The role of minimal data models like OSIRIS to simplify RWE

Connect to Win - DIGICORE
November 4th, 2021

Julien Guérin
Head of Health Data Factory
Data Office, Institut Curie

Data sharing : a necessity inherent to cancer disease

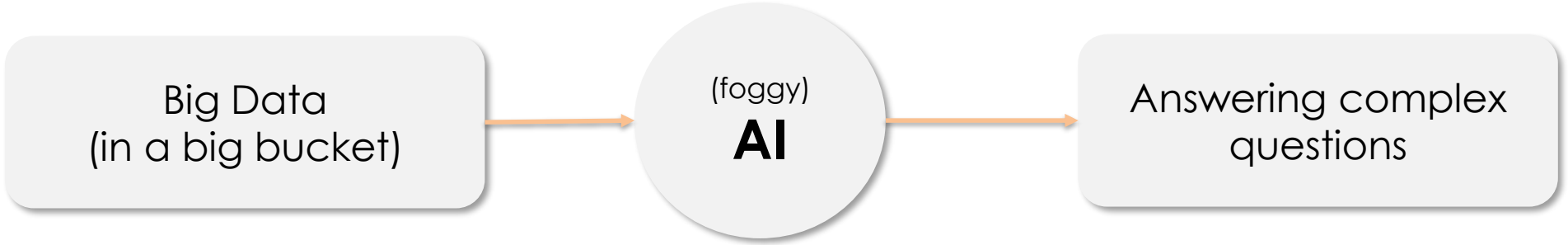
"The molecular uniqueness of each cancer and the number of genetic variants present in an individual's genome makes precision oncology not only challenging from a clinical and biological perspective but also from a computational perspective".
NhanDo et al. Seminars in Oncology 2019

« The real value of genomic data will be realized only when they are linked to high-quality, longitudinal, computationally amenable clinical information, allowing researchers to identify precise genotype phenotype associations.
If we don't concentrate our efforts (and dedicate substantial resources) to robustly improve data sharing, we risk undermining precision oncology's capacity to deliver substantive advances for people with cancer ». *NEJM*, May 2017

We need data before AI

- New technologies has emerged offering tremendous opportunities to revolutionize patient care and cancer research
- Storage and compute costs have dropped dramatically the last two decades
- Great, we can do almost…anything

Wrong paradigm



- **We need qualitative data first, more likely small (less data but understandable) and smart (updated regularly and analysable on the fly) than big.**
- **Models are required to speak the same language in oncology**

