A SPARQL-based i2b2 Endpoint in the EHR4CR project

Boris De Vloed[2], Sajjad Hussain[1], Christel Daniel[1], Eric Lepage[1], Kristof Depraetere[2], Dirk Colaert[2]

[1]Assistance Publique-Hôpitaux de Paris (AP-HP)

[2]Agfa Healthcare

European i2b2 Workshop, Erlangen

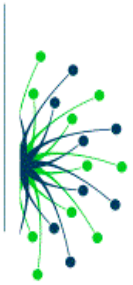
March 25-26, 2013

EHR4CR project: Electronic Health Records for Clinical Research

The EHR4CR (Electronic Health Records for Clinical Research) project aims to improve the efficiency and reduce the cost of conducting clinical trials, through better leveraging of routinely collected clinical data in the trial design and execution life-cycle

The EHR4CR platform will implement 4 use cases

- clinical protocol feasibility
- patient identification and recruitment
- clinical trial execution
- adverse event reporting



EHR4CR project: Health Institutions

- The 4 use cases will be demonstrated by 11 pilots in 5 European countries

  - Germany (WWU, FAU)
  - France (AP-HP, U936)
  - UK (UoD, UoG, UoM, UCL)
  - Switzerland (HUG)
  - Poland (MuW)

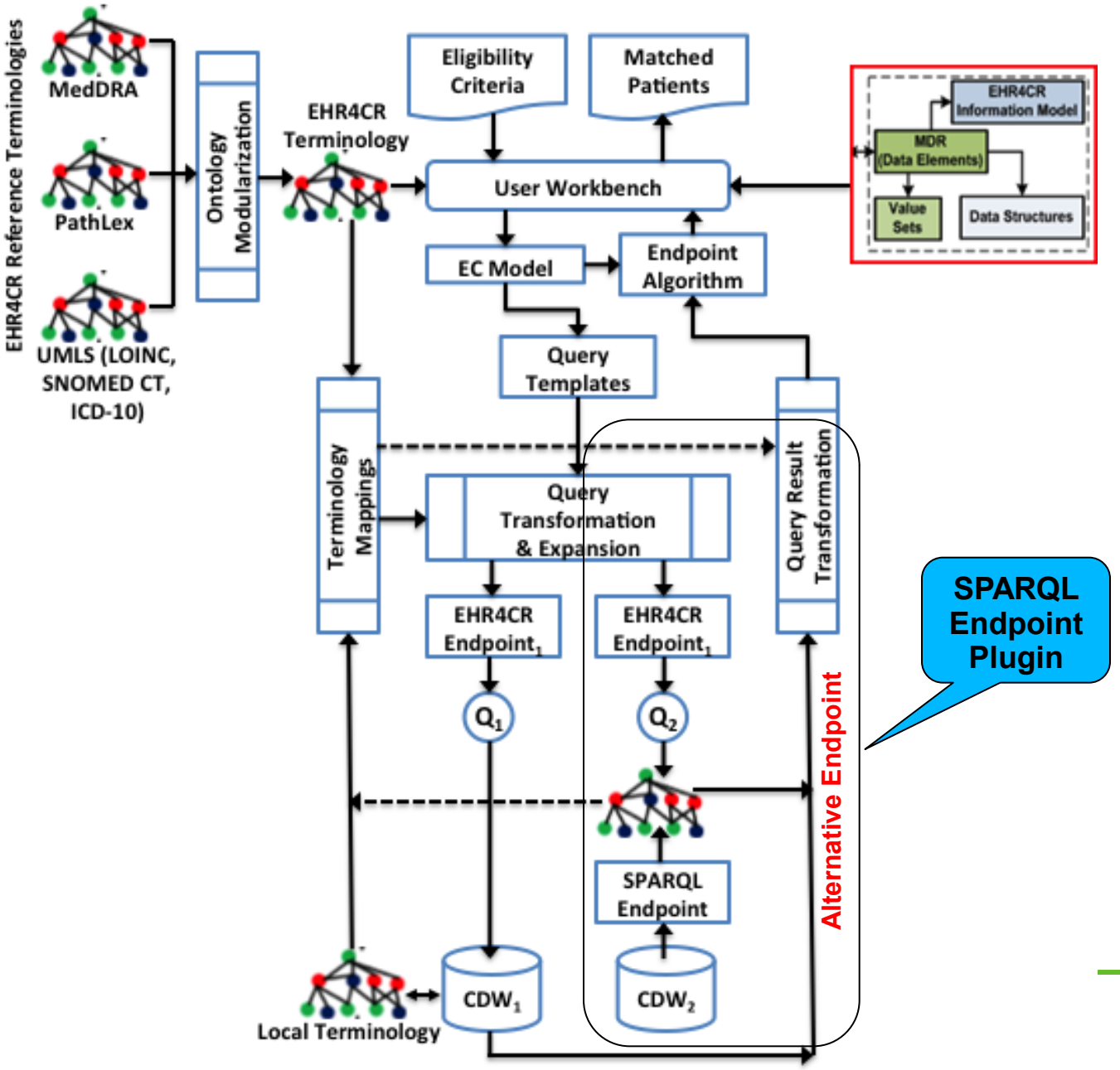
EHR4CR Consortium

11 Pharmaceutical Companies (members of EFPIA)
22 Public Partners (Academia, Hospitals and SMEs)
5 Subcontractors (Advisory Board)

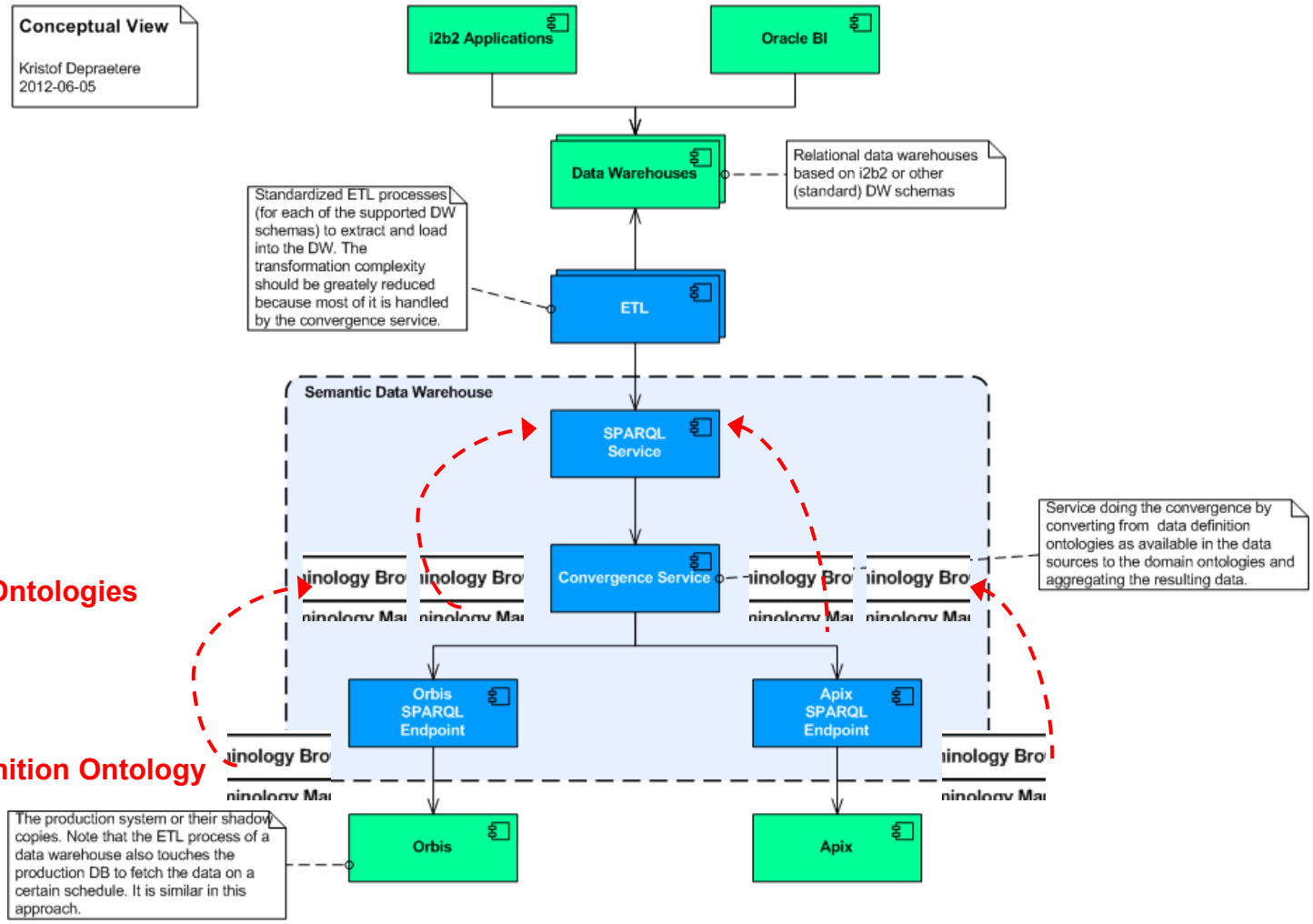
Westfälische Wilhelms-Universität Münster

