



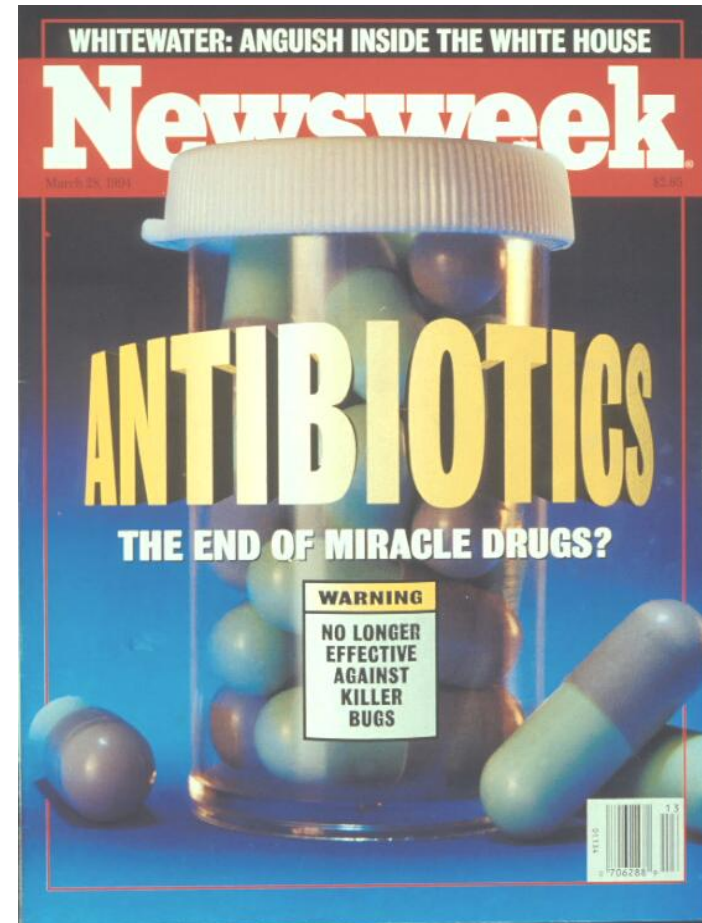
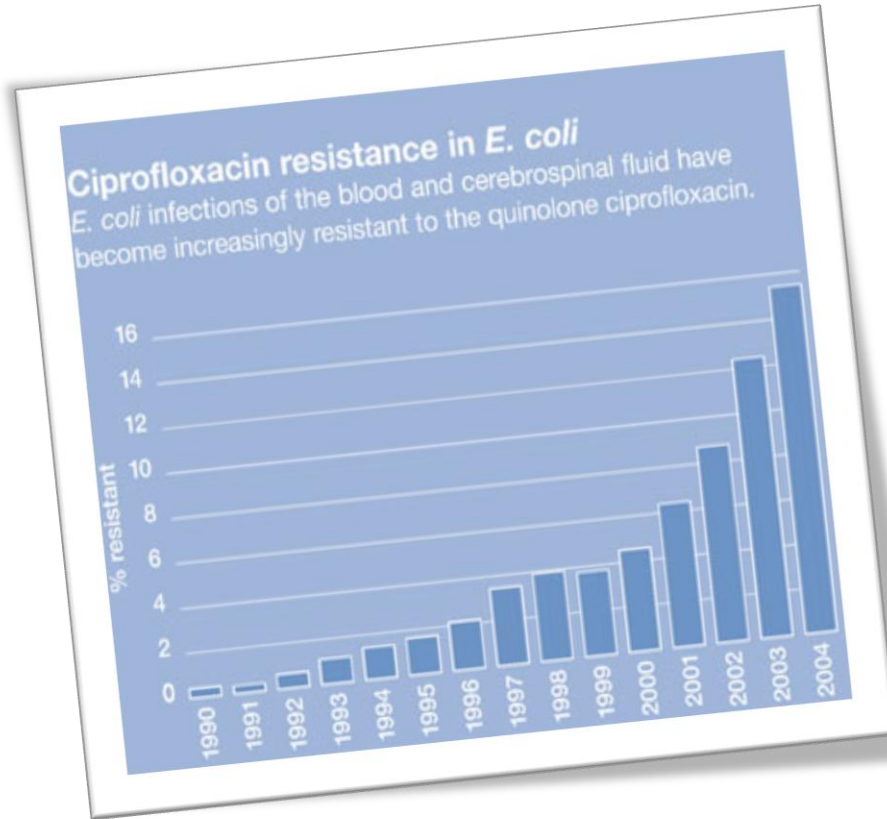
The DebugIT Core Ontology - Semantic integration of antibiotics resistance patterns

Medinfo 2010

Daniel Schober, Martin Boeker, University Medical Center Freiburg



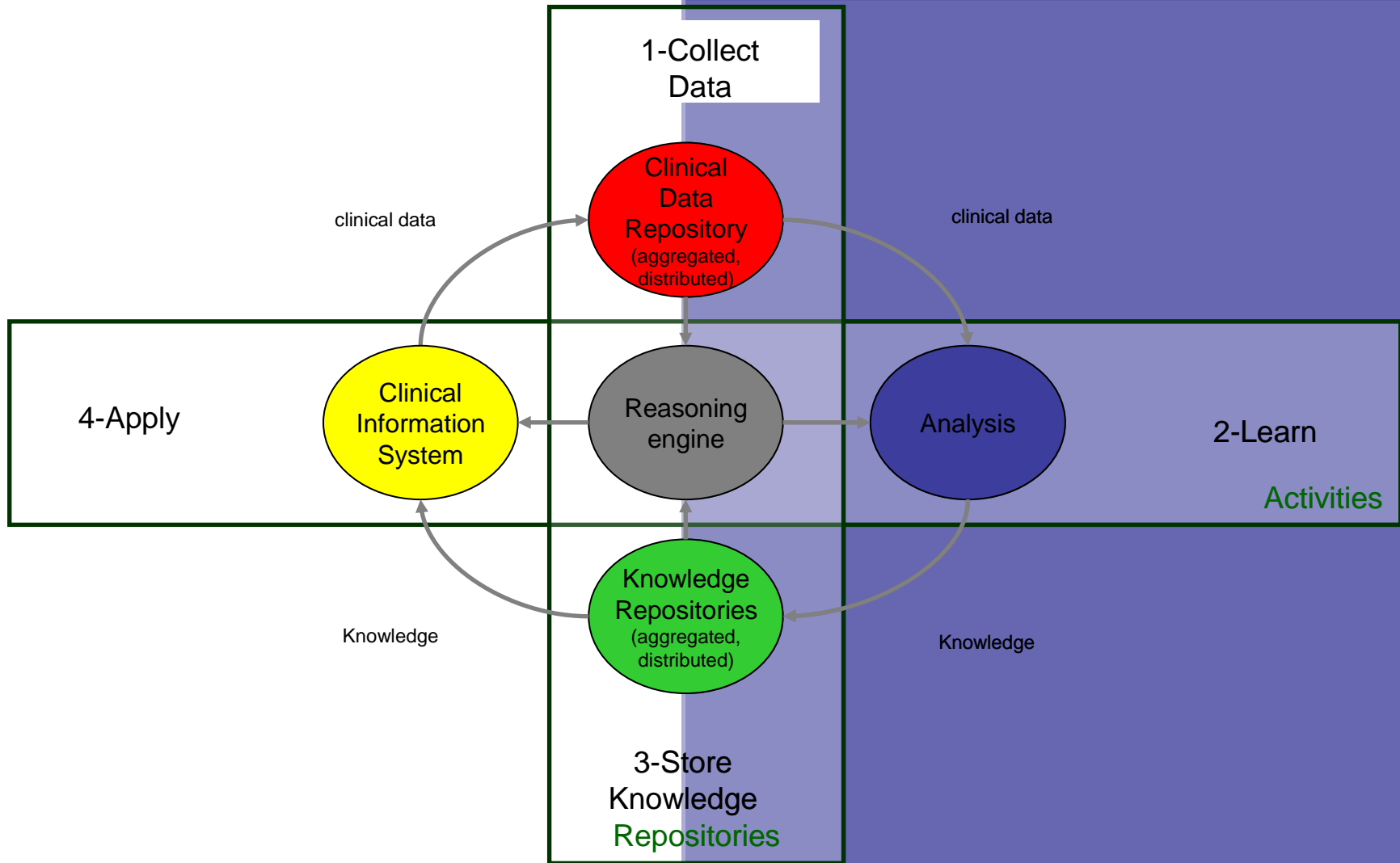
Project Healthcare Context



A need for 'IT-biotics'

- **DebugIT**
 - Detecting and Eliminating Bacteria Using Information Technology
 - Using 'semantic linked data' to exploit distributed clinical data
 - Acquire **new knowledge**
 - Through advanced data mining
 - Apply knowledge in **decision support**
 - Antibiotics choice, drug dosages
 - Apply knowledge in **monitoring**
 - Analyze current & predict future outcomes / trends
 - Discover patient safety patterns

The DebugIT project: general architecture



Using Ontologies in DebugIT

- Provide semantic identifiers
 - Allow crosstalk within Interoperability Platform
 - Used in SPARQL querying to express research questions
 - Provide formal meaning exploitable by logic reasoners
- Integrate access to heterogeneous CIS

Competency question

“What percentage of **Escherichia coli cases**, **cultured** from urine samples, is resistant to the combination of **trimethoprim/sulfamethoxazol (TMP/SMX)** or **trimethoprim** over a period x?”

