Using Electronic Healthcare Record for Clinical Research (EHR4CR)

Dr Christel DANIEL

INSERM (UMR_S 872, Eq. 20)

AP-HP (CCS Domaine Patient)

Paris, France

Acknowledgments: David OUAGNE, Sajjad HUSSAIN, Eric SADOU, Marie-Christine JAULENT

Integrating clinical research to the Healthcare Enterprise - Why is it needed?

- Protocol feasibility & Patient recruitment [eClinical Forum 06, Kahn 07]
  - Only 7% of eligible patients enroll in a clinical trial
  - 86% of all trials fail to enroll on time
  - Women, minorities, children and special populations under-represented
- Study execution: reduce redundant data capture & improve data quality
  - 5-35% of clinical research data are collected in patient records [Bleicher 06]
  - Investigators feel that they duplicate 70-100% of clinical research data
  - Monitoring : 7 € per transcription error
- Adverse Event Reporting: improve reporting through process simplification

Kahn, Michael G. MD, PhD; Kaplan, David MD. Implementing Translational Research Policies in Electronic Medical Records. Academic Medicine. 82(7):661-669, July 2007. Draft version 0.1, March 3, 2006;The eClinical Forum and PhaRMA EDC. The Future Vision of Electronic Health Records as eSource for Clinical Research

Integrating clinical research to the Healthcare Enterprise - Why is it needed?

- Trends in Clinical Research Informatics (17/51 sites (33%))

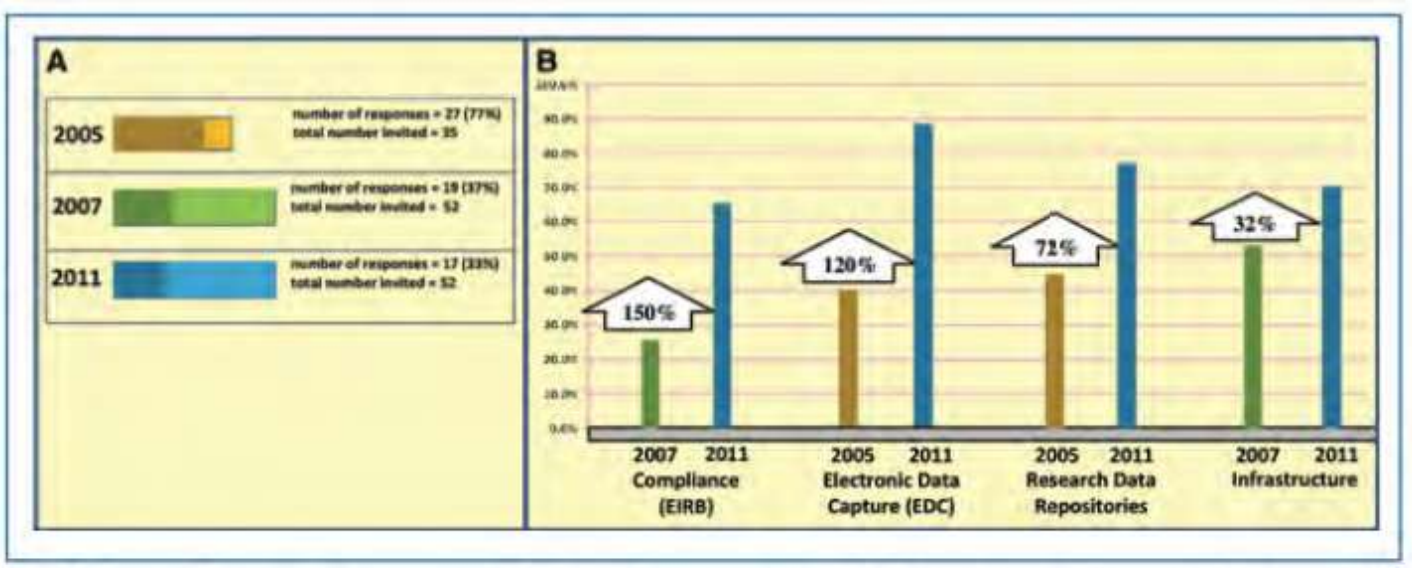


Figure 1. Comparison of response rates and responses regarding adoption of major categories of research IT infrastructure between the current (2011) and previous (2005 and 2007) surveys. (**A**) It demonstrates the response rate difference. (**B**) It depicts percentage increases for each category.

Murphy SN, et al. Current State of Information Technologies for the Clinical Research Enterprise across Academic Medical Centers" Clin Trans Sci. 2012

Integrating clinical research to the Healthcare Enterprise - Why is it difficult?