institut Curie

We need data before AI

In 2016, AlphaGo (Deep Mind) beat Lee Sedol, world champion and master of Go

Google DeepMind

Compute
- 1200 CPU
- 300 GPU

Power consumption
- 20.000 watts

Compute
- 100 billions of neurons

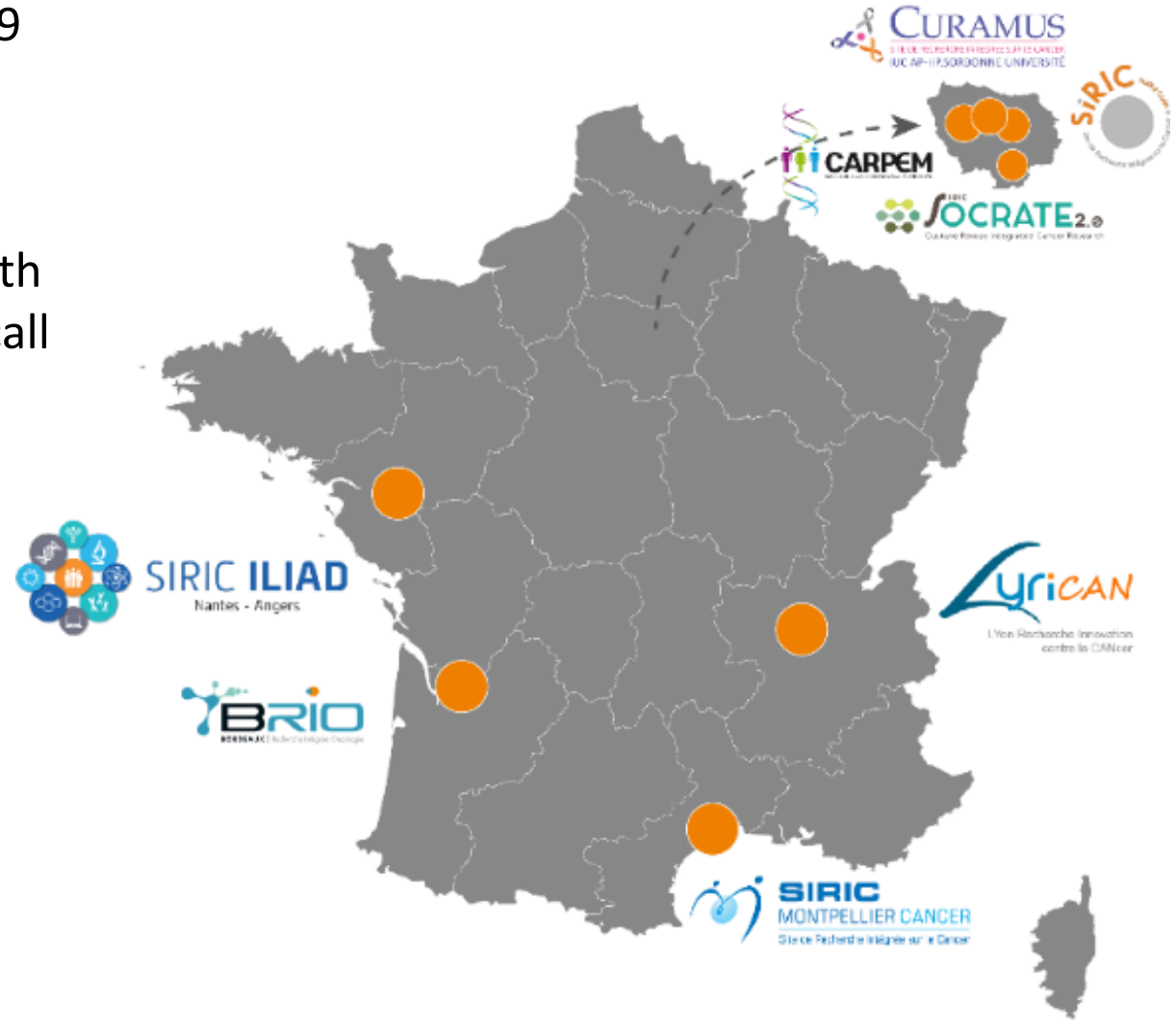
Power consumption
- 20 watts

Too much energy to play…

institut Curie

The SiRIC network

- Launched during Plans Cancer 2011 – 2019
- A national network
- A label awarded by INCa in partnership with Inserm and DGOS through a competitive call based on scientific research programs
- OSIRIS Inter-SIRIC group born in 2014



INSTITUT NATIONAL DU CANCER

Inserm

DGOS

institut Curie

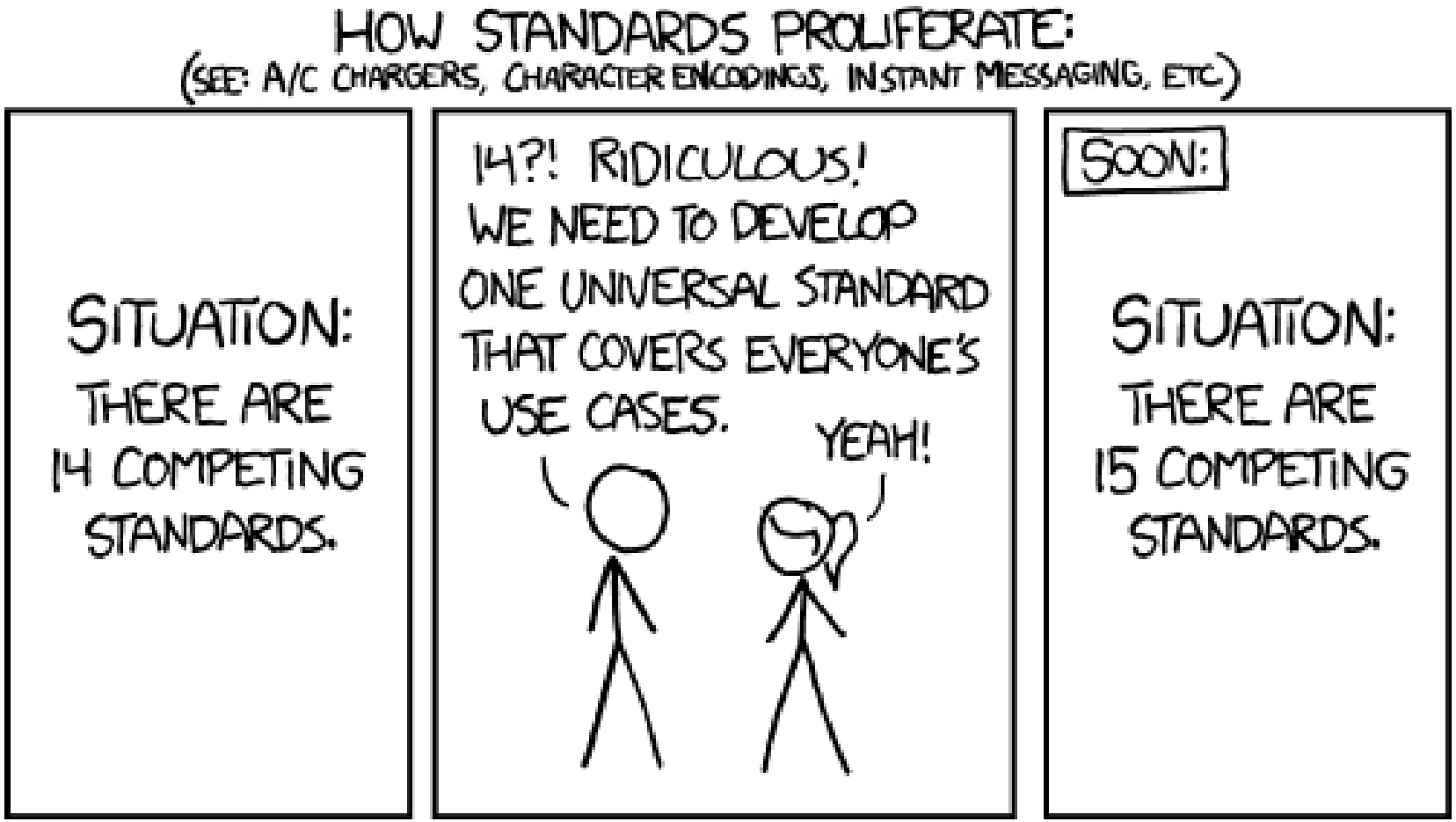
OSIRIS partners

Hôpital européen Georges-Pompidou

OSIRIS : why we need common data models?