Implied Classes

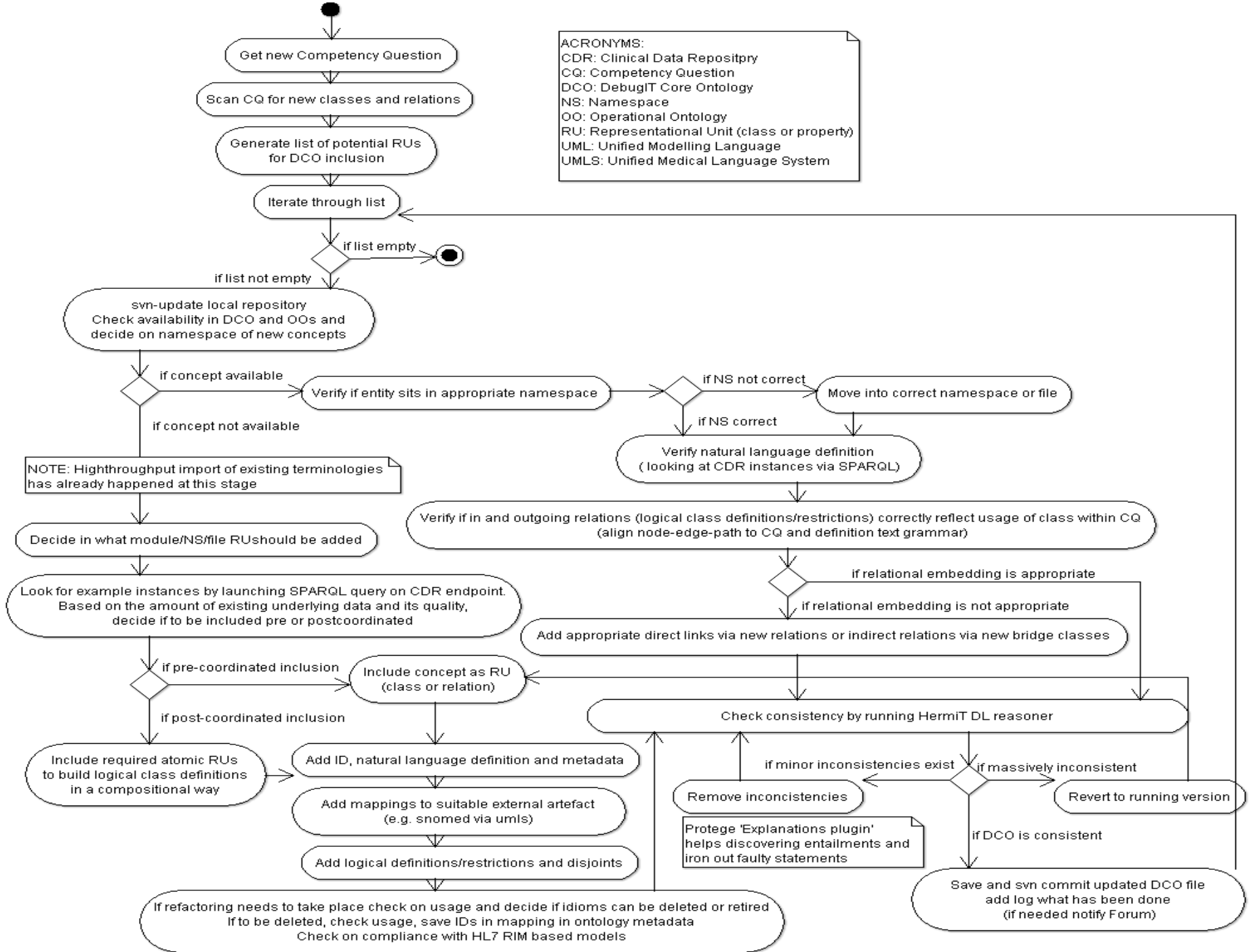
- Patient
- **Case**
- Disease
- **Microorganism**
- **Antimicrobial drug**
- Treatments
- **Microbial Culture**
- **Sampling**
- Diagnostic procedure
- Time

DCO ontology

- OWL-DL (description logics)
 - Reasoner for autoclassification & consistency checks during OE
 - Reasoner infers multiple parenthood
- Reusing BioTop
 - Ensure a rigid modeling view
 - Provides reuseable constraints (bridges to
- Concepts harvested from
 - Hospital CDR schemata
 - Competency questions from clinical use case
 - Datadriven *bottom up*
 - Domain terminologies in use
 - Via UMLS or OLS
 - Ontology modularisation tools (A.Rector)
 - HL7 v3 based



ACRONYMS:
 CDR: Clinical Data Repository
 CQ: Competency Question
 DCO: DebugIT Core Ontology
 NS: Namespace
 OO: Operational Ontology
 RU: Representational Unit (class or property)
 UML: Unified Modelling Language
 UMLS: Unified Medical Language System



SKOS mappings

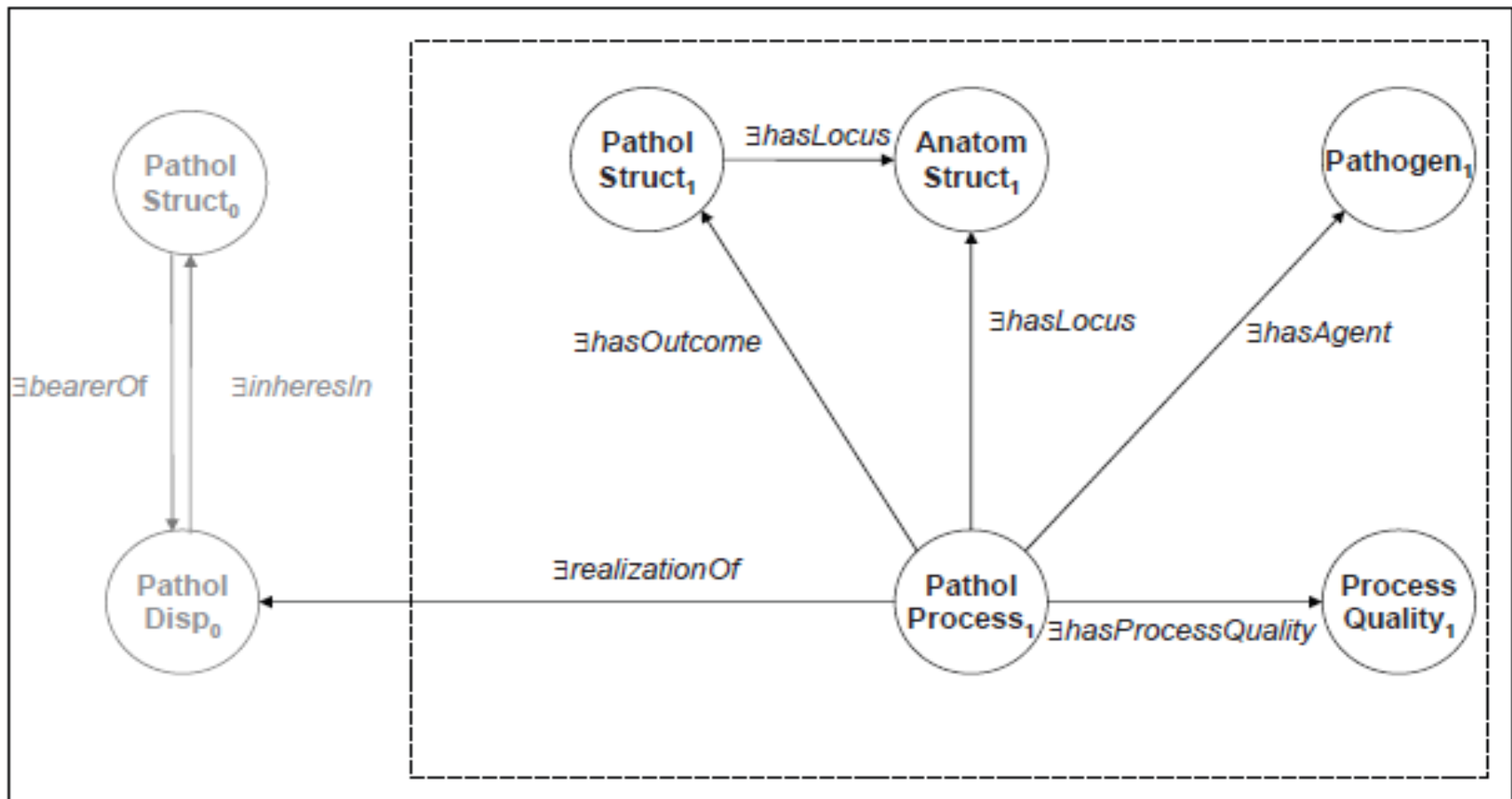
for

to

- **Diseases** → SNOMED CT findings
- **Human anatomy** → Foundational Model of Anatomy
- **Bacteria** → Uniprot NEWT taxonomy
- **Drugs, antibiotics** → WHO ATC codes

DCO disease-pattern example

Distinction between pathological disposition, process & structure



DCO content (statistics)

Ontology elements & axioms	Overall	DCO	BioTop
Classes	1281	965	375
Object Properties (relations)	78	3	74
Datatype Properties	11	10	0
Subclass Axioms	1494	1050	444
Equivalent Class Axioms	197	98	99
Disjoint Axioms	76	1	75

Artificial Intelligence helping DCO modelling

- **Reasoner infers new facts from given statements**
 - E.g. enrich 'Sample' Class hierarchy by inference
 - discover and classify 'Samples' that are 'BodyLiquidSamples'
- **Reasoner alerts on faulty class definitions**
 - Inconsistent class definitions are highlighted
 - DL semantics used to spot and avoid modelling errors

... two examples ...