- Organisational issues
  - Clinicians, epidemiologists, biostatisticians, ergonomists
- Ethical & regulatory issues
- Technical issues

EHR4CR project

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 at key points in the trial design and execution life-cycle**

EHR4CR technical platform
4 scenarios…

- Harmonized access to multiple heterogeneous and distributed EHR or Clinical Data Warehouses (CDW)
- 4 scenarios

1. – clinical protocol feasibility
2. – patient identification and recruitment
3. – clinical trial execution
4. – adverse event reporting

EHR4CR technical platform
4 scenarios…

- 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 technical platform
4 scenarios & the same need of semantic interoperability

EHR

CDMS

Question 1	Hemoglobin A1c
Question 2	Age
Question 3	SBP (mmHG)
Question 4	Weight (kg)n
Question 5	Tabacco ? yes/no
Question 6	Cardio-vascular disease -HBP -Angor -Cardiac failure

EC 1	Between ages 18 and 56
EC 2	Hemoglobin A1c value within the diabetic range
EC 2	At least 2/3 systolic blood pressure measurements ≥ 140 mmHg

Question 1	Last known weight (kg)
Question 2	Systolic Blood Pressure (mmHG)
Question 3	Hypertension? Yes/No
Question 4	Smoking / Non smoking ?

4 scenarios, only one rule
« *The Best Use of Standards* » (1/3)

Health Level Seven

http://wwww.hl7.org

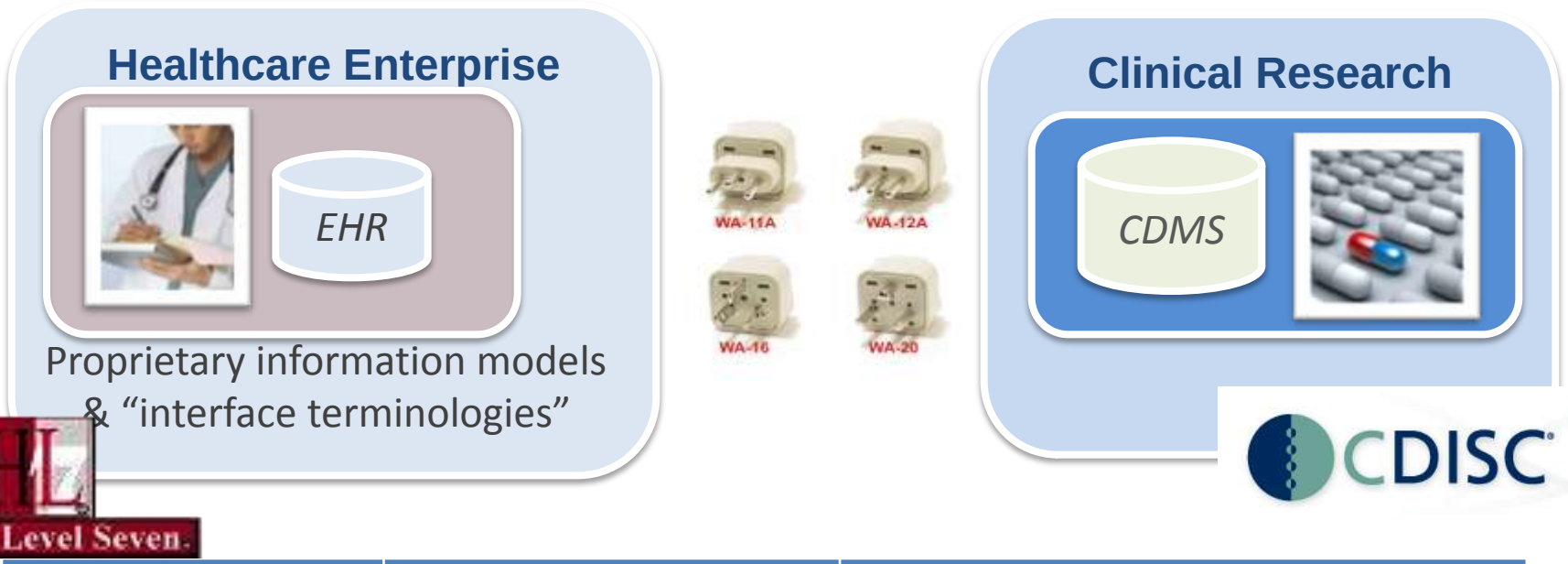
- Organisme de développement de standards (ANSI) créé en 1987
- 500 organisations membres, 1500 inscrits, 15 affiliés internationaux
  - France depuis 2004