- To accelerate data sharing between health institutes
- To allow translational studies on large datasets
- To support clinical studies as well as real world data studies
- To provide qualitative datasets for Open Data and AI approaches

Interoperability pitfall



https://xkcd.com/927/

OSIRIS model rules

- Keep it simple
  - Less than 100 items
- Technology agnostic
  - Can be deployed in any type of database or device
- Aligned with standards
  - International scope
  - Providing guidelines for the data collect
- Compatible with other initiatives
  - Defining mapping procedures or automatic data transformation from OSIRIS to another model

OSIRIS building process

SIRICs network

An existing set of data structures

National datawarehouses (Ex: Consore)

Molecular clinical trials

European tumor databases (Ex: Conticabase)

Extraction of Relevant Data Elements

SIRICs multidisciplinary scientific meeting (clinicians, bioinformaticians, data managers, researchers ,etc.)

Every week/16 weeks

Weekly meetings to compare and extract valuable Common Data Elements (CDE)

Set OSIRIS version v1

Validation of the set OSIRIS v1 with a set of scientific questions

National Scientific Board including clinicians and researchers

Technical Board Including bioinformaticians and data managers

Every week/32 weeks

Weekly meetings to map the OSIRIS set on international standards and develop interoperability tools (i.e Definitions of a set of structured flat-files)