Inference of new facts (BloodSample is a BodyLiquidSample)

Stated Facts

BodyLiquidSample =

● Sample
and derivesFrom some BodyLiquid

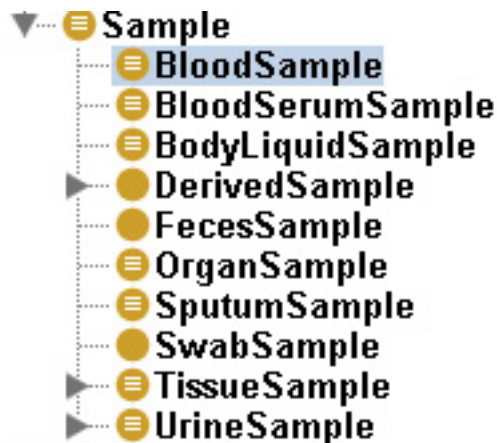
BloodSample =

● Sample
and derivesFrom some Blood

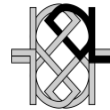
BodyLiquid =

● Blood
● BloodSerum
● CerebrospinalFluid
● Sputum
● Urine

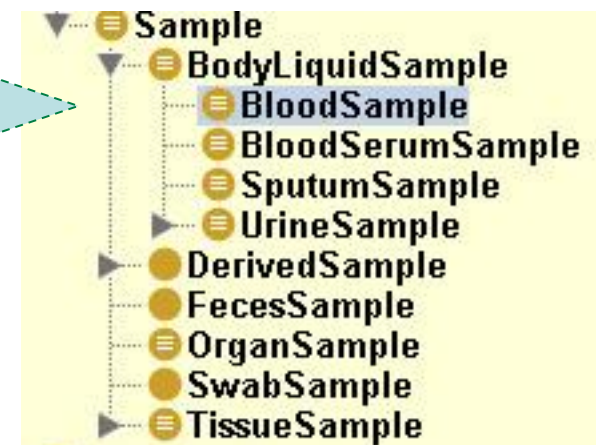
Asserted Hierarchy (flat list)



Logics Reasoner



Inferred Hierarchy (more structure)



Autodetecting inconsistent classes

The screenshot displays a software interface with two main panels. The left panel, titled 'Asserted class hierarchy', shows a tree view of classes. The right panel, titled 'Class Annotations', shows details for the selected class 'disk diffusion susceptibility test'.

Asserted class hierarchy:

- concentration
 - 'drug concentration'
 - 'minimum inhibitory concentration'
- 'drug dose'
- frequency
- 'inspired oxygen fraction'
- 'laboratory analysis finding'
- pressure
- size
- temperature
- weight
- 'route of administration'
- sex
- 'test result'
- 'time stamp'
- 'quality region'
 - 'canonicity region'
 - 'developmental region'
 - 'taxon region'
 - 'bacteria growth region'
 - 'clinical score region'
 - 'culture condition region'
 - 'diagnosis evidence region'
 - 'laboratory procedure'
 - 'measurement of substance in sample'
 - 'urine dipstick test'
 - 'urine leukocyte esterase test'
 - 'urine nitrite test'
 - 'microbiology procedure'
 - 'antimicrobial susceptibility test'
 - 'disk diffusion susceptibility test'
 - 'minimum inhibitory concentration suscept'
 - 'bacterial serotyping'
 - 'microbial stain analysis'
 - 'microbiological culture procedure'
- 'microbiological susceptibility region'
 - intermediate

Class Annotations:

Annotations: 'disk diffusion susceptibility test'

Annotations (some annotations are hidden) +

- CID: "C0201180"
- SID: "359872008"
- label: "disk diffusion susceptibility test"

Description: 'disk diffusion susceptibility test'

Equivalent classes +

- 'antimicrobial susceptibility test' and 'has outcome' some (result and encodes some 'inhibitory zone size')
- Nothing

Superclasses +

Inherited anonymous classes

- 'has participant' some sample
- 'laboratory procedure' and has_participant some 'bacterial cell'
- 'has participant' some antibiotic
- 'preceded by' some 'microbiological culture procedure'
- has_participant some (sample and has_proper_part some 'bacterial cell')
- has_participant some (amount-of-matter and bearerOf some 'antibiotic role')
- 'has outcome' some antibiogram

- Contents
- All Ontologies
 - All Resources
 - All Classes (1204)
 - All Object Properties (87)
 - All Data Properties (8)
- dco
- Classes (861)
 - Object Properties (26)
 - Data Properties (8)
- biotop
- Classes (343)
 - Object Properties (61)

- 'adapted empirical therapy protocol'
- 'admission to hospital department'
- 'admission to hospital'
- 'admission to intensive care unit'
- 'admitter role'
- 'adult organism'
- 'adult region'
- 'adverse event diagnosing'
- 'adverse event'
- 'aerobic bacterial culture procedure'
- 'age quality'
- 'agent in'
- 'allergy to amoxicillin'
- 'allergy to anti-infective agent'
- 'allergy to antibacterial drug'
- 'allergy to ciprofloxacin'
- 'allergy to drug'
- 'allergy to fosfomycin'
- 'allergy to lomefloxacin'

Class: 'antimicrobial susceptibility test'

http://www.debugit.eu/ontology/1.0/dco.owl#AntimicrobialSusceptibilityTest

Asserted Class Hierarchy

- + 'diagnostic procedure'
 - + 'laboratory procedure'
 - + 'microbiology procedure'
 - + **'antimicrobial susceptibility test'**
 - 'disk diffusion susceptibility test'
 - 'minimum inhibitory concentration susceptibility test'

Annotations (8)

- CID: "C1697858"
- definition: "'Test methods designed to determine a microorganism`s susceptibility to being killed by an antibiotic' (HL7 Version 3.0)"
- pl: "AntimicrobialSusceptibilityTests"
- sg: "AntimicrobialSusceptibilityTest"
- SID: "14788002"
- synonym: "antibiogram procedure"
- comment: "ToDo: Link to SampleCollection and to SampleSource !"
- label: "antimicrobial susceptibility test"

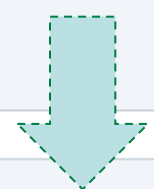
Superclasses (5)

- 'microbiology procedure'
- 'has outcome' some antibiogram
- 'has participant' some antibiotic
- 'has participant' some ('amount of substance' and ('bearer of some 'reagent role'))
- 'preceded by' some 'bacterial culture procedure'

Usage (11)

- /* acetext: "Every AntimicrobialSusceptibilityTest hasOutcomes an Antibiogram." */
- /* acetext: "Every AntimicrobialSusceptibilityTest hasParticipants an AmountOfSubstance that bearerOfs a ReagentRole." */
- /* acetext: "Every AntimicrobialSusceptibilityTest hasParticipants an Antibiotic." */

Use CNL for Ontology Evaluation



DCO usage in SPARQL querying

“ Select all UrineSample antibiograms for E.Coli on Trimethoprim with result = susceptible for patients with Urinary Tract Infection ”

PREFIX **dco:** <<http://www.debugit.eu/ontology/1.0/dco.owl#>>

CONSTRUCT {

 _:therapy a **dco:TreatingUrinaryTractInfection**;
 biotop:hasPatient _:patient.

 _:urineSampling a **dco:UrineSampleCollection**;
 biotop:hasParticipant _:patient;
 dco:hasOutcome _:urineSpecimen.

 _:culturing a **dco:MicrobiologicalCultureProcedure**;
 biotop:hasParticipant _:urineSpecimen;
 dco:hasOutcome [a **dco:Result**; biotop:encodes _:bacteriaName].

 _:bacteriaName biotop:qualityLocated [a biotop:SpeciesEscherichiaColiRegion].

 _:susceptibilityTest1 a **dco:AntimicrobialSusceptibilityTest**;
 biotop:precededBy _:culturing;
 dco:hasParticipant [a **dco:Trimethoprim**];
 dco:hasOutcome [a **dco:Result**; biotop:encodes ?susceptibility1].

 ?susceptibility1 a **dco:MicrobiologicalSusceptibility**;
 biotop:qualityLocated [a ?result1].

[...] }

WHERE { DDO mappings go here }

Conclusion

- Semantically rich Ontology used in practice
 - Practical SPARQL query building
 - Covering use case requirements
- DL-reasoning helps ontology engineering
 - DL limitation justified for smaller ontologies
 - For larger models use rule-based reasoning
- Some domain modeling issues
 - Time modeling
 - TimeQuality classes vs. xsd dateTime datatype properties
 - Intervals: Episode of care, Patient Stay
 - Qualities & quality regions

Resources & Acknowledgements

Resources

- **Ontology sources**
 - <http://www.imbi.uni-freiburg.de/~schober/dco/>
- **TermBrowser**
 - http://www.imbi.uni-freiburg.de/~schober/dco_owlDoc/
- **Simplifications & reviews on**
<http://www.imbi.uni-freiburg.de/~schober/Simplifications/>

Acknowledgements

- Hans Cools, Martin Boeker, Csaba Huszka, Kristof Depraetere, Douglas Teodoro, Remy Choquet, Stefan Schulz, Ilinca Tudose, Maren Kechel, Djamila Raufie, Jessica Bullenkamp, the DebugIT team
- Funded by grant agreement ICT-2007.5.2-217139



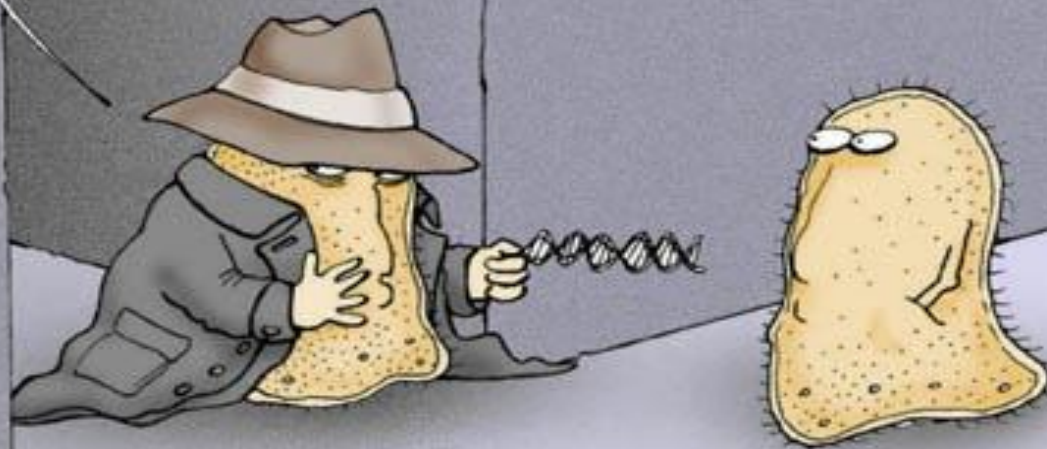
DEBUGIT

Detecting and Eliminating Bacteria Using Information Technology
Large-scale integrating project (FP7, 2008 - 2011)



