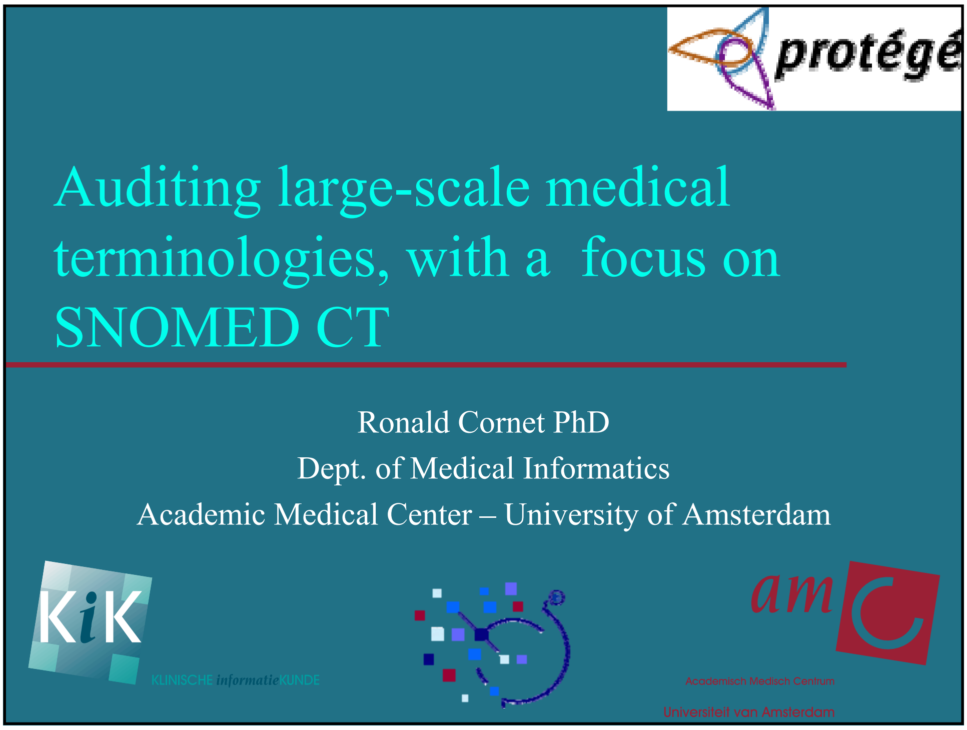
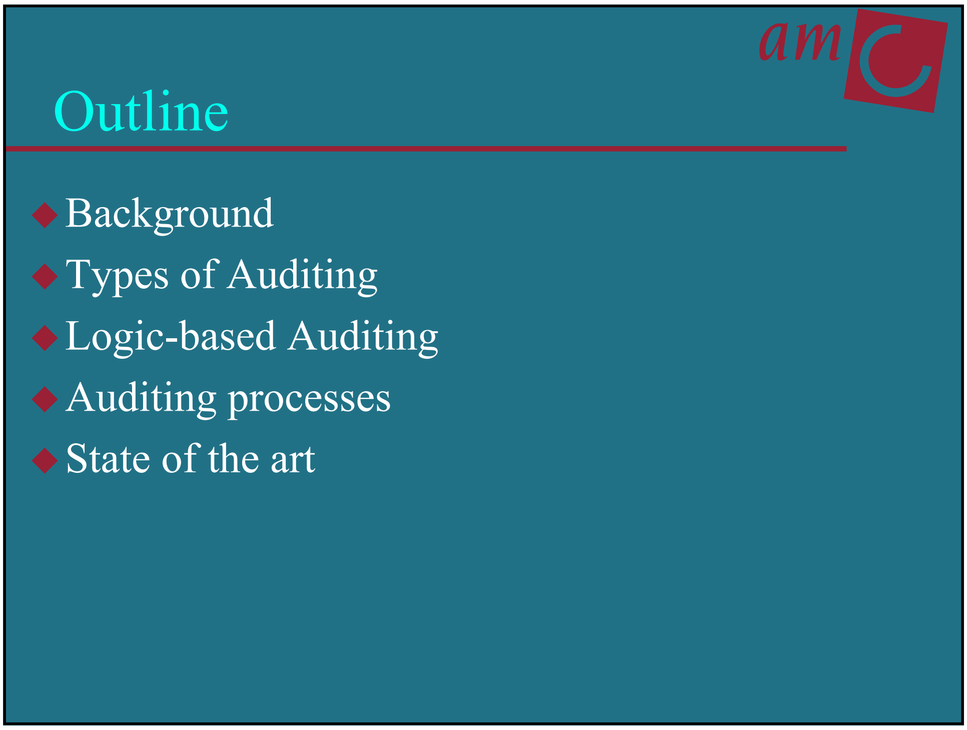


Auditing large-scale medical terminologies, with a focus on SNOMED CT

Ronald Cornet PhD
Dept. of Medical Informatics
Academic Medical Center – University of Amsterdam

KLINISCHE informatieKUNDE

Academisch Medisch Centrum
Universiteit van Amsterdam



Outline

- Background
- Types of Auditing
- Logic-based Auditing
- Auditing processes
- State of the art

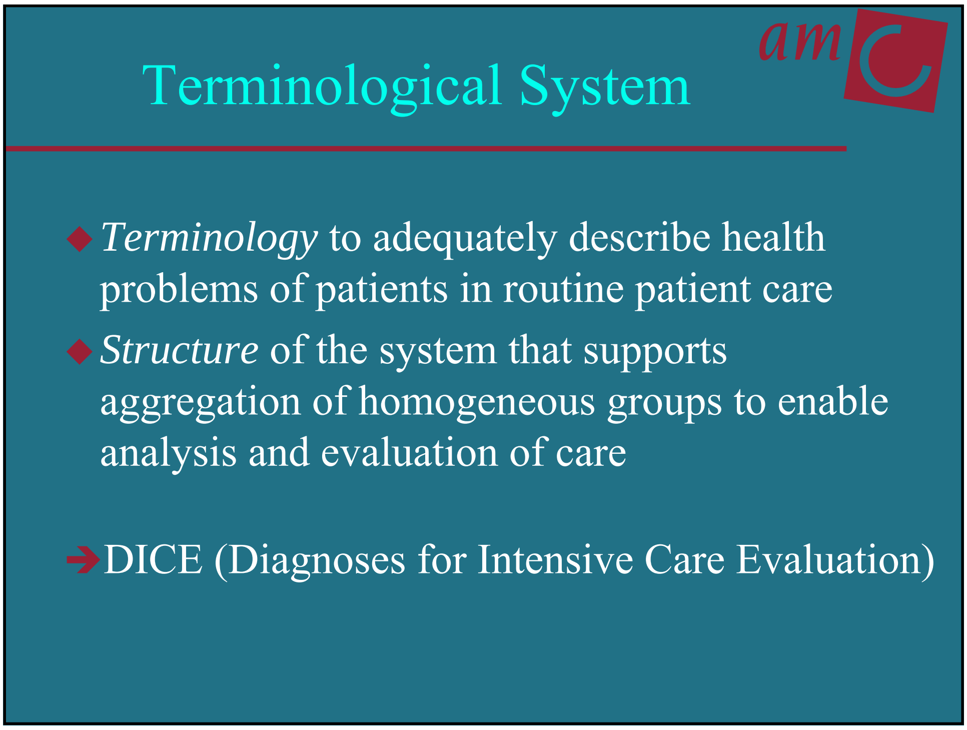
Outline

- **Background**
- Types of Auditing
- Logic-based Auditing
- Auditing processes
- State of the art

Practical Use: Intensive Care

- Young discipline, large development, expensive

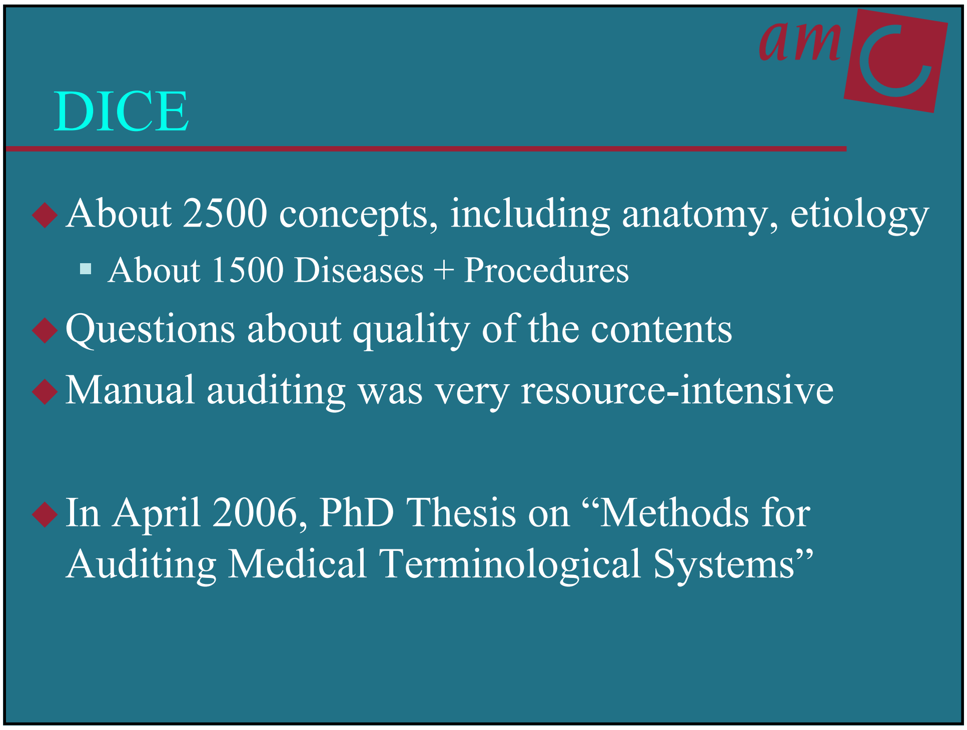
- Need for:
  - High quality registration of patient information
  - Quality assessment and improvement
  - Epidemiology of (rare) diseases on ICU



Terminological System

- *Terminology* to adequately describe health problems of patients in routine patient care
- *Structure* of the system that supports aggregation of homogeneous groups to enable analysis and evaluation of care

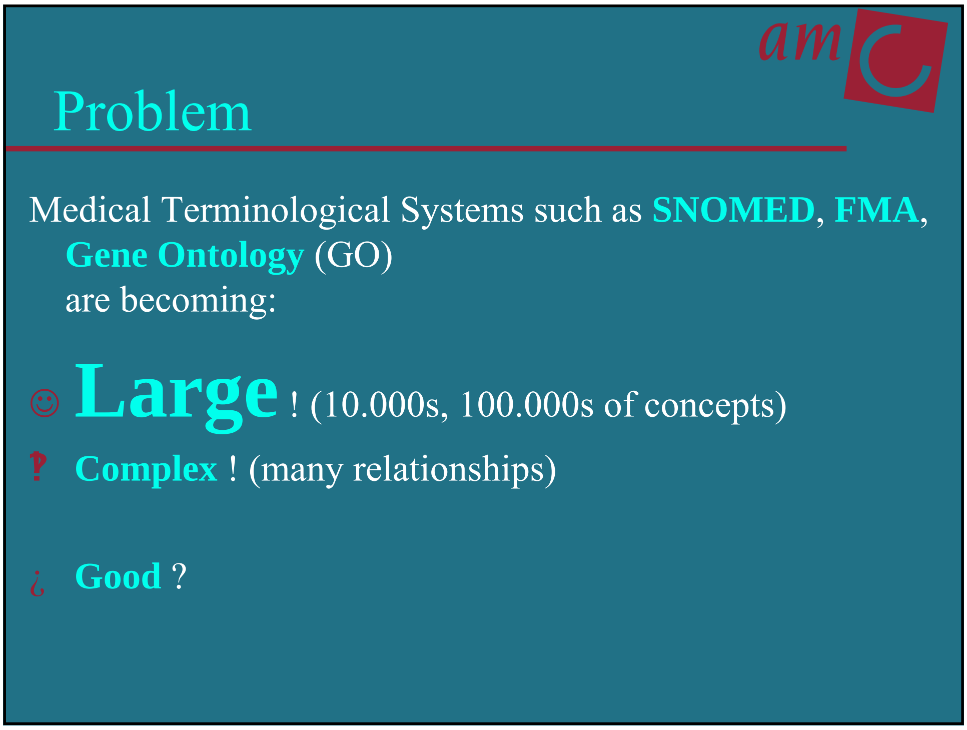
→DICE (Diagnoses for Intensive Care Evaluation)



DICE

- About 2500 concepts, including anatomy, etiology
  - About 1500 Diseases + Procedures
- Questions about quality of the contents
- Manual auditing was very resource-intensive

- In April 2006, PhD Thesis on “Methods for Auditing Medical Terminological Systems”



Problem

Medical Terminological Systems such as **SNOMED**, **FMA**, **Gene Ontology** (GO) are becoming:

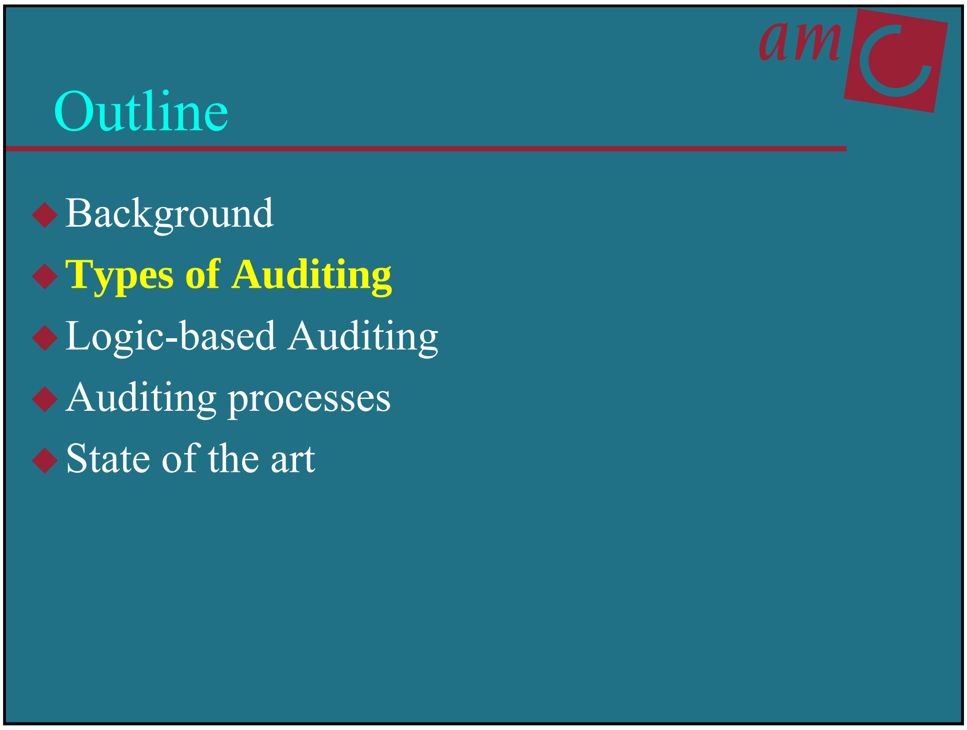
☺ **Large** ! (10.000s, 100.000s of concepts)

❢ **Complex** ! (many relationships)

¿ **Good** ?

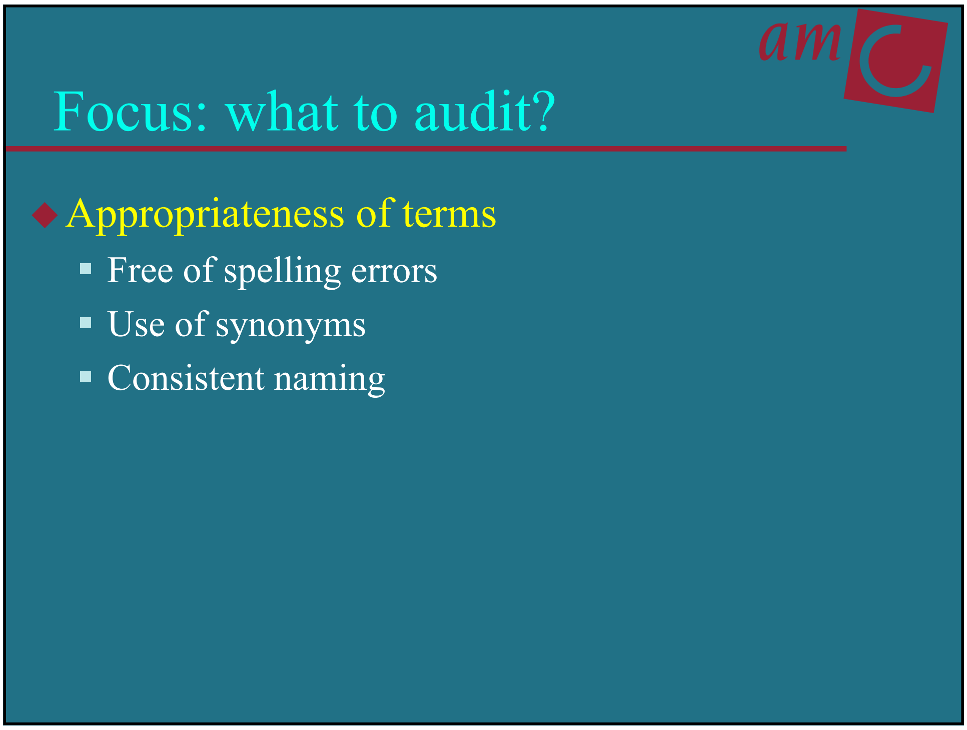
Current activities

- Member of IHTSDO Quality Assurance Committee
  - QA of SNOMED prevails over expansion
  - Development of a QA framework



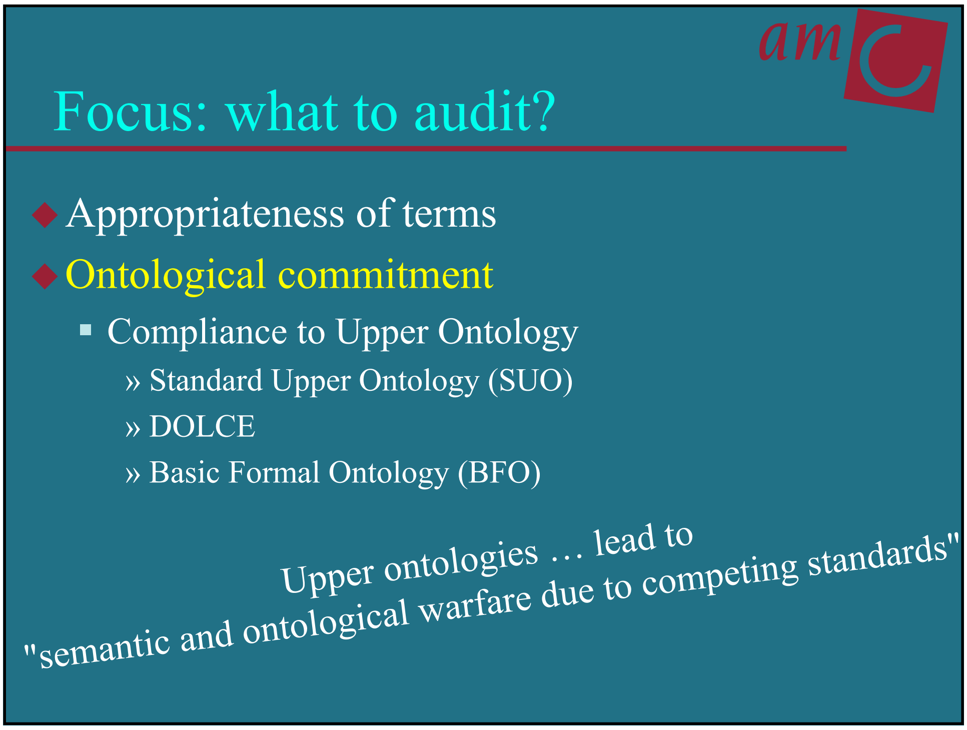
Outline

- Background
- **Types of Auditing**
- Logic-based Auditing
- Auditing processes
- State of the art



Focus: what to audit?

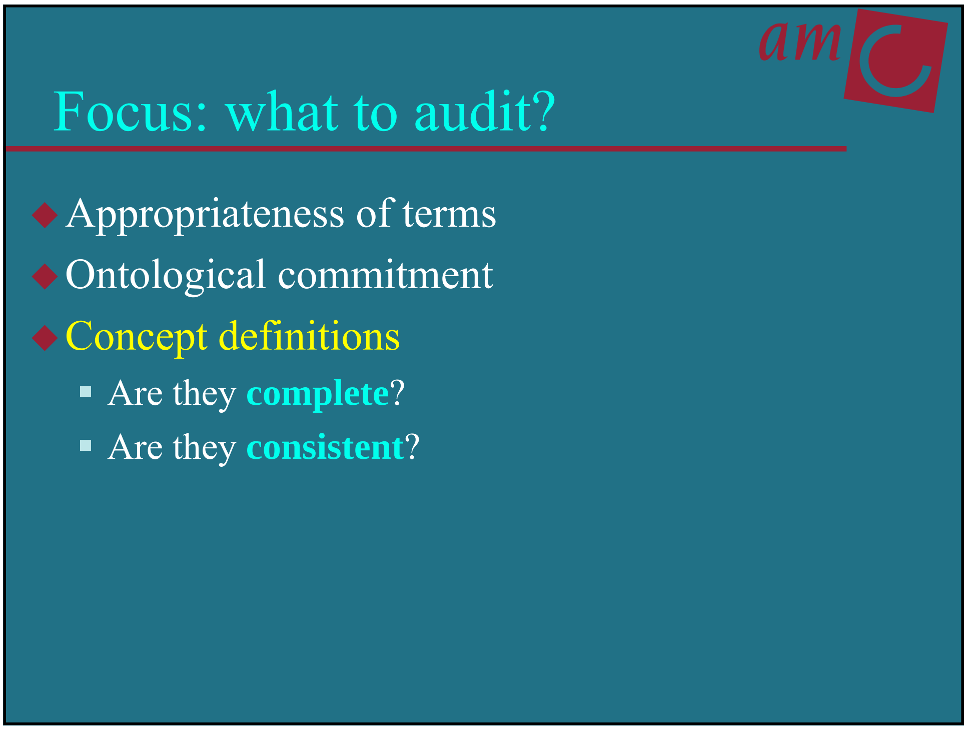
- Appropriateness of terms
  - Free of spelling errors
  - Use of synonyms
  - Consistent naming



Focus: what to audit?

- Appropriateness of terms
- Ontological commitment
  - Compliance to Upper Ontology
    - » Standard Upper Ontology (SUO)
    - » DOLCE
    - » Basic Formal Ontology (BFO)

Upper ontologies … lead to "semantic and ontological warfare due to competing standards"

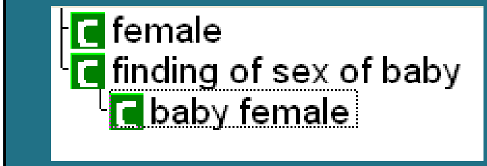


Focus: what to audit?

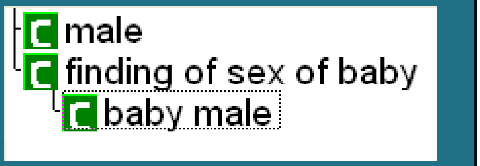
- Appropriateness of terms
- Ontological commitment
- Concept definitions
  - Are they **complete**?
  - Are they **consistent**?

Mutual consistency in SNOMED

- Version July 2007:

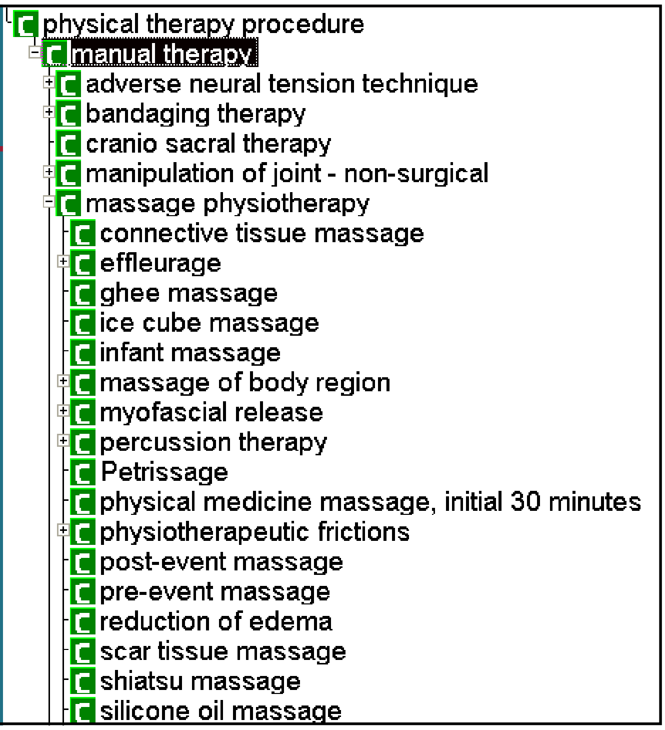


female
finding of sex of baby
baby female



male
finding of sex of baby
baby male

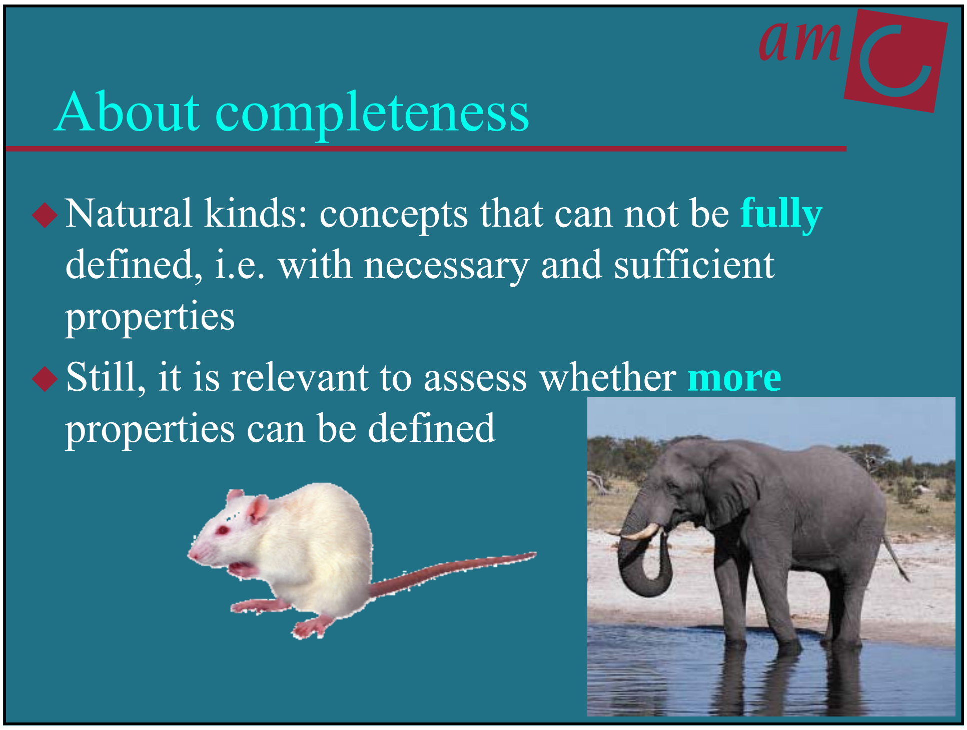
Version January 2008!



- physical therapy procedure
  - manual therapy
    - adverse neural tension technique
    - bandaging therapy
    - cranio sacral therapy
    - manipulation of joint - non-surgical
    - massage physiotherapy
      - connective tissue massage
      - effleurage
      - ghee massage
      - ice cube massage
      - infant massage
      - massage of body region
      - myofascial release
      - percussion therapy
      - Petrissage
      - physical medicine massage, initial 30 minutes
      - physiotherapeutic frictions
      - post-event massage
      - pre-event massage
      - reduction of edema
      - scar tissue massage
      - shiatsu massage
      - silicone oil massage

Outline

- Background
- Types of Auditing
- **Logic-based Auditing**
- Auditing processes
- State of the art

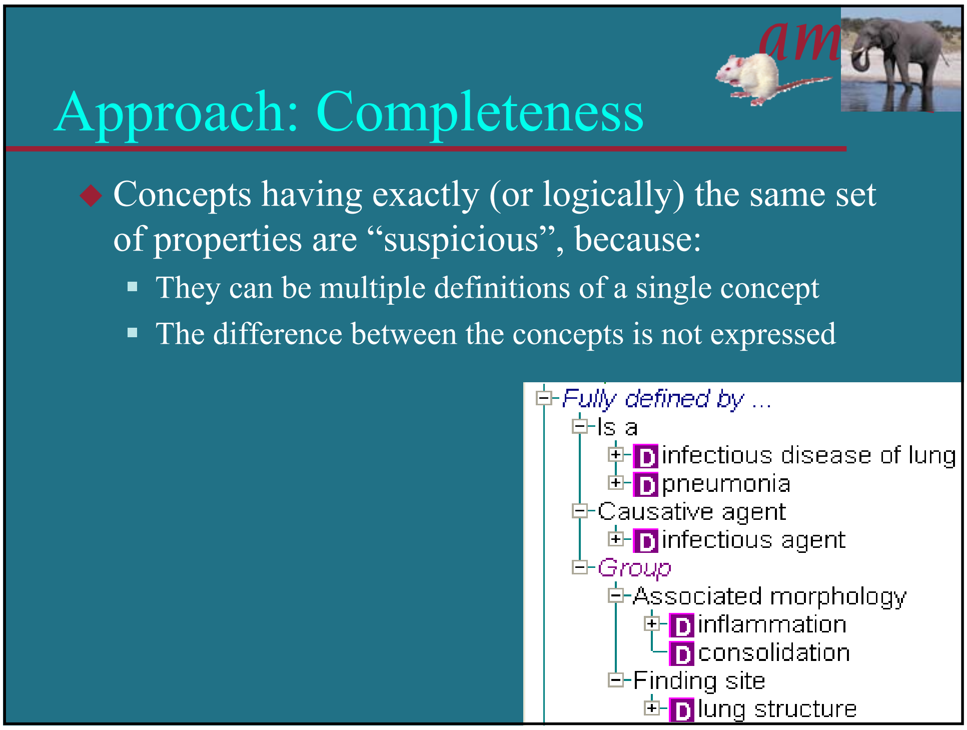


About completeness

- Natural kinds: concepts that can not be **fully** defined, i.e. with necessary and sufficient properties
- Still, it is relevant to assess whether **more** properties can be defined

About consistency

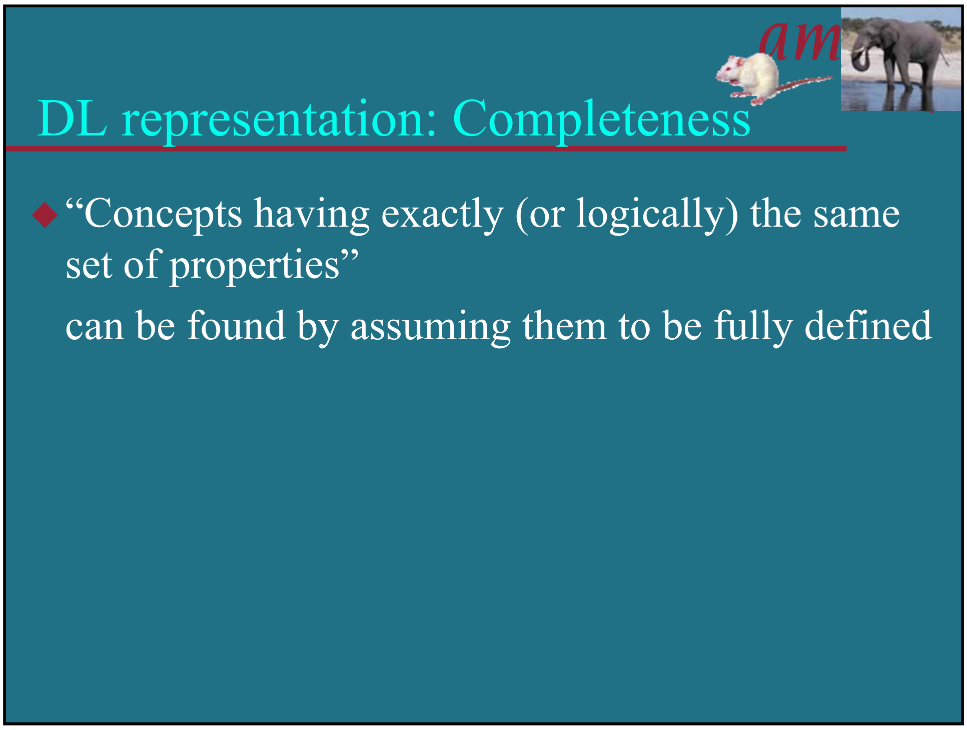
- Properties of a concept should be consistent with the properties of super-ordinate concepts
- Consistency depends on semantics



Approach: Completeness

- Concepts having exactly (or logically) the same set of properties are “suspicious”, because:
  - They can be multiple definitions of a single concept
  - The difference between the concepts is not expressed

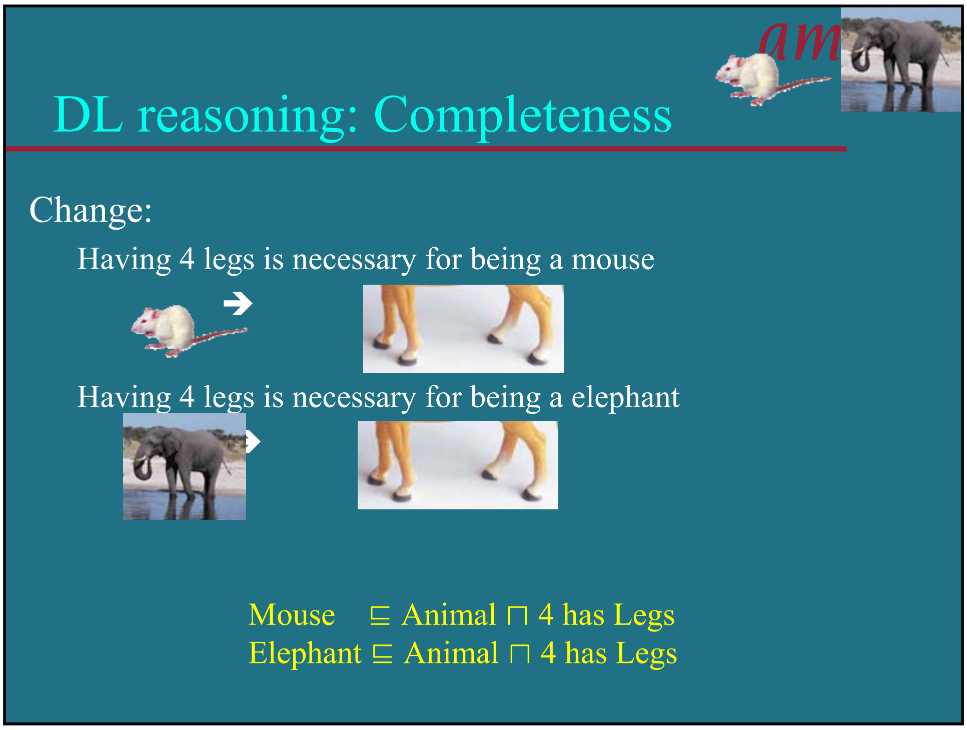
```
⊟-Fully defined by ...
   ⊟-Is a
      ⊞-D infectious disease of lung
      ⊞-D pneumonia
   ⊟-Causative agent
      ⊞-D infectious agent
   ⊟-Group
      ⊟-Associated morphology
         ⊞-D inflammation
          └D consolidation
      ⊟-Finding site
         ⊞-D lung structure
```



DL representation: Completeness

- “Concepts having exactly (or logically) the same set of properties”

  can be found by assuming them to be fully defined



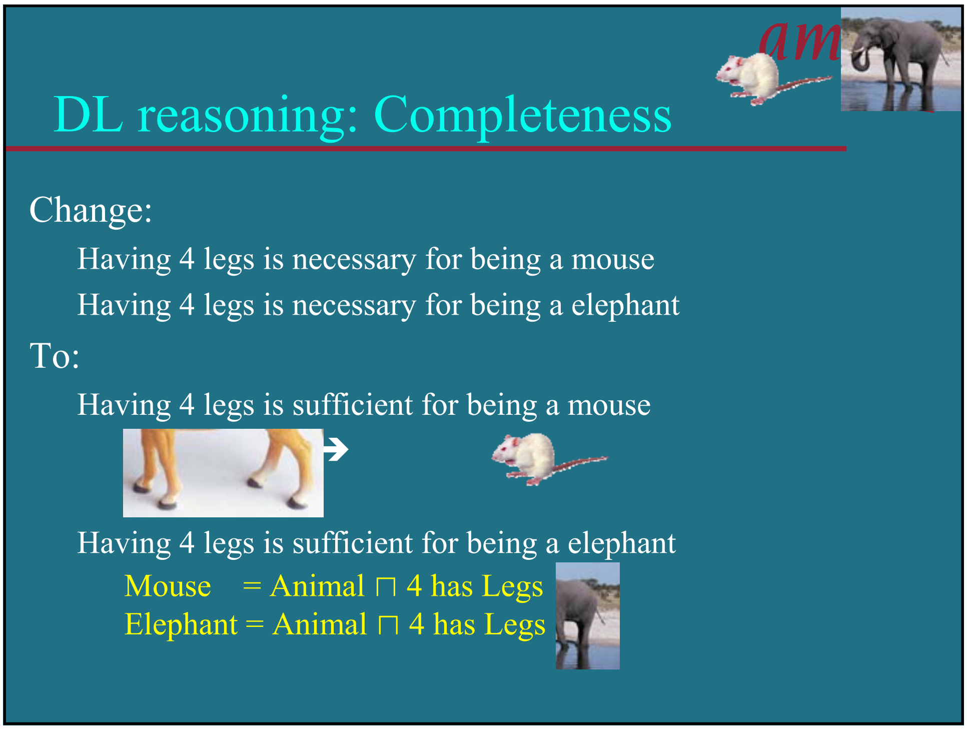
DL reasoning: Completeness

Change:

Having 4 legs is necessary for being a mouse

Having 4 legs is necessary for being a elephant

Mouse $\sqsubseteq$ Animal $\sqcap$ 4 has Legs
Elephant $\sqsubseteq$ Animal $\sqcap$ 4 has Legs



DL reasoning: Completeness

Change:

Having 4 legs is necessary for being a mouse

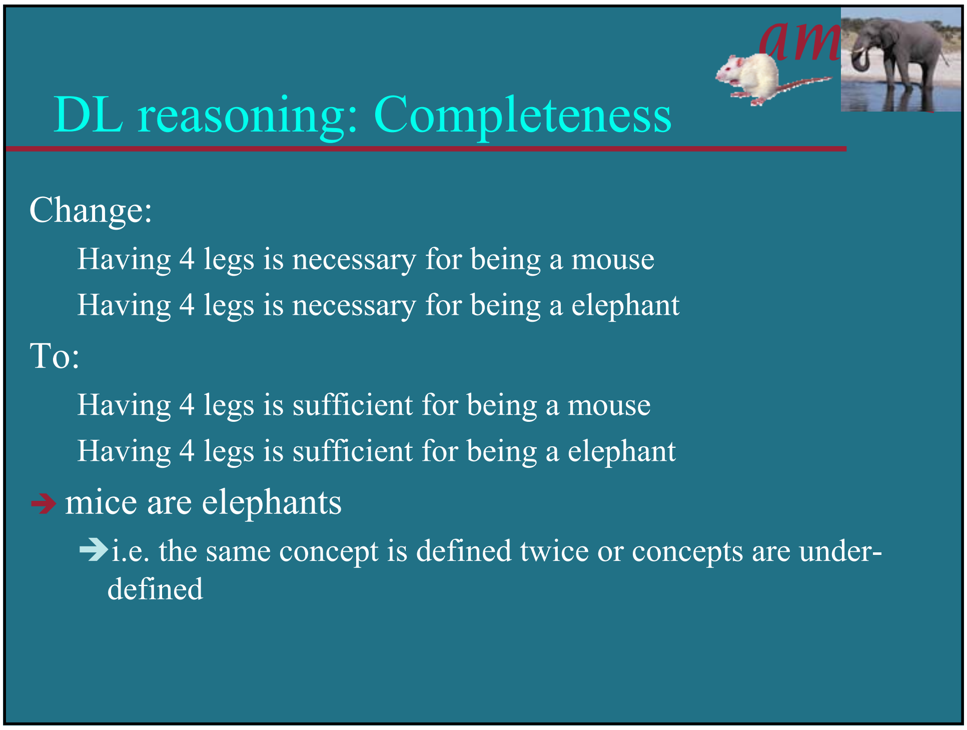
Having 4 legs is necessary for being a elephant

To:

Having 4 legs is sufficient for being a mouse

Having 4 legs is sufficient for being a elephant

Mouse = Animal ⊓ 4 has Legs
Elephant = Animal ⊓ 4 has Legs



DL reasoning: Completeness

Change:

Having 4 legs is necessary for being a mouse

Having 4 legs is necessary for being a elephant

To:

Having 4 legs is sufficient for being a mouse

Having 4 legs is sufficient for being a elephant

➔ mice are elephants

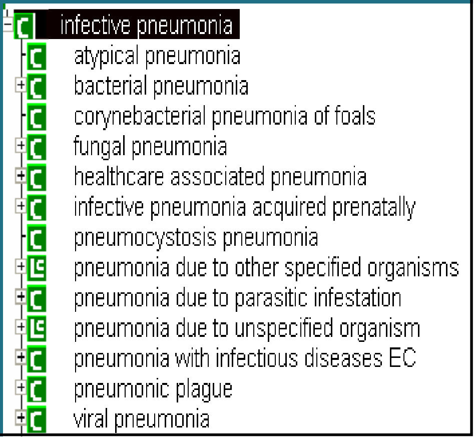
➔i.e. the same concept is defined twice or concepts are under-defined

Approach: Consistency

- “Properties of a concept should be consistent with the properties of super-ordinate concepts”
- Maximize the possibilities for finding potential inconsistencies by “closing the world”

DL representation: Consistency

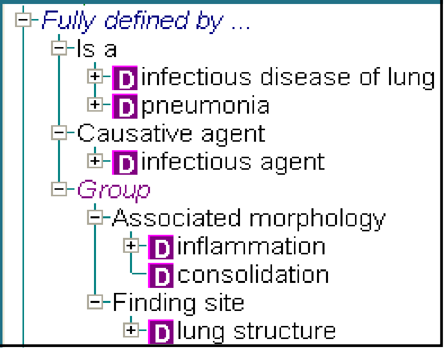
- Assume maximal restriction (closure axioms)
  - Siblings are disjoint



infective pneumonia
- atypical pneumonia
- bacterial pneumonia
- corynebacterial pneumonia of foals
- fungal pneumonia
- healthcare associated pneumonia
- infective pneumonia acquired prenatally
- pneumocystosis pneumonia
- pneumonia due to other specified organisms
- pneumonia due to parasitic infestation
- pneumonia due to unspecified organism
- pneumonia with infectious diseases EC
- pneumonic plague
- viral pneumonia

DL representation: Consistency

- Be maximally restrictive (closure axioms)
  - Siblings are disjoint
  - No other values than those mentioned are allowed



- *Fully defined by ...*
  - Is a
    - D infectious disease of lung
    - D pneumonia
  - Causative agent
    - D infectious agent
  - *Group*
    - Associated morphology
      - D inflammation
      - D consolidation
    - Finding site
      - D lung structure

DL reasoning: Consistency

- Viral pneumonia

  Is a: Infective pneumonia

  Causative agent: Virus , nothing else
- Staphylococcal pneumonia

  Is a: Viral pneumonia

  Causative agent: Staphylococcus
- Staphylococcus

  Bacterium

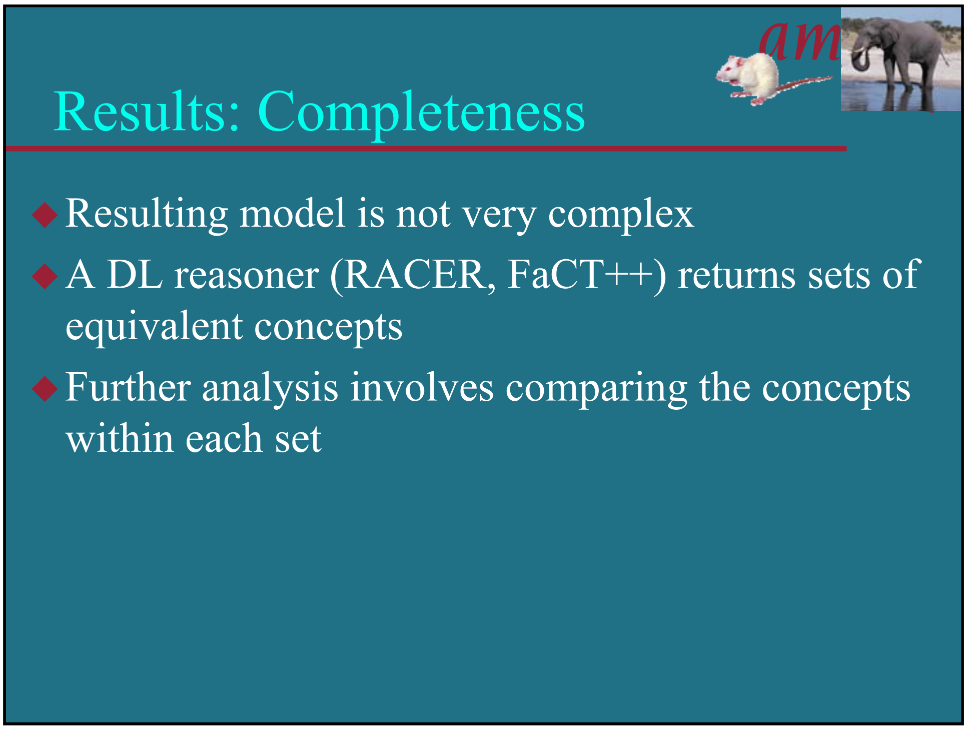
  - Virus ≠ Bacterium

    Vir_P ⊑ Inf_P ⊓ ∃ cause Virus *⊓ ∀ cause Virus*

    Staph_P ⊑ Vir_P ⊓ ∃ cause Staph