Production set OSIRIS version v1.1.05

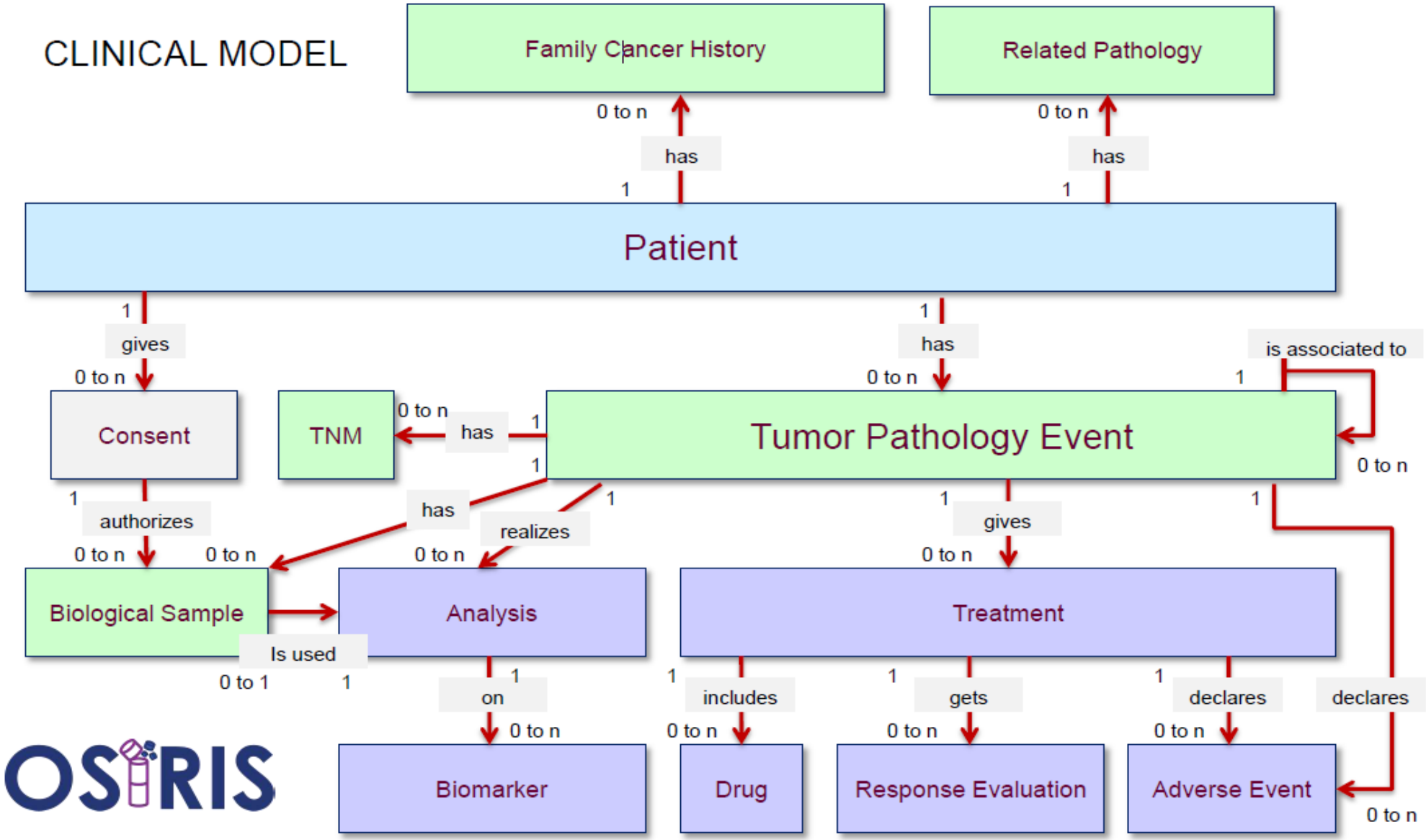
institut Curie

OSIRIS overview

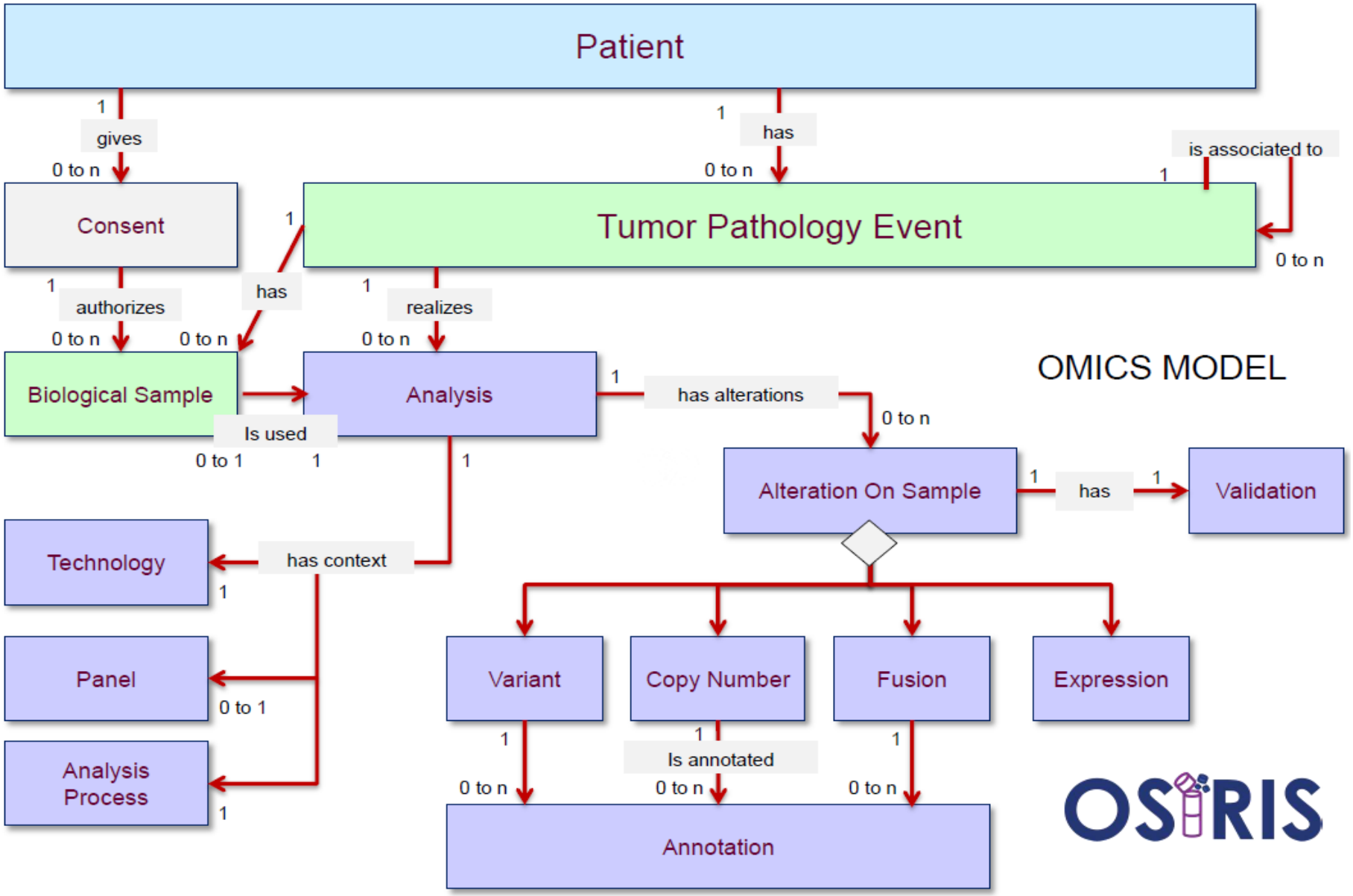
- **First stable release - version 1.1 (6 septembre 2018)**
  - 67 clinical items (13 objects)
  - 62 omics items (10 objects)
  - Over 40 value sets

- **Mapping with international standards**
  - FHIR
  - UMLS
  - LOINC
  - ICD-10
  - ICD-O-3 (Topo/Morpho)
  - ATC
  - MedDRA/CTCAE
  - TNM
  - RECIST
  - HGVS
  - …