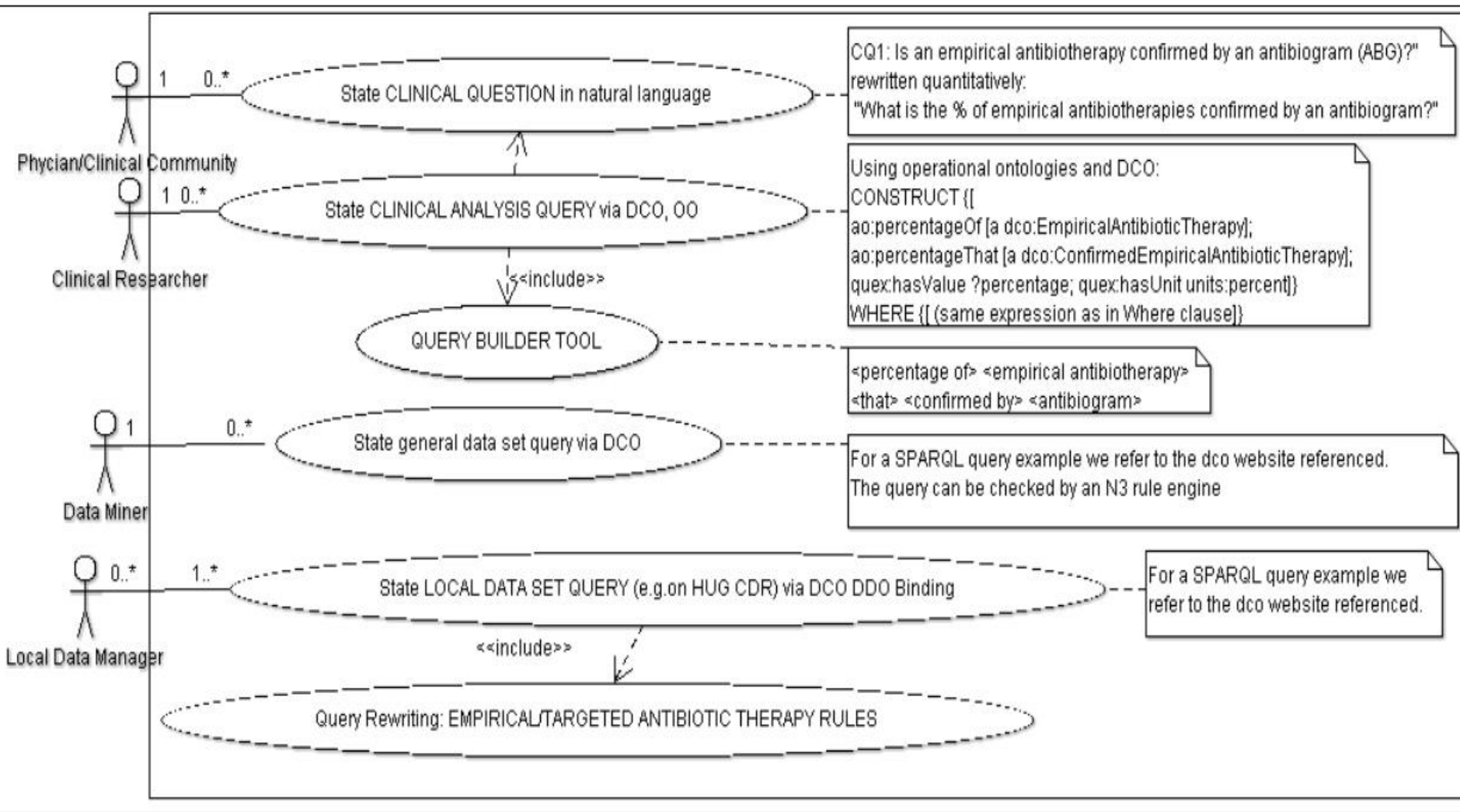
In the Hospital kitchen I was approached by a member of the feared 'Antibiotics Resistance' ...

Pssst! Hey kid! Wanna be a Superbug...?
Stick some of this into your genome...
Even penicillin won't be able to harm you...!

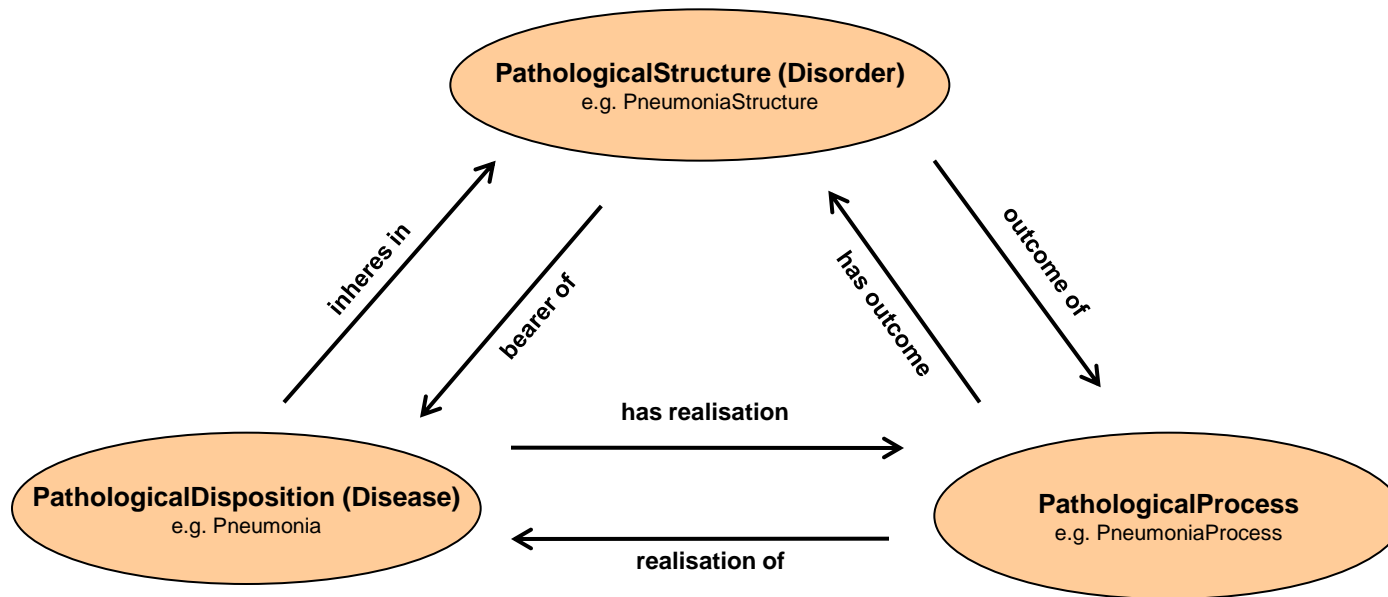


WCK

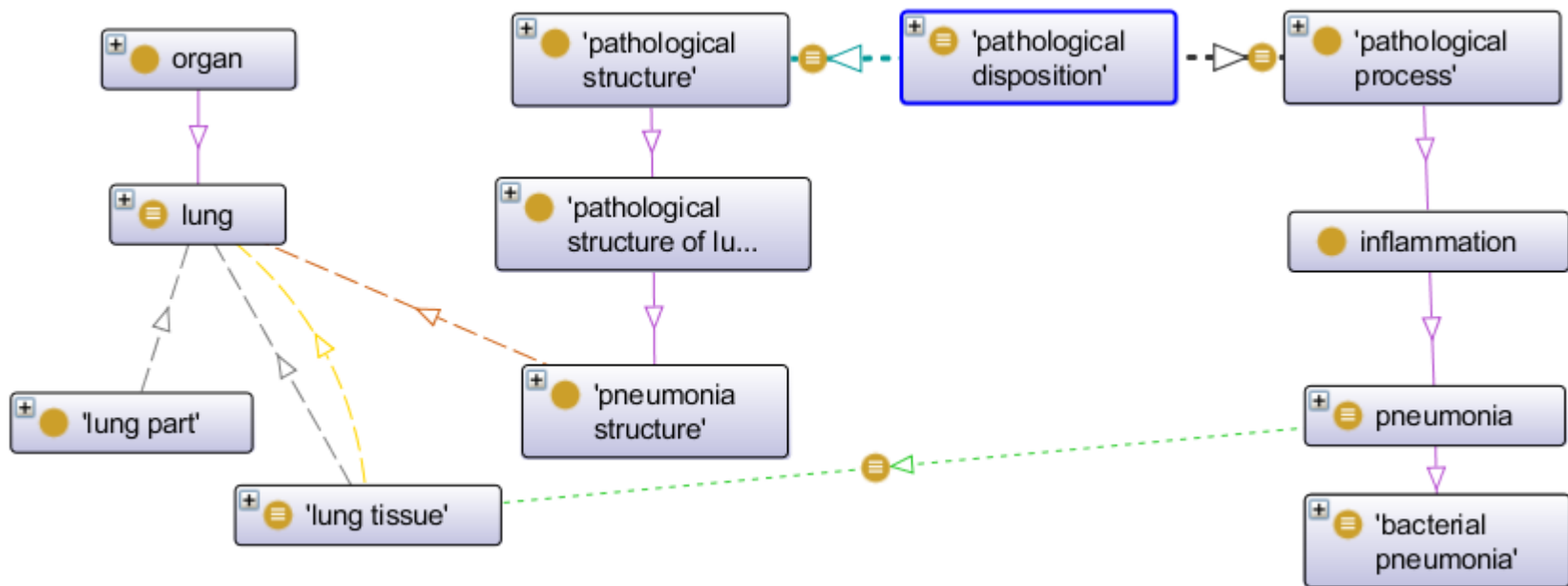
UML Use Case diagram illustrating Ontology usage in formalisation steps for competency question



