachoff ¹, Alban Gaignard ², Lise Bellanger ³, Philippe Bijlenga ⁴, Pacôme Constant Dit Beaufils ¹, Vincent L'Allinec ⁵, Olivier Levrier ⁶, Pierre Aguetzaz ⁷, Jean-Philippe Desilles ⁸, Caterina Michelozzi ⁹, Gaultier Marnat ¹⁰, Anne-Clémence Vion ², Gervaise Loirand ², Hubert Desal ¹¹, Richard Redon ², Pierre-Antoine Gourraud ¹, Romain Bourcier ¹²; ICAN Investigators

Collaborators, Affiliations + expand

PMID: 33097563 DOI: [10.1136/jnnp-2020-324371](https://doi.org/10.1136/jnnp-2020-324371)



(controlled) access and sharing

INEX-MED data hub ≡

GENERAL

- Dashboard
- Admin

SEARCH

- Centers
- Gender Distribution

ANEURYSMS

- Aneurysms Position
- Genomic Variants

AUTHENTICATION

- Sign-in
- Sign-out

INEX-MED data hub

Exploring clinical, genomics, and imaging data to better understand congenital myopathies and intracranial aneurysms

See the [official INEX-MED wiki page](#) for more information.

SUBJECTS	CLINICAL OBSERVATIONS	BIFURCATIONS	GENETIC VARIANTS
2,974	2,613	9,490	1,549,005

FACTS

143,745,265

Web demonstrator → <https://inexmed-api.univ-nantes.fr>

Related works

Sample knowledge graphs : SPHN network

nature > scientific data > articles > article

Article | [Open Access](#) | Published: 10 March 2023

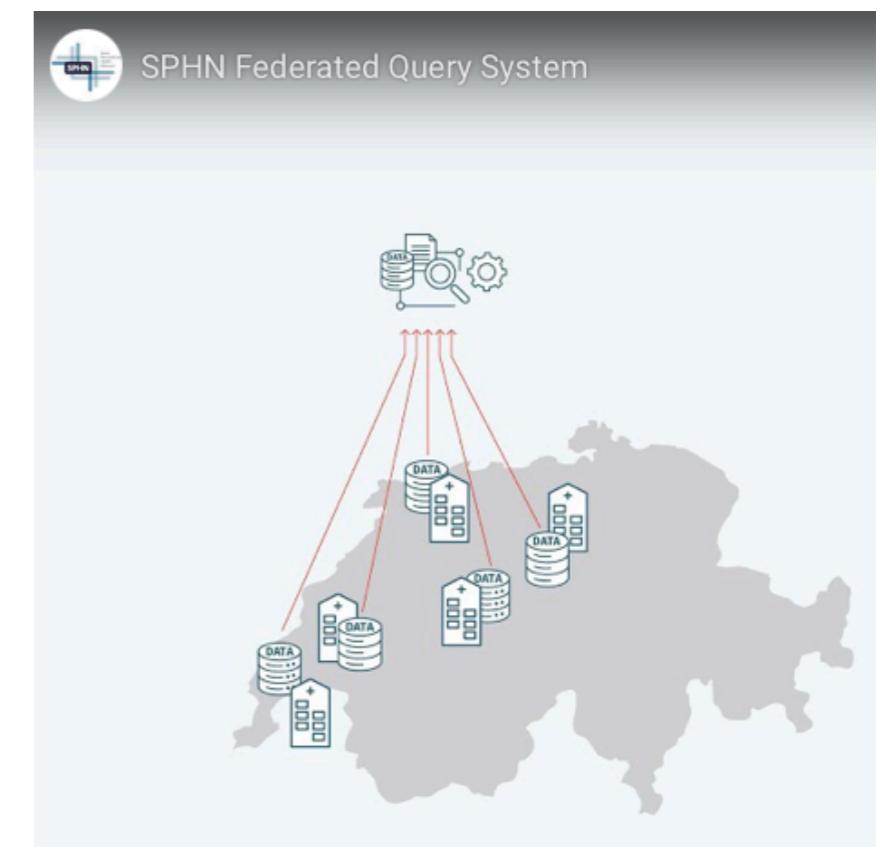
FAIRification of health-related data using semantic web technologies in the Swiss Personalized Health Network