OSIRIS clinical model



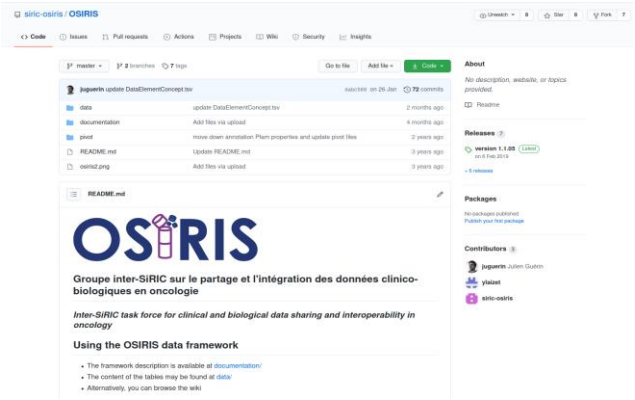
OSIRIS omics model



institut Curie

OSIRIS documentation

https://github.com/siric-osiris/OSIRIS



Patient

Property	Definition EN
Id	Patient ID in the center providing the information
Gender	Gender of the patient
Ethnicity	Ethnicity of the patient
BirthDate	Date of birth of the patient
DeathDate	Date of death of the patient
ProviderCenterId	Center ID of the center providing the information
OriginCenterId	Center ID of the center of origin of the patient
CauseOfDeath	Description of the cause of death of the patient
LastNewsDate	Date of last news (= date of last patient visit if there is no other information available) (day is set to the 15th)
LastNewsStatus	Indicator of the patient's last known vital status (alive or deceased)

Specifications

Attribute	Value
FormatConceptualDomain	String
Required	mandatory

Usage

Attribute	Value
Examples	

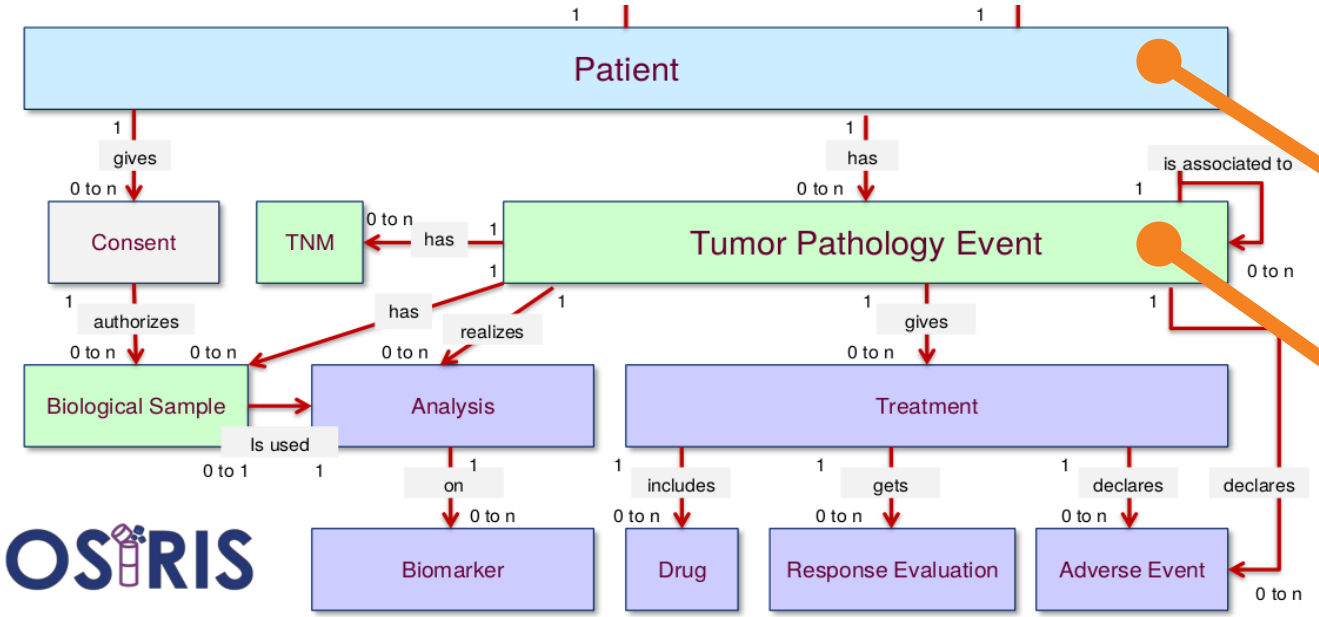
ConceptualDomain

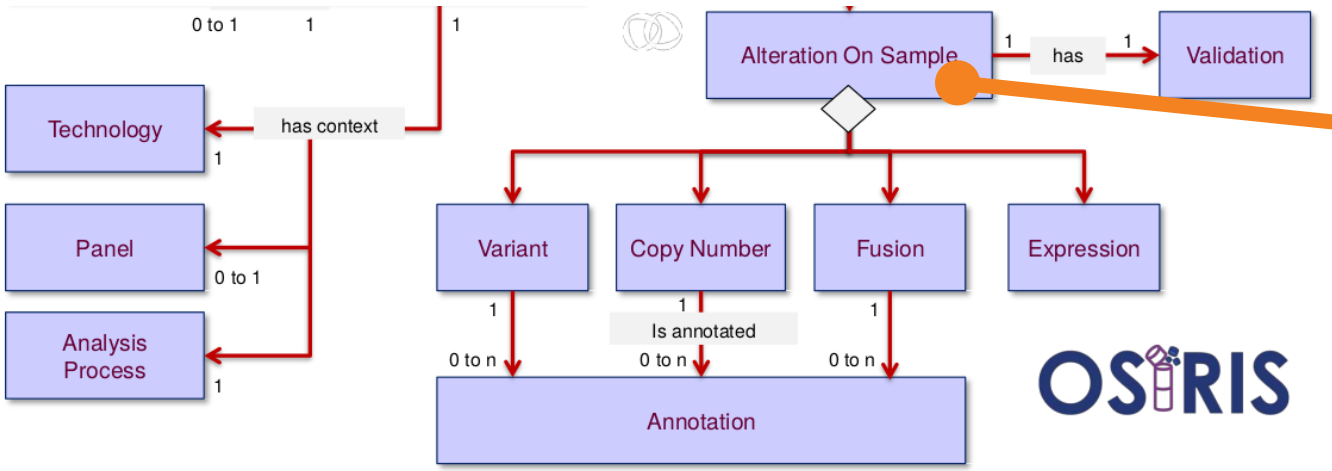
ValueMeaning	LabelValueMeaning
HL7:M	Male
HL7:F	Female
HL7:UN	Undifferentiated
UMLS:C0439673	Unknown