A tripartite granular disease model



A tripartite disease model (detail)



Next steps

- Enhance coverage
- Refinement of DCO structure
 - Addressing drugs dosages & disease therapies
 - Use rectors Snomed CT modularisation algorithm to extract relevant SNOMED CT IDs form DCO-provided seed list
- Evaluation
- Publish and distribute
 - E.g. on Bioportal

UML Diagram Standard for Visualizing Processes

- UML Activity diagram
 - DCO Engineering
 - Kickoff phase
 - C:\Programme\ArgoUML\initial_phase_onto-dev_debug-it.emf
 - DCO OE, when receiving new clinical question
 - C:\Programme\ArgoUML\DCO_OEactivityUML.gif
- UML Use Case diagram
 - Ontology usage in CQ 8 querying
 - <C:\Programme\ArgoUML\HansExamplesUseCase8UML.gif>

.... Find all these diagrams in the DCO Design principle documents !

query

Formal World

GUI

OO: CAO, AO
DO: DCO, MEO

SPARQL query/ies

• Ontologies

CASQ by Clinitian

result set(s)

• rules

DO: DCO, MEO
(OO: CAO, AO)

DSQ by Dataminer

DDO

D2R

SPARQL end point

SQL query



Types of ontologies within DebugIT

DebugIT Core Ontology (DCO)

- Clinical domain of discourse
 - medical domain of infectious diseases
- OWL-DL



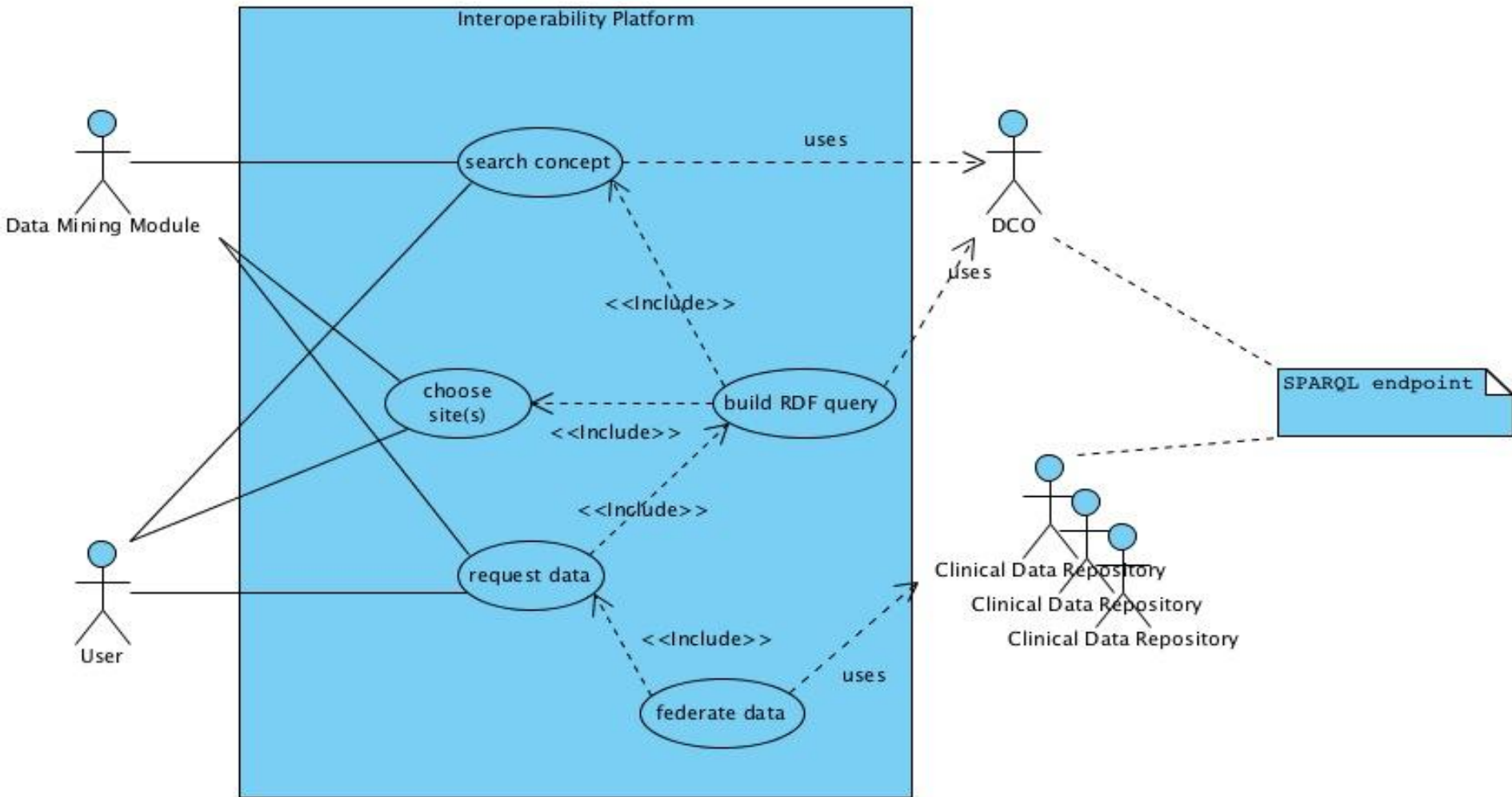
Operational ontologies (OO)

- Implementation & module crosstalk
 - query building, statistics, analysis, evidences, maths, units, ...
- OWL-Full

Data Definition Ontologies (DDO)

- Describing hospital specific CIS Data model

Generating SPARQL queries for clinical data integration



DCO coverage

- Administrative patient care level
 - provider institutions, clinical episode
- Individual patient level
 - personal attributes
- Clinical level
 - diseases, findings, adverse reactions, anatomical entities
- Microbiological level
 - pathogens, culture, antibiograms, specific treatment
- Geospatial level
 - bedroom, surgery room, common shared facilities
- Epidemics
- Molecular level both on pro- and eukaryotes
 - genes, proteins
- Evolution
 - mutagenesis, prokaryote gene transfer

Reasoning example

- Given statements:

Blood: Superclass

BodyLiquid

^

→gets exploited by reasoner

BodyLiquidSample: Equivalent

Sample and derivesFrom some `BodyLiquid` →gets exploited by reasoner

BloodSample: Equivalent

Sample and derivesFrom some `Blood`

→gets autoclassified by reasoner

- Inferred** statement:

→**BloodSample** is a **BodyLiquidSample**

BloodSample: Equivalent

Sample and derivesFrom some `Blood`

Superclass

BodyLiquidSample

DL reasoner maintains multiple parenthood

Enriching the *Sample* hierarchy by auto-classifying
BodyLiquidSamples

E.g. from the given facts ...

- $Blood \sqsubseteq BodyLiquid$
- $BodyLiquidSample \equiv Sample \sqcap \exists \text{ derivesFrom}.BodyLiquid$
- $BloodSample \equiv Sample \sqcap \exists \text{ derivesFrom}.Blood$

... the reasoner infers that ...