EHR4CR

EHR4CR Semantic Interoperability Framework

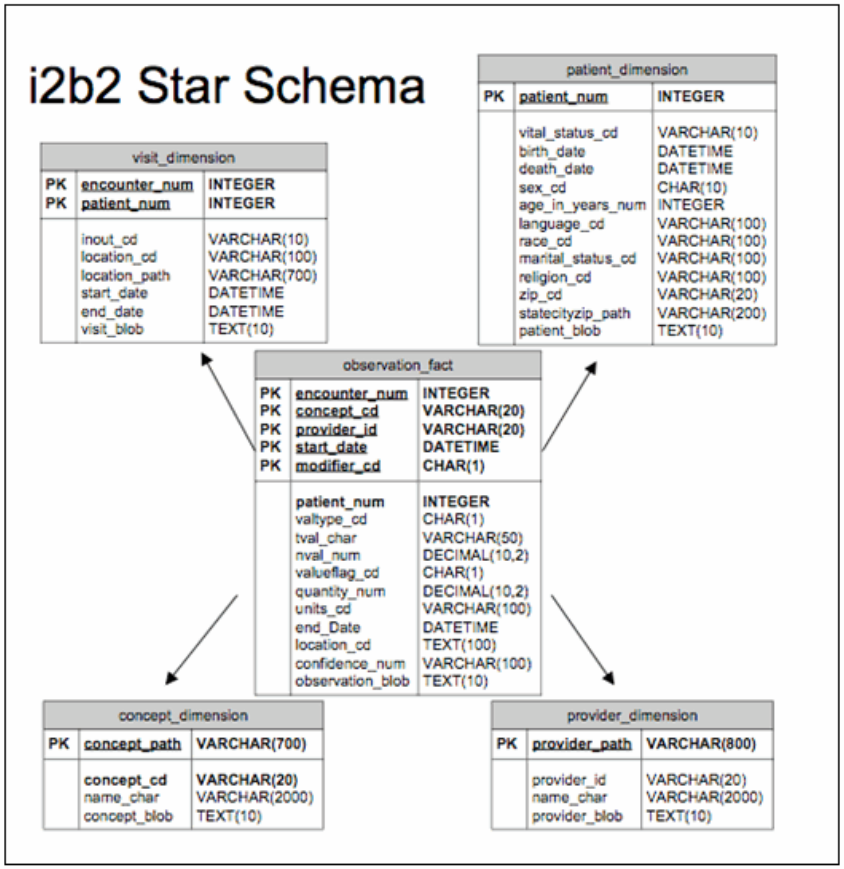


Building Semantic CDW at AP-HP (by Agfa Healthcare): An Overview



Target i2b2 application

- Data level
  - Star schema
    - Observation facts
    - Patient
    - Visit
    - Provider
    - Concepts
  - Lookup tables
- Metadata level
  - I2b2 “ontologies” (not RDF/OWL semantic web ontologies)
  - Hierarchy of concepts
  - Used by i2b2 GUI



SDW Web API

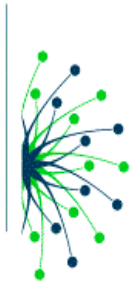
http://wopeg.he.agfa.be:2020/ehr4cr/portal/**observationfacts?start=2000-01-01T00:00:00&end=2012-09-26T23:59:59**