Value set

institut Curie

OSIRIS to FHIR





HL7® FHIR®

Level 1 Basic framework on which the specification is built

Foundation	Base Documentation, XML, JSON, Da

Level 2 Supporting implementation and binding to external specifications

Implementer Support	Security & Privacy	Conformance
Downloads, Version Mgmt, Use Cases, Testing	Security, Consent, Provenance, AuditEvent	StructureDefinition, CapabilityStatement, ImplementationGuide, Profiling

Level 3 Linking to real world concepts in the healthcare system

Administration	Patient, Practitioner, CareTeam, Devi

Level 4 Record-keeping and Data Exchange for the healthcare process

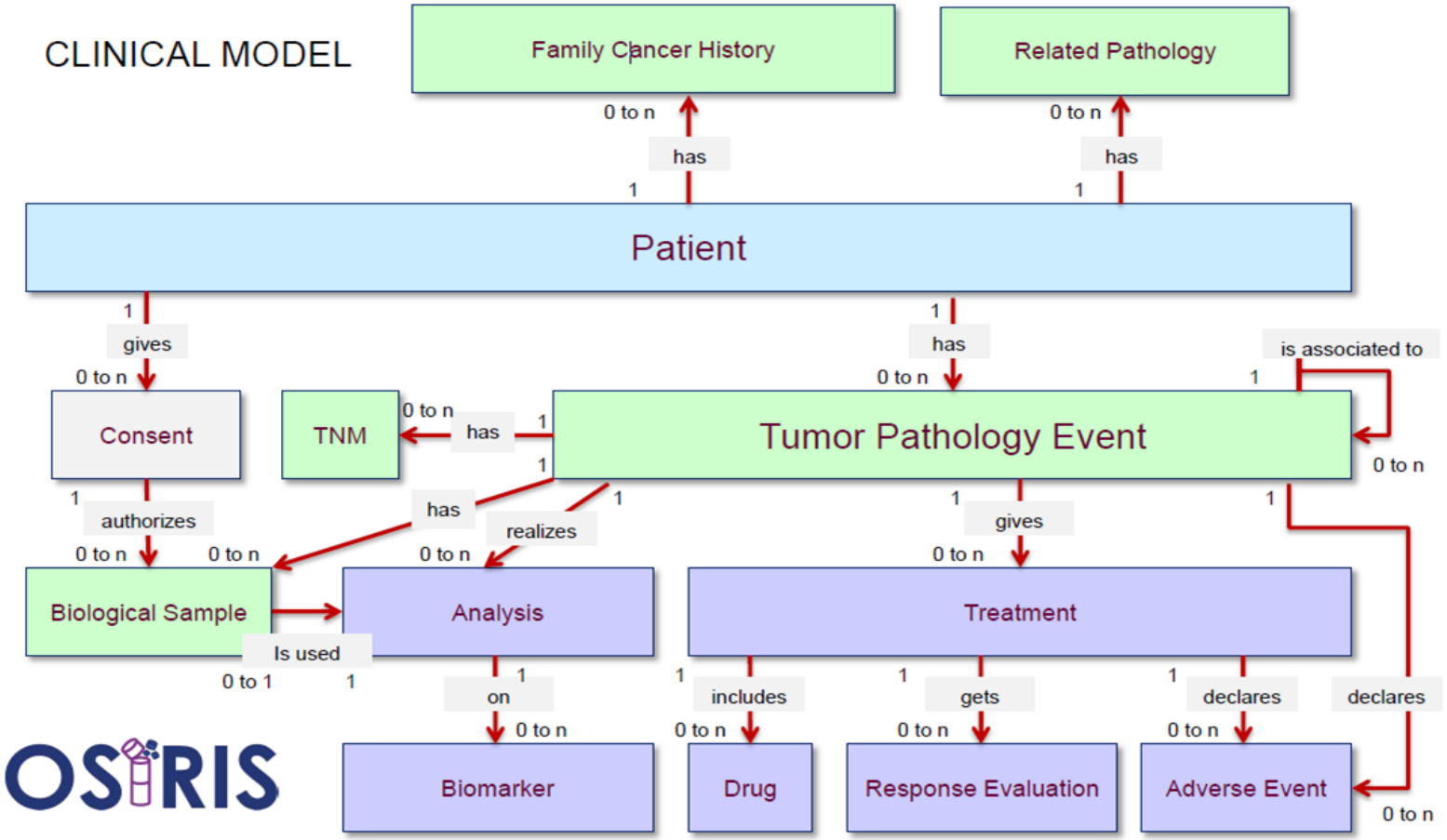
Clinical	Diagnostics	Medications
Allergy, Problem, Procedure, CarePlan/Goal, ServiceRequest, Family History, RiskAssessment, etc.	Observation, Report, Specimen, ImagingStudy, Genomics, Specimen, ImagingStudy, etc.	Medication, Request, Dispense, Administration, Statement, Immunization, etc.