- $BloodSample \sqsubseteq BodyLiquidSample$

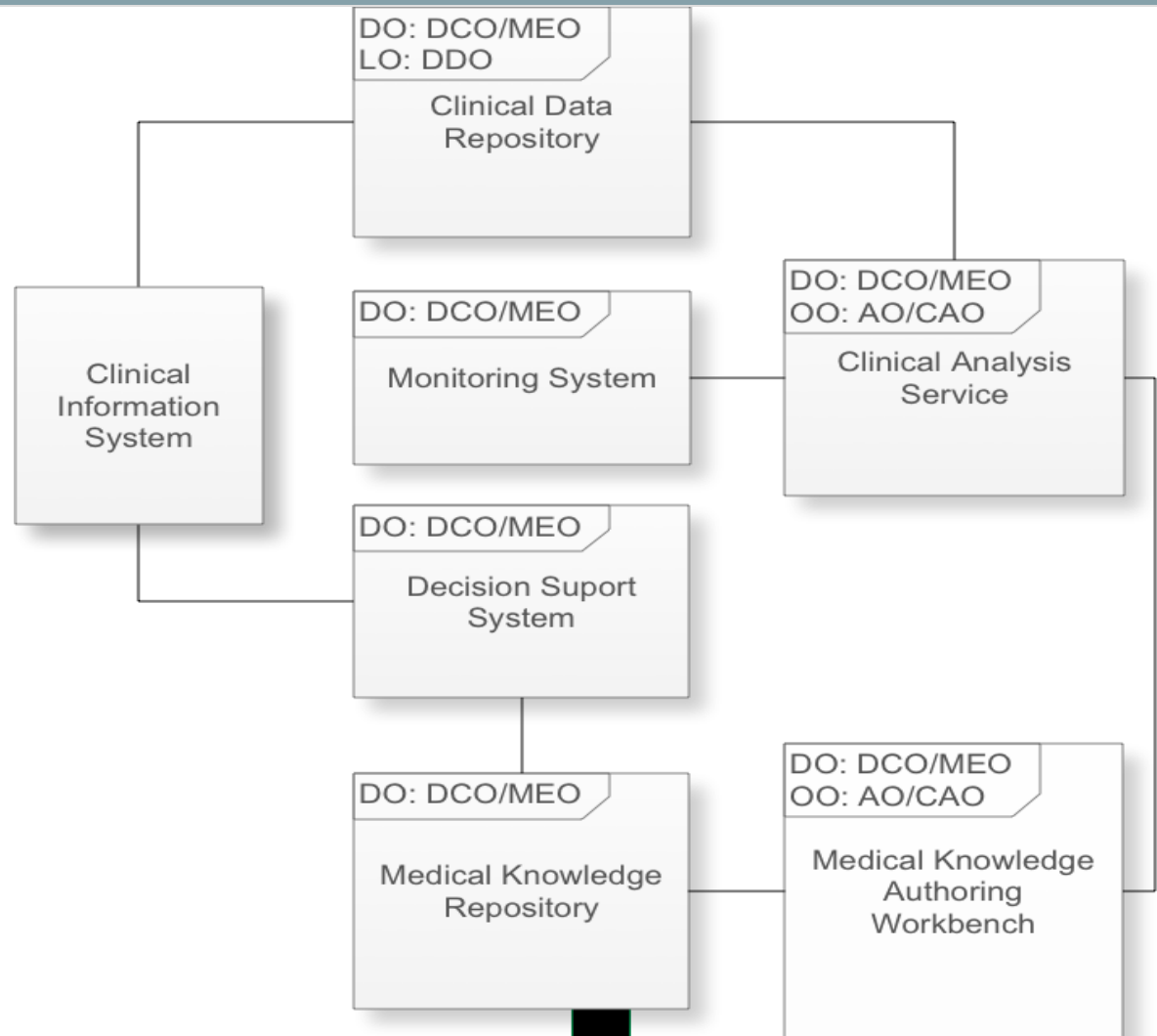
... enriching the taxonomy

This is also done for all other samples

DCO evaluation

- Ultimate overall evaluation
 - Can clinicians
 - run the overall system ?
 - build queries and understand results ?
 - Can data miner
 - create data set results ?
 - do data mining and formalize quality criteria for results ?
- DCO internal evaluation
 - Fitness for use tested by ability to answer CQ
 - Evaluate validity of assertions by
 - Reasoners
 - Graphical and textual representations to domain experts
 - Serialization of modules into **Constrained Natural Languages** (CNL)

The use of the DebugIT ontologies in overall IP

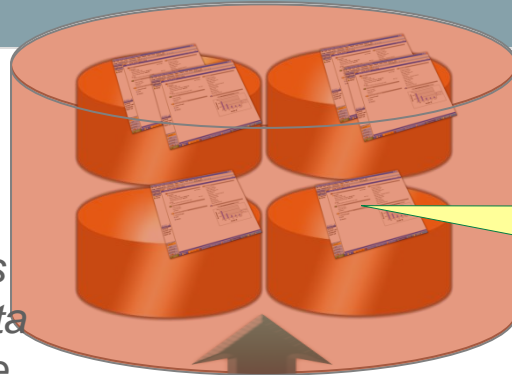


<http://www.sign.ac.uk/pdf/sign50.pdf>

Ontology role in UKLR view

WP2 Clinical Data Repository

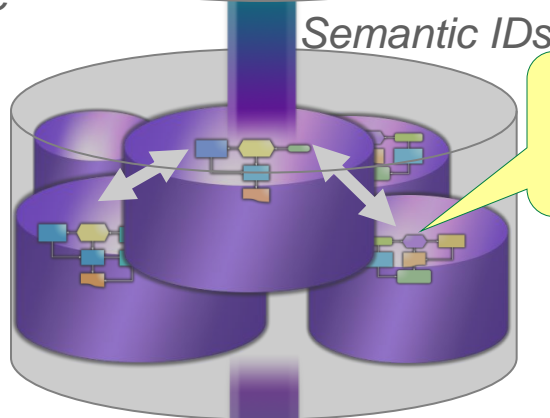
*represents
patient data
knowledge
& beliefs*