Specify

- Table, e.g. Observationfacts
- Scope, e.g. Start and end of period
- Extra options: e.g. No cache

Results ETL 2000 - Today:

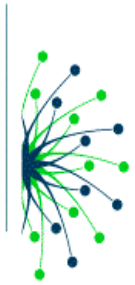
- Observation fact (~220 records)
- Patient dimension: (~39 000 records)
- Visit dimension (~66 000 records)
- Provider dimension (~12 300 records)
- Concept dimension (~17 000 records)
- Cim 10 ontology (~17 000 records)



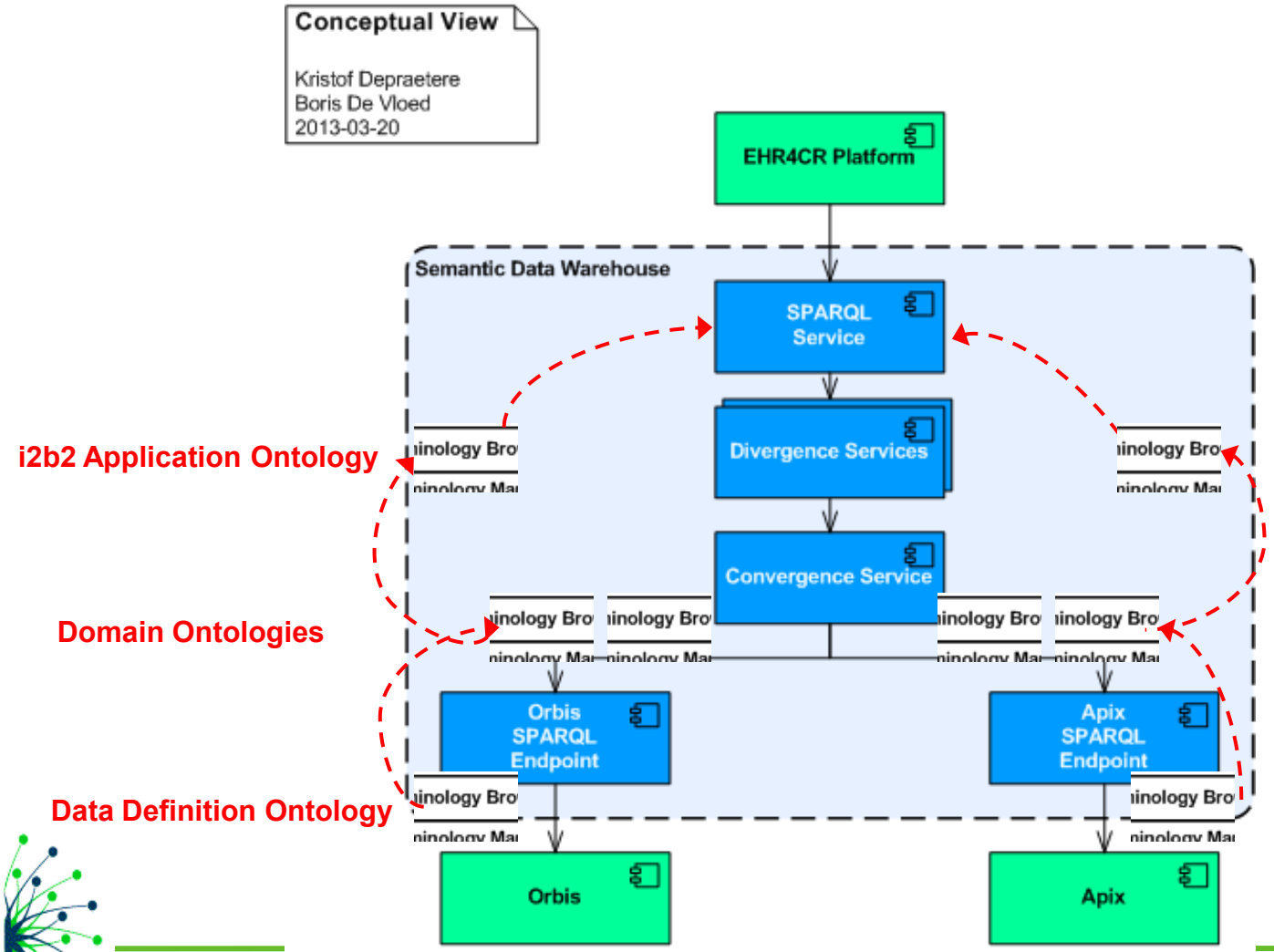
ETL Run time

Talend Job

- Call the SDW Web API
  - Download DSV files for the different columns
- Delta of previous versions
- Mapping for patient, visit and provider dimension
- Syntactic transformations (e.g. Character escaping)
- Load data into i2b2 database