Vasundra Touré, Philip Krauss, Kristin Gnodtke, Jascha Buchhorn, Deepak Unni, Petar Horki, Jean Louis Raisaro, Katie Kalt, Daniel Teixeira, Katrin Cramer & Sabine Österle [✉](#)

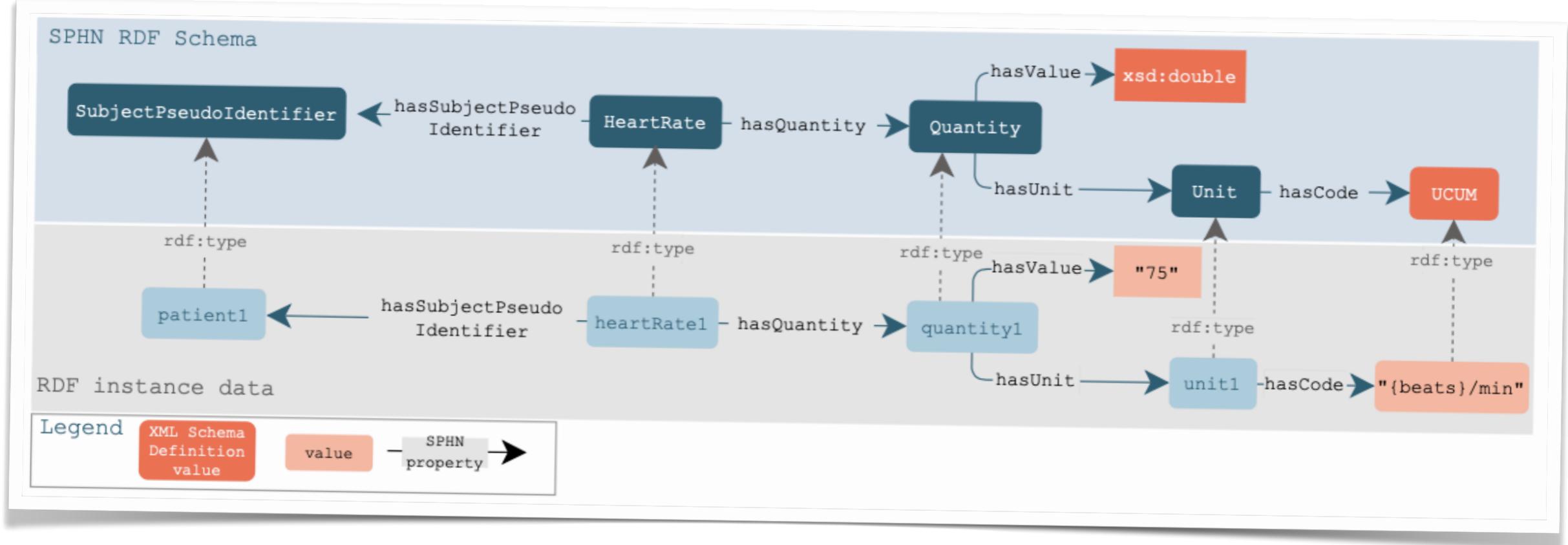
[Scientific Data](#) 10, Article number: 127 (2023) | [Cite this article](#)

Abstract

The Swiss Personalized Health Network (SPHN) is a government-funded initiative developing federated infrastructures for a responsible and efficient secondary use of health data for research purposes in compliance with the FAIR principles (Findable, Accessible, Interoperable and Reusable). We built a common standard infrastructure with