CDISC

www.cdisc.org

- Organisation internationale créée en 1997
  - liée à l'ISO TC 215
  - accord avec HL7 depuis 2001

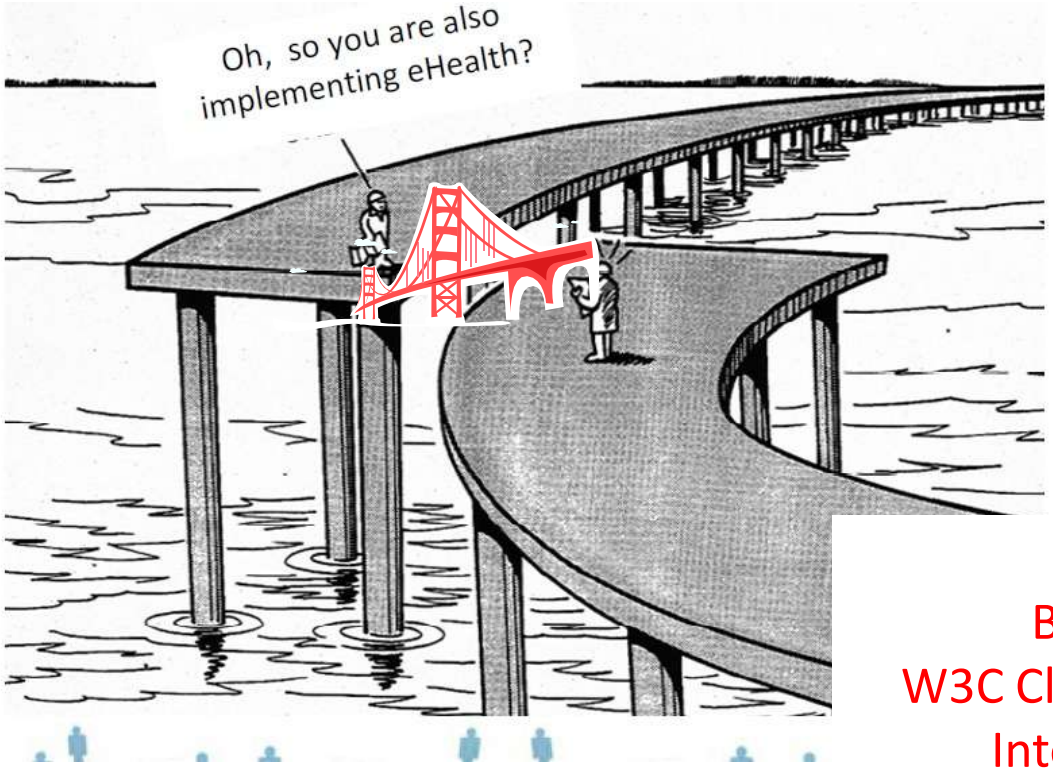
Semantic integration challenges



	Soin	Recherche clinique
Information models	HL7 (CDA) CEN/TC 251 (EN13606) ISO 21090 (data types)	Operational Data Model (ODM) - Clinical Data Acquisition Standards Harmonization (CDASH)
Reference terminologies	LOINC, SNOMED CT, ICD-10, etc	MedDRA (adverse events)

4 scenarios, only one rule
« *The Best Use of Standards* » (2/3)

- Is becoming easier: current trend of «Standardizing the standards»



HL7 RCRIM[1]
BRIDG[2] model
W3C Clinical Observations Interoperability (COI)

[1] *Regulated Clinical Research Information Management*
[2] *Biomedical Research Integrated Domain Group*

4 scenarios, only one rule « The Best Use of Standards » (3/3)

- Key organization: Integrating the Healthcare Enterprise (IHE)
  - Connectathons
  - 7 domains including
    - QRPH (Quality Research & Public Health)
      - Information exchange relevant to quality improvement in patient care, clinical research and public health monitoring.
    - ITI (Information Technology Infrastructure)
      - Security, confidentiality
      - Document sharing