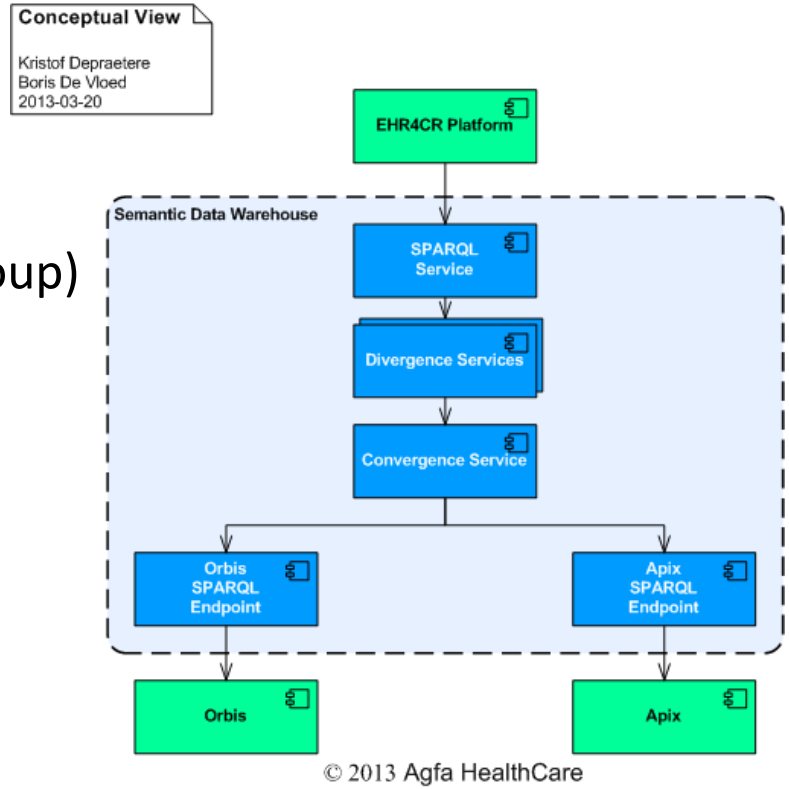


Building Semantic CDW → i2b2-based SPARQL Endpoint



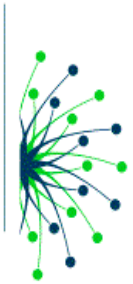
Building Semantic CDW at AP-HP: Data Sources

- Orbis Sparql Endpoint on relational tables: (DONE)
  - Demographic information
  - Hospital stays
  - iDRG tables (DRG – Diagnosis Related Group)
- Sparql Endpoint on APIX (PLANNED)



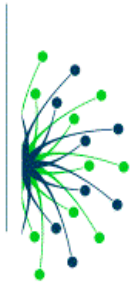
Convergence: Source Data → Common Domain Ontologies

- Source data expressed in Orbis Data Definition Ontology terms
  - Formal representation of data as stored in DB
  - DDO is a formal representation of a data source database schema/tables
- Domain ontologies
  - FOAF for names, addresses etc
  - SKOS for coding systems
  - Agfa Domain Ontologies
  - …
- Convergence
  - Apply N3 mapping rules on the source data
  - Converting & aggregating source data into domain ontology entities



Divergence: Common Domain Ontologies → Application Ontology (i2b2)

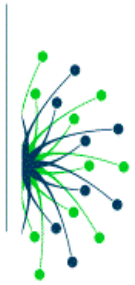
- Entity data expressed in Application Ontology (AO)
  - An AO is a formal representation of an Application Database Structure. i.e. the i2b2 DB structure
  - Partitioning on date, patient id, …
  - Examples entities:
    - Different entities for the different i2b2 dimensions
    - Multiple kinds of fact entities (Lab results, diagnoses, …)
    - I2b2 ontology entities (i.e. CIM and CCAM hierarchies generated from data in Orbis)



