

FAIRification of health-related data in the Swiss Personalized Health Network

Dr. sc. ETH Sabine Österle & Dr. Vasundra Touré

Personalized Health Informatics group,
SIB Swiss Institute of Bioinformatics

A project of



MILA Seminar, University of Greifswald

24-Jan-22



Outline

- Introduction to SPHN
- The SPHN Ecosystem components
 - The SPHN semantics
 - The standard exchange format
 - Terminology Service
 - Data validation framework
 - Teaching and training
- Future perspectives

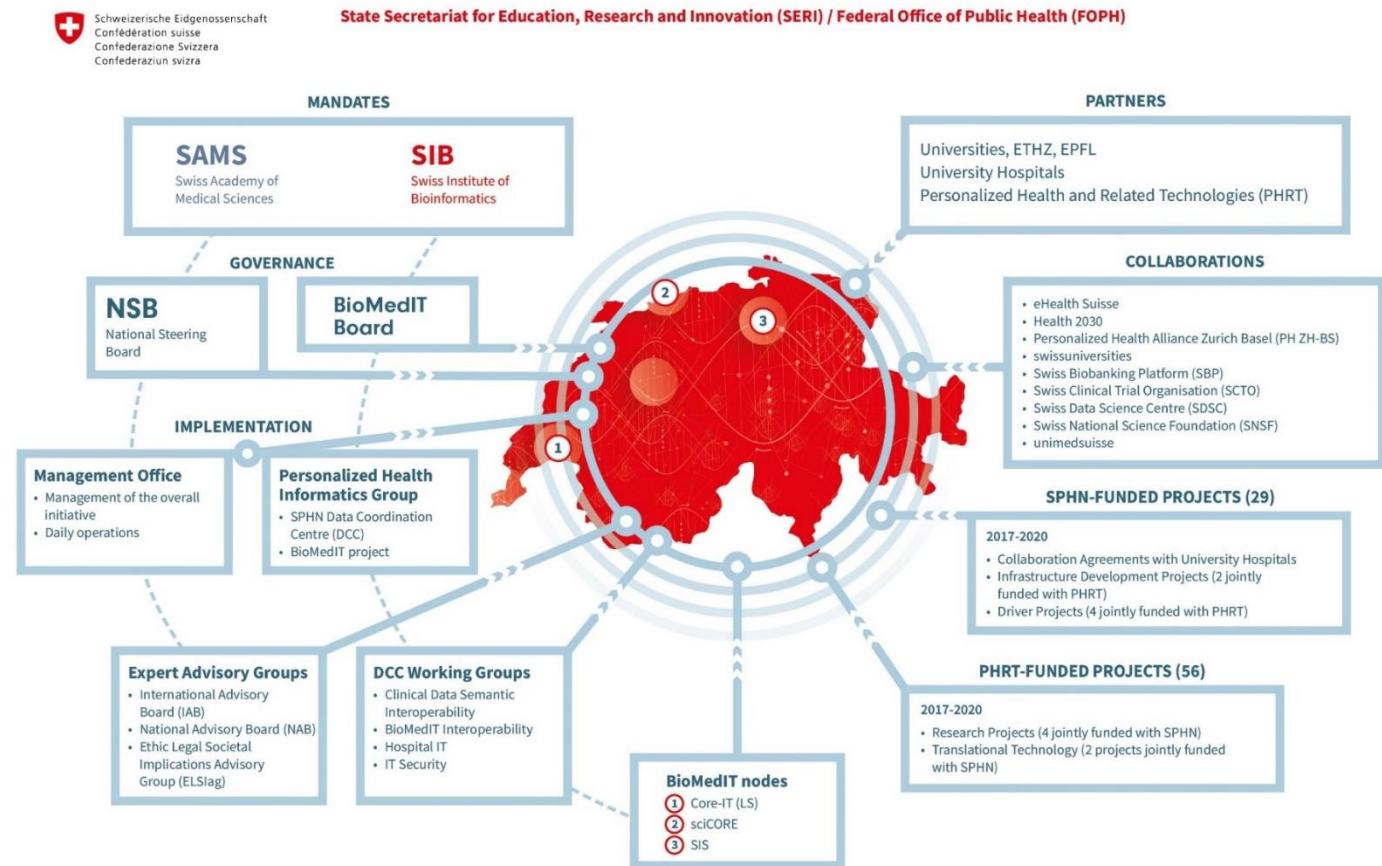
The Swiss Personalized Health Network (SPHN)

Government funded initiative

- 2017-2020: focus on 5 Swiss university hospitals, bottom-up PoC, and BioMedIT (CHF 68 Mio)
- 2021-2024: consolidate and expand the network, and BioMedIT (CHF 67 Mio)

Aim

Promote the development, implementation, and validation of coordinated infrastructures to make health data interoperable and shareable for research in Switzerland



SPHN network and SPHN mission

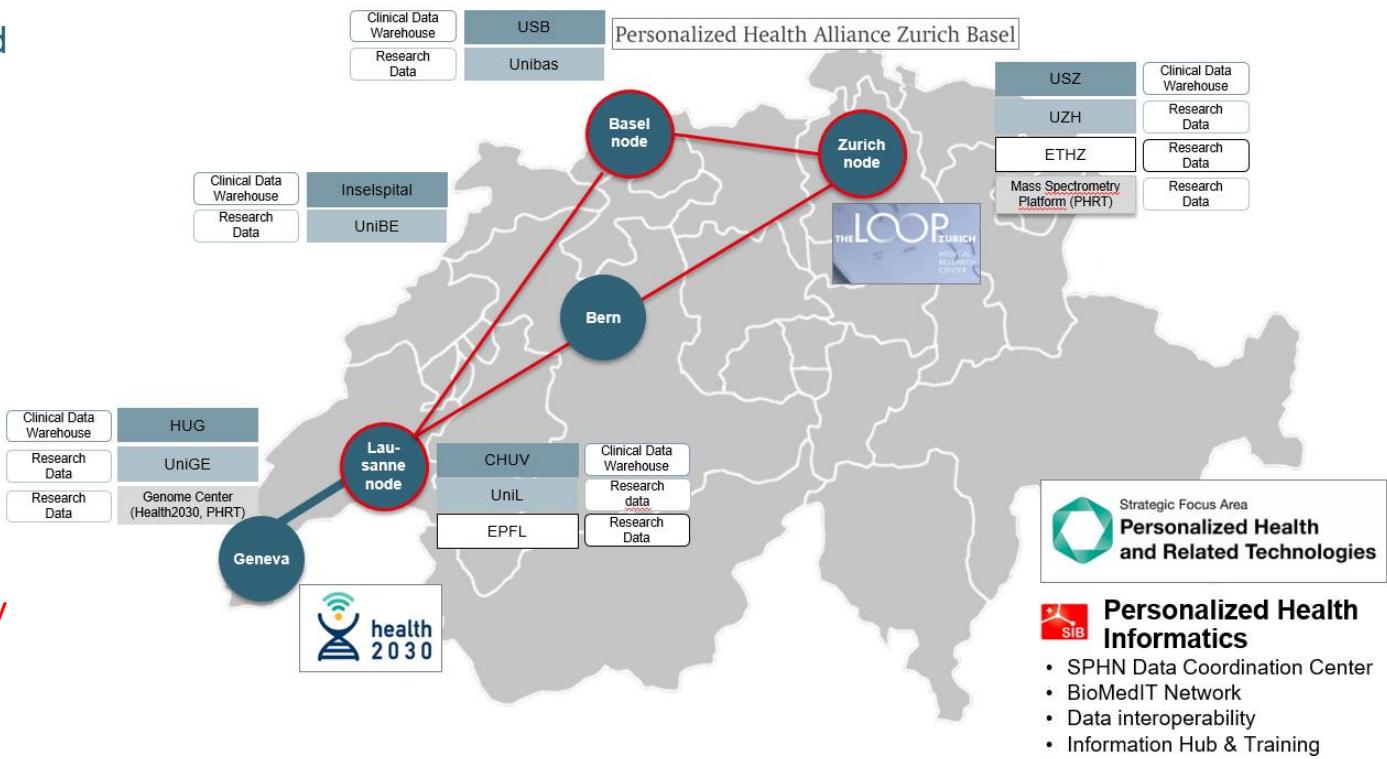
A decentralized approach: No central data lake. **Data remain with data providers**, only gathered for specific projects; federated query, distributed learning

Getting data out of silos: Working towards a **FAIR use of health data** for research; democratic access to data

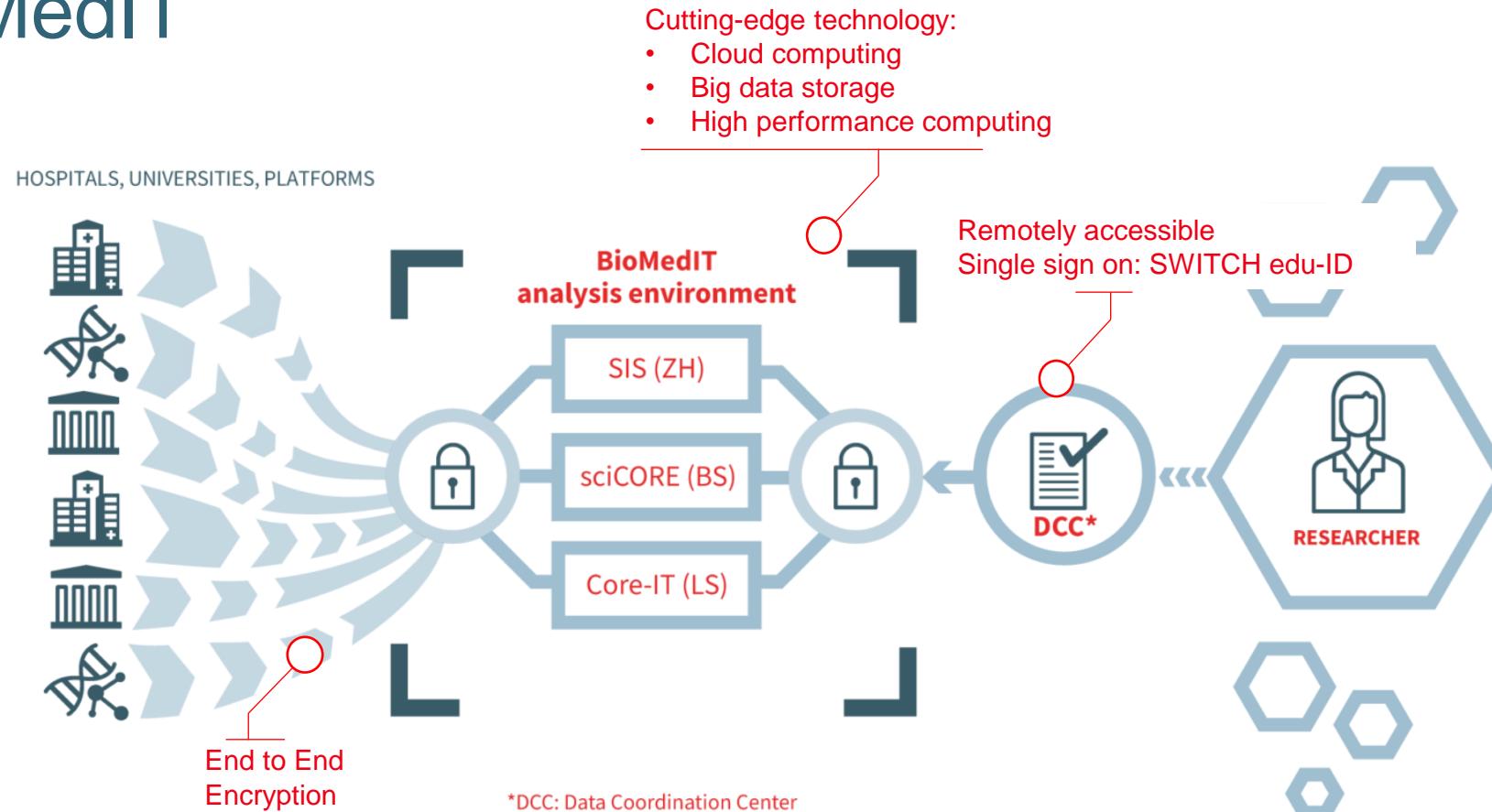
Building, connecting and aligning systems: Establishing a national infrastructure network and a **harmonized data governance strategy**

Interoperability framework: **Semantic strategy** and knowledge representation

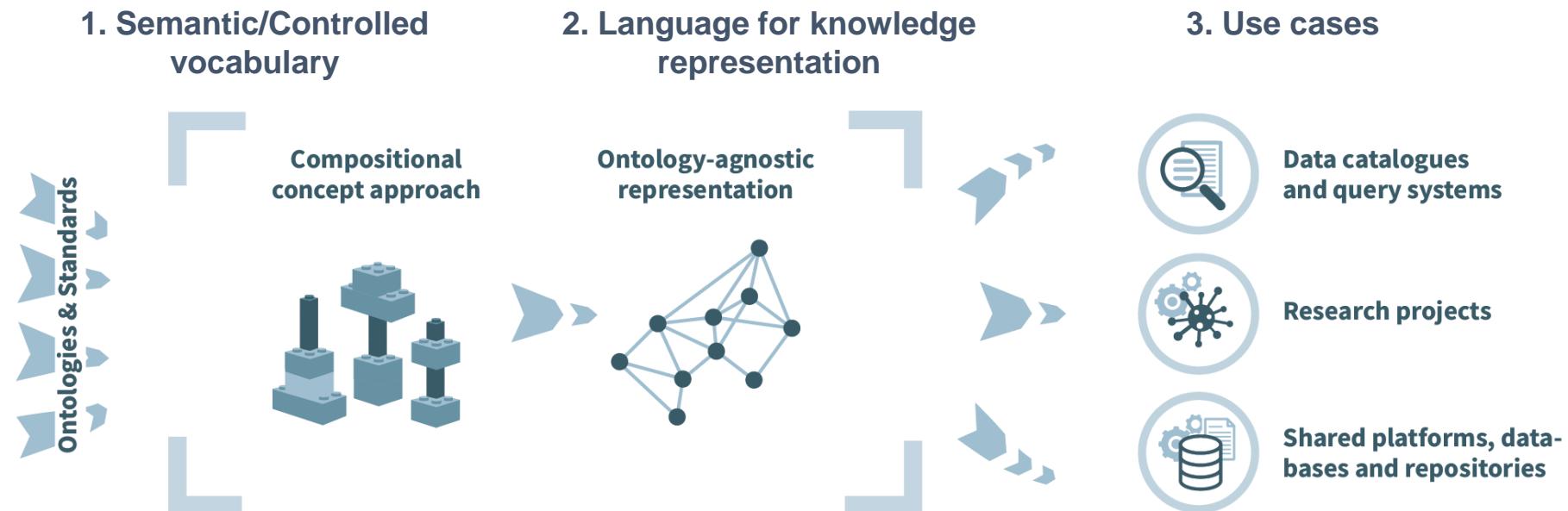
BioMedIT: **Secure HPC infrastructure** for data analysis



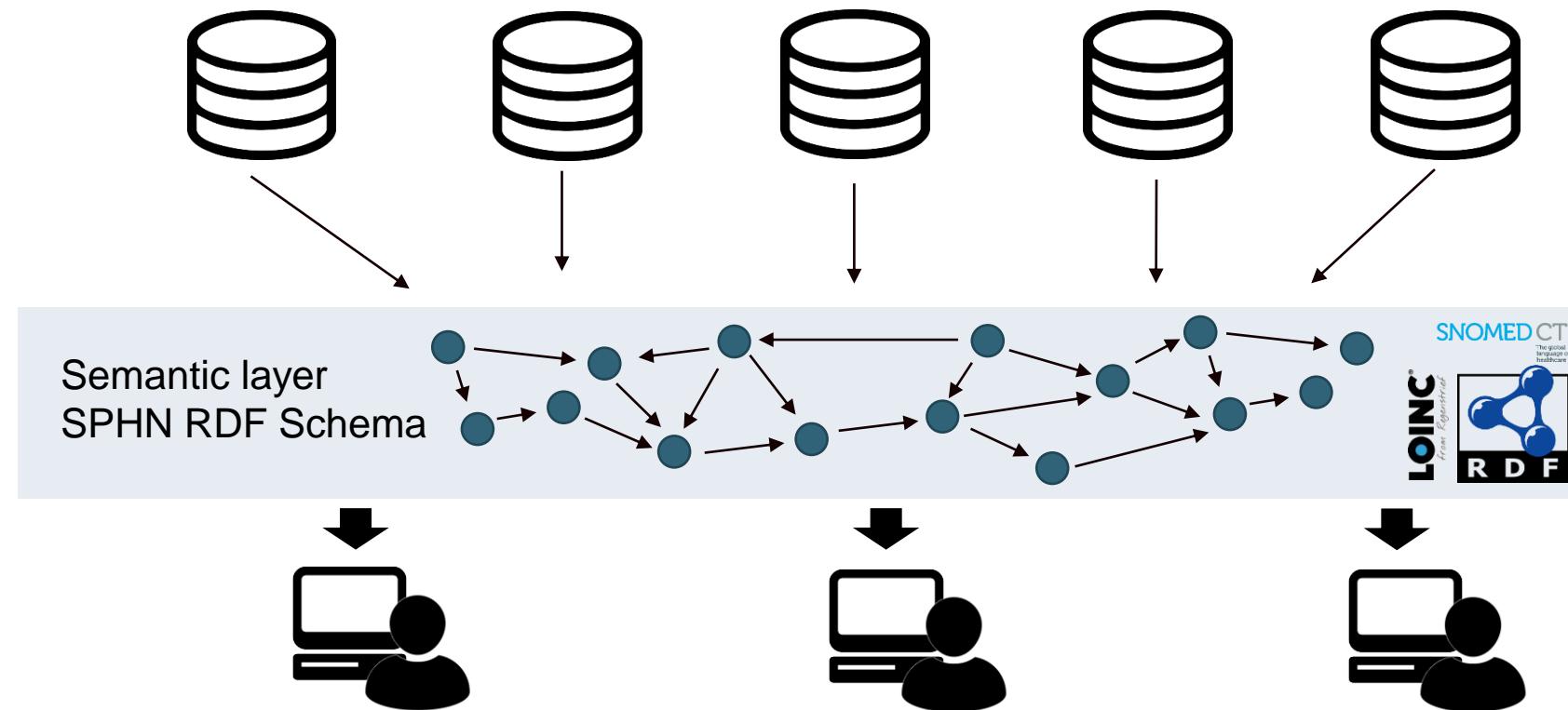
BioMedIT



Semantic strategy (three pillars)



How to bring data together in a FAIR manner?

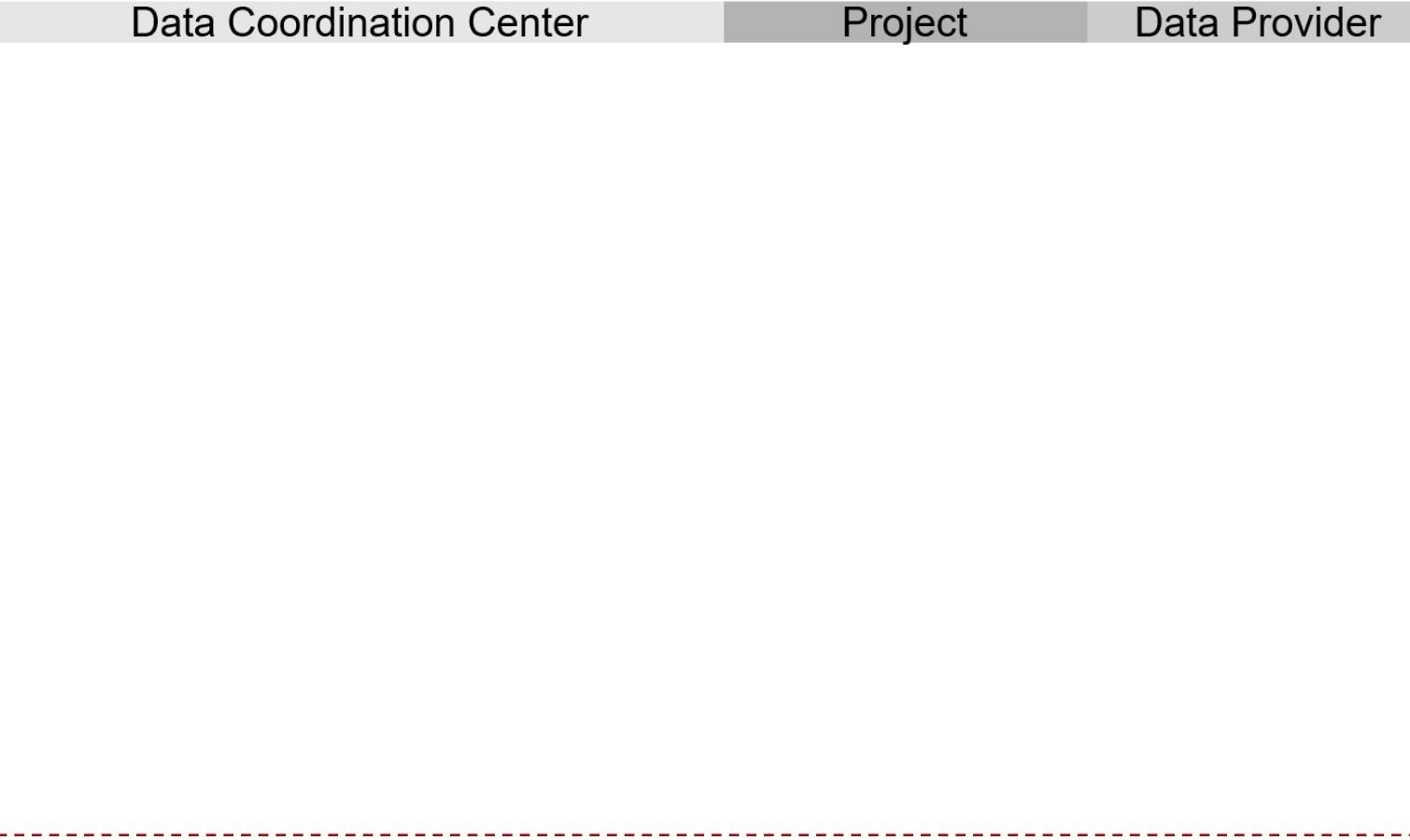


SPHN Ecosystem for FAIR data

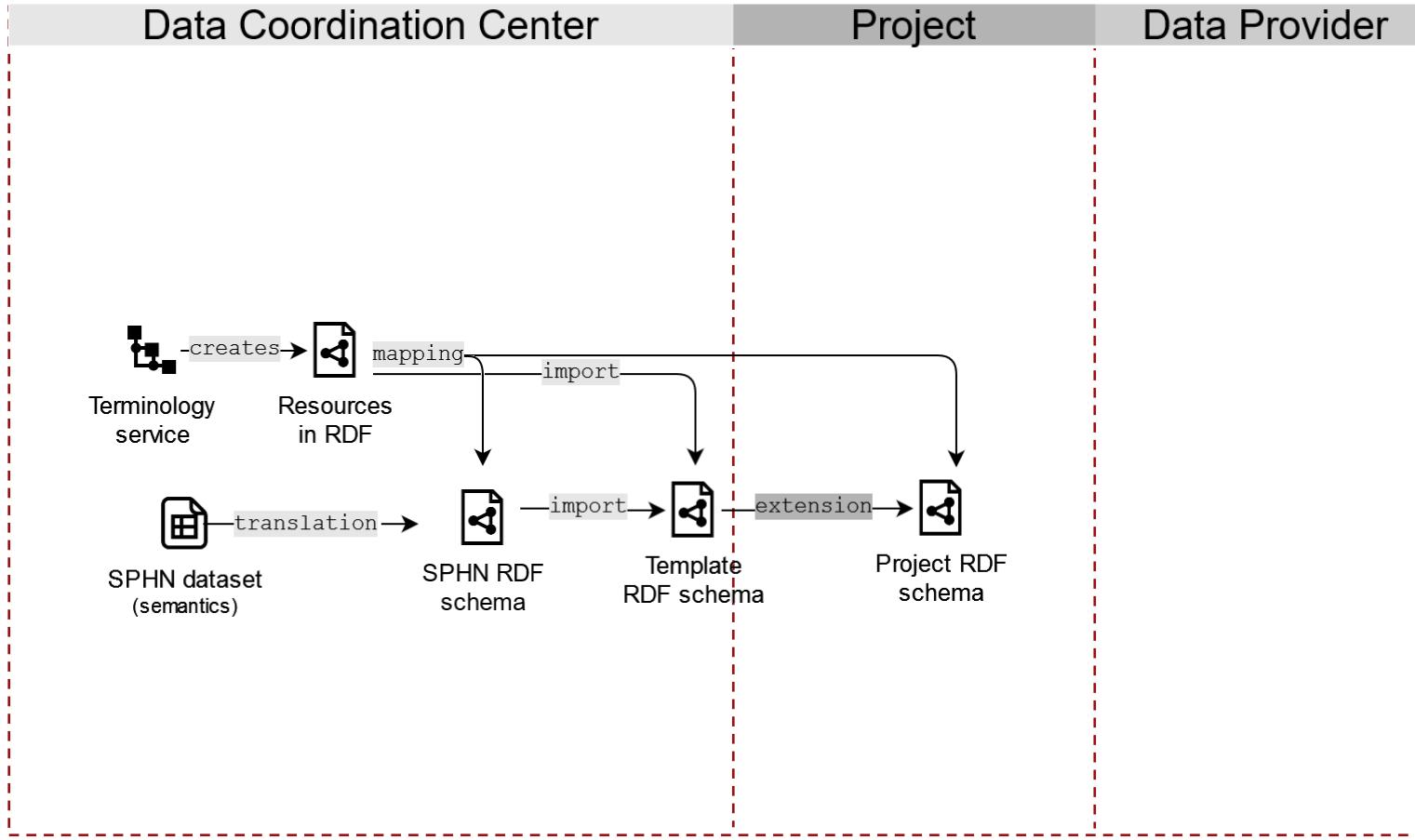
Data Coordination Center

Project

Data Provider



SPHN Ecosystem for FAIR data



SPHN Dataset: Semantic interoperability for clinical data

- Semantics of basic clinical concepts of interest with descriptions and value sets
- Fixation of meaning by pointers to semantic standards (e.g. SNOMED CT, LOINC)
- Collaborative approach - driven by research project needs
- Regular releases of the **SPHN Dataset** (58 compositional concepts in 2021 release)

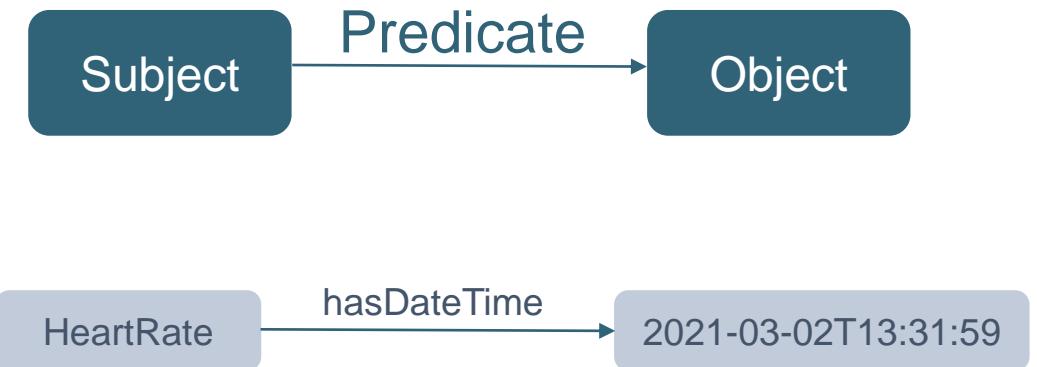
Composed of

Concept name	Description	Type	Meaning binding SNOMED CT LOINC Value set
Heart Rate	frequency of the heart beats, i.e. the number of time a heart beats per unit of time		364075005 Heart rate (observable entity) 8867-4 Heart rate
rate	measured heart rate, and time unit	Frequency	
datetime	datetime of measurement	temporal	
body site	body site where the heart rate was measured	Body Site	
method	method of heart rate measurement (e.g. palpation, stethoscope, ECG)	Measurement Method	
regularity	regular or irregular heart rate	Code	SNOMED CT: 271636001 Pulse regular (finding) ; 61086009 Pulse irregular (finding) ; 261665006 Unknown (qualifier value)

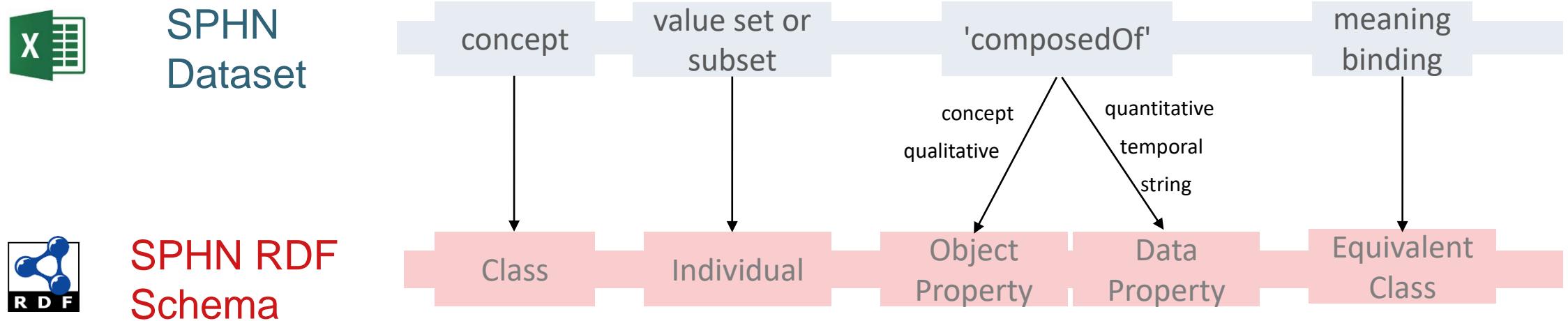
RDF as a standard exchange data format

Resource Description Framework

- Semantic Web technology
- Graph-based structure
 - Connectivity
 - Flexibility
 - Unique Identifiers
 - SPARQL querying language
 - Inference capabilities (reasoning)



From the semantics to RDF



The SPHN RDF schema

Accessible (under CC BY-NC-SA 4.0) at:

<https://biomedit.ch/rdf/sphn-ontology/sphn/2021/2>

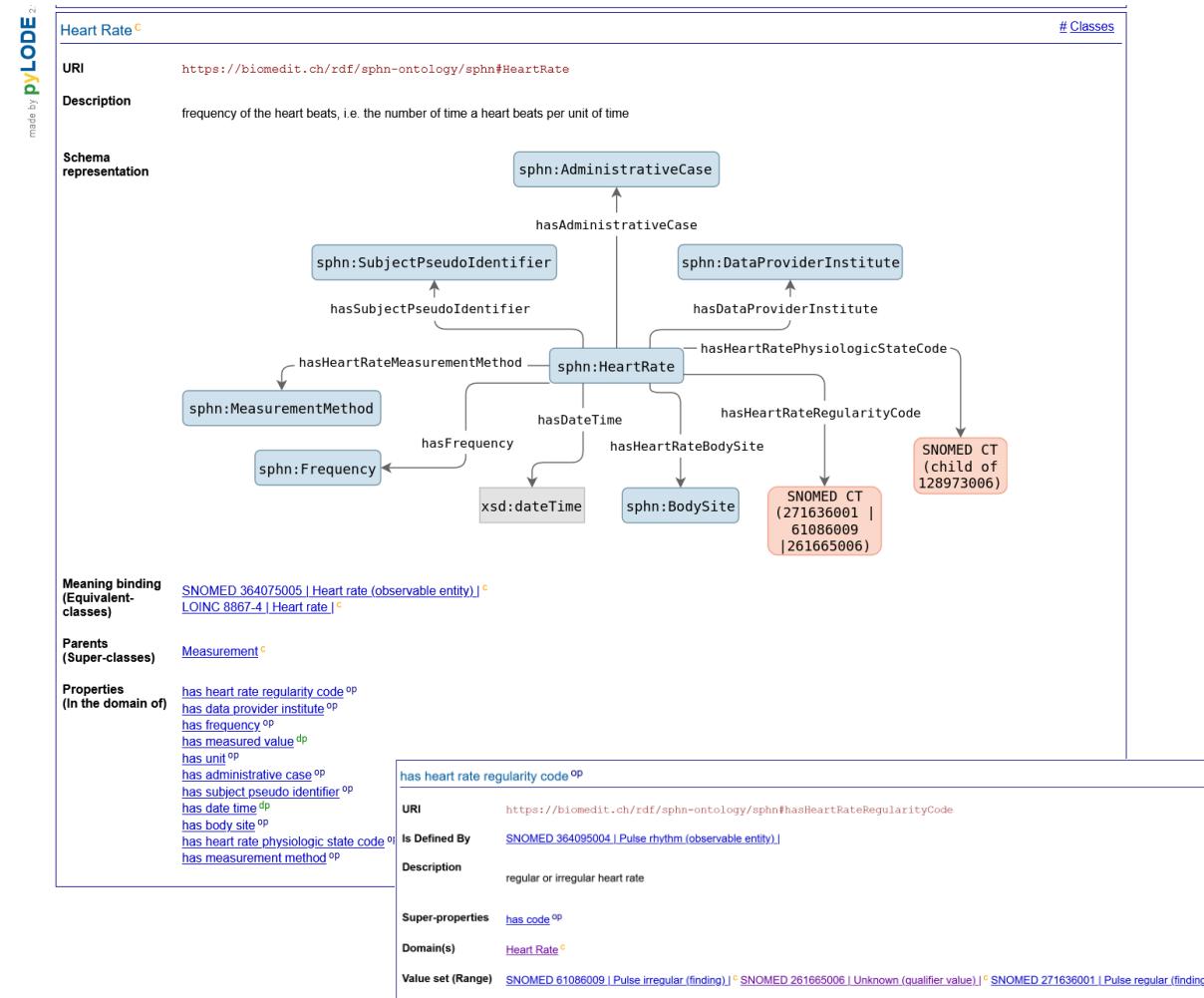
Stack tools used

Built from Protégé desktop

Collaborative editing possible on WebProtégé

Some numbers (release 2021.2)

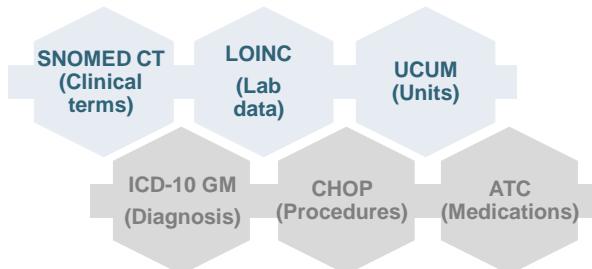
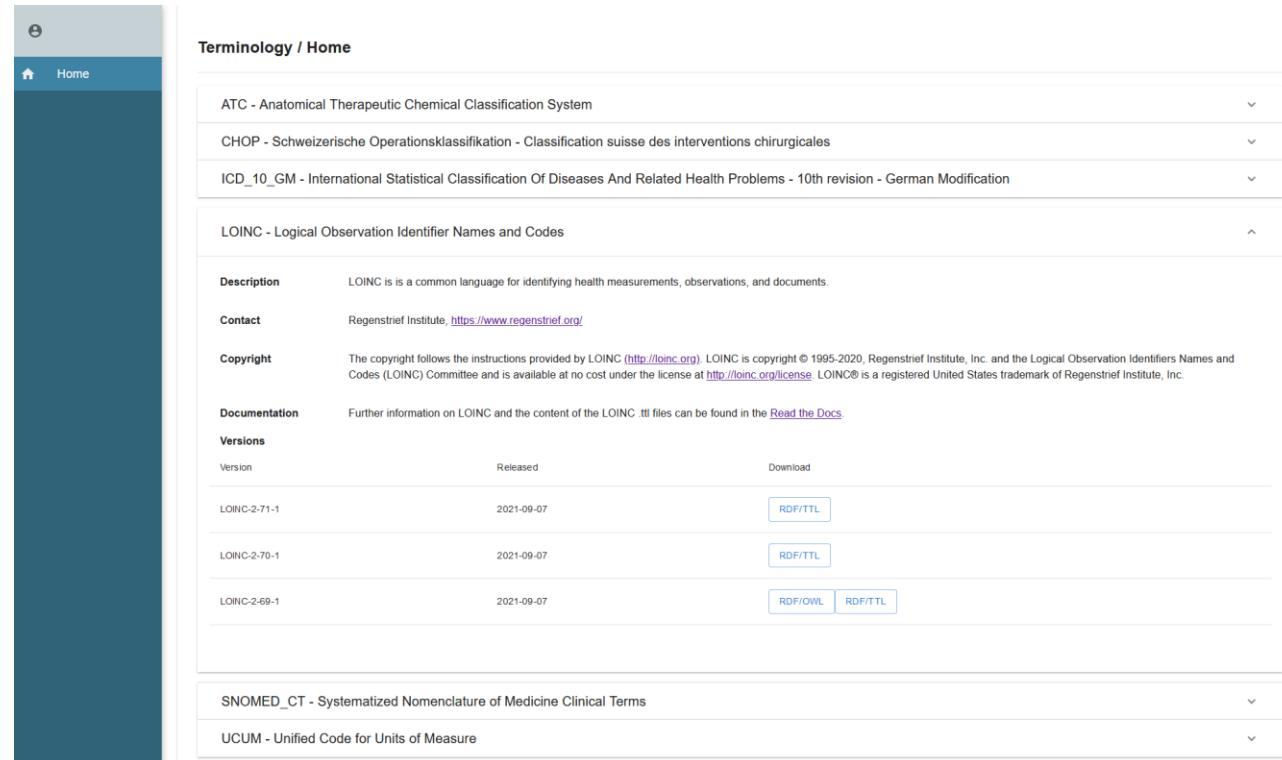
87 Classes & 183 properties



Terminology service: Bringing standards to SPHN

Automated pipeline for converting clinical resources in RDF

- RDF's made accessible in BioMedIT portal for data users (individual researchers)
- Also as a MinIO service for downloading a bundle for data providers (hospitals and service providers)

Terminology / Home

ATC - Anatomical Therapeutic Chemical Classification System

CHOP - Schweizerische Operationsklassifikation - Classification suisse des interventions chirurgicales

ICD_10_GM - International Statistical Classification Of Diseases And Related Health Problems - 10th revision - German Modification

LOINC - Logical Observation Identifier Names and Codes

Description LOINC is a common language for identifying health measurements, observations, and documents.

Contact Regenstrief Institute, <https://www.regenstrief.org/>

Copyright The copyright follows the instructions provided by LOINC (<http://loinc.org>). LOINC is copyright © 1995-2020, Regenstrief Institute, Inc. and the Logical Observation Identifiers Names and Codes (LOINC) Committee and is available at no cost under the license at <http://loinc.org/license>. LOINC® is a registered United States trademark of Regenstrief Institute, Inc.

Documentation Further information on LOINC and the content of the LOINC ttl files can be found in the [Read the Docs](#).

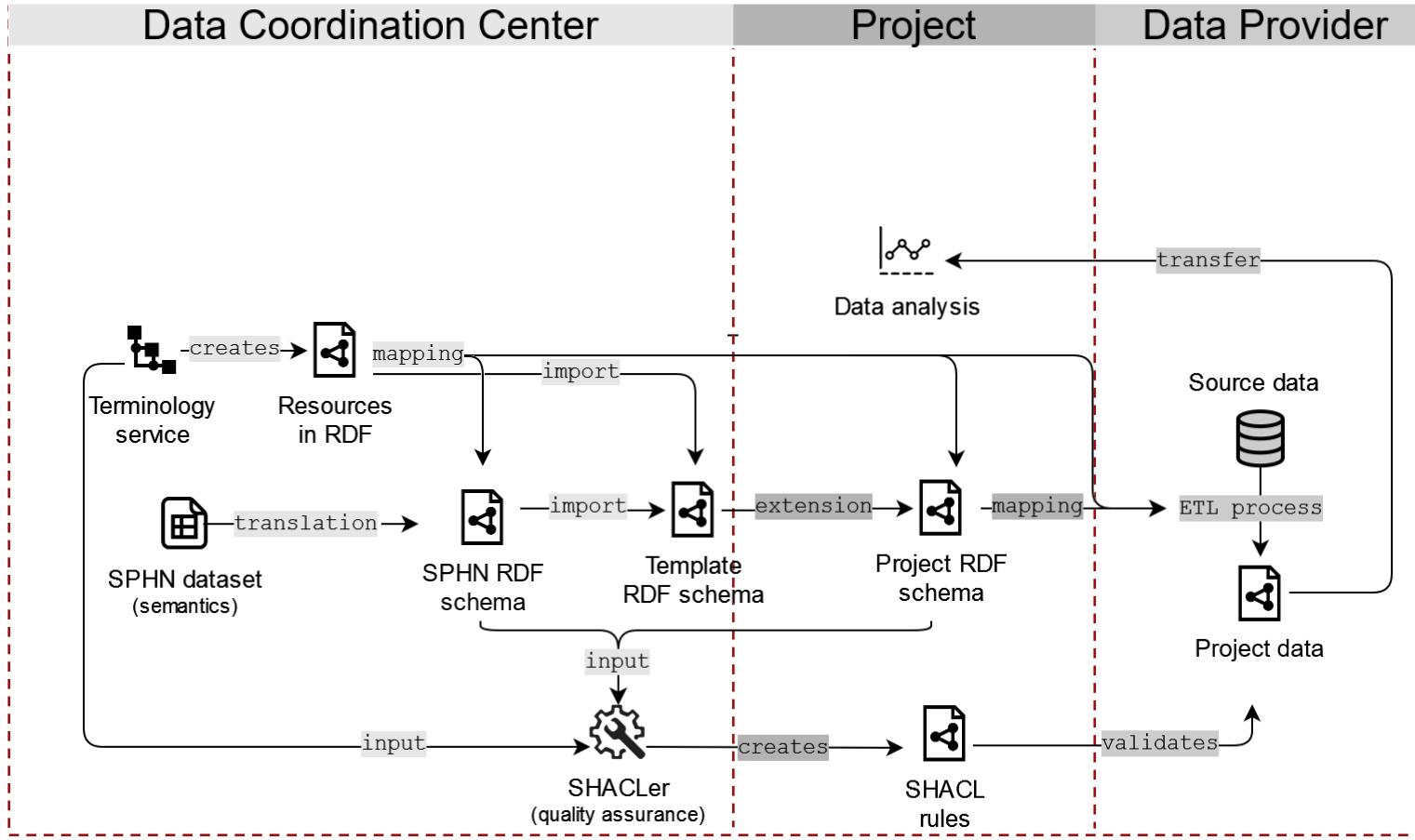
Versions

Version	Released	Download
LOINC-2-71-1	2021-09-07	RDF/TTL
LOINC-2-70-1	2021-09-07	RDF/TTL
LOINC-2-69-1	2021-09-07	RDF/OWL RDF/TTL

SNOMED_CT - Systematized Nomenclature of Medicine Clinical Terms

UCUM - Unified Code for Units of Measure

SPHN Ecosystem for FAIR data



Quality control framework

SHACLeR

Automatically generates SHACL rules for a given schema to validate data against SPHN RDF schema



Examples of SHACL core constraints currently in SPHN

Constraint Type	Constraints (namespace sh: is omitted for readability)	SPHN application example
Value Type	class, datatype, nodeKind, targetClass	hasBodySiteCode property points to either an instance or subinstance of snomed:123037004
Cardinality	minCount, maxCount	Only one SubjectPseudoidentifier is connected to a BodyHeight
Values	node, in, hasValue, path	Death_status is only allowed to be instantiated by the individuals of sphn:Death or sphn:Unknown

SHACL exceptions implemented in SPHN

- Not instantiable classes

```
{
  "type" : "notInstantiableClass",
  "class" : "https://biomedit.ch/rdf/sphn-ontology/sphn#Measurement"
}
```

- Range extension

```
{
  "type" : "rangeExtension",
  "property" : "https://biomedit.ch/rdf/sphn-ontology/sphn#hasLabResultLabTestCode",
  "class" : "https://biomedit.ch/rdf/sphn-ontology/sphn#LabResult",
  "extendedRange" : "https://biomedit.ch/rdf/sphn-ontology/sphn#Code"
}
```

- Not instantiable properties

```
{
  "type" : "notInstantiableProperty",
  "property" : "https://biomedit.ch/rdf/sphn-ontology/sphn#hasValue",
  "class" : null
}
```

Quality control framework

Statistical queries

Count patients per provider

```
SELECT DISTINCT ?dataProvider ?dataProviderName (COUNT(?patient) AS ?nbOfPatients)
WHERE {
  ?patient a sphn:SubjectPseudoidentifier .
  ?dataProvider a sphn:DataProviderInstitute .
  ?patient sphn:hasDataProviderInstitute ?dataProvider .
  ?dataProvider sphn:hasDataProviderInstituteCode/sphn:hasCodeName
    ?dataProviderName .
} GROUP BY ?dataProvider ?dataProviderName
```

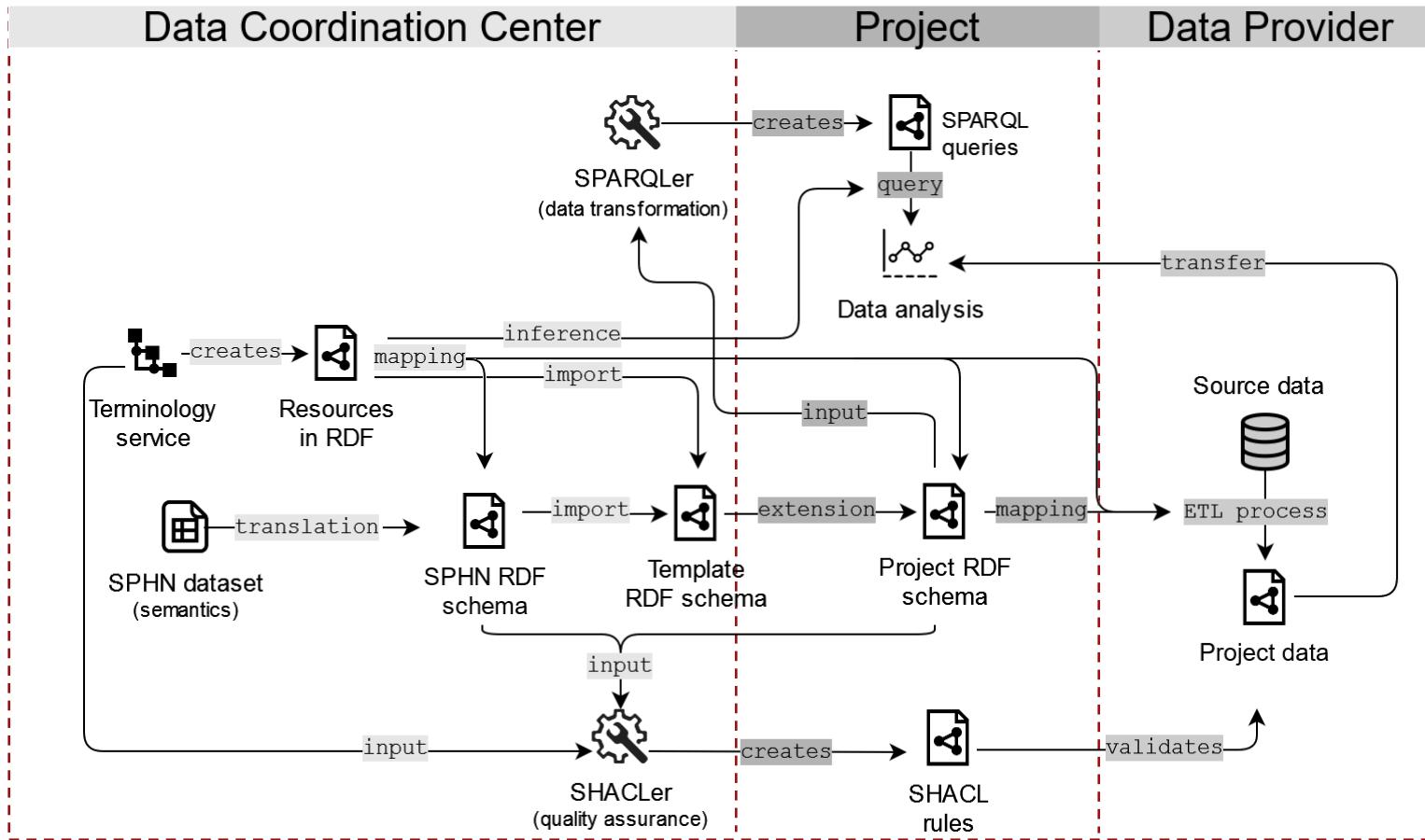
	dataProvider	dataProviderName	nbOfPatients
1	resource:CHE-108_907_884-DataProviderInstitute	"Les hôpitaux universitaires de Genève"	"2033"^^xsd:integer
2	resource:CHE-108_904_325-DataProviderInstitute	"University Hospital Zurich"	"1977"^^xsd:integer

Count of concepts' instances & min/max values

```
SELECT ?concept ?attribute ?range (COUNT(distinct ?value) AS ?sphn_objects_count)
(MIN(?value) AS ?min_value) (MAX(?value) AS ?max_value) (AVG(?value) AS ?avg_value)
WHERE {
  optional { ?attribute rdfs:range ?range }
  ?concept rdfs:subClassOf+ sphn:SPHNConcept .
  ?attribute rdfs:subPropertyOf sphn:SPHNAttributeDatatype .
  ?resource a ?concept .
  ?resource ?attribute ?value.
} group by ?concept ?attribute ?range order by ?concept ?attribute ?range
```

	concept	attribute	range	sphn_objects_count	min_value	max_value	avg_value
1	sphn:AdministrativeCase	sphn:hasIdentifier	xsd:string	"120"^^xsd:integer	"case1"	"case99"	
2	sphn:Code	sphn:hasCodeCodingSystemAndVersion	xsd:string	"1"^^xsd:integer	"UID"	"UID"	
3	sphn:Code	sphn:hasCodeName	xsd:string	"5"^^xsd:integer	"Fondation CHUV"	"University Hospital Zurich"	

SPHN Ecosystem for FAIR data



SPARQLer:
Data exploration
&
Data validation

Quality control framework

SPARQLer

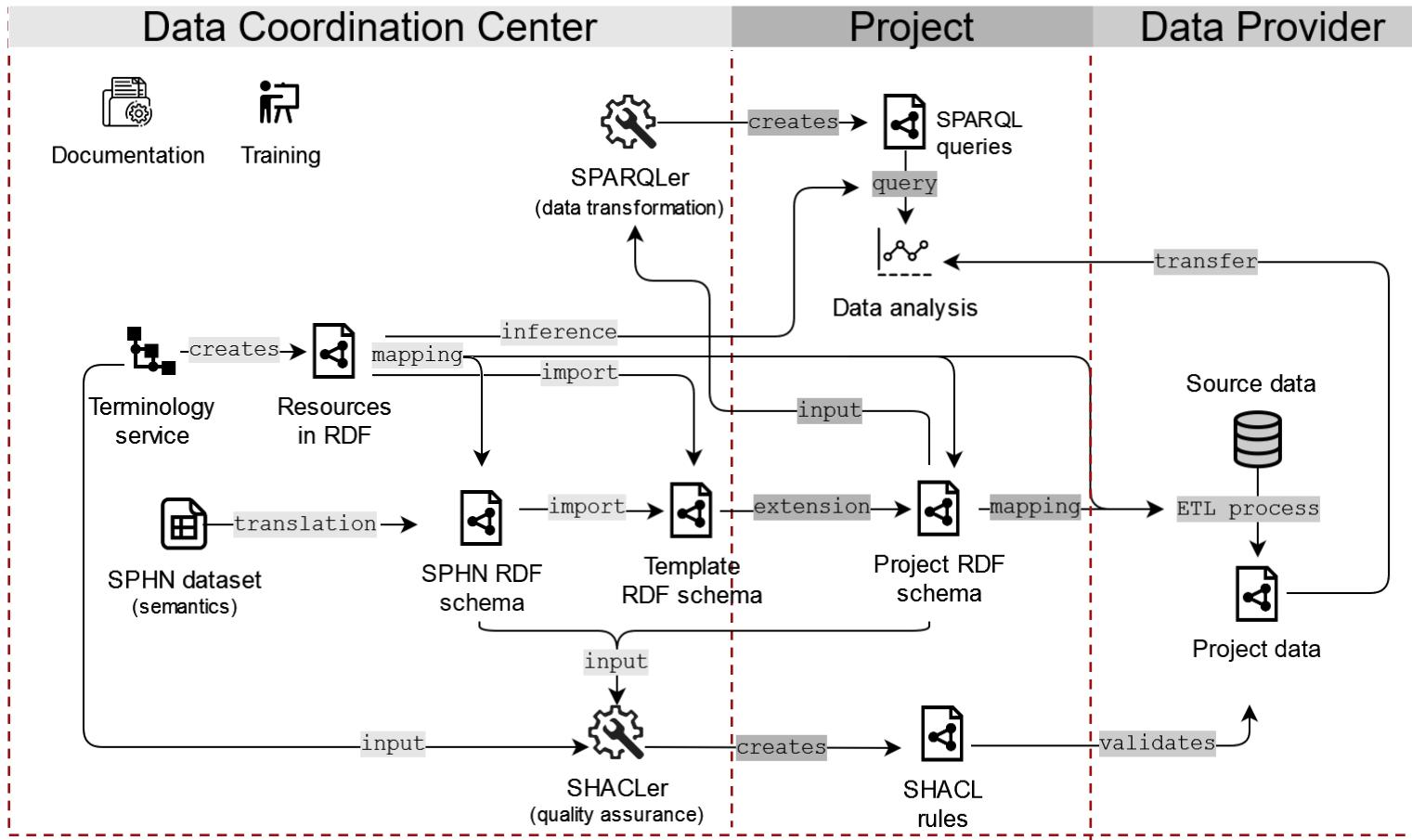
Automatically generate SPARQL queries for
‘data flattening’

```
PREFIX sphn:<https://biomedit.ch/rdf/sphn#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX lab: <http://purl.org/lab/vocab#>
PREFIX ucum: <http://unitsofmeasure.org/ucum#>
PREFIX loinc: <http://loinc.org/>
PREFIX che: <http://purl.org/che/ontology/>
PREFIX labresult: <http://purl.org/che/ontology/labresult#>

SELECT *
WHERE {
  ?resource a sphn:LabResult .
  optional{ ?resource sphn:hasSubject ?subjectPseudoidentifier }
  optional{ ?resource sphn:hasLabTest ?labResultLabTestCode }
  optional{ ?resource sphn:hasLabResultValue ?labResultValue }
  optional{ ?resource sphn:hasLabResultAnalysisDateTime ?labResultAnalysisDateTime }
  optional{ ?resource sphn:hasLabResultNormalRange ?labResultNormalRange }
  optional{ ?resource sphn:hasLabTestCode ?labResultLabTestCode }
  optional{ ?resource sphn:hasLabTestUnit ?labResultUnit }
  optional{ ?resource sphn:hasLabTestCode ?labResultLabTestCode }
  optional{ ?resource sphn:hasLabResultValue ?labResultValue }
  optional{ ?resource sphn:hasLabResultAnalysisDateTime ?labResultAnalysisDateTime }
  optional{ ?resource sphn:hasLabResultNormalRange ?labResultNormalRange }
  optional{ ?resource sphn:hasLabTestUnit ?labResultLabTestUnit }
  optional{ ?resource sphn:hasLabTestCode ?labResultLabTestCode }
  optional{ ?resource sphn:hasLabResultValue ?labResultValue }
  optional{ ?resource sphn:hasLabResultAnalysisDateTime ?labResultAnalysisDateTime }
  optional{ ?resource sphn:hasLabResultNormalRange ?labResultNormalRange }
}
```

	resource	subjectPseudoidentifier	labResultLabTestCode	labResultUnit	labResultValue	labResultAnalysisDateTime	labResultNormalRange
	_hasIdentifier	_hasIdentifier	hasLabResultLabTestCo	_hasUnitCode	hasLabResultValu	hasLabResultAnalysi	hasLabResultNormalRan
1	resource:CHE-108_904_325-LabResult-labresult1495	"patient1495"	resource:Code-LOINC-6690-2	resource:ucum	"23"	"2020-02-15 10:34:34"^^xsd:dateTime	
2	resource:CHE-108_904_325-LabResult-labresult8050	"patient8050"	resource:Code-LOINC-6690-2	resource:ucum	"36"	"2020-11-01 16:34:17"^^xsd:dateTime	
3	resource:CHE-108_904_325-LabResult-labresult9729	"patient9729"	resource:Code-LOINC-26464-8	resource:ucum	"115"	"2020-12-22 04:46:01"^^xsd:dateTime	

SPHN Ecosystem for FAIR data



Importance of teaching and training

Comprehensive documentation ([readthedocs](#))

- SPHN Ecosystem
- Dataset + RDF schema implementation
- External Terminologies
- User guide
- Examples of implementation in hospitals

Training events ([webinars](#))

- SPHN Data Ecosystem for FAIR Data
- Semantic Standards
- Training Primer (RDF and SPARQL)
- Expanding the SPHN RDF Schema
- RDF Schema and Data Visualization
- Querying Data with SPARQL
- How to use Python and R with RDF Data
- Validate Graph Data with SHACL

SPHN Ecosystem for FAIR data

Getting data out of silos by linking data: **FAIR health data** for research

- F1: Unique Identifier → URI in the RDF
- F2: Rich meta data → SPHN concepts and general meta-data in RDF
- A1: Universal query language → SPARQL
- I1: Language for knowledge representation → RDF
- I2: Controlled vocabulary → LOINC, SNOMED CT ...
- I3: Linked data → RDF via common URIs or controlled vocabulary

SPHN Ecosystem for FAIR data

Getting data out of silos by linking data: FAIR health data for research

- F1: Unique Identifier → URI in the RDF
- F2: Rich meta data → SPHN concepts and general meta-data in RDF
- F3: Link data and meta-data
- F4: Register meta-data in a reachable resources
- A1: Universal query language → SPARQL
- I1: Language for knowledge representation → RDF
- I2: Controlled vocabulary → LOINC, SNOMED CT ...
- I3: Linked data → RDF via common URIs or controlled vocabulary
- R1: Reuse of data

Future perspectives

- FAIR meta-data endpoint/ Meta-data catalogue and repository for datasets
- SPHN HospFAIR
- National data streams

Acknowledgements

The PHI Group:

Katrin Crameri, Shubham Kapoor, Julia Maurer,
 Michael Müller-Breckenridge, Kristin Gnadtke, Jan
 Armida, Petar Horki, Martin Fox, Simone Guzzi,
 Patricia Fernandez Pinilla, Christian Ribeaud

The Hospital workforces @ USZ, USB, CHUV, Insel, HUG

Philip Krauss and Martin Zablocki @ Trivadis part
 of Accenture

-  <https://bit.ly/3K7VH0Q>
-  <https://bit.ly/31ng7RW>
-  <https://bit.ly/34mdNfh>
-  [@SPHN_ch](#)
-  sabine.oesterle@sib.swiss
vasundra.toure@sib.swiss
-  www.sphn.ch | www.sib.swiss/phi