



# Semantic Web Technology for Assessing Clinical Trials Eligibility

Vipul Kashyap, Eric Prud'hommeaux, Helen Chen,  
Jyotishman Pathak, Rachel Richesson, Holger Stenzhorn, M.  
Scott Marshall and Joshua Phillips

CSHALS 2010, Cambridge, MA

February 24, 2010



# Outline



- Developers of this Demonstration
- The Healthcare and Lifesciences Ecosystem
- Use Cases and Functional Requirements
- What is the Semantic Web?
  
- Demo
- Conclusions and Next Steps



# Developers of this Demonstration

- Clinical Observation Interoperability (COI) Task Force

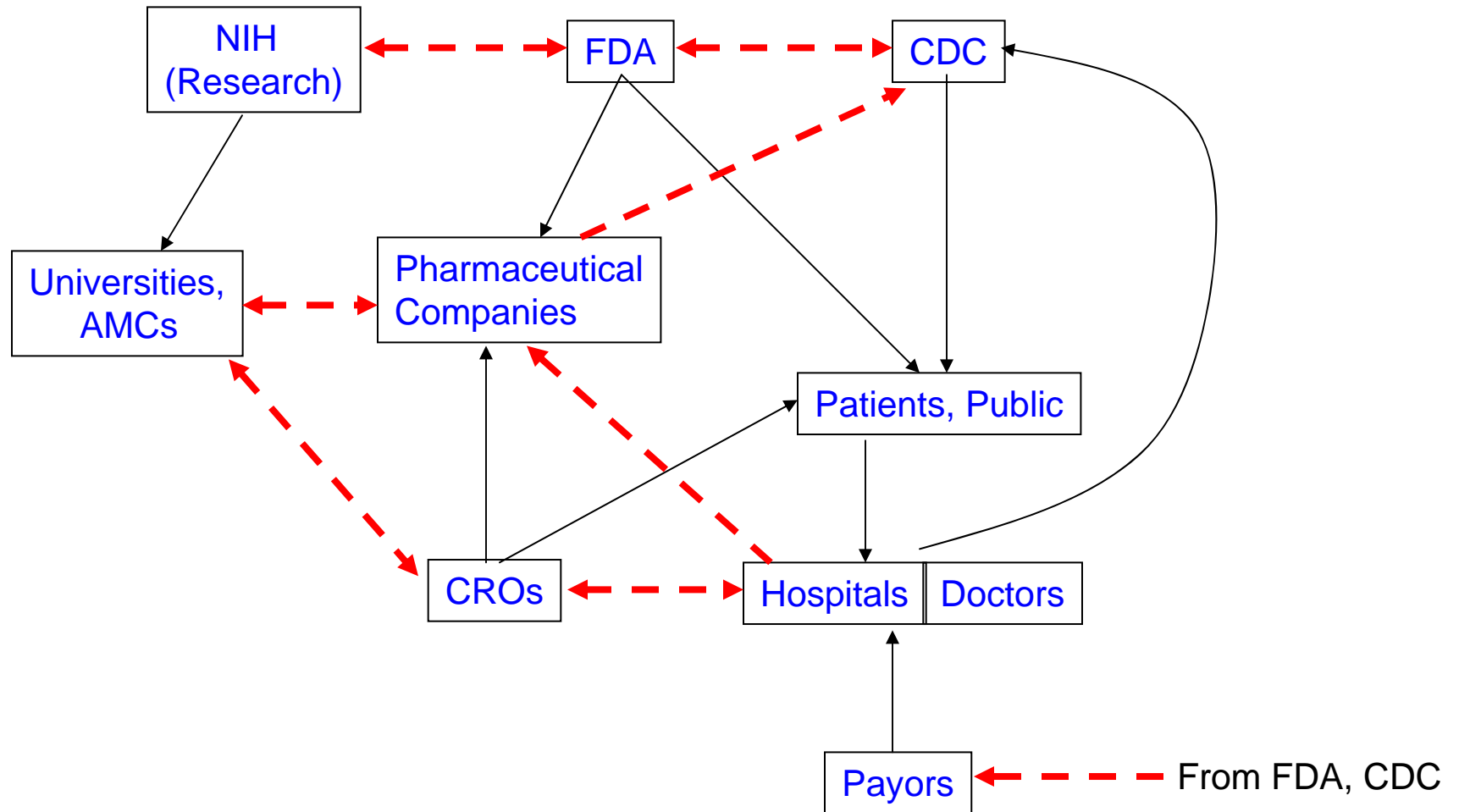
<http://esw.w3.org/topic/HCLS/ClinicalObservationsInterop>



Semantic Web Health Care and Life Sciences (HCLS) Interest Group

- Members from CDISC, clinical trial researchers and

# Healthcare and Life Sciences Ecosystem



- The ability to share and exchange clinical observations is a critical enabler
- Critical to bring down the cost of healthcare in the US!