*Patient #123 suspected
to be infected by MRSA
after heart transplant*

WP1 Ontologies

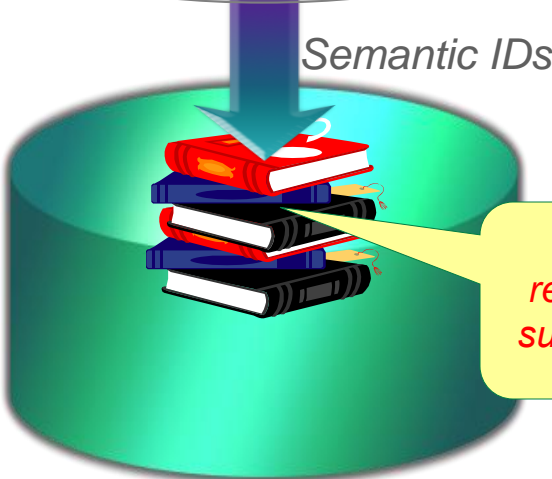
*represent
invariant
meaning*



*MRSA is defined as SA
for which methicillin has
no toxic effect",*

WP4 Medical Knowledge Repository

*basis for
clinical
decisions*

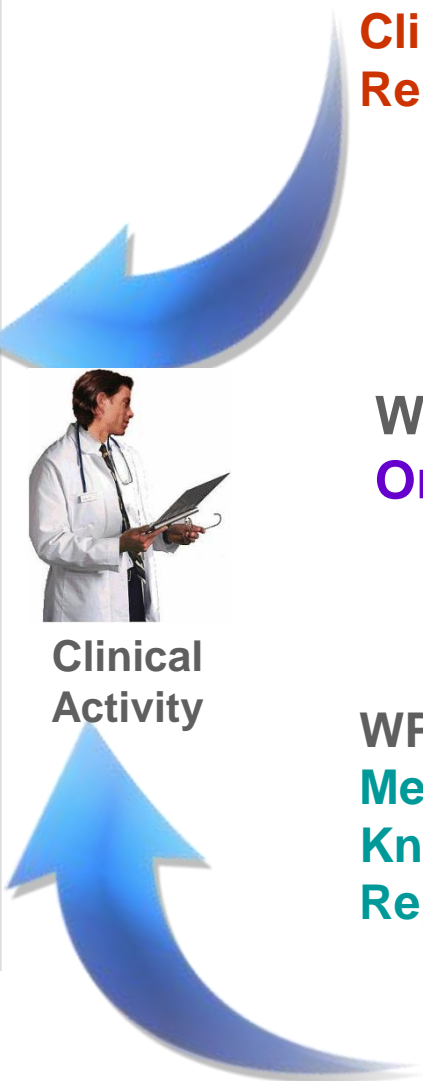


*"The risk of antibiotic
resistance after thoracic
surgery is estimated at x"*

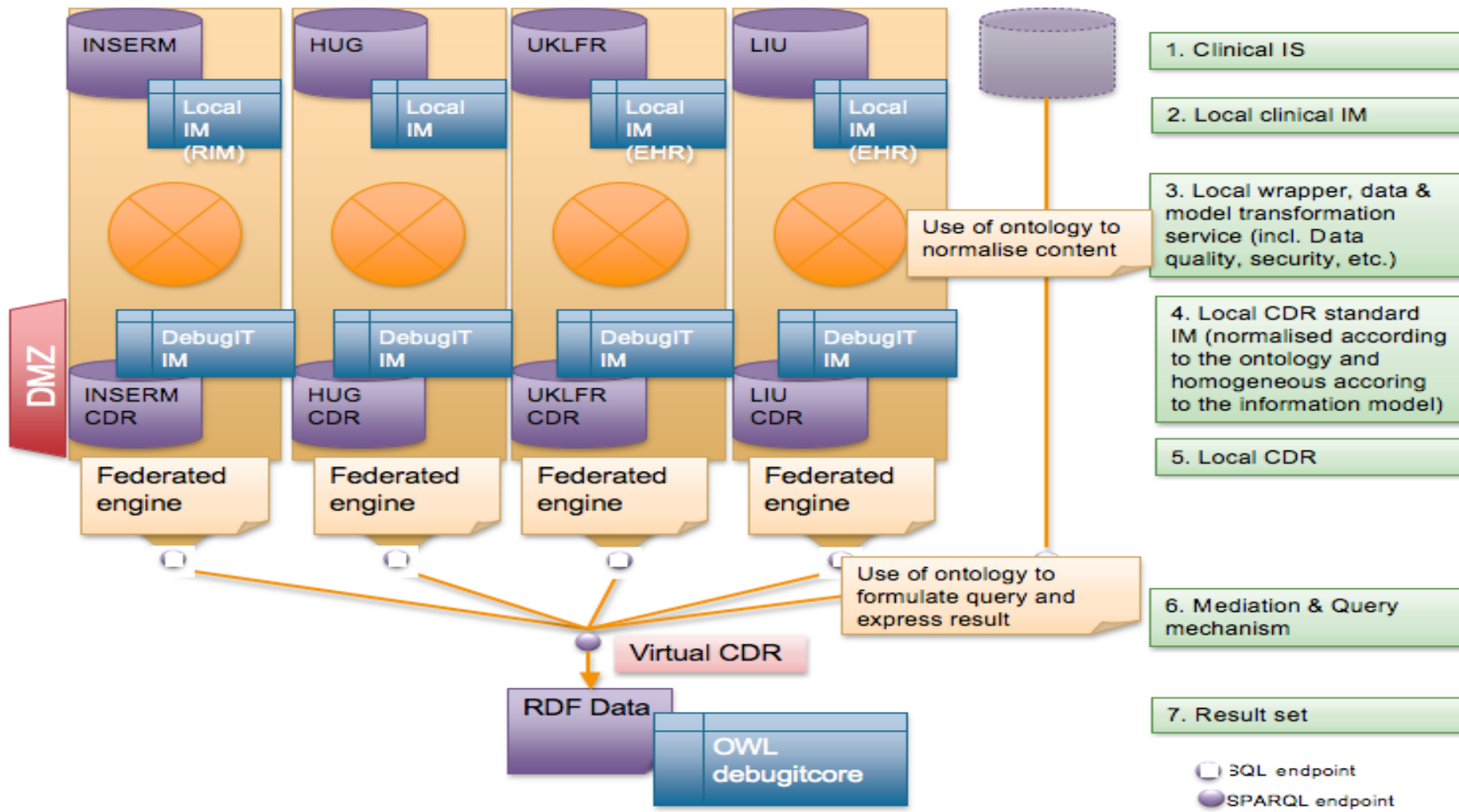
WP 3: Analysis



**Clinical
Activity**

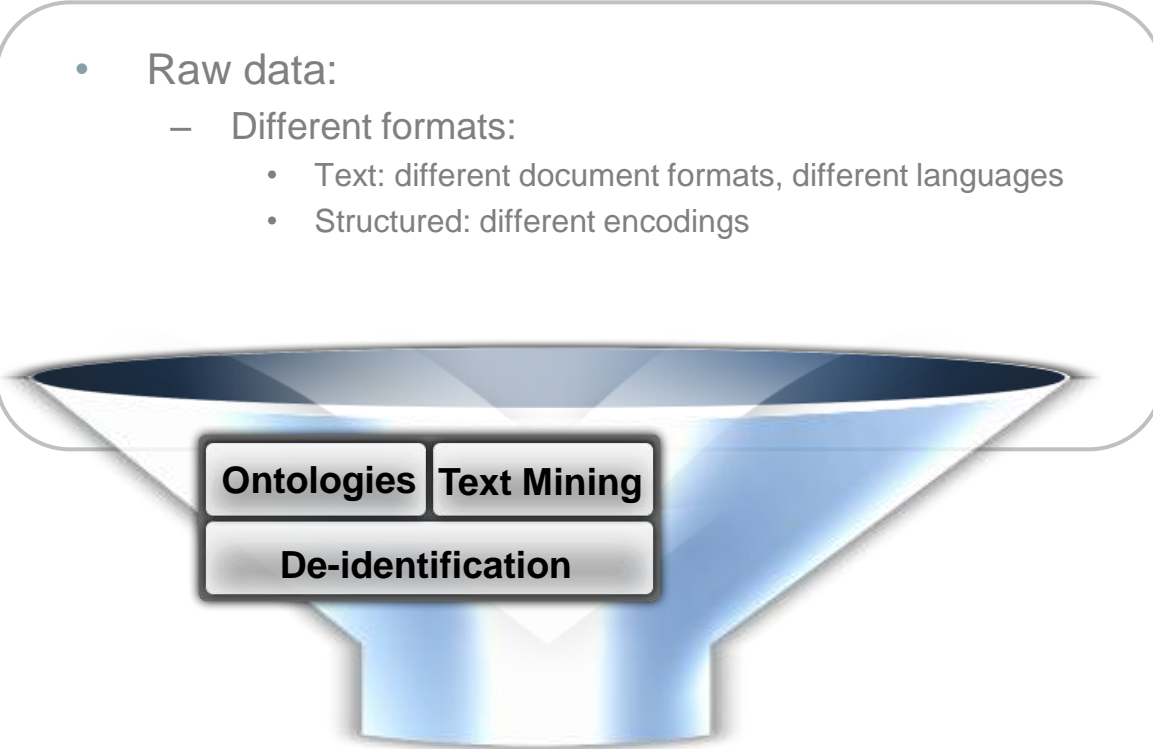


Integrating heterogeneous clinical data



Ontologies within the Interoperability platform

- Raw data:
 - Different formats:
 - Text: different document formats, different languages
 - Structured: different encodings

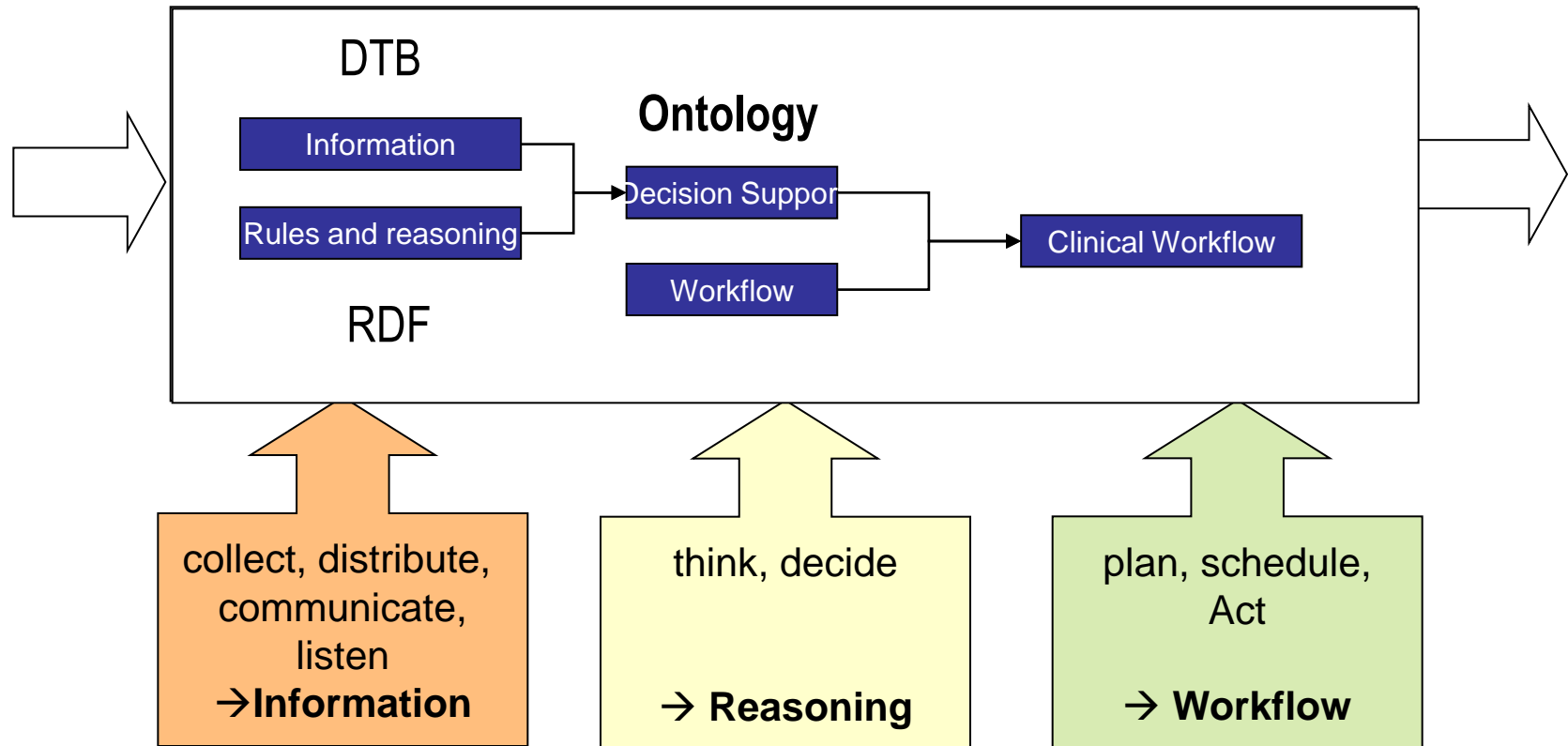


Ontologies **Text Mining**
De-identification

- Refined data
 - Texts mapped to Structure
 - uniform data format
 - uniform semantics
 - pseudonymized

Decision Support & Monitoring

- Decision support tools apply generated knowledge (antibiotics choice, dose)
- Monitoring tools analyze current & predict future outcomes



OO usage – in a **Clinical SPARQL Query**

“Percent of Cases with E.Coli resistant to Trimethoprim ?”

PREFIX **dco:** <<http://www.debugit.eu/ontology/1.0/dco.owl#>>

PREFIX **units:** <<http://sweet.jpl.nasa.gov/2.0/sciUnits.owl#>>

PREFIX **ao:** <<http://www.agfa.com/w3c/2009/AnalysisOntology#>>

PREFIX **cao:** <<http://www.agfa.com/w3c/2009/ClinicalAnalysisOntology#>>

PREFIX **quex:** <<http://eulersharp.sourceforge.net/2003/03swap/quantitiesExtension#>>

CONSTRUCT {

```
[   a quex:Percentage;
    ao:percentageOf [ a cao:EColi];
    ao:percentageThat [ #           a cao:EColi;
    cao:resistantTo [ a dco:Trimethoprim]];
    quex:hasValue ?percentage;
    quex:hasUnit units:percent].
```