Level 5 Providing the ability to reason ab

Clinical Reasoning	Li

ARKHN
UNLOCKING THE WEALTH OF DATA

2021 extensions



Imaging / Radiomics

Radiotherapy

Epidemiology

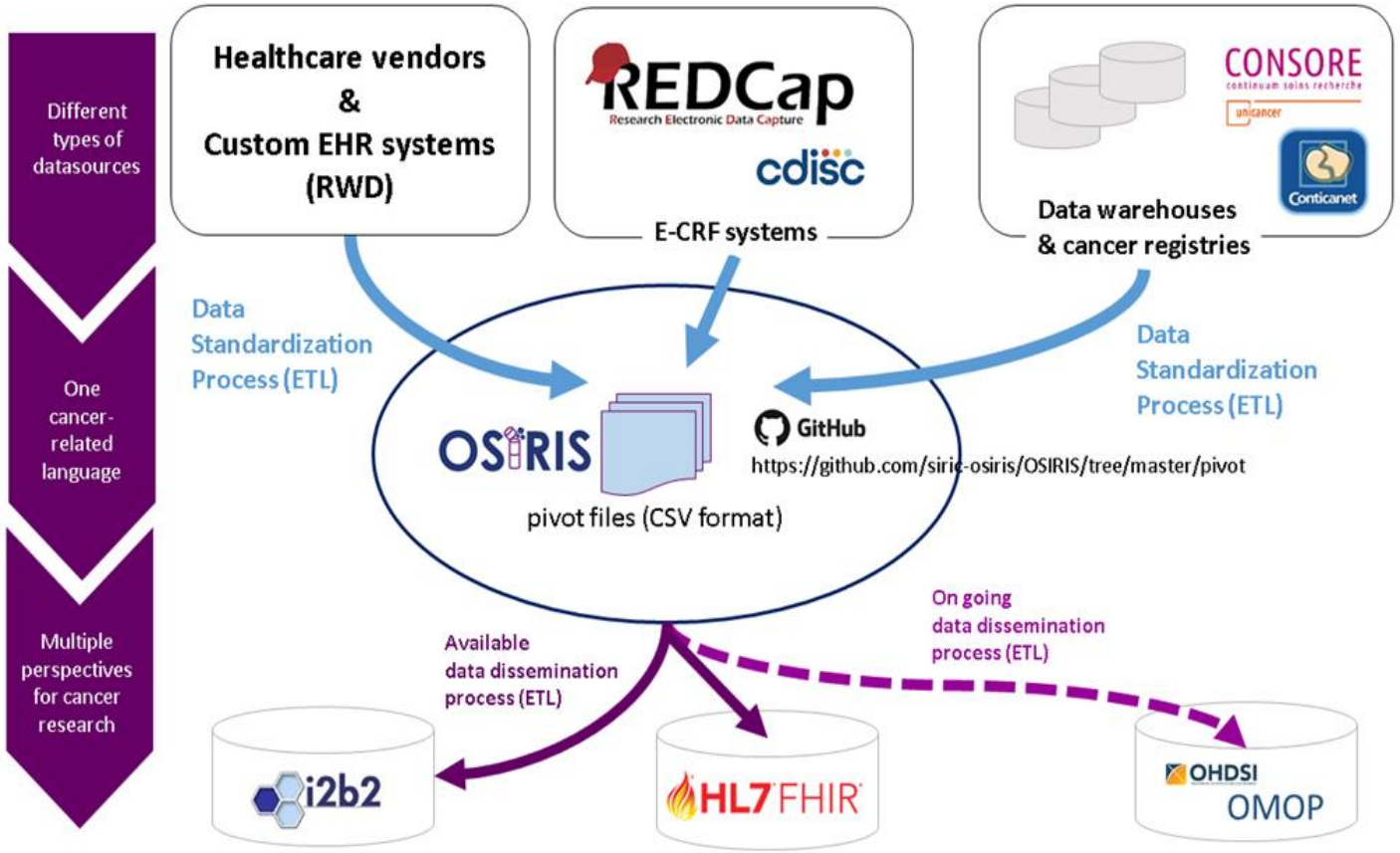
(Epidemium, data challenge, season 3)

institut Curie

OSIRIS main assets

- A national consensus (INCa, Unicancer, SIRICs, CHU …)
- A minimal dataset for data sharing
- Two connected facets : clinical and omics
- A temporal representation
- An *open source* initiative
- Terminologies based on international standards

JCO CCI publication



OSIRIS: A Minimum Data Set for Data Sharing and Interoperability in Oncology

Guerin J, Laizet Y, Le Texier V, Chanas L, Rance B, Koeppel F, Lion F, Gourgou S, Martin AL, Tejeda M, Toulmonde M, Cox S, Hess E, Rousseau-Tsangaris M, Jouhet V, Saintigny P.