Objectives

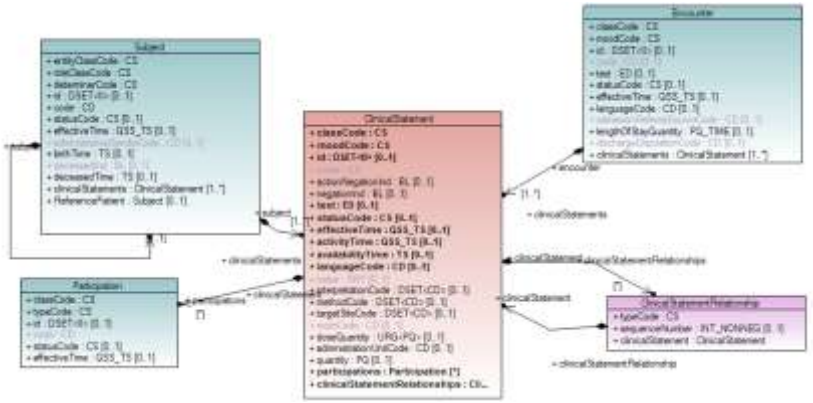
- To propose a **standard-based expressive and scalable Semantic Interoperability Framework**
  - Dynamic mappings and consistent interpretation of clinical data between data structures and semantics of varying data sources.
- Step 1 : Implementing the core elements of the EHR4CR semantic interoperability framework
  - Shared conceptual reference model (EHR4CR information model) & terminology (EHR4CR terminology)
- Step 2: Evaluating the EHR4CR semantic interoperability framework
  - Semantic interoperability services for eligibility determination

Methods - Building semantic resources EHR4CR information model

- A Shared Model of Meaning acting as a global as view model to correlate the schemas and concepts from different sources
  - Used by both EHR and EDC system vendors as the basis for enabling cross-vendor semantic interoperability
- Material : “source” models
  - BRIDG model & HL7 v3 models (HL7 RCRIM WG)
    - «StudyDesign» and «A_SupportingClinicalStatementUniversal» models
  - I2b2 model
- Method: Model-driven engineering
  - Transforming HL7 v3 models in UML models and adapt these models to the purpose and scope of the EHR4CR project.
- Tooling: Open Medical Development Framework (OMDF) [Ouagne10]

Methods - Building semantic resources EHR4CR information model

- Multidimensional EHR4CR Information Model
  - A fact class: ‘Clinical Statement‘
  - A set of dimensions
    - class ‘Subject’
    - class ‘Encounter’
    - class ‘Participation’
  - ‘ClinicalStatementRelationship‘



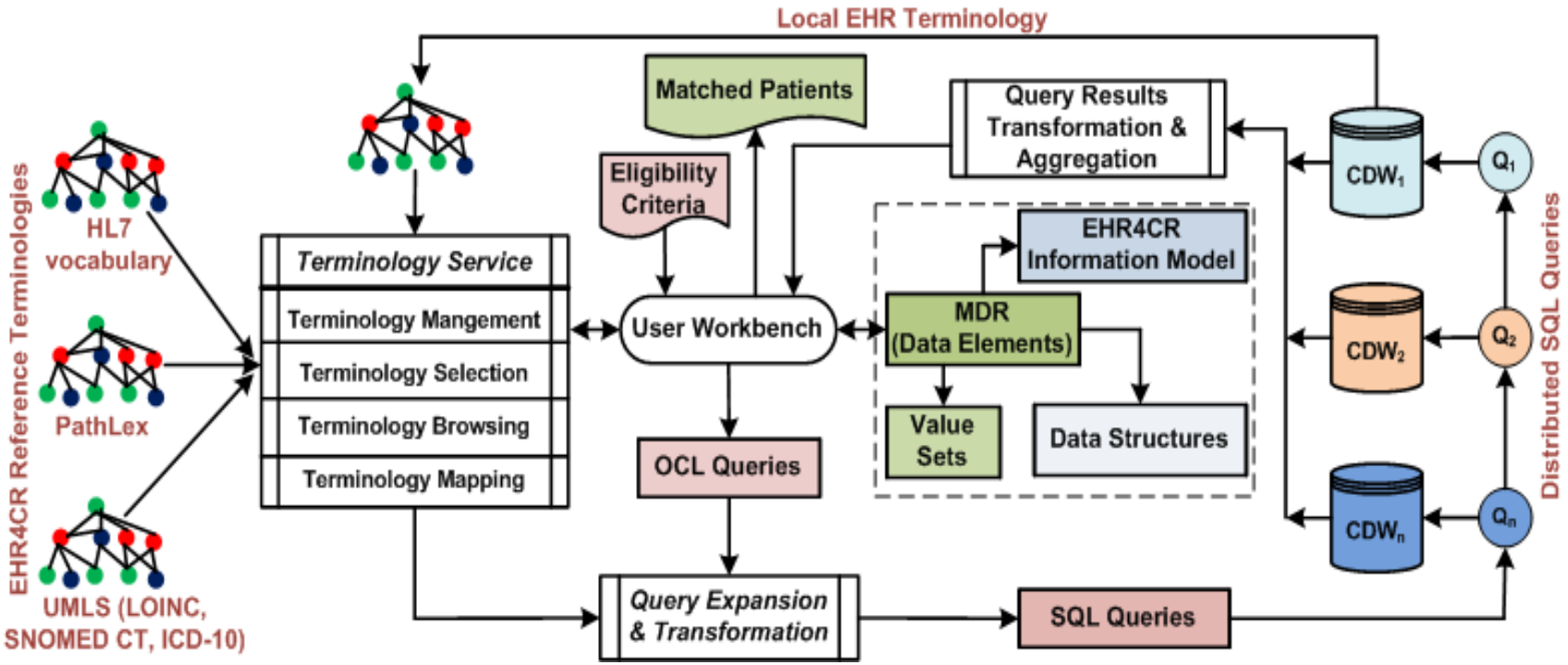
Methods - Building semantic resources EHR4CR core data elements

