Norwegian Centre for
E-health Research

# Towards National Scale Clinical Data Warehousing: Experiences and Perspectives on Data Reuse for the Learning Healthcare System

Luis Marco Ruiz, PhD [1,2]

[1]Norwegian Centre for Integrated Care and Telemedicine

[2] Medizinischen Hochschule Hannover

# Background

- Computer Scientist, MSc Applied Statistics, PhD Health Science (Medical Informatics)
- Since 2013 working for the Norwegian Centre for E-health Research (NSE)
- Postdoctoral research fellow and advisor for clinical information standardization
- Semantic Interoperability Specialist at the Hannover Medical School (HiGHmed project)

# Agenda

- The Learning Healthcare System
- Conceptualization of clinical data warehousing (DW)
- Computational models involved
- Information representation for clinical DW
- Knowledge representation for clinical DW
- Experiences in Norway and Germany
- Organization and challenges
- Acknowledgements

# Agenda

- The Learning Healthcare System
- Conceptualization of clinical data warehousing (DW)
- Computational models involved
- Information representation for clinical DW
- Knowledge representation for clinical DW
- Experiences in Norway and Germany
- Organization and challenges
- Acknowledgements

# The Learning Healthcare System

Evidence Based Medicine paradigm is facing many challenges (Djulbegovic 2009, Greenhalgh 2014). Among others, it is needed to decrease the 17 years that currently elapse since knowledge is generated until it is exploited at the bedside (Balas 2000, Friedman 2010).

The IOM proposed in 2007 the Learning healthcare system as a paradigm to overcome these challenges. The LHS heavily relies on the use of technologies for (IOM 2007, ESF 2016):

- Implementing that knowledge to exploit latest evidence at several levels (clinicians, patients, citizens, populations).
- Providing communication channels and tools across the participants in care processes.
- **Facilitating secondary use of data to generate new knowledge.**
  **->data reuse networks and Clinical Data Warehousing (DW)**

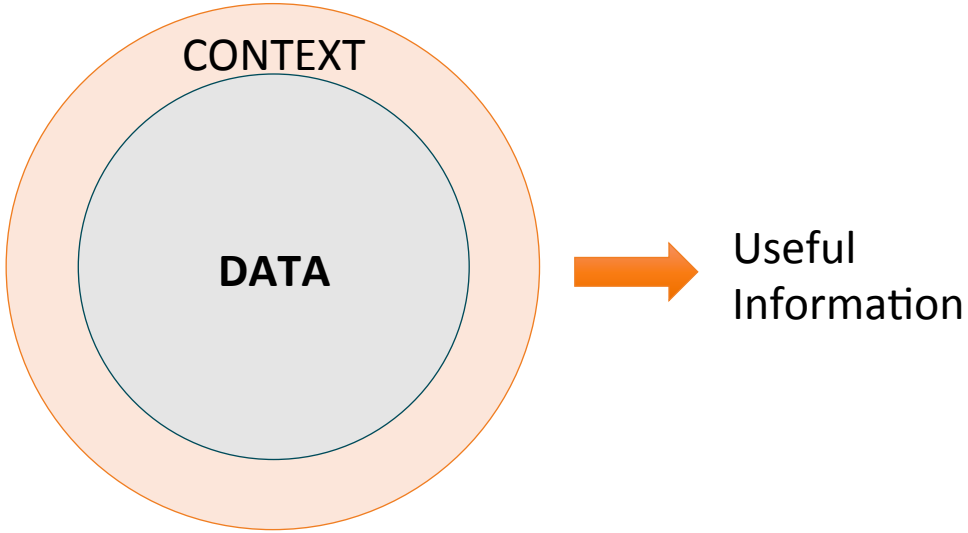
Budrionis A, Bellika JG. The Learning Healthcare System: Where are we now? A systematic review. J Biomed Inform 2016;64:87–92. doi:10.1016/j.jbi.2016.09.018.

# Agenda

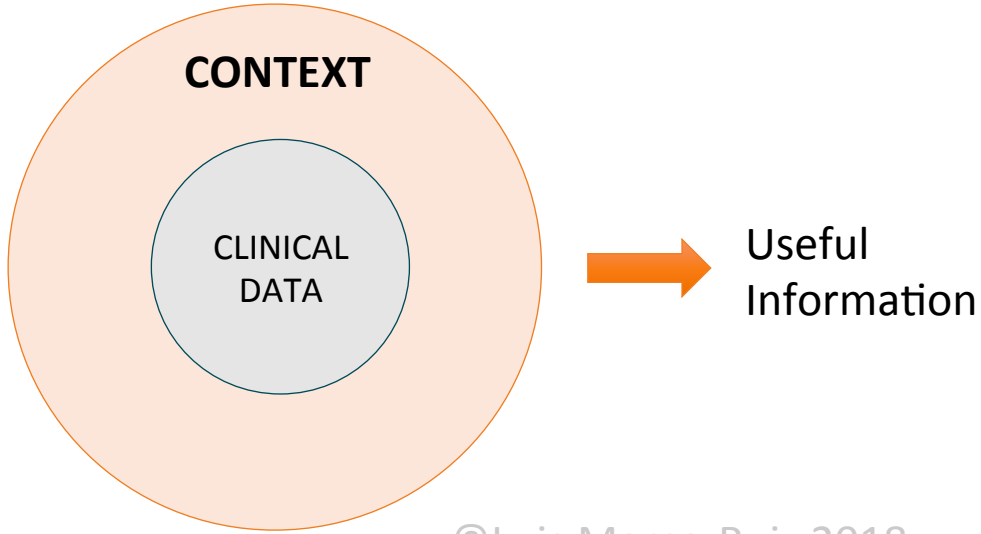
- The Learning Healthcare System
- Conceptualization of clinical data warehousing (DW)
- Computational models involved
- Information representation for clinical DW
- Knowledge representation for clinical DW
- Experiences in Norway and Germany
- Organization and challenges
- Acknowledgements

# Clinical Data Warehousing

- In the recent years the reuse of clinical data has received lot of attention for implementing the LHS.
- This has led to the birth of many data reuse networks and the Clinical Data Warehousing (DW) discipline.
- Clinical DW has significant differences with respect to traditional enterprise DW.
- While in enterprise DW the focus is on dealing with large amounts of data, clinical DW focuses on preserving the meaning of data.




Enterprise DW

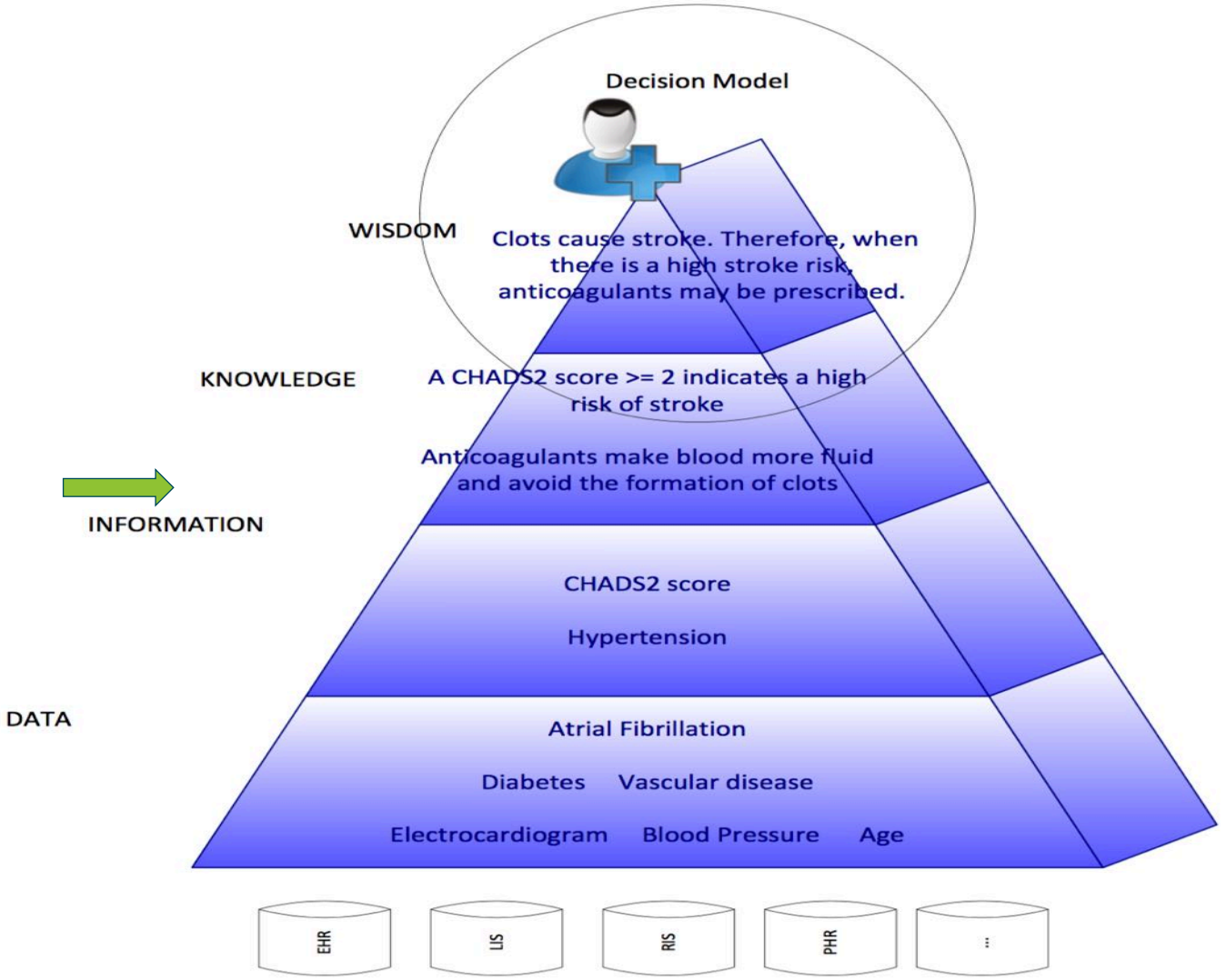



Clinical DW

# Generating knowledge from data

Semantics (concepts)

Data Analytics

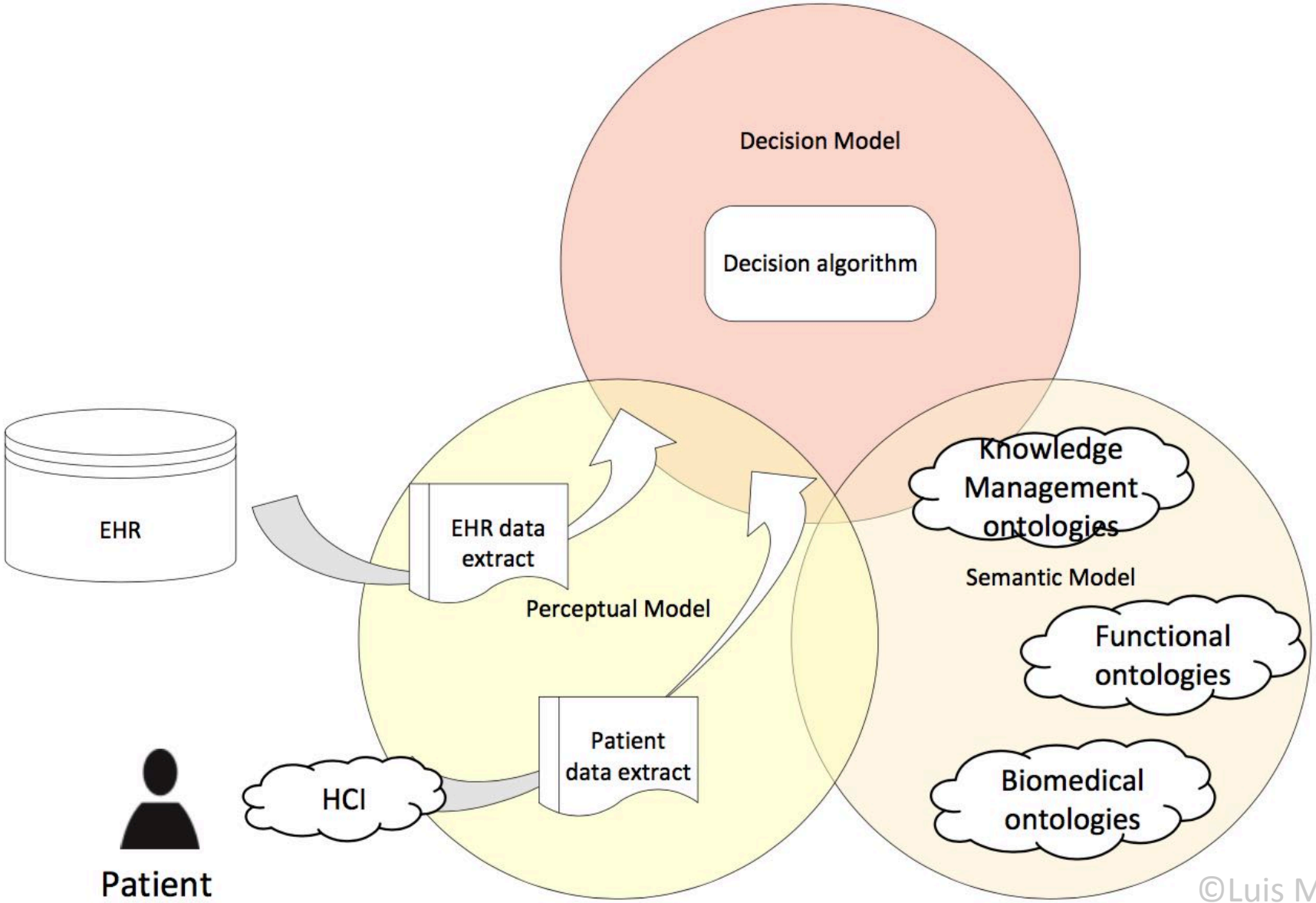



(Adapted from Sheth et al. 2013)

# Agenda

- The Learning Healthcare System
- Conceptualization of clinical data warehousing (DW)
- Computational models involved
- Information representation for clinical DW
- Knowledge representation for clinical DW
- Experiences in Norway and Germany
- Organization and challenges
- Acknowledgements

# Computational models involved



# Representing information and knowledge

**Data access is necessary but not enough…**

- The meaning of information when accessing it across different institutions must be preserved.
- Data users(e.g. researchers, health indicators, etc.) need to be able to understand the meaning of the information accessed.
- Data users may belong to different organizations.
- A lack of precision in communicating the meaning of a dataset may result in inaccurate or erroneous conclusions in quality studies irrespectively of the algorithm used.

**AVOID AMBIGUITY! MAXIMIZE DATA QUALITY!**

# Representing information and knowledge

- In order to preserve the meaning of information across disparate organizations, representation mechanisms (languages) for both the information and the meaning are needed.
- Clinical Information standards are used to represent information structures.
- Description Logics are used to define ontologies that are formal (mathematical) representations of the semantics conveyed by clinical information models.

**Information representation**

openEHR

CDA Clinical Document Architecture

HL7® FHIR®

**Knowledge representation**

SNOMED CT The global language of healthcare

OMIM®

hpo

# Expressivity needs of phenotyping algorithms

| Feature | Ratio of algorithms using the feature |
|---|---|
| Simply boolean | 12/14 |
| Negation | 8/14 |
| Cardinality | 6/14 |
| Temporal proximity | 10/14 |
| Complex temporal | 10/14 |
| External codes | 14/14 |
| Collection of codes, results etc. | 12/14 |
| Meta knowledge | 0/14 |

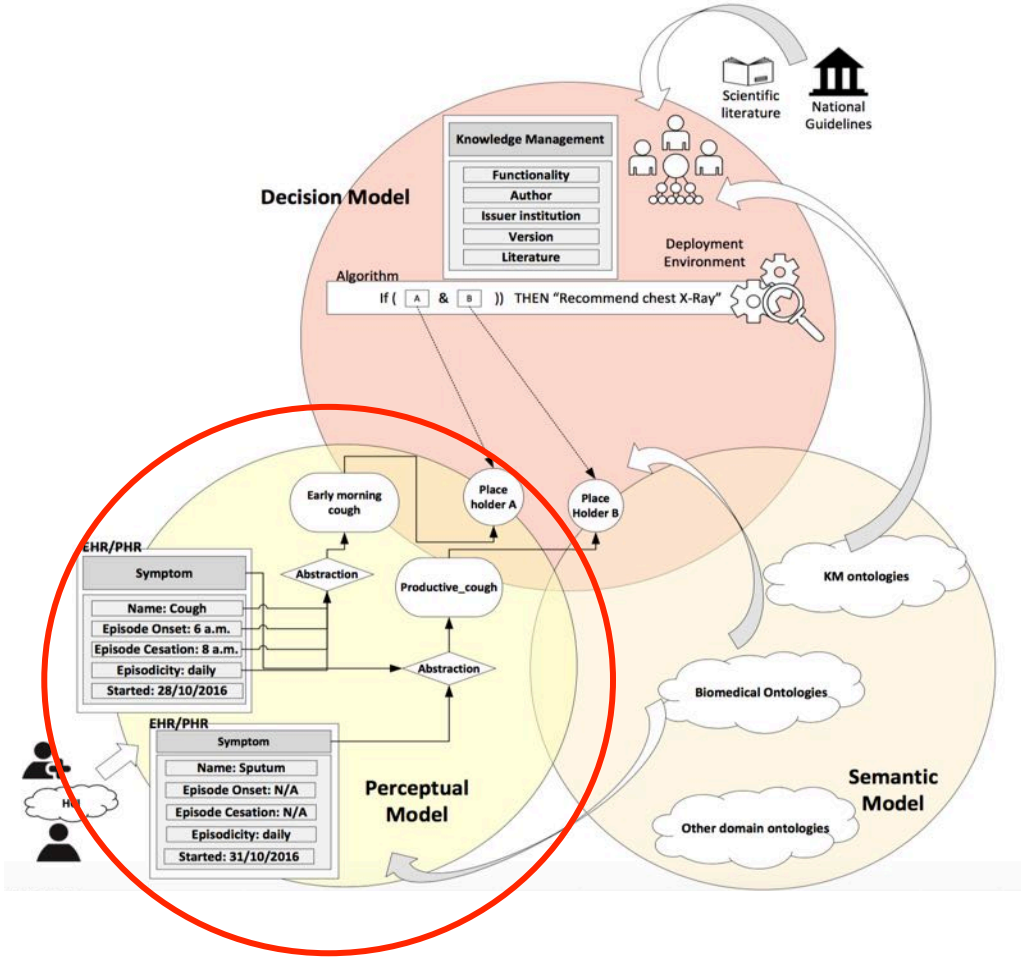
(*)Conway M, Berg RL, Carrell D, Denny JC, Kho AN, Kullo IJ, et al. Analyzing the Heterogeneity and Complexity of Electronic Health Record Oriented Phenotyping Algorithms. AMIA Annu Symp Proc 2011;2011:274–83.

# Agenda

- The Learning Healthcare System
- Conceptualization of clinical data warehousing (DW)
- Computational models involved
- Information representation for clinical DW
- Knowledge representation for clinical DW
- Experiences in Norway and Germany
- Organization and challenges
- Acknowledgements

# Information representation for clinical DW

# Building clinical information models for DW



# Building clinical information models for DW

- **OpenEHR** has been **selected as standard** for representing clinical information in EHRs by **3 out of 4 health regions**.

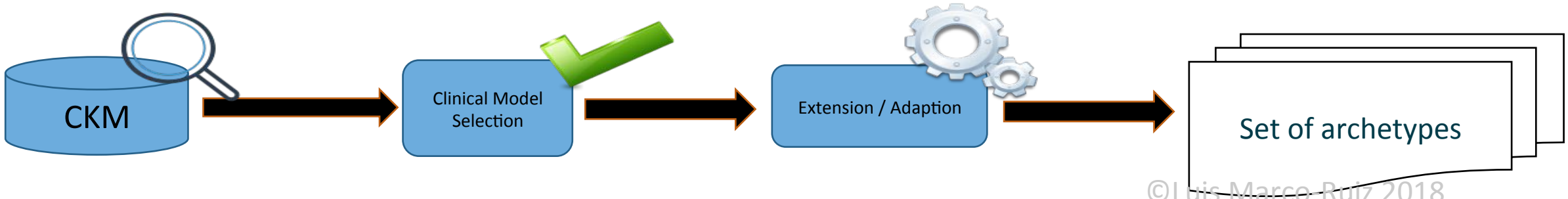
- In 2013 the National ICT Health Trust created the **National Editorial Group for Archetypes (NRUA)** for coordinating the development of the national repository of clinical models (archetypes).

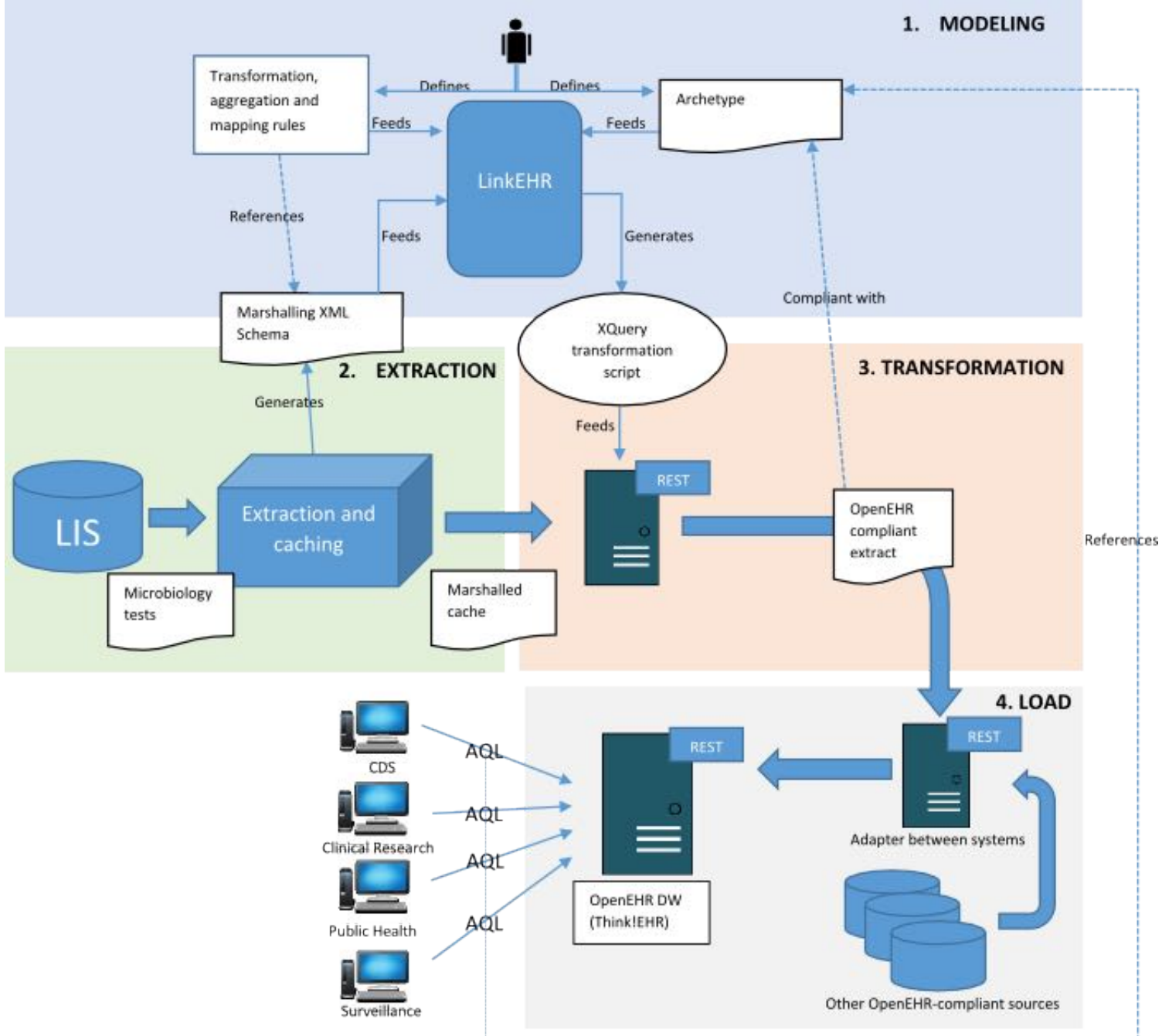
- In November 2015 several projects started exploring the mapping of SNOMED-CT towards the most commonly used Electronic Medical Record (EMR) functionalities.

- Among the suggestions of the directorate is the elicitation of clinical information models (archetypes) and terminology assets for building the patient summary for continuity of care (SNOMED-CT value sets).

# Building clinical information models for DW

- Archetype reuse must be attempted checking national and international repositories.
- Often, CKM archetypes need to be extended to accommodate data reuse requirements.
- The set of archetypes chosen must guarantee the highest level of reusability.
- Archetypes should not be influenced by a particular reuse scenario.
- Keep any new or extended Archetypes unconstrained as much as possible.





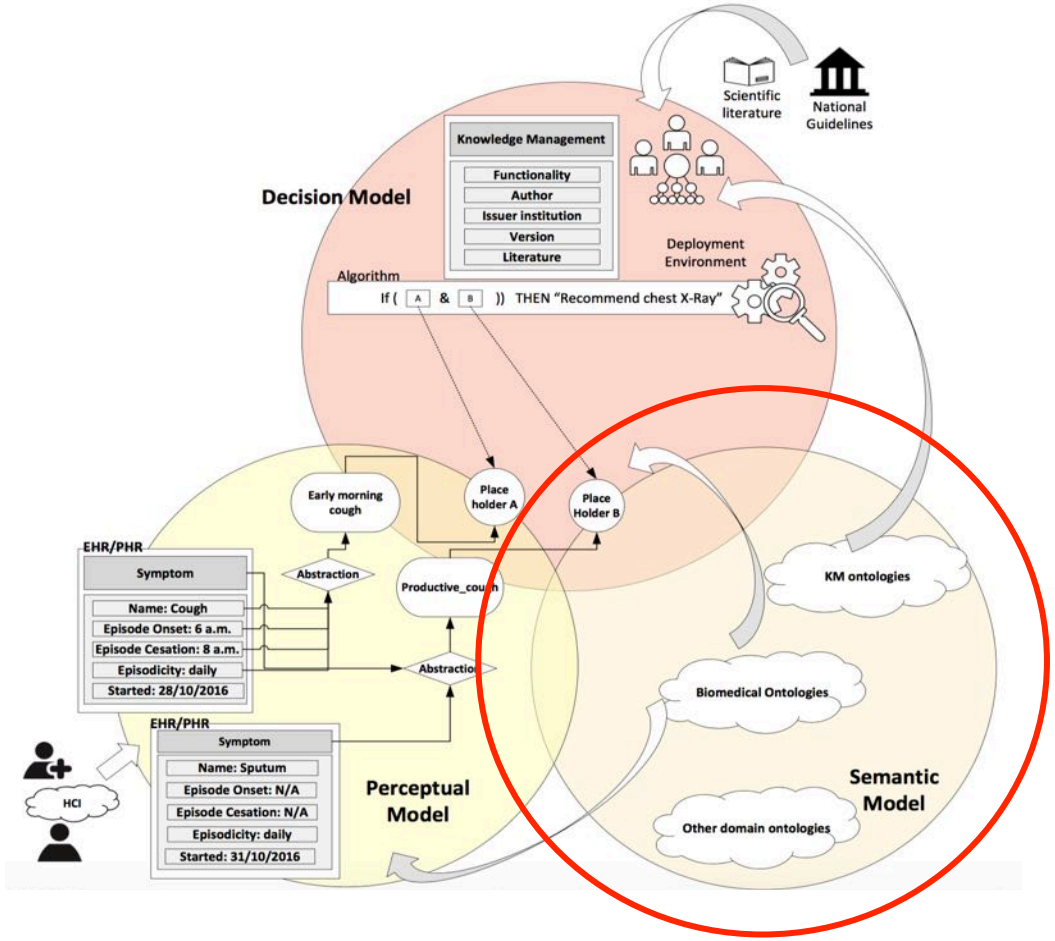

Marco-Ruiz L, Moner D, Maldonado JA, Kolstrup N, Bellika JG. Archetype-based data warehouse environment to enable the reuse of electronic health record data. Int J Med Inf 2015;84:702–14. doi:10.1016/j.ijmedinf.2015.05.016.

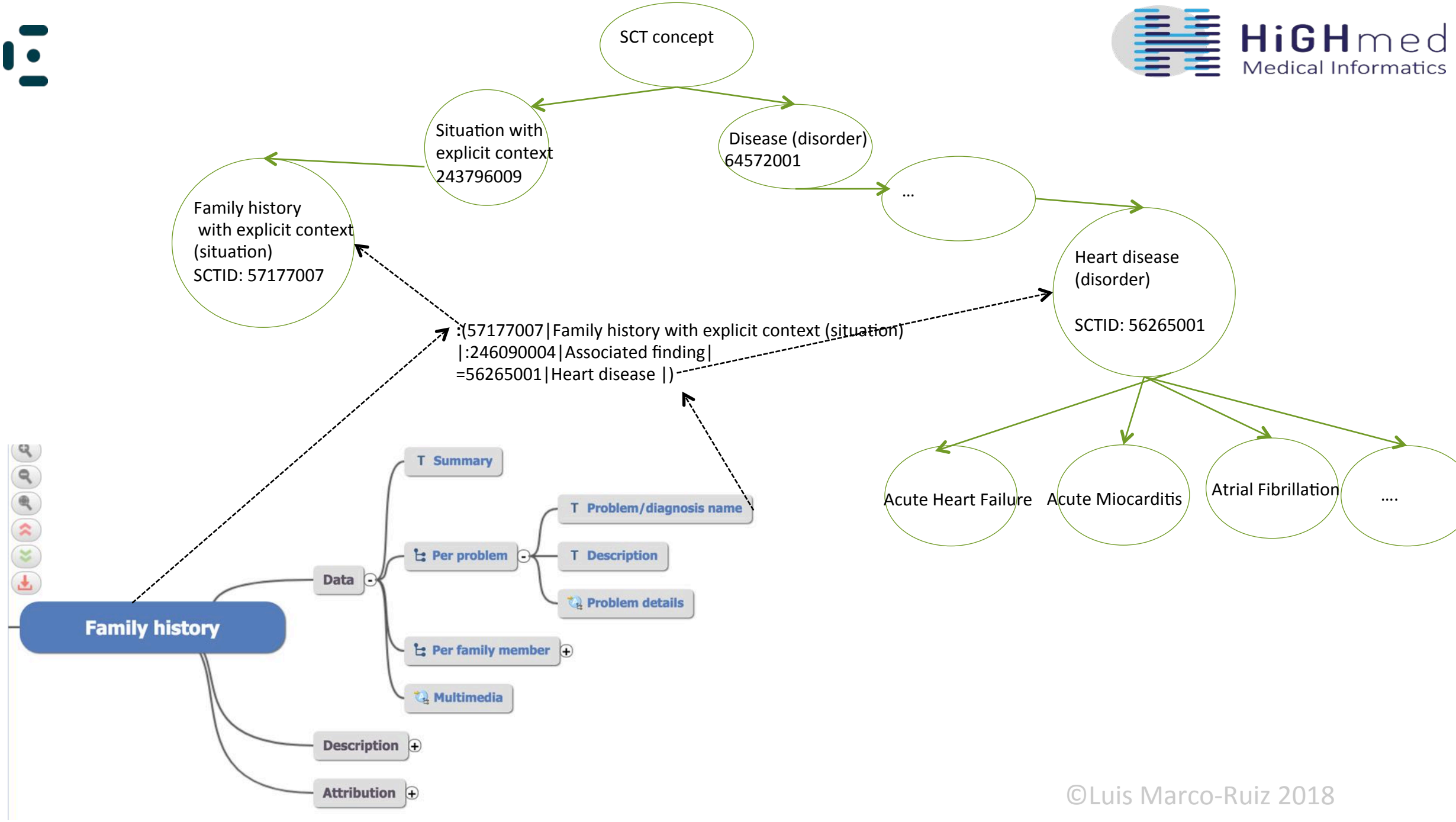
# Agenda

- The Learning Healthcare System
- Conceptualization of clinical data warehousing (DW)
- Computational models involved
- Information representation for clinical DW
- Representations of semantics in clinical DW
- Experiences in Norway and Germany
- Organization and challenges
- Acknowledgements

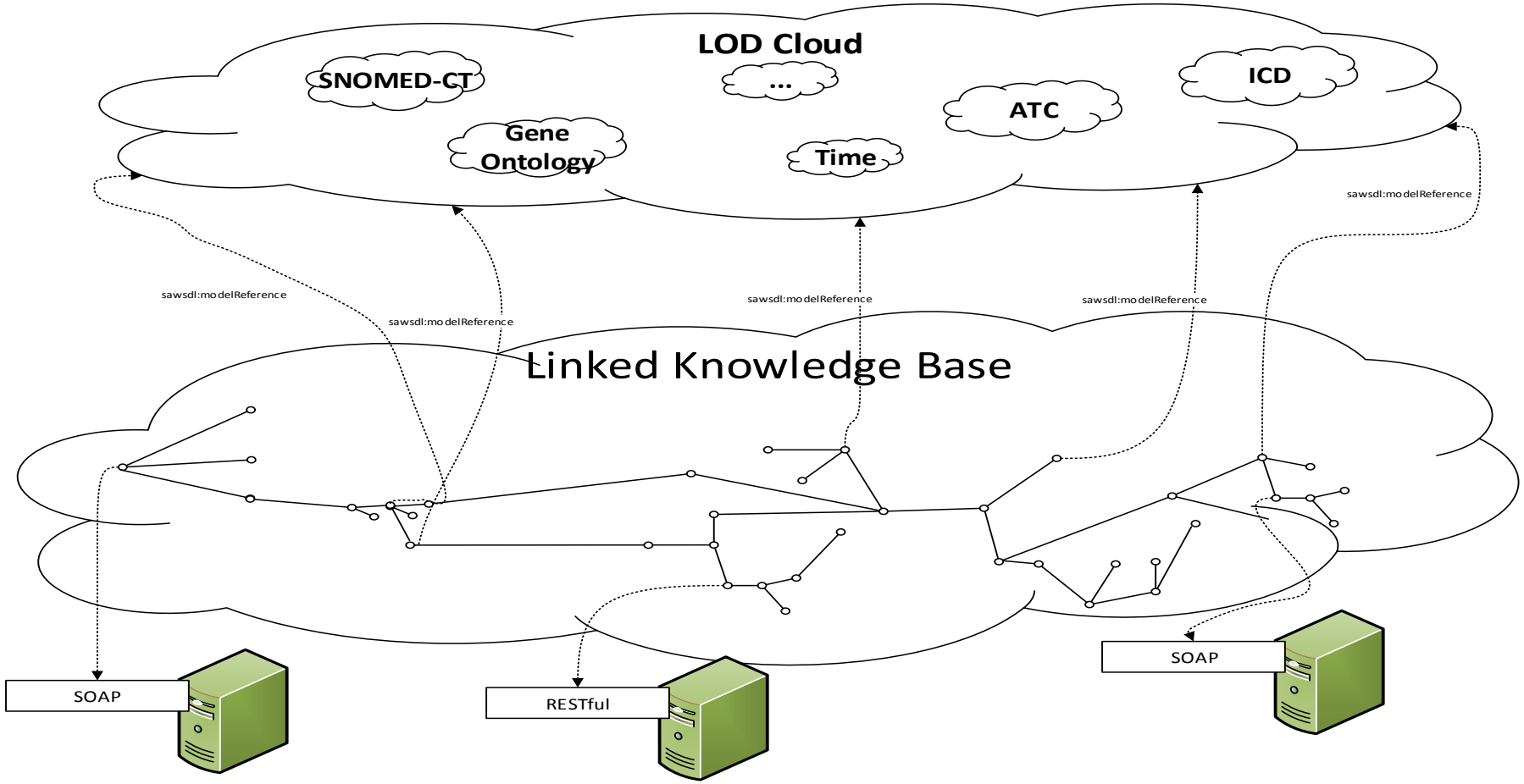
# Representation of semantics in clinical DW

# Building clinical information models for DW





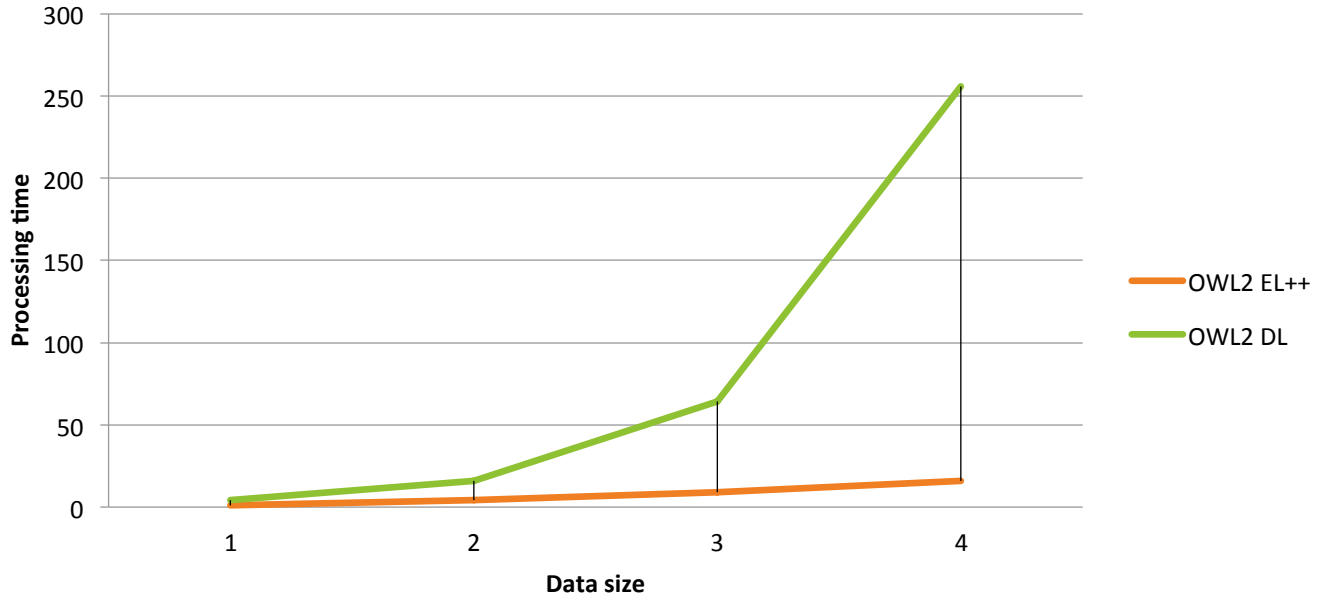
# Binding Clinical Information Models to domain ontologies




Marco-Ruiz L, Pedrinaci C, Maldonado JA, Panziera L, Chen R, Bellika JG. Publication, discovery and interoperability of Clinical Decision Support Systems: A Linked Data approach. J Biomed Inform 2016;62:243–64. doi:10.1016/j.jbi.2016.07.011.

# Overdose of formal semantics

- Computational restrictions
- Human restrictions


Marco-Ruiz L, Nicolaisen K, Pedersen R, Makhlysheva A, Bakkevoll PA. Ontology-based terminologies for healthcare - Impact assessment and transitional consequences for implementation. vol. 1. 1st ed. Tromsø: Norwegian Centre for E-health Research; 2017.

# Agenda

- The Learning Healthcare System
- Conceptualization of clinical data warehousing (DW)
- Computational models involved
- Information representation for clinical DW
- Knowledge representation for clinical DW
- Experiences in Norway and Germany
- Organization and challenges
- Acknowledgements

# Nation wide clinical data warehousing initiatives

## LHS-toolbox (Norway) & HiGHmed (Germany)

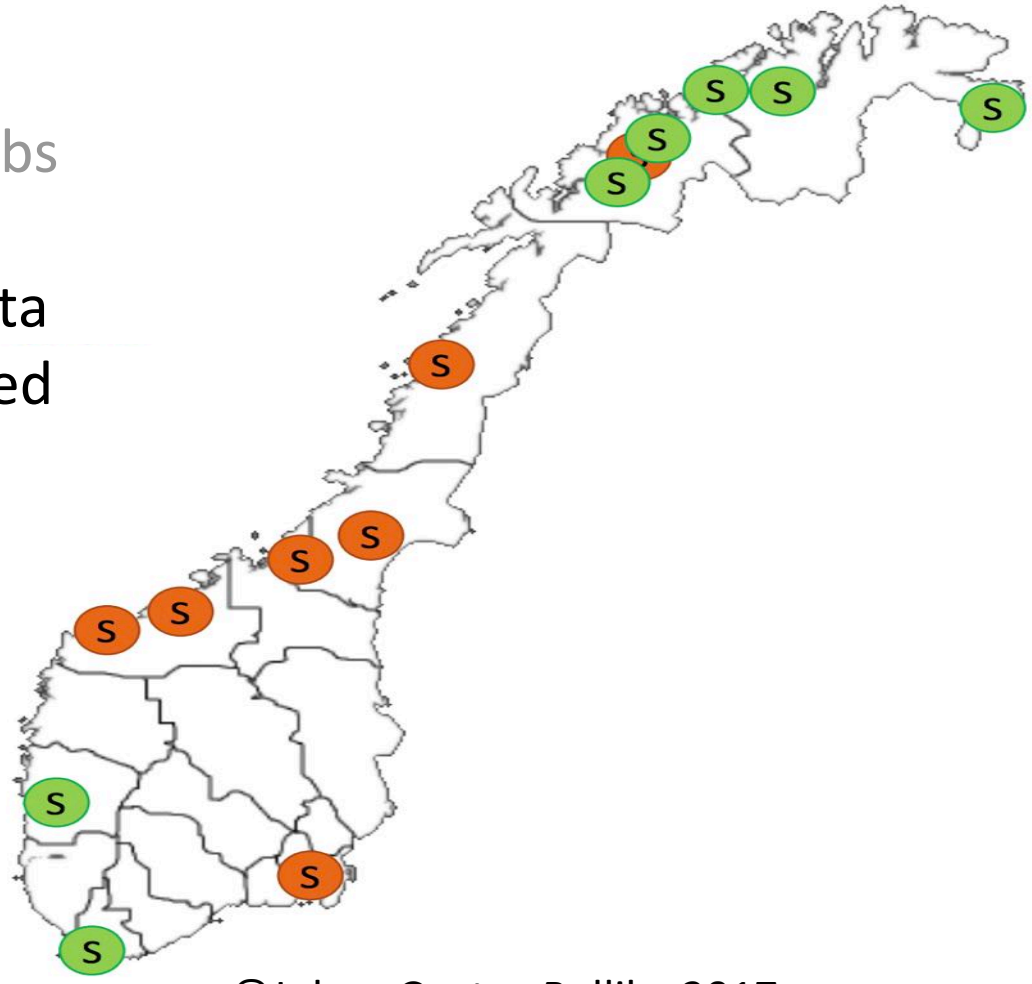
# Clinical DW in Norway: the LHS-Toolbox

The LHS-toolbox project is funded by the Norwegian Research Council with 12,000,000 NOK. The project is based on the evolution of the SNOW system for distributed computations.

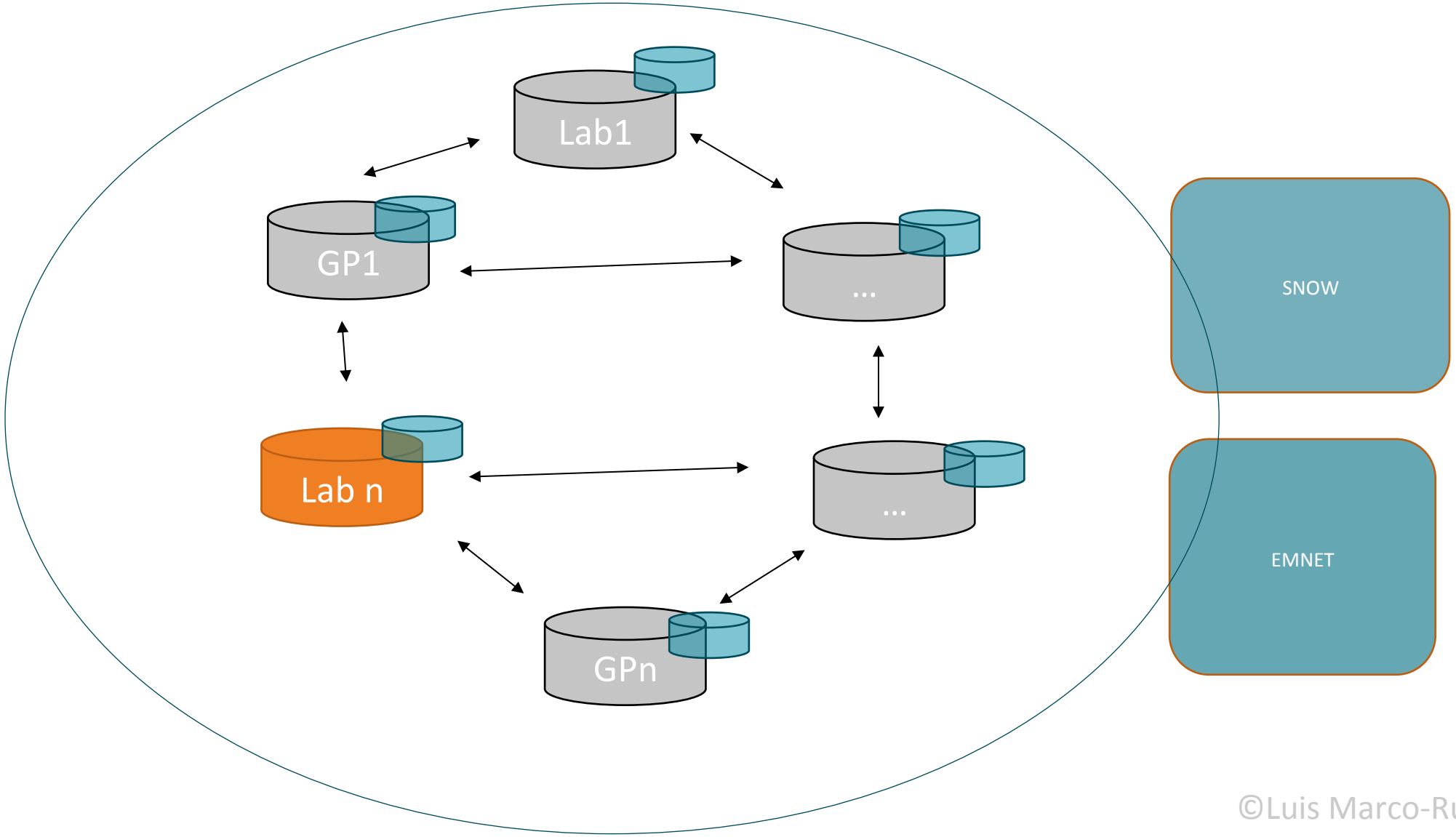
# LHS and SNOW

Deployed Snow system nodes

- A common system for extraction of data from labs and health institutions
- A common system for presentation of health data
- A common system for computation of distributed health data



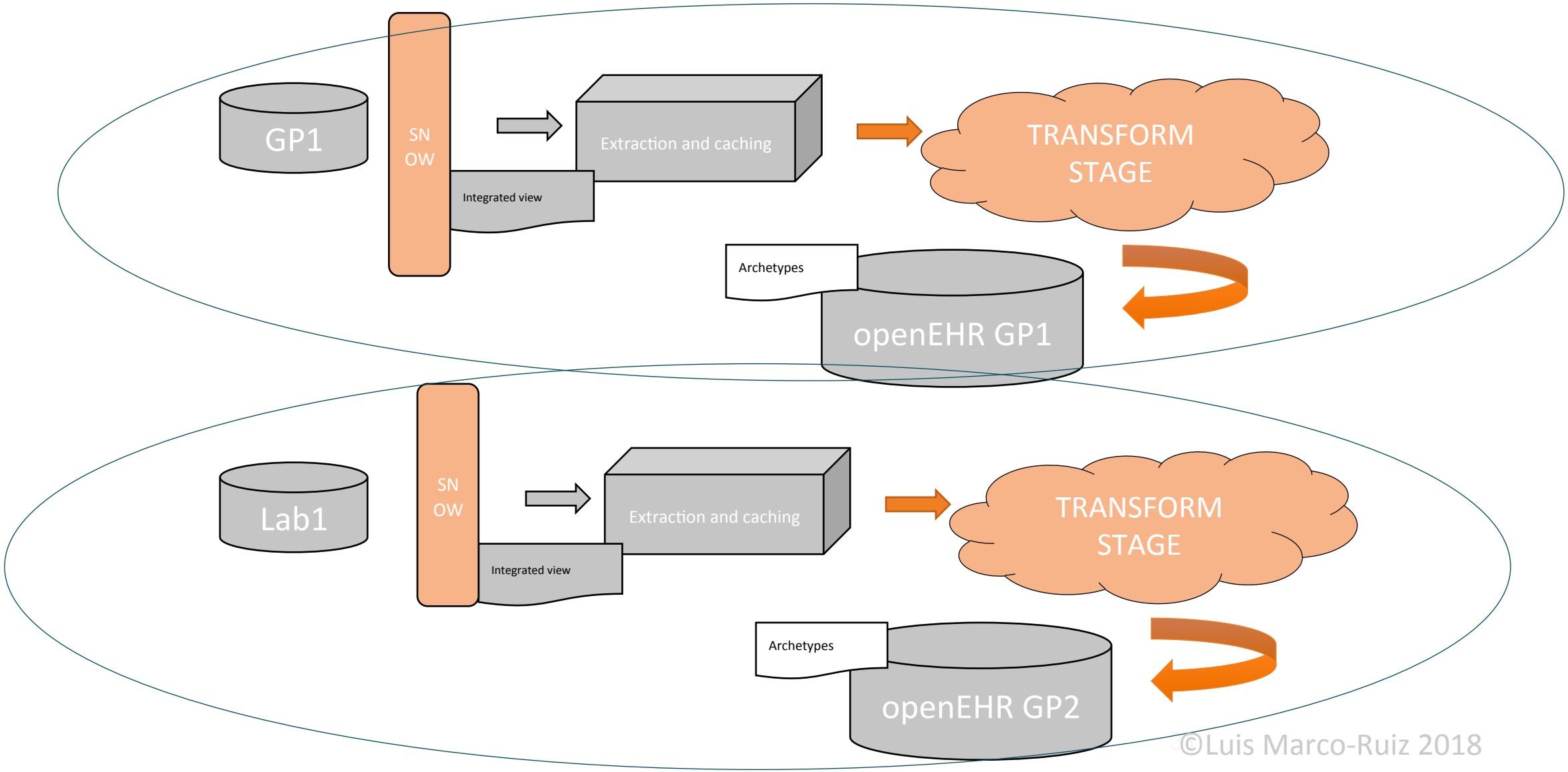
# LHS-toolbox (Snow and Emnet)



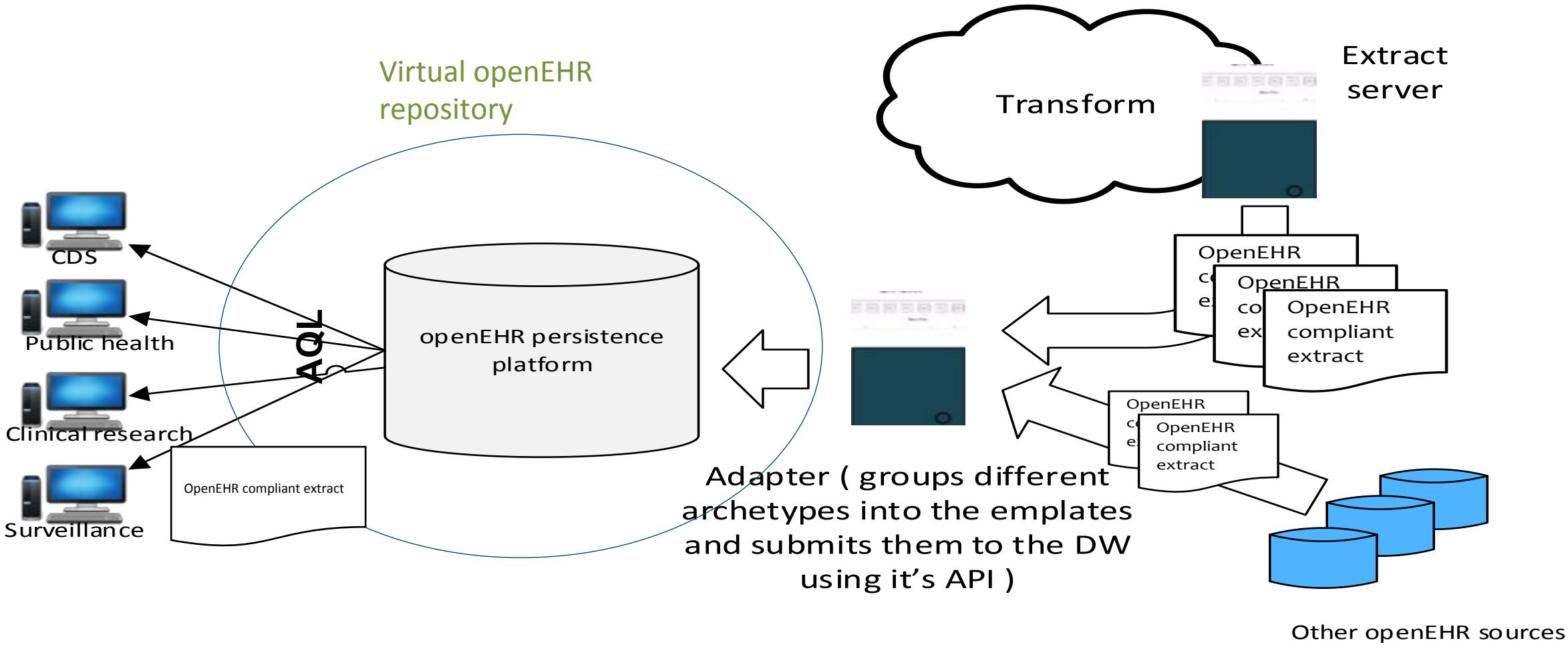

SNOW

EMNET

# Transformation into openEHR extracts



# Virtual openEHR-based view



HiGHmed
Medical Informatics

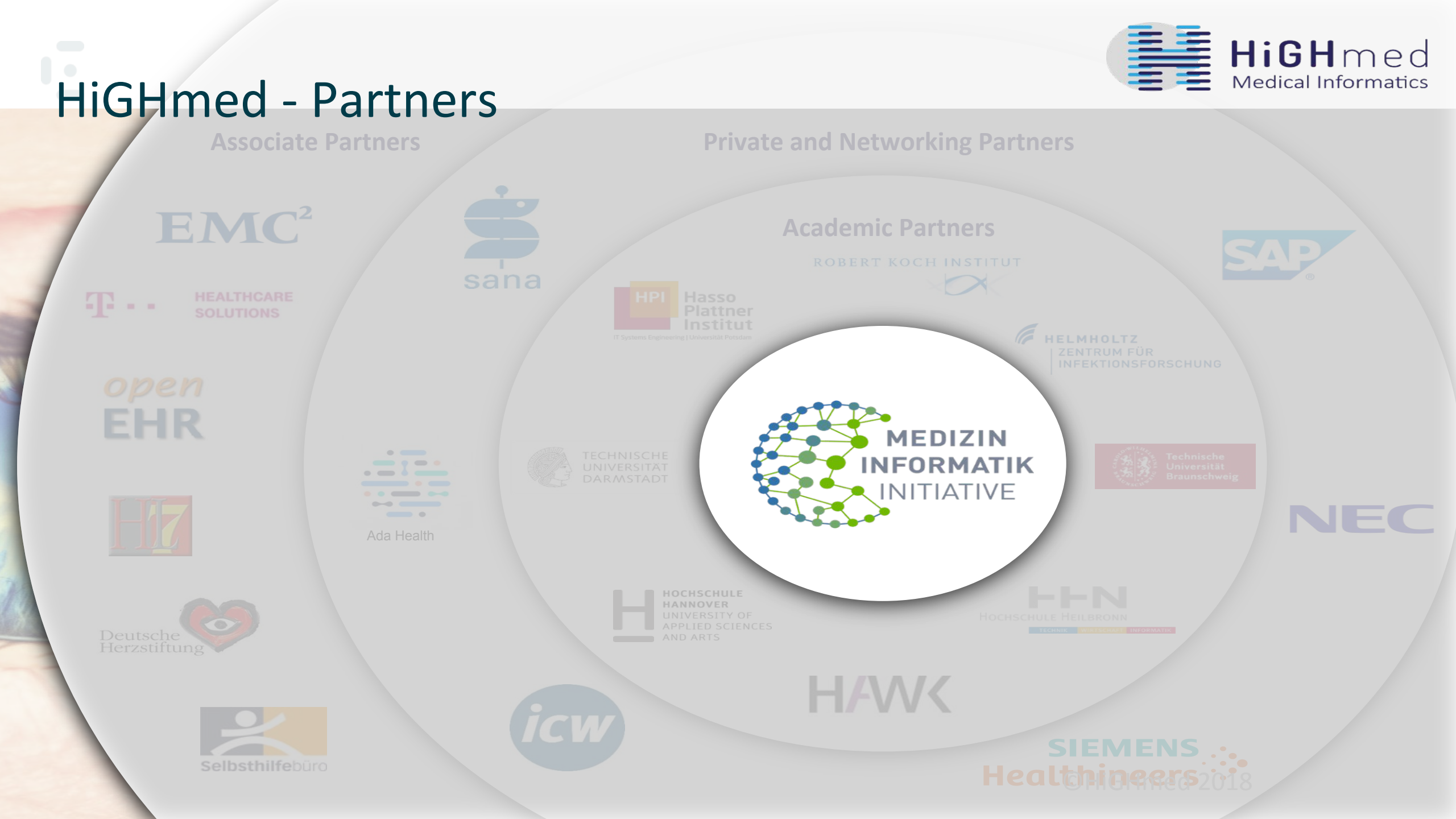
# HiGHmed Summary

# Clinical DW in Germany (HiGHmed)

The HiGHmed project is funded by the Federal Ministry of Education and Research (BMBF) with 30,000,000. Euro. This funding scheme was designed to fund initiatives that will foster the digitalization in the field of medicine

Haarbrandt B, Schreiweis B, Rey S, Sax U, Scheithauer S, Rienhoff O, et al. HiGHmed – An Open Platform Approach to Enhance Care and Research across Institutional Boundaries. Methods Inf Med 2018;57:e66–81. doi:10.3414/ME18-02-0002.

# HiGHmed - Partners



## Das Konsortium — HiGHmed

**Göttingen**
*Universitätsmedizin Göttingen*
**Hannover**
*Medizinische Hochschule Hannover*
**Heidelberg**
Ruprecht-Karls-Universität Heidelberg
*Universitätsklinikum Heidelberg*
*Deutsches Krebsforschungszentrum (DKFZ)*


**Berlin**
*Robert Koch-Institut*
**Berlin**
*Ada Health GmbH*
**Braunschweig**
*Helmholtz-Zentrum für Infektionsforschung GmbH*
**Braunschweig**
*Technische Universität Braunschweig*
**Darmstadt**
*Technische Universität Darmstadt*
**Erlangen**
*Siemens Healthcare GmbH*
**Hannover**
*Hochschule Hannover*
**Heidelberg**
*NEC Europe Laboratories*
**Heilbronn**
*Hochschule Heilbronn*
**Hildesheim**
*Hochschule für angewandte Wissenschaft und Kunst (HAWK)*
**Ismaning**
*Sana Kliniken AG*
**Potsdam**
*Hasso-Plattner-Institut für Softwaresystemtechnik GmbH*
**Walldorf**
*InterComponentWare AG*
**Walldorf**
*SAP SE*

# Requirements

- **Health Information Exchange**: To gain a holistic view on patients' disease progresses, information has to follow patients across organizational boundaries.
- **Cross-Enterprise Data Analysis:** To support all kinds of analytics based on structured and unstructured data from clinical care, research and trial databases, and quality registries in a distributed environment, we want to support **complex queries on a granular patient-record data level**.
- **Cross-Enterprise Knowledge Management:** There are still barriers to practice-based evidence and evidence-based practice in healthcare. Therefore, we need to iteratively work on the establishment of a *learning healthcare system*, in which the latest results of research and data analysis are readily available to caregivers at the bedside, and, in turn, health outcomes of patients help researchers to gain new knowledge.

# Basic Principles

- **Patients first**: In addition to fine-grained use and access control, patients will be able to view and obtain their health data through user friendly tools.

- **Data Safety and Privacy**: Data safety and privacy and the patients' right of self-determination are highest priorities. Data access is regulated unambiguously in a transparent and traceable process.

- **Clinically led Data Modelling**: Healthcare professionals and researchers alike need to be actively engaged to establish semantic interoperability. This creates a new clinico-technical role, the so-called Data Steward, responsible for the management and fitness of data elements within the clinicians.

- **Scalability**: Technical solutions must be optimized to handle high volumes of complex, constantly changing information and clinical workflows.

# Technical Concept

- Every HiGHmed Site has established a **Data Integration Center** according to common HiGHmed standards

- Establish **shared services and data models**

- Use **IHE XDS Cross-Enterprise Document Sharing** to establish vendor-neutral archives and to support **continuity of care processes**

- Working on joint semantics through **openEHR**; with common Archetypes and AQL query language interface

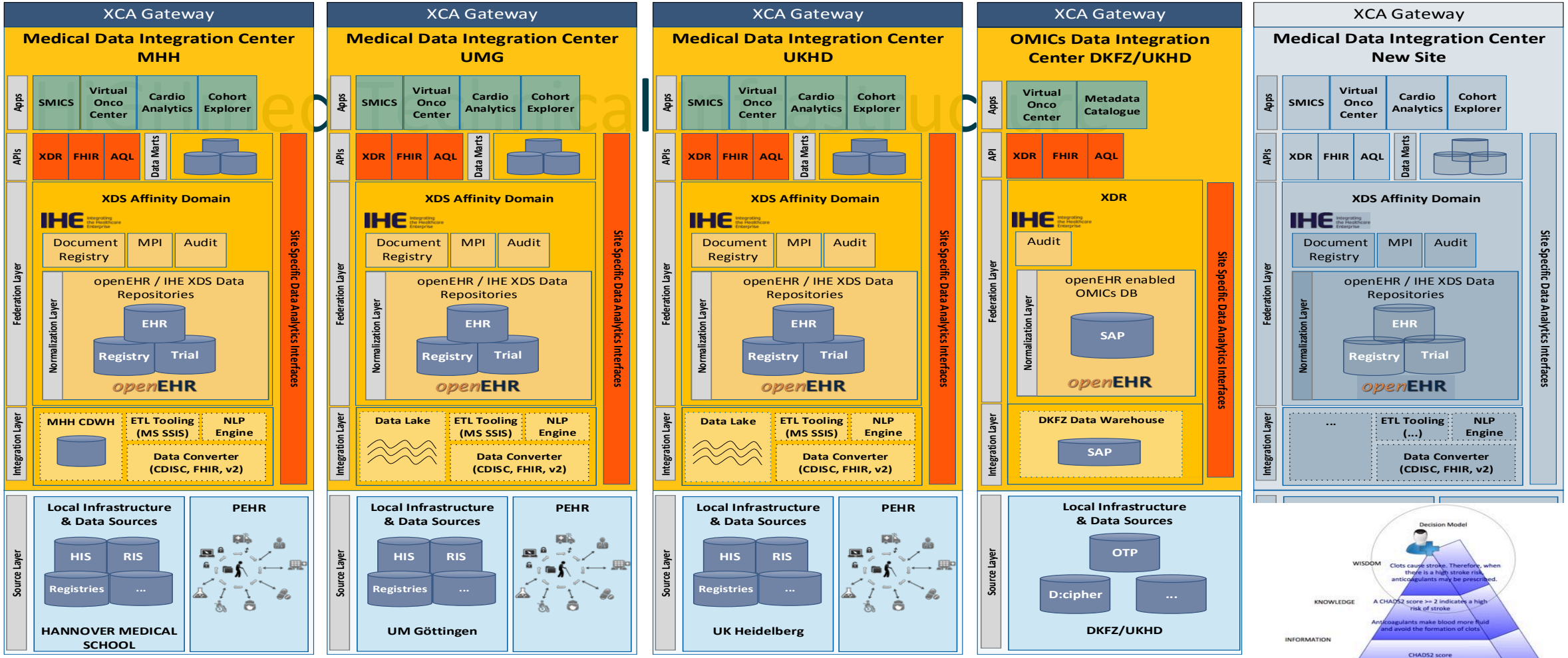
- Implement **FHIR interfaces** to enhance data exchange capabilities (vendor neutral)

HiGHmed Medical Informatics

| FHIR Terminology Service | Clinical Knowledge Manager | Knowledge Service | Pseudonymization Service | Management Service | Cross-Consortia Service | OMICS Services | ... |
|---|---|---|---|---|---|---|---|

Authentication Service

Cross Community Access (XCA, XCPD)



# HiGHmed Use Cases

- **Use Case 1 Oncology**: The oncology use case will help us to address challenges in Medical Informatics when integrating **omics data** from genome sequencing and radiology into clinical practice. New, mobile diagnostic devices are expected to change current medical practice and research by contributing to the long-term monitoring of personal health data at an unprecedented level.

- **Use Case 2 Cardiology**: Within the cardiology use case, we will systematically explore and address IT challenges related to **the integration of data from wearable and connected devices** into our IT architecture.

- **Use Case 3 Infection Control**: The infection control use case will develop an automated early warning and cluster analysis system to support **the algorithmic detection of pathogen clusters**. It will include multidrug- resistant organisms within and across university hospitals, the verification of whether clusters represent outbreaks, and the identification of possible causes of outbreaks.

# Agenda

- The Learning Healthcare System
- Conceptualization of clinical data warehousing (DW)
- Computational models involved
- Information representation for clinical DW
- Knowledge representation for clinical DW
- Experiences in Norway and Germany
- Organization and challenges
- Acknowledgements

# LHS-toolbox and HiGHmed share requirements and challenges

# Technical requirements of clinical DW

- Integration of heterogenous data sources in **structure** and **meaning**
- **Common representation formats** are needed (openEHR, HL7 FHIR, *IHE XSD*)
- **Semantic Interoperability** is required for the consistency and meaning
- **Expressive queries** (e.g. Some kind of…)

# High level organization of teams

**Shared for all the project**

- 1 Enterprise architect
- 1 Project manager
- 2-4 developers
- 1 semantic interoperability specialist
- Others: Data stewards, PhDs, administration staff etc.

**Internal to each institution (data integration center)**

- The staff depends depending on the needs and the role in the data reuse network

# Organizational challenges in clinical DW

- Need for more flexibility managing project funds.
- Excess of regulation in public funded projects
  -> burocracy consumes huge amount of resources, slow procurement process, slow hiring processes etc.
- The complexity of procurement processes may affect small-medium size companies.
- Collaboration among different health regions may be complex.
- Recruit highly skilled IT professionals for public projects is difficult. We are not competitive when compared with private companies. This may affect quality of our developments -> we try to alleviate this by partnering with vendors that are interested in our use cases/developments.

# Organizational challenges in clinical DW

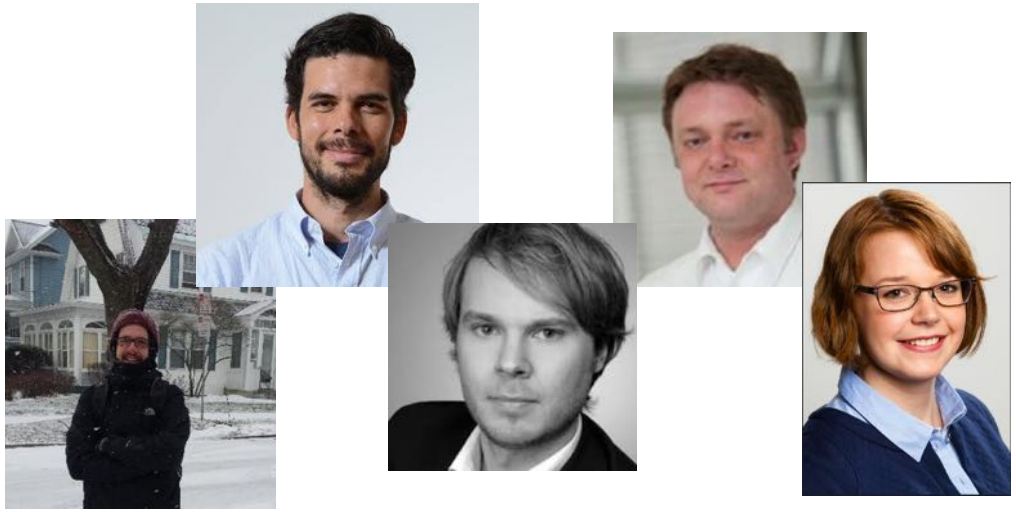
- GDPR-> currently the regulation is ambiguous and no clear answers are provided
  -> this affects the whole data reuse infrastructure.
- We need more efficient mechanisms to enable patient control.
- The border between a research study and a study measuring health indicators is not clear
  -> uncertainty on how to proceed

# Acknowledgements

## LHS-toolbox team


Nasjonalt senter for e-helseforskning

## HiGHmed team



PLRI PETER L. REICHERTZ INSTITUT FÜR MEDIZINISCHE INFORMATIK