JCO Clinical Cancer Informatics, March 2021.

institut Curie

The OSIRIS-LUNG project

Federated architecture with OSIRIS-standardised data

Technical goal :

- Assessing the benefit of interoperability
- Replicable statistical analysis
- Federated learning approach

institut
Curie

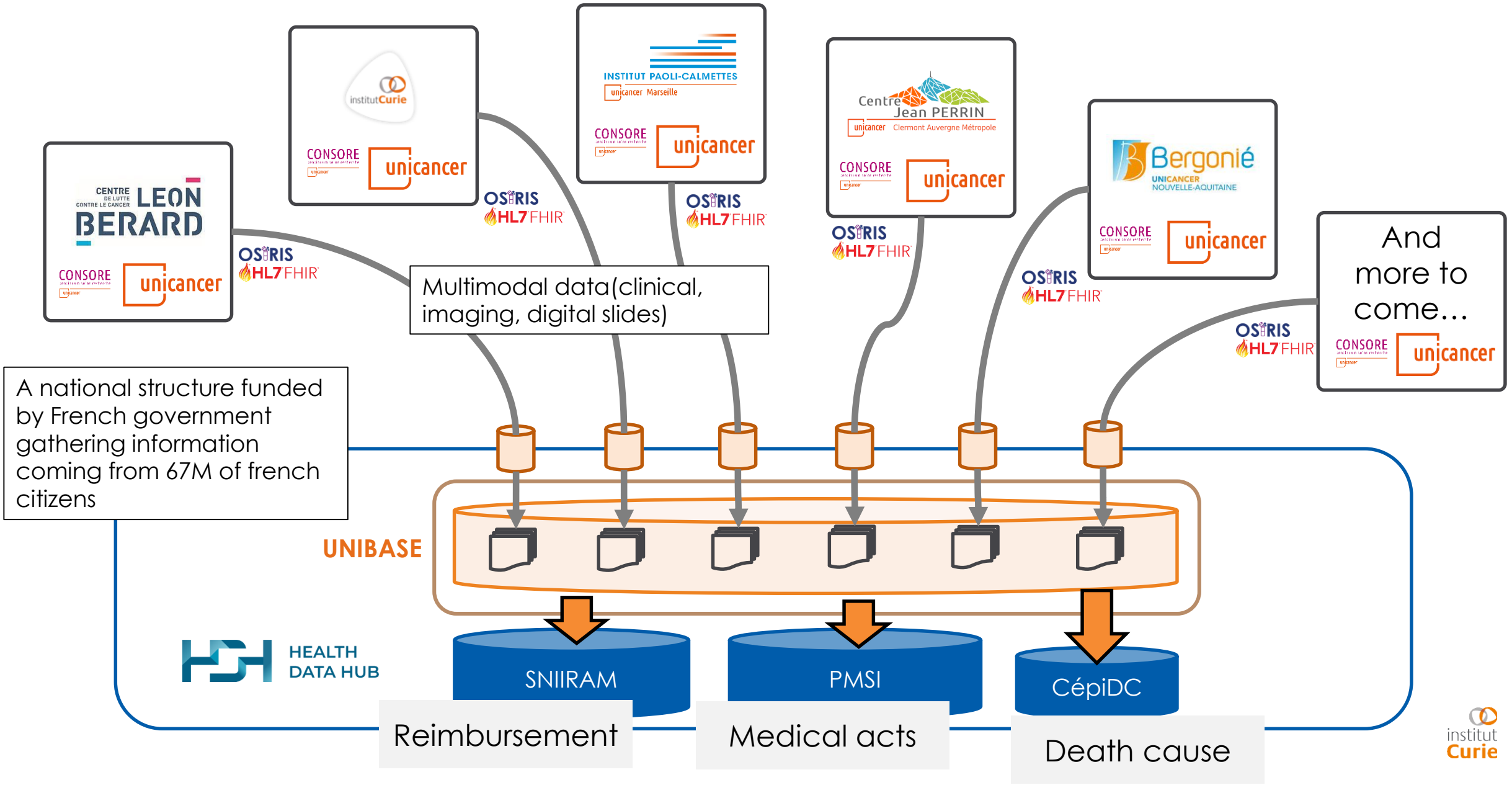
The UNIBASE project

CONSORE
continuum soins recherche
unicancer

- **Started in 2013 at Institut Curie**
- **10 deployed centers :**
  - Institut Curie (Paris)
  - Lyon
  - Marseille
  - Lille
  - Montpellier
  - Clermont-Ferrand
  - Dijon
  - Bordeaux
  - Nancy
  - Caen

institut Curie

The UNIBASE project



Thank you!

institut Curie

institutCurie

INSTITUT NATIONAL DU CANCER

INVESTISSEMENTS D'AVENIR

DigiCore
DIGITAL INSTITUT FOR CANCER OUTCOMES RESEARCH EUROPE

SiRIC Institut Curie
Site de Recherche Intégrée sur le Cancer

unicancer

Cancéropôle Île-de-France