- Library of agreed clinical data structure definitions
  - Based on generic reference models for representing clinical data (e.g. ISO/HL7 RIM) and on standard data types (ISO 21090)
  - Explicitly bound to reference terminologies/ontologies (e.g. LOINC, SNOMED-CT, ICD-10) through value sets
- Material
  - Data elements extracted from
    - From free-text eligibility criteria
    - Ongoing projects: caDSR (NCI) , eMERGE, SHARE (CDISC), etc
- Method
  - Solving the semantics gap between medical terminologies, ontologies and information models [Shulz10, Sahay11]

Methods - Building semantic resources EHR4CR terminology

- A network of reference terminologies for clinical findings, test results, labs, or medications, etc. (UMLS, Bioportal, etc)
- Material
  - UMLS (SNOMED CT, LOINC, ICD-10 codes, etc.)
  - Non UMLS sources (e.g for ATC, PathLex, epSOS terminology)
- Method & Tooling
  - Generating a networked knowledge-base from available medical ontologies using Semantic Web technologies [Ghazvinian09]
  - LexEVS for uploading terminologies from UMLS

Methods – Developping semantic Interoperability Services for patient eligibility determination



Methods – Developping Terminology Services at the workbench

- For constructing the user-defined eligibility criteria at the Workbench
  - Terminology Selection/Browsing Service
    - Selecting the preferred terminology and terms in which the user wants to define the eligibility criteria (attributes and values from value sets)

Methods – Developping semantic Interoperability Services at the endpoint

- Query Expansion & Transformation Service
  - Based on the pre-defined mappings, we transform the OCL queries into SQL queries based on the local CDW terminology, which can then be executed across different clinical data warehouses
  - Query Expansion & Transformation Service invokes Terminology Mapping Services—for mapping between central and local terminology codes.

Methods – Developping semantic Interoperability Services at the endpoint

- Result Transformation and Aggregation Service:
  - Translating back the query-results obtained from various CDWs into an integrated result format
  - Obtaining the list of all matched patients from the various CDWs that satisfy the initially given eligibility criteria in one uniform and standard view.

Results

- First version of the Semantic Interoperability Framework
- Use of the EHR4CR platform to perform protocol feasibility study queries running on heterogeneous CDWs
  - CDWs developed on purpose for the project (EHR4CR CDWs)
  - i2b2 systems independently developed (AP-HP, FAU, HUG, U936).

Results

10 clinical trials – 11 pilot sites

Internal Study ID	EFPIA Partner	Disease Area	AP-	FAU	HUG	KCL*	MU W	U936	UCL	Univ	UoG	uom	WW	Total
11899	**Bayer**	**Cardiovascular**	X		X								X	**3**
20050182	**Amgen**	**Oncology**		X	X									**2**
27919	**Merck**	**Nervous system disorders**			X				X		X			**3**
BIO111482	**GSK**	**Oncology**								X			X	**2**
CENA713B2315	**Novartis**	**Neurology**						X						**1**
COU-AA-301	**Janssen**	**onco**	X						X					**2**
D3191C00009	**AstraZeneca**	**CV/Arrhythmias**					X							**1**
D4320C00015	**AstraZeneca**	**Oncology**	X			X					X		X	**4**
EFC11785	**Sanofi**	**Oncology**		X		X					X			**3**
NC25113	**Roche**	**Cardiovascular and Metabolic**								X		X		**2**
Total			**3**	**2**	**3**	**2**	**1**	**1**	**2**	**2**	**3**	**1**	**3**	**23**

Result

- 10 clinical trials - 269 eligibility criteria.
  - 99/269 free-text eligibility criteria were manually pre-processed and translated into 186 elementary queries consistently with the template-based approach adopted in the project and represented into a human readable format
  - Elementary queries were formally represented in OCL as constraints on the **EHR4CR information model**
  - Medical concepts of the queries were encoded using the **EHR4CR terminology**
- The OCL were distributed to endpoints in pilot sites and transformed into SQL statements to be executed on heterogeneous information models of legacy CDWs.

Example of free-text eligibility criteria

Free text: Male and female between the ages of 30 to 80 years at screening with diagnosis of idiopathic Parkinson's Disease of *more than 5 years duration*, with a Hoehn and Yahr stage of I-IV during an "off" phase. The diagnosis should be based on medical history and neurological examination.

Example of eligibility criteria in OCL

```
def: getECT01Size(dateStart: TS, dateEnd: TS, dateLimit: TS): Integer =
    Subject.allInstances()->select(sbj Subject |
        sbj.classCode.code = 'PAT'
        and sbj.moodCode.code = 'INSTANCE'
        and sbj.birthTime.greaterOrEqual(dateStart)
        and sbj.birthTime.lessOrEqual(dateEnd)
        and sbj.clinicalStatements->exists(cs ClinicalStatement |
            cs.classCode.code = 'OBS'
            and cs.code.code = 'idiopathic Parkinson''s Disease'
            and cs.effectTime->exists(ts TS | ts.greaterOrEqual(dateLimit))
            )
        and sbj.clinicalStatements->select(cs ClinicalStatement |
            cs.code.code = 'Hoehn and Yahr stage'
            and cs.value.oclIsTypeOf(INT)
            )->exists(cs ClinicalStatement |
                cs.value.oclAsType(INT).value >= 1
                and cs.value.oclAsType(INT).value <= 4
                )
        )->size()
```

Discussion - Conclusion

- In EHR4CR, eligibility criteria were successfully represented and executed over heterogeneous CDWs using
  - An expressive language to define executable eligibility rules
    Formal representation of free-text eligibility criteria [Tu11,Weng10].
  - A standard-based patient information model based on HL7 «StudyDesign» model
    Multidimensional, well suited for querying CDWs.
  - Appropriate reference clinical terminologies

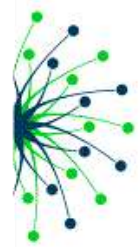
to facilitate mapping from eligibility concepts to patient data

Discussion - Conclusion

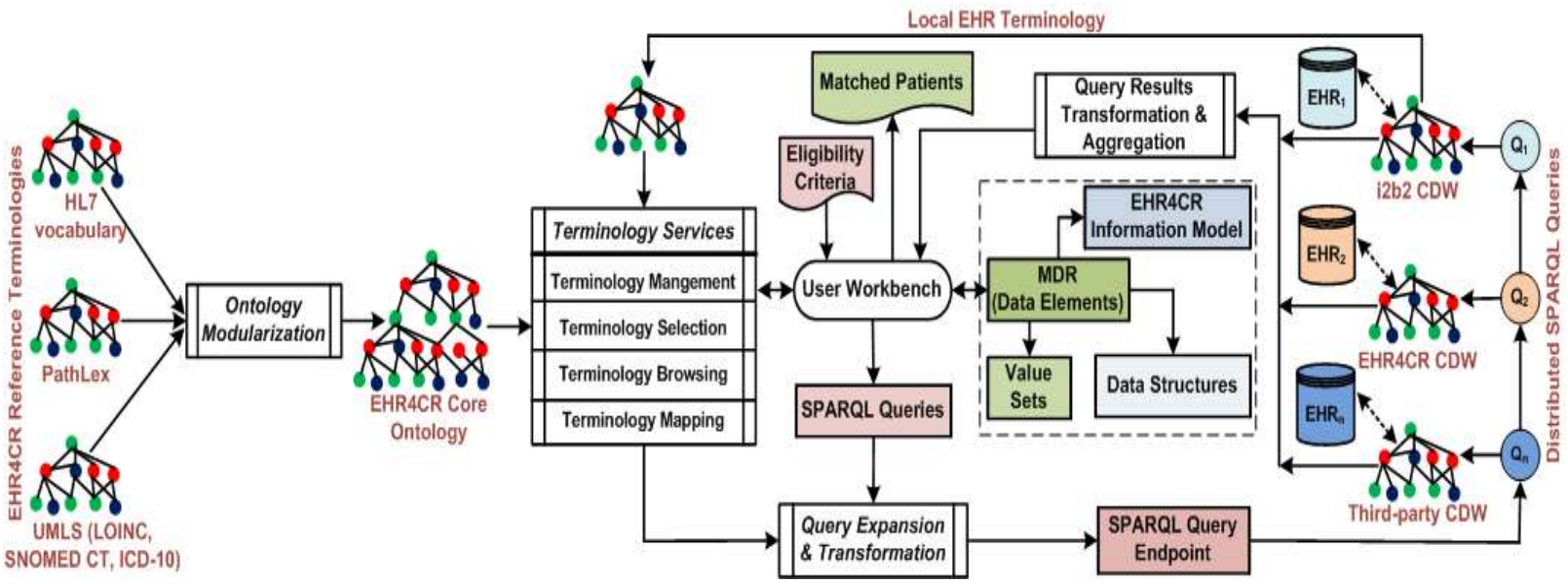
- Issues and imitations
  - Labor-intensive manual task of transforming free-text eligibility criteria in formal queries
  - Expressiveness of the query language (including the possibility of query expansion and temporal constraints)
  - Creation and maintenance of the shared controlled terminology as well as of the mapping between the EHR4CR/local information models and terminologies. In our approach, we deal with the above mentioned issues.

Discussion - Conclusion

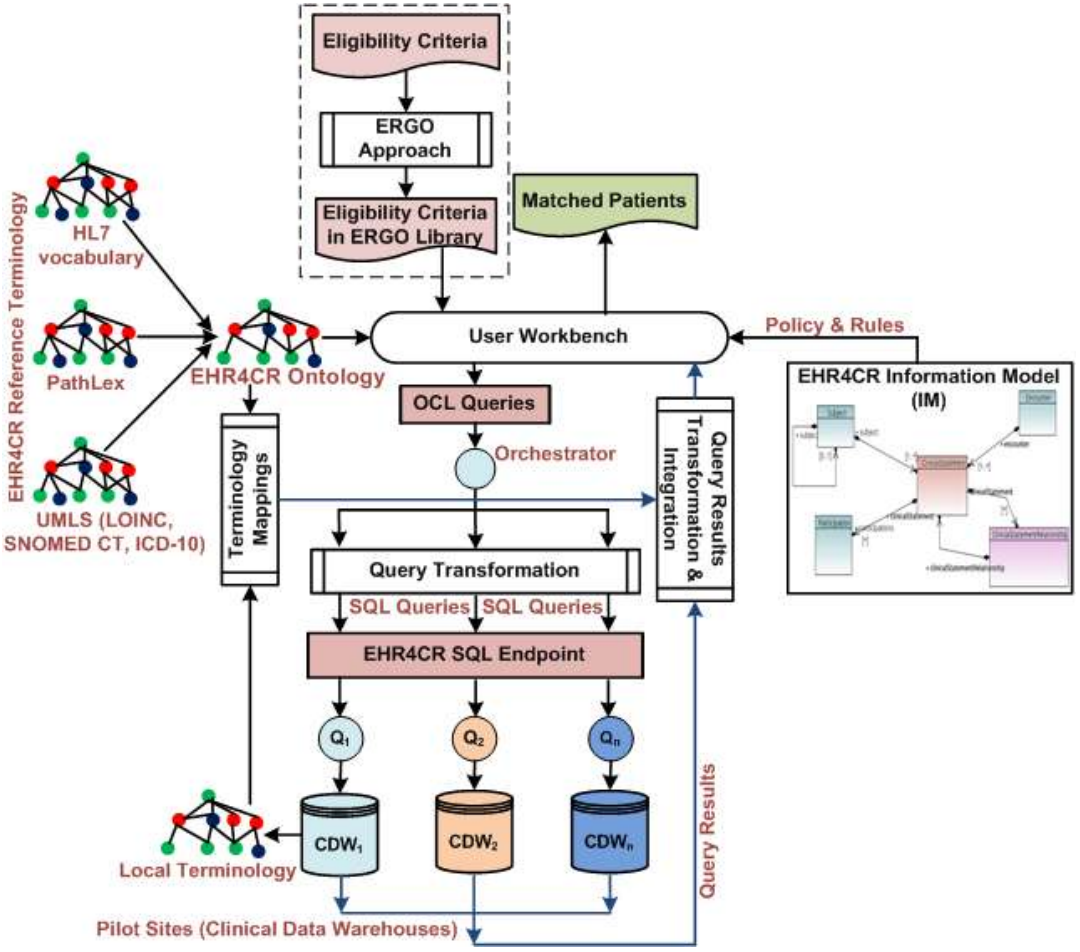
- Short term perspectives
  - To extend information model and OCL expressions in order to better represent temporal information and clinical context
  - To extend the terminology server
    - Develop specific loaders for LexEVS for terminologies that are not yet in UMLS (such as HL7, IHE, CDA and PathLex vocabularies, etc)
    - Represent value sets defined for core data elements



Perspectives
Connecting SPARQL endpoints



Perspectives
Considering the Eligibility Rule Grammar and Ontology (ERGO)



Thank you for your attention
Any question?

Electronic Healthcare Record for Clinical Research (EHR4CR) Online]. Available: http://www.ehr4cr.eu/

El Fadly A, Rance B, Lucas N, Mead C, Chatellier G, Lastic PY, Jaulent MC, Daniel C. Integrating clinical research with the Healthcare Enterprise: From the RE-USE project to the EHR4CR platform. J Biomed Inform. 2011 Dec;44 Suppl 1:S94-S102.

Schulz, S., Schober, D., Daniel, C., and Jaulent, M. (2010). Bridging the semantics gap between terminologies, ontologies, and information models. Stud Health Technol Inform, 160(Pt 2), 1000–1004.

Sahay, R., Zimmermann, A., Fox, R., Polleres, A., and Hauswirth, M. (2011). Interoperability in Healthcare Information Systems: Standards, Management, and Technology, chapter A Formal Investigation of Semantic Interoperability of HCLS Systems. IGI Global.

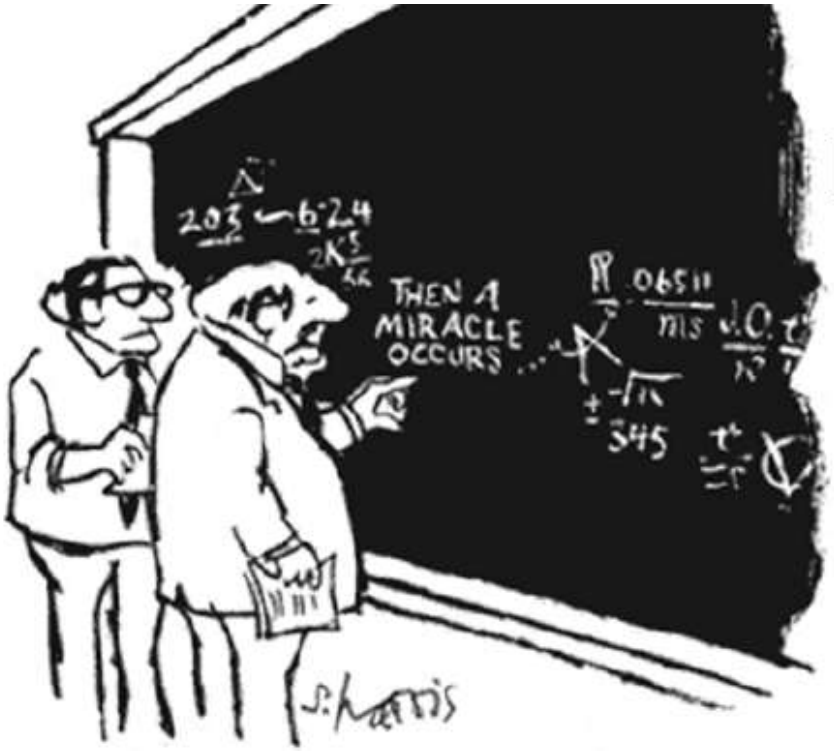
Ghazvinian, A., Noy, N. F., Jonquet, C., Shah, N., and Musen, M. A. (2009). What four million mappings can tell you about two hundred ontologies. In Proceedings of the 8th International Semantic Web Conference, pages 229–242. Springer-Verlag.

Tu SW, Peleg M, Carini S, Bobak M, Ross J, Rubin D, Sim I. A practical method for transforming free-text eligibility criteria into computable criteria. J Biomed Inform. 2011 Apr;44(2):239-50.

Weng C, Tu SW, Sim I, Richesson R. Formal representation of eligibility criteria: a literature review. J Biomed Inform. 2010 Jun;43(3):451-67.

Object Constraint Language, Version 2.0. Online]. Available: http://www.omg.org/spec/OCL/2.0/PDF/

Ouagne D, Nadah N, Schober D, Choquet R, Teodoro D, Colaert D, Schulz S, Jaulent MC, Daniel C. Ensuring HL7-based information model requirements within an ontology framework. Stud Health Technol Inform. 2010;160(Pt 2):912-6.



"I THINK YOU SHOULD BE MORE EXPLICIT HERE IN STEP TWO."