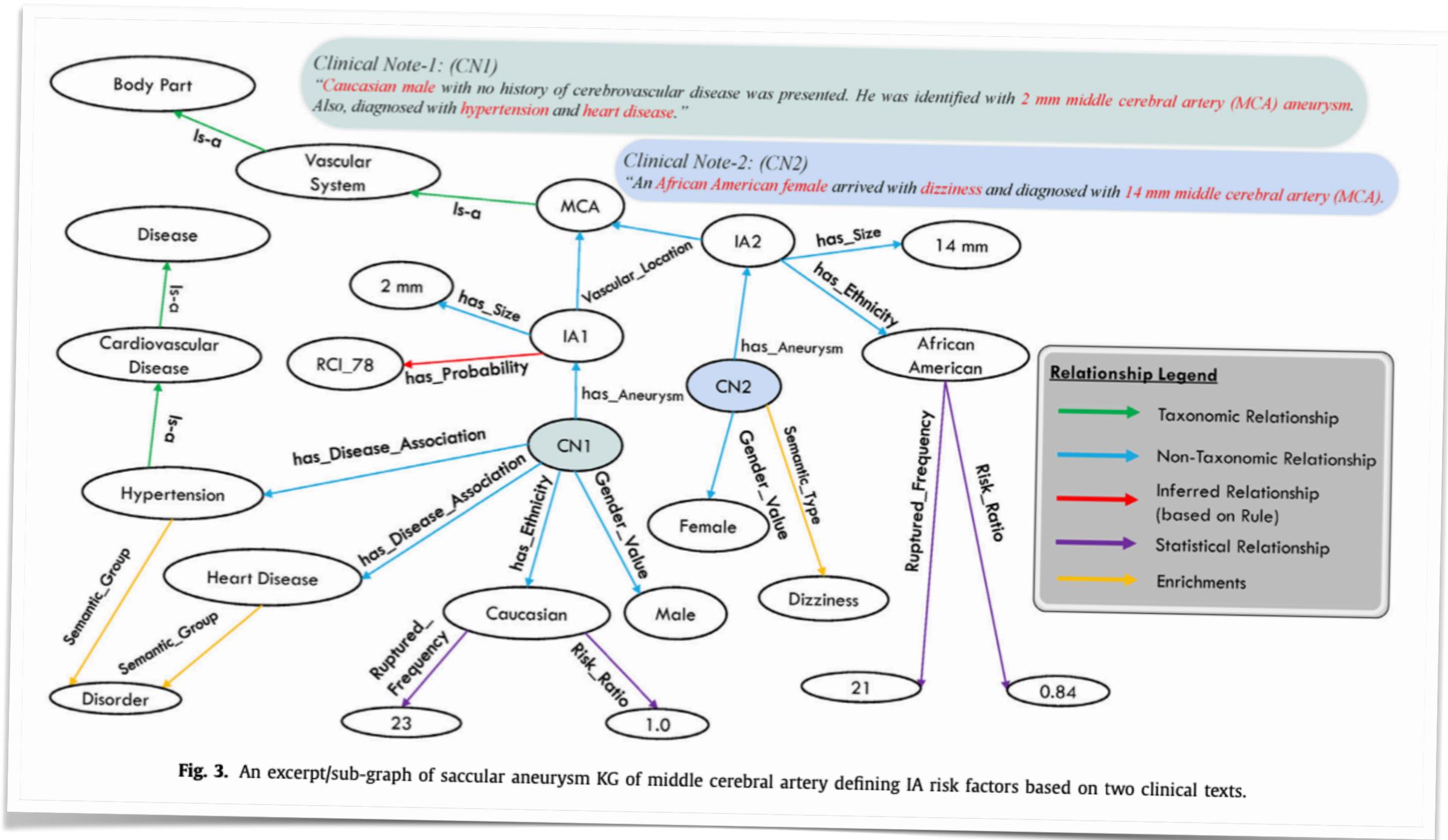


Sample knowledge graphs : SPHN network



- + Allergy
- + Biobank
- + Consent
- + Radiology diagnosis
- + Genomic variants position
- + ECG

Sample knowledge graphs: Malik et. al. 2020



Unruptured IA: 51 "core" common data elements / 1254

Age value, Birth date, Birth sex assigned type, Body system category, Race USA category, Sex genotype type, Ethnicity USA category, Gender identity type, Gender type,

Aneurysm diameter maximum measurement, Aneurysm height measurement, Aneurysm morphology type, Aneurysm width measurement,

Hypertension indicator,

Tobacco current use indicator, Tobacco prior use indicator, Tobacco use started age value, Tobacco use stopped age value,

World Federation of Neurological Surgeons (WFNS) - grading system subarachnoid hemorrhage scale

Behavioral history assessment date and time,

Biospecimen collection other text, Biospecimen collection type,

Cerebrospinal fluid collection anatomic site, Cerebrospinal fluid collection anatomic site other text, Cerebrospinal fluid collection method other text, Cerebrospinal fluid collection method type,

Data collected date and time, Data source,

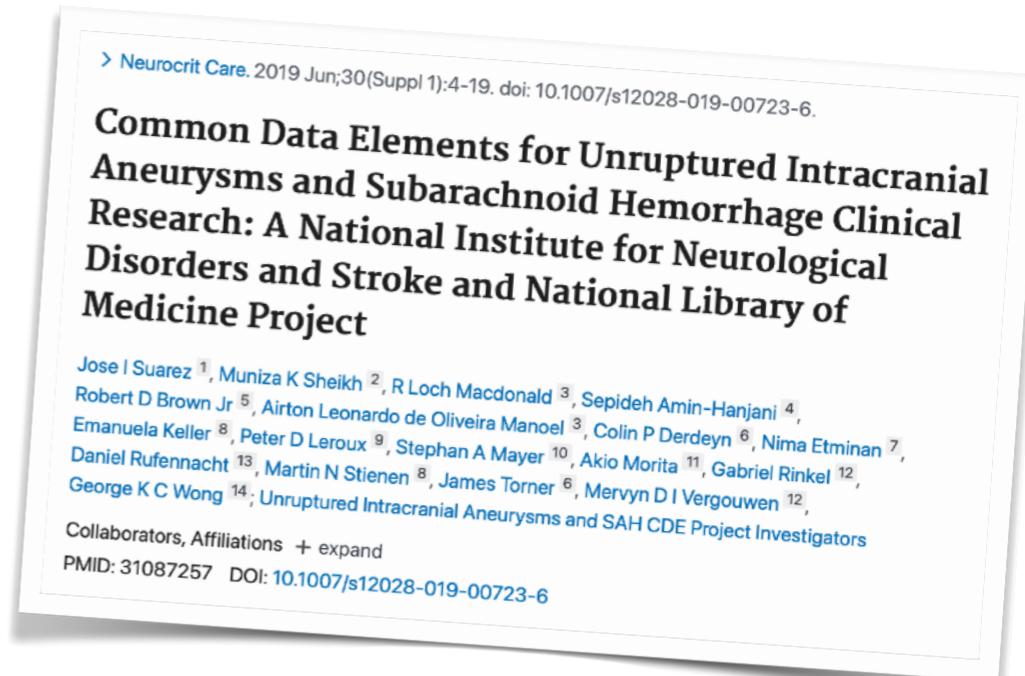
History data not obtained reason,

Imaging modality other text, Imaging modality type, Imaging modality vessel imaging angiography type, Imaging vessel angiography arterial anatomic site, Imaging vessel angiography arterial result,

Intervention endovascular status, Intervention surgical status, Lab specimen collection date and time,

Medical history condition SNOMED CT code, Medical history condition text, Medical history taken date and time,

Specimen acquisition anatomical site, Specimen baseline collection indicator, Specimen biologic source type, Specimen centrifugation duration, Specimen centrifugation revolutions per minute value, Specimen centrifugation temperature value, Specimen collection tube other text, Specimen collection tube type, Specimen not collected reason, Specimen storage temperature value, Specimen time collection process interval,



✓ Large community effort

✓ Consensual definitions

⚠ Flat list of terms

⚠ No recommendation data structures

Future works

Wrap up & next steps

- ▶ Gather **consensual expert knowledge** with computational **Ontologies**
- ▶ Knowledge Graphs are "**machine-processable**"
- ▶ Link **multi-scale data** with Knowledge Graphs
- ▶ Query **multi-source** data



Future works

- ▶ Include familial relations to query IA **familial forms**
- ▶ **Share** and enrich data models with **international** initiatives
- ▶ Further develop **predictive approaches** (rupture risk prediction)
- ▶ Develop a national **data hub** on intracranial aneurysms (France 2030 PEPR Santé Numérique funding)

Acknowledgments

Florent Autrusseau,
Raphaël Blanchet,
Romain Bourcier,
Hubert Desal,
Christian Dina,
Maxime Folschette,
Matilde Karackachoff,
Vincent L'Allinec,
Fabien Laporte,
Anass Nouri,
Richard Redon,
Olivia Rousseau,
Hala Skaf-Molli,



Kirsley Chennen,
Jocelyn Laporte,
Olivier Poch,
Norma Romero,
Julie Thompson

ICAN consortium

BiRD facility,
NeurlInfo facility



Backup slides