# Use Cases and Functional Requirements

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	
1	Req#	Requirement Description						Use Cases						Systems			
2			Patient Recruitment	Clinical Trial Data Collection	Drug Safety Surveillance	Clinical Trial Data Collection	Epidemiology	Document Management for Clinical Trials	Research Outcomes	Support Regulatory Approval	Virtual Phase IV Trials	Audit Medication Workflow	Post Launch Drug Use	Adaptive Clinical Trials	Electronic Medical Record (EMR)	Clinical Trials Management System (CTMS)	
3	1	Data Capture and Interfaces															
4	2	Identity and Confidentiality															
5	3	Data Content		X												X	X
6	4	Data Exchange and Interoperability		X												X	X
7	5	Protocol Specification		X													X
8	6	Data Retrieval		X												X	X
9	6	Data Access/Authentication/Audit & Logging															
10	7	Data Quality/Data Integrity															
11	8	Data Storage															
12																	
13																	
14																	
15																	
16																	

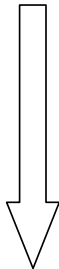
- X identifies the Use Cases, Systems and Functional Requirement under consideration of the COI Task Force
- Based on the Functional Requirements Specification developed by EHRVA/HIMSS



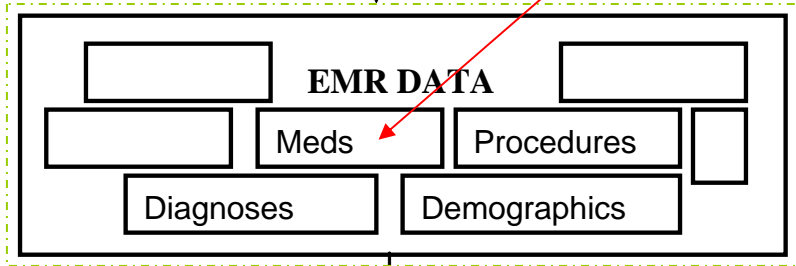
Research Coordinator selects protocol for patient screening:

**Clinical Research Protocol**  
Eligibility Criteria:  
- Inclusion  
- Exclusion

**PROBLEM:**  
*Same construct in 2 different representations....*



Research Coordinator views list of patients and selects which ones to approach in person for evaluation and recruitment.



Clinical Evaluation and Recruitment

Patient MR #	Potentially Eligible for Protocol	# Criteria Met / Total Criteria in Protocol	Criteria #1 (Pass/Fail/Researcher Needs to Evaluate)	No Criteria #2 (Pass/Fail/Researcher Needs to Evaluate)	Criteria #3 (Pass/Fail/Researcher Needs to Evaluate)	...
0011111	Yes	6/8 criteria met	Pass	Pass	Pass	...
0022222	No	3/8 criteria met	Pass	Fail	Pass	...
0033333	Yes	5/8 criteria met	Pass	Pass	Fail	...
...	...	...	...	...	...	...



# Examples: Drug Class in Research Protocols

- monotherapy with **metformin**, **insulin secretagogue**, or **alpha-glucosidase inhibitors** and a low dose combination of all
- Long term **insulin** therapy
- Therapy with **rosiglitazone** (Avandia) or **pioglitazone** (Actos), or **extendin-4** (Byetta), alone or in combination
- **corticosteroids**
- **weightloss drugs** e.g., Xenical (orlistat), Meridia (sibutramine), Acutrim (phenylpropanol-amine), or similar medications
- **nonsteroidal anti-inflammatory** drugs
- Use of **warfarin** (Coumadin), **clopidogrel** (Plavix) or other **anticoagulants**
- Use of **probenecid** (Benemid, Probalan), **sulfinpyrazone** (Anturane) or other **uricosuric** agents

# Prescription Information in EMR



- "132139","131933","98630","GlipiZIDE-Metformin HCl 2.5-250 MG Tablet","54868079500",  
",98630,"2.5-250","TABS","","MG","","15","GlipiZIDE-Metformin HCl",  
",","","GlipiZIDE-Metformin HCl 2.5-250 MG Tablet"
- "132152","131946","98629","GlipiZIDE-Metformin HCl 2.5-500 MG Tablet","54868518802",  
",98629,"2.5-500","TABS","","MG","","15","GlipiZIDE-Metformin HCl",  
",","","GlipiZIDE-Metformin HCl 2.5-500 MG Tablet"
- "132407","132201","98628","GlipiZIDE-Metformin HCl 5-500 MG Tablet","54868546702",  
",98628,"5-500","TABS","","MG","","15","GlipiZIDE-Metformin HCl",  
",","","GlipiZIDE-Metformin HCl 5-500 MG Tablet"
- "132642","132436","C98630","GlipiZIDE-Metformin HCl TABS","54868079500",  
",98630","","TABS","","","","15","GlipiZIDE-Metformin HCl",  
",","","GlipiZIDE-Metformin HCl TABS"

**NDC Code**



# Clinical Observations Interoperability



Construct:

Clinical Trial  
Eligibility

Patient  
Characteristics

Data/Knowledge  
source:

Research  
Protocols

EMR

Semantic Model:

SDTM

DCM/RIM

Standards  
Development  
Organization:

CDISC

HL7



# Semantic Web Technologies



- RDF (Resource Description Framework)
- OWL (Web Ontology Language)
- RIF (Rule Interchange Format)
- N3 (Notation 3)
- SPARQL (Query Language for RDF)




# Methods




- Developed semantic models for:
  - clinical trial based upon SDTM
  - clinical practice based upon RIM/DCM
- Encoded Eligibility queries using:
  - The SDTM model
  - SPARQL queries
- Storage of Clinical Data from a real world clinic in a relational database
- Mappings
  - Mappings between clinical trials and clinical practice constructs
  - Use of drug ontology to facilitate mappings on drug concepts
- Mapping of RIM/DCM model to a relational database schema
- Query Transformation:
  - Translation of an SDTM SPARQL Query into DCM/RIM SPARQL query
  - Translation of DCM/RIM query into SQL query
  - Execution of the SQL query against the relational database



# COI Demo – Clinical Trial Eligibility


 Clinical Observations Interoperability (COI) Demo : Clinical Trial Eligibility Criteria Selection

---

**INCLUSION CRITERIA** 

Domain	Inclusion Category	Constraint	
--------	--------------------	------------	--

---

**EXCLUSION CRITERIA** 

Domain	Exclusion Category	Constraint	
--------	--------------------	------------	--

---

**SPARQL**

**Clinical Trial Eligibility: Inclusion Criteria:**

- Ages Eligible for Study: 18 Years - 75 Years,
- Genders Eligible for Study: Both
- Type 2 diabetes on diet and exercise therapy or monotherapy with metformin, insulin secretagogue, or alpha-glucosidase inhibitors, or a low-dose combination of these at  $\leq 50\%$  maximal dose (see Appendix). Dosing is stable for 8 weeks prior to randomization.

**Clinical Trial Eligibility: Exclusion Criteria:**

- Therapy with rosiglitazone (Avandia) or pioglitazone (Actos), or extendin-4 (Byetta), alone or in combination in the previous 6 months
- Use of weight loss drugs [e.g., Xenical (orlistat), Meridia (sibutramine), Acutrim (phenylpropanol-amine), or similar over-the-counter medications] within 3 months of screening or intentional weight loss of  $\geq 10$  lbs in the previous 6 months
- Use of warfarin (Coumadin), clopidogrel (Plavix) or other anticoagulants
- Use of probenecid (Benemid, Probalan), sulfinpyrazone (Anturane) or other uricosuric agents

**Query Patients** **Reset Criteria**

# COI Demo – Selecting Inclusion Criteria

W3C Semantic Web Clinical Observations Interoperability (COI) Demo : Clinical Trial Eligibility Criteria Selection

**INCLUSION CRITERIA**

Domain	Inclusion Category	Constraint	
sdm	Metformin		x

Inclusion in SDTM ontology

**EXCLUSION CRITERIA**

Domain	Exclusion Category	Constraint	
--------	--------------------	------------	--

SPARQL

**Demographic Information**

Male  Female

Minimum Age  Maximum Age

**Medication:**

- LaboratorySpecimenConditionior
- LaboratorySpecimenType
- Medication
  - Acutrim
  - Alpha-Glucosidase Inhibitors
  - Anticoagulant
  - BetaBlocker
  - Corticosteroid
  - ExtendIn-4
  - InsulinSecretagogue
  - Metformin
  - NSAID
  - Orlistat
  - Phenylpropanolaamine

SDTM clinical trial ontology

# Criteria in SPARQL



```
?medication1 sdtm:subject ?patient ;  
  spl:activeIngredient ?ingredient1 .  
?ingredient1 spl:classCode 6809 .
```

metformin

```
OPTIONAL {  
  ?medication2 sdtm:subject ?patient ;  
  spl:activeIngredient ?ingredient2 .  
  ?ingredient2 spl:classCode 11289 .  
} FILTER (!BOUND(?medication2))
```

anticoagulant

Exclusion Criteria



# COI Demo – Drug Ontology Inference

W3C Semantic Web Clinical Observations Interoperability (COI) Demo : Clinical Trial Eligibility Criteria Selection

**INCLUSION CRITERIA**

Domain	Inclusion Category	Constraint	
sdtm	Metformin		x

**EXCLUSION CRITERIA**

Domain	Exclusion Category	Constraint	
do	Anticoagulant		x

Exclusion in Drug ontology

**Demographic Information**

**Drug**

- argatroban
- Indanedione
- Phenindione
- anisindione
- Direct acting anticoagulant
- Pentosan Polysulfate
- Pentosan Polysulphate Sodium
- Selective factor Xa inhibitor
- fondaparinux
- Fondaparinux sodium
- Ancrod
- Heparin
- Unfractionated heparin
- Enoxaparin
- Enoxaparin sodium
- Dalteparin
- Dalteparin Sodium
- tinzaparin
- Tinzaparin sodium
- reviparin
- Reviparin sodium
- ardeparin

*Subclasses of "anticoagulant"*

SPARQL

Apply Criteria

Query Patients Reset Criteria

# COI Demo – Selecting Mapping Rules

W3C Semantic Web Clinical Observations Interoperability (COI) Demo : Clinical Trial Eligibility Criteria Selection

### INCLUSION CRITERIA

Domain	Inclusion Category	Constraint	
sdtm	Metformin		x
sdtm	Male	true	x
sdtm	Female	true	x
sdtm	AgeMin	18	x
sdtm	AgeMax	75	x

### EXCLUSION CRITERIA

Domain	Exclusion Category	Constraint	
do	Anticoagulant ⚠		x

### Demographic Information

Male       Female

Minimum Age       Maximum Age

### Medication:

SDTM Query

SDTM - HL7 Mapping Rule

Database - HL7 Mapping Rule

Drug Ontology Mapping Rule

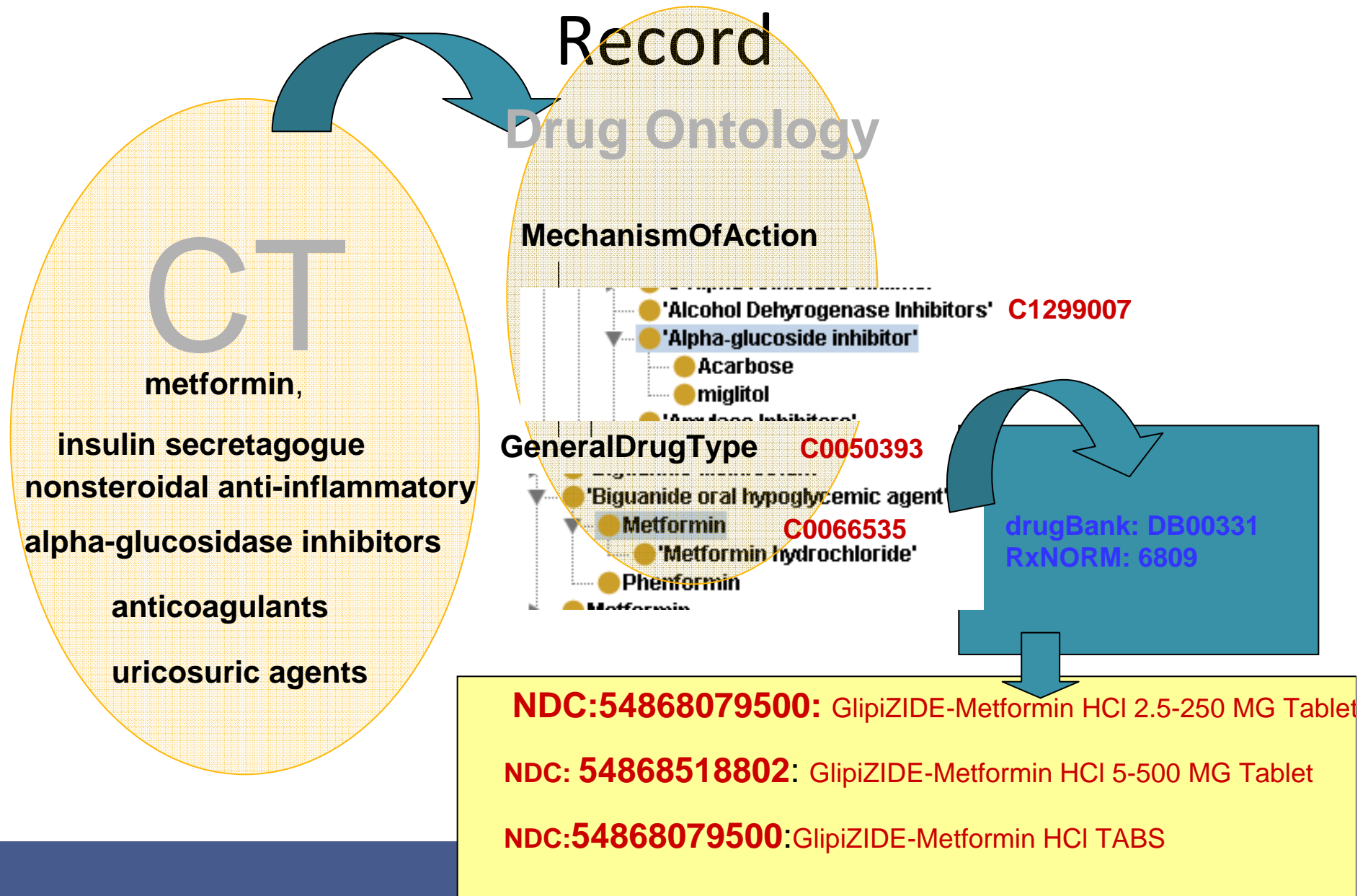
**#check all drugs that "may\_treat obese"**

```
{?A rdfs:subClassOf ?B; rdfs:label ?D.  
?B a owl:Restriction;  
    owl:onProperty :may_treat;  
    owl:someValuesFrom :C0028754}  
=>  
{?D a :WeightLoseDrug}.
```

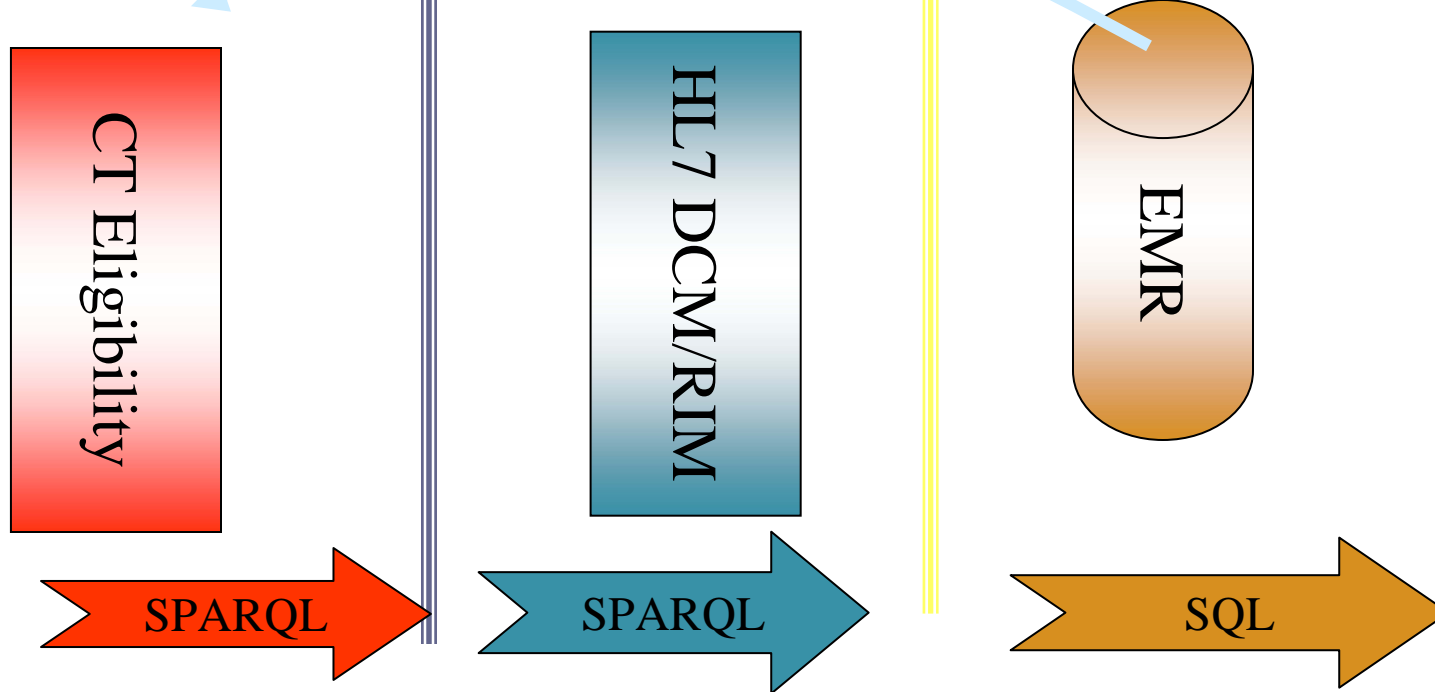


# Mapping Between CT and Patient Record

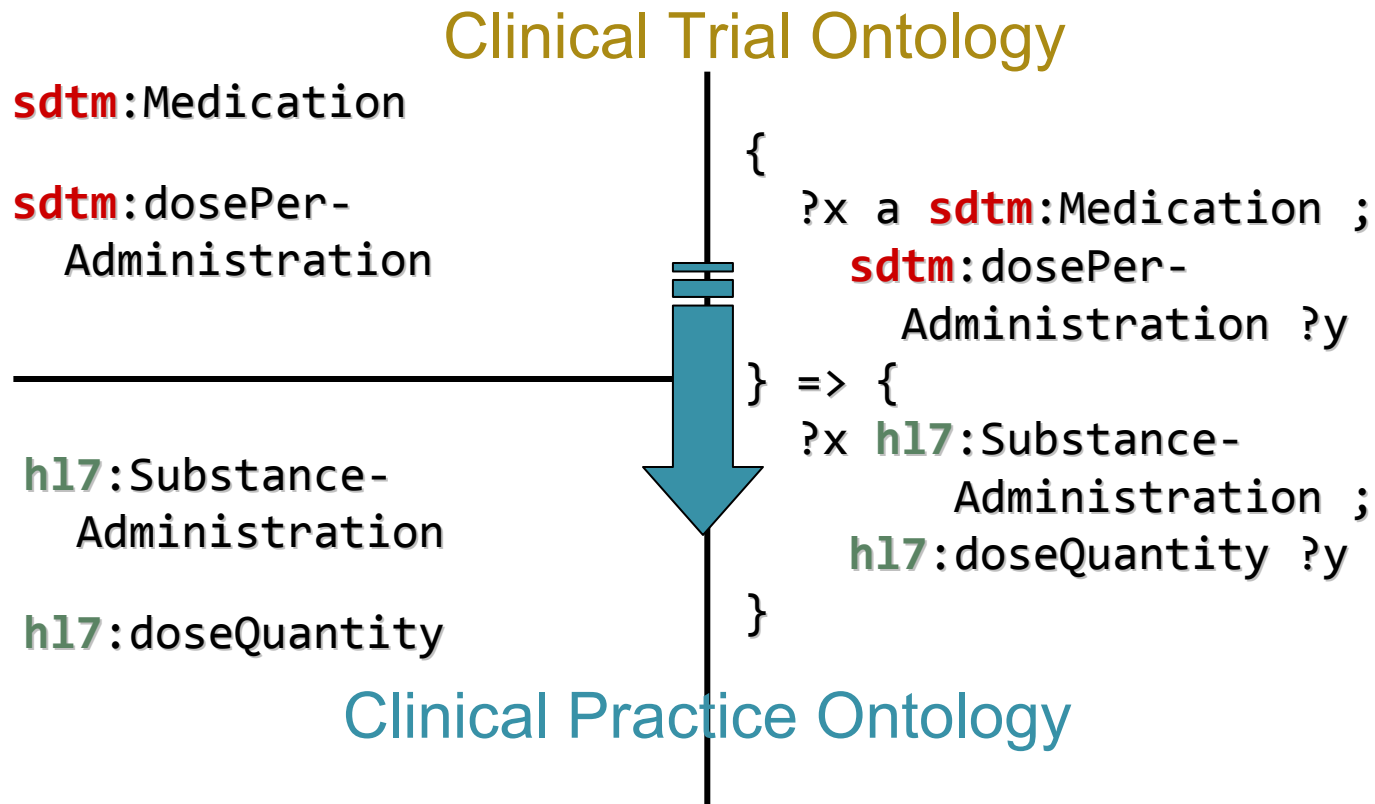


# Pushing Query to Database

- SPARQL in SDTM ontology to SPARQL in HL7 ontology
- SPARQL in HL7 ontology to SQL in EMR database

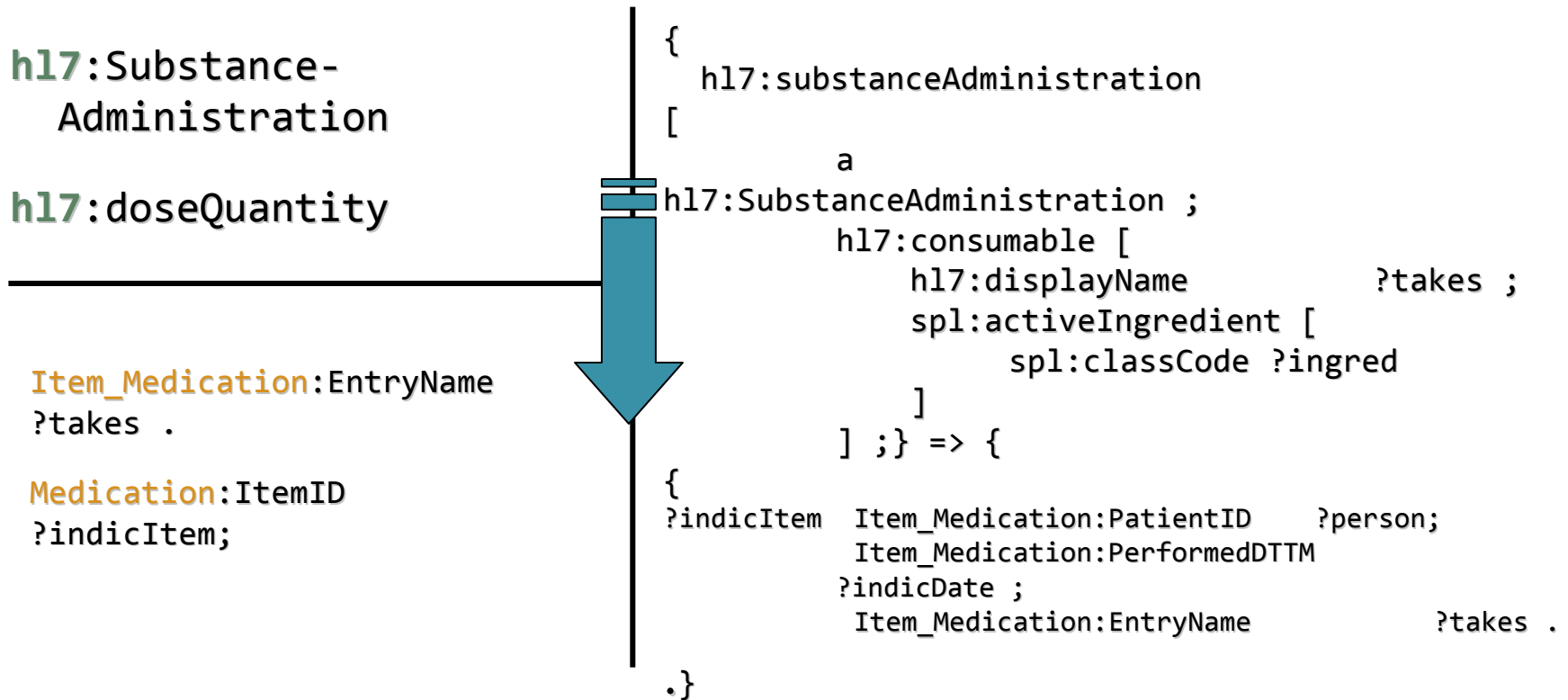


# SDTM to HL7 Transformation



# HL7 to EMR Database Transformation

## SPARQL in Clinical Practice Ontology



SQL to EMR Database

# SQL to Database



**SELECT** patient.id AS patient, patient.DateOfBirth AS dob, sexEntry\_gen0.EntryName AS sex, indicItem\_gen1.EntryName AS takes, indicItem\_gen1.PerformedDTTM AS indicDate

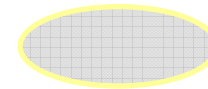
**FROM** Person AS patient

**INNER JOIN** Sex\_DE AS sexEntry\_gen0 ON sexEntry\_gen0.id=patient.SexDE

**INNER JOIN** Item\_Medication AS indicItem\_gen1 ON indicItem\_gen1.PatientID=patient.id

**INNER JOIN** Medication AS indicMed\_gen2 ON indicMed\_gen2.ItemID=indicItem\_gen1.id

**INNER JOIN** Medication\_DE AS indicDE\_gen5 ON  
indicDE\_gen5.id=indicMed\_gen2.MedDictDE



**INNER JOIN** NDCcodes AS indicCode\_gen6 ON indicCode\_gen4.ingredient=6809 AND  
indicCode\_gen6.NDC=indicDE\_gen5.NDC



# COI Demo – Getting Right Patients

## Clinical Observations Interoperability (COI) Demo : Patient List

Patient	Birthday	Gender	Medication	PrescribDate
1517441	1964-09-01 00:00:00	Female	Glucophage TABS	2007-07-23 00:00:00
1517441	1964-09-01 00:00:00	Female	Glucophage TABS	2007-07-23 00:00:00
1517000	1958-01-01 00:00:00	Male	Metformin HCl 500 MG Tablet	2007-06-29 00:00:00
1517000	1958-01-01 00:00:00	Male	Metformin HCl 500 MG Tablet	2007-06-29 00:00:00
1517403	1981-05-11 00:00:00	Male	Metformin HCl 500 MG Tablet	2008-01-03 00:00:00
1505583	1950-02-25 00:00:00	Female	Metformin HCl 500 MG Tablet	2008-01-14 00:00:00
1234561	1983-01-02 00:00:00	Male	GlipiZIDE-Metformin HCl 2.5-250 MG Tablet	2007-09-28 00:00:00
1234561	1983-01-02 00:00:00	Male	GlipiZIDE-Metformin HCl 2.5-250 MG Tablet	2007-09-28 00:00:00
1234562	1963-12-27 00:00:00	Female	GlipiZIDE-Metformin HCl 2.5-250 MG Tablet	2007-09-28 00:00:00
1234562	1963-12-27 00:00:00	Female	GlipiZIDE-Metformin HCl 2.5-250 MG Tablet	2008-07-28 00:00:00
1234563	1983-02-25 00:00:00	Male	GlipiZIDE-Metformin HCl 2.5-250 MG Tablet	2007-07-28 00:00:00
1517441	1964-09-01 00:00:00	Female	Glucophage TABS	2007-07-23 00:00:00
1517441	1964-09-01 00:00:00	Female	Glucophage TABS	2007-07-23 00:00:00
1516986	1995-06-13 00:00:00	Male	Metformin HCl 1000 MG Tablet	2007-12-04 00:00:00

# COI Work - Evolving



- coi svn:
  - <http://code.google.com/p/coi/source/checkout>
- Public access:
  - <http://hcls.deri.org/coi/demo/>
- Application to the NCI – caBIG Platform – Led by Joshua Phillips
- Presentation in the Electronic Data for Clinical Trials Track at SCOPE 2010, Philadelphia, March 2010



# caBIG Outreach



- caBIG program is actively investigating use of Semantic Web
- Semantic Infrastructure Initiatives [1] include applying W3C Semantic Web technology.
- Identified HCLS/caBIG Collaboration Opportunities [2]
  - Initial focus
    - Applying COI approach to BreastCancerTrials.org data.
    - Reaching out to Clinical Trials Reporting Program (CTRP) [3] to develop Structured Eligibility Criteria representation.
  - Future
    - Evaluating security and privacy policy

1. <http://tinyurl.com/ya8qkbp>
  2. <http://tinyurl.com/ygcsgsx>
  3. <http://www.cancer.gov/clinicaltrials/ctrp/>
- 

# Conclusions



- Benefits of Semantic Web Approach:
  - Unambiguous conceptual model for separate domains without early commitment to a common model.
  - Reusable/Configurable mapping rules
  - Late binding of coding systems, models and database schema.
  - Query Transformation approach reflecting real time discovery and integration needs
- Need to design and instantiate interoperability architecture for multiple cross-industry use cases
- Need to align with industry standards, e.g., information models, vocabularies
- Imperfection in information models and vocabularies needs to be accepted and improved iteratively. Not a good idea to wait for perfection! Let's try to demonstrate incremental value ..



# Acknowledgements



- Major developers:
  - Helen Chen
  - Holger Stenzhorn
  - Eric Prud'hommeau
- Other supporters
  - Jennifer Fostel
  - Bo Anderssen
  - Kerstin Forsberg
  - M. Scott Marshall
  - Tom Oniki
- Dr. Parsa Mirhaji,  
University of Texas Health  
Science Center at Houston,  
Center for Biosecurity and  
Public Health Informatics  
Research (sample data)
- W3C Interest Group on  
the Semantics for the  
Healthcare and Life  
Sciences (HCLS)