}

WHERE {

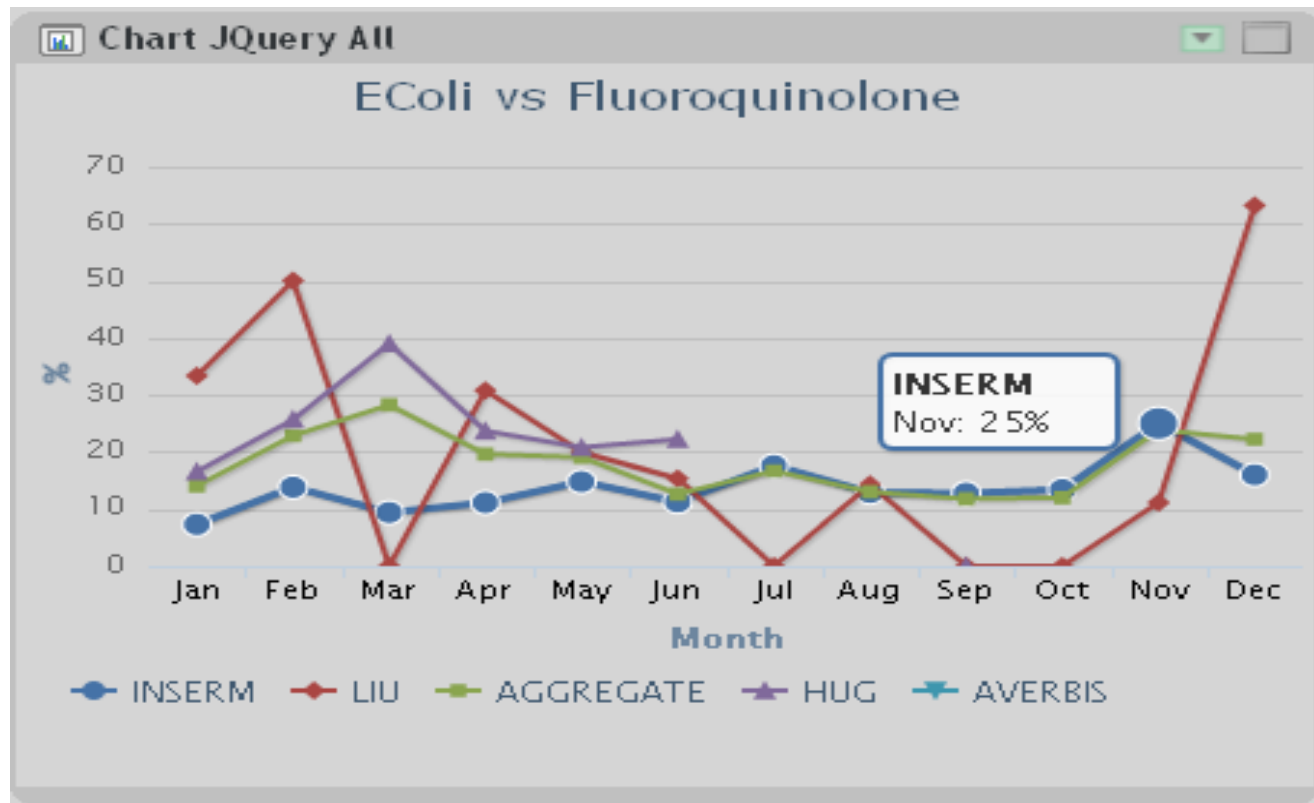
```
[ // same as in construct clause
```

}

DCO in decision support and monitoring

Rule-based knowledge generation

- IF DCO fact set (proposition p1 & p2)
THEN proposition p3 (conclusion)



Execution of a data query on a local CDR

- Execution of a dataset query on a local CDR consists of the following activities (Figure 9):
- The flow starts when a dataset query is received.
- The dataset query is executed.
- Local CDR Data policies are applied to the result set of

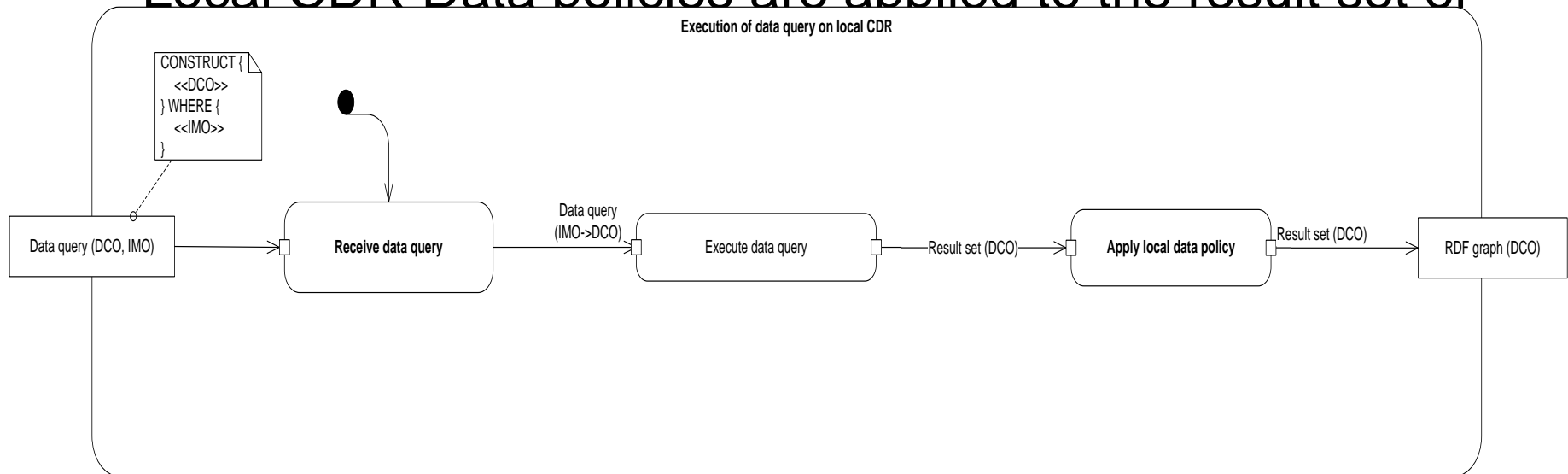


Figure 9 – Execution of a data query on a local CDR Activity Diagram

Using Bayesian Belief probabilistics in rules to bind analog (probability) to boolean values

T-Rule #1

```
{ (dco:UrinaryTractInfection) event:indicationBelief ?belief. ?belief math:equalTo 1.0 } => { ({dco:UrinaryTractInfection e:boolean e:T} {dco:Fever e:boolean e:T} ) e:conditional 0.88}
```

```
{ (dco:UrinaryTractInfection) event:indicationBelief ?belief. ?belief math:equalTo 1.0 } => { ({dco:UrinaryTractInfection e:boolean e:T} {dco:Fever e:boolean e:F} ) e:conditional 0.12}
```

T-Rule #2

Models are written in a form of **T-rules** which simplifies the modelling technique to expressing relationships with **True** or **False** structures and attaching **believes** to them. Furthermore the rules expressing medical knowledge make use of built-ins of the eye reasoner.

Decision Support and Monitoring example

- **Scenario** <requesting prediction of future trends of resistance>
- **Goal:** Dr Larou, a physician requests prediction on future trends of resistance
- **Precondition:** Information on the selected bacterium & antibiotic is available
- **MainFlow:** Dr. Larou selects a bacterium whose resistance should be evaluated, the target antibiotic, the decision support outcome (prediction on outcome of therapy), time frame as basis for projection
- He sets conditions (constraints) for decision support (e.g. patient's condition and / or current other therapies)
- He executes decision support
- **Postcondition:** An analytic result (prediction) is compiled and displayed

N3 query result set

```
_:sk0 quex:hasNumerator _:sk1.  
_:sk1 fos:hasMember _:sk2.  
_:sk2 a cao:EColi. _:sk2 cao:resistantTo _:sk3.  
_:sk3 a dco:Trimethoprim.  
_:sk2 cao:resistantTo _:sk4.  
_:sk4 a dco:Co-Trimoxazole.  
_:sk2 cao:resistantTo (_:sk5 _:sk6).  
_:sk5 a dco:Trimethoprim. _:sk6 a dco:Sulfamethoxazole.  
_:sk1 fos:counts 6.  
_:sk0 quex:hasDenominator _:sk7.  
_:sk7 fos:hasMember _:sk8. _:sk8 a cao:EColi.  
_:sk7 fos:counts 21.  
_:sk0 quex:hasPercentage _:sk9.  
_:sk9 quex:hasValue 28.6. _:sk9 quex:hasUnit _:sk10.  
_:sk10 units:hasSymbol "%".
```

≡ E.Coli resistant to TMP, Cotri, TMP/SMX = 6; E.Coli total = 21 → 28,6%

Ontological issues

- Time
 - TimeQuality classes vs. xsd dateTime datatype properties
 - Intervals: Episode of care, Patient Stay
- ‘Adapted therapy’ (process modifications)
- Intervals
 - Episode of care, Patient Stay
- Complexity
 - Quality & quality regions (not applicable to time)
- Infectious disease → state vs. condition/disposition
- Compound drugs & prescribed dosages
- Functions vs. Process
- Roles

Ontology publication

- Latest DCO available in svn
 - <http://www.greeninghealthcare.org/repository/debugit/trunk/src/ontology/dco.owl>
- TermBrowser (owlDoc generated HTML serialisation)
 - http://www.imbi.uni-freiburg.de/~schober/dco_owlDoc/
- TaxBrowser
 - <http://infoserv1.ukl.uni-freiburg.de/DebugITBrowser/>

... put in ols, OBO lib, TONES
(<http://owl.cs.manchester.ac.uk/repository/browser>)... ?

Increasing human readability of DCO class definitions

Omit logics-specific symbolism:

From

HeparinBiosynthesis \sqsubseteq
(HeparinMetabolism \sqcap (Biosynthesis \sqcap
 \exists *acts_on*. Heparin))

to Manchester OWL Syntax

HeparinBiosynthesis
SubClassOf HeparinMetabolism
SubClassOf (Biosynthesis AND *acts_on* SOME Heparin)

or Attempto Controlled English (CNL)

“Every HeparinBiosynthesis *is a* HeparinMetabolism.
Every HeparinBiosynthesis *is a* Biosynthesis *that acts_on a* Heparin.”

Definition :

The interpretation of the Antibiogram. The interpretation of the result of the AntimicrobialSusceptibilityTest

Formal Definition: (Conceptual Lego™ in Protege 4.x)

Superclasses 

- ☰ **'diagnostic procedure'**
- **'has participant' some antibiogram**
- **'preceded by' some 'antimicrobial susceptibility test'**
- **'processual part of' some 'episode of care'**
- **'has outcome' some (result
and encodes some 'microbiological susceptibility')**

Inferred anonymous superclasses

- **'has locus' some ('bearer of some 'service delivery location role')**
- **hasResultDateTime exactly 1 dateTime**
- **'bearer of' some 'test result'**
- **'health care activity'
and 'bearer of' some 'diagnostic role'**
- **'processual entity'
and 'has agent' some ('immaterial nonphysical entity'
or 'immaterial physical entity'
or 'material entity')**