Accessing Semantic CDW: i2b2 Endpoint

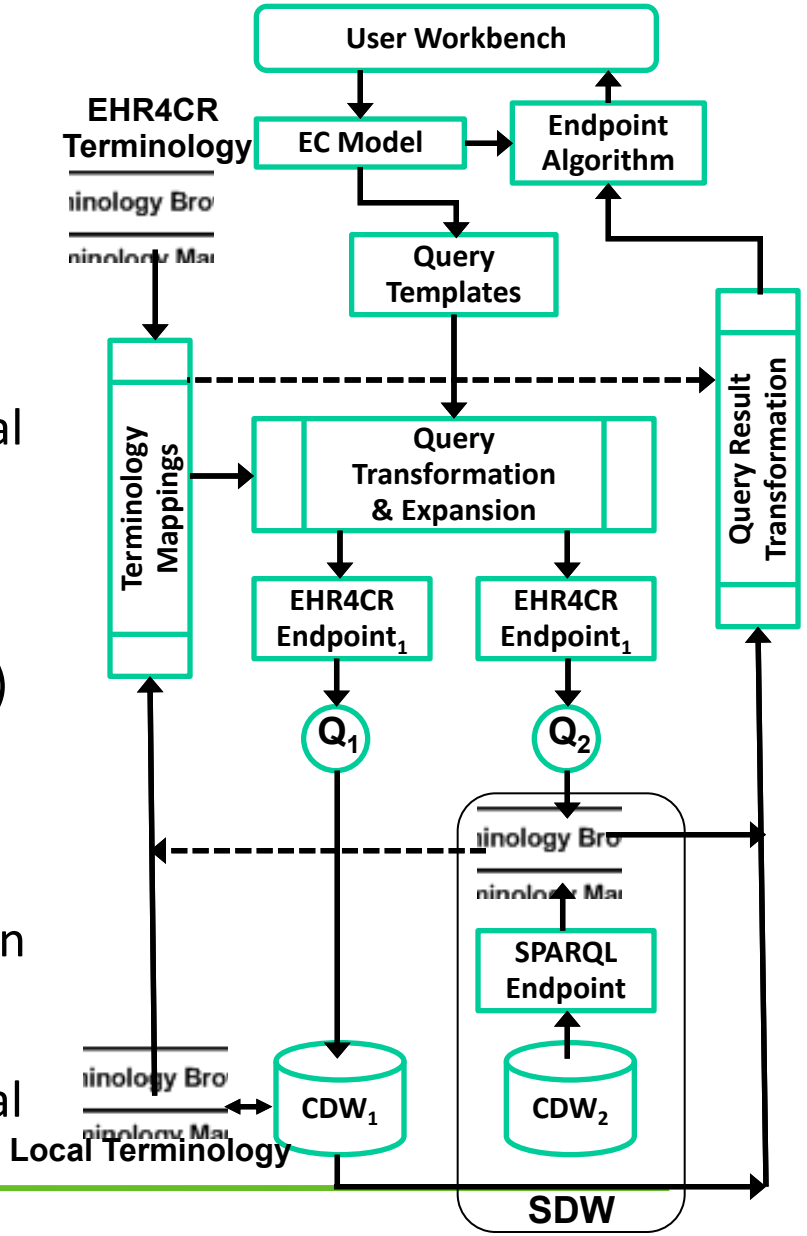
- A Generic SPARQL endpoint to query the different entities

- SPARQL queries on the SDW i2b2 entities could also be executed on a D2R SPARQL endpoint on top of an i2b2 DB.



Pluging i2b2-based SPARQL Endpoint into EHR4CR Platform

- Query Transformation: EC Model → SPARQL queries
  - Query Language Transformation
  - Terminology Mapping: central ←→ local codes
- Entity selection
  - Select any AO entity (lab result facts, ...)
  - Optionally: choose a partition based on date or code
- Query Result Transformation and Aggregation
  - Syncronizing query result ino EC model
  - Terminology Mapping: local ←→ central codes



Innovative Medicines Initiative

Thank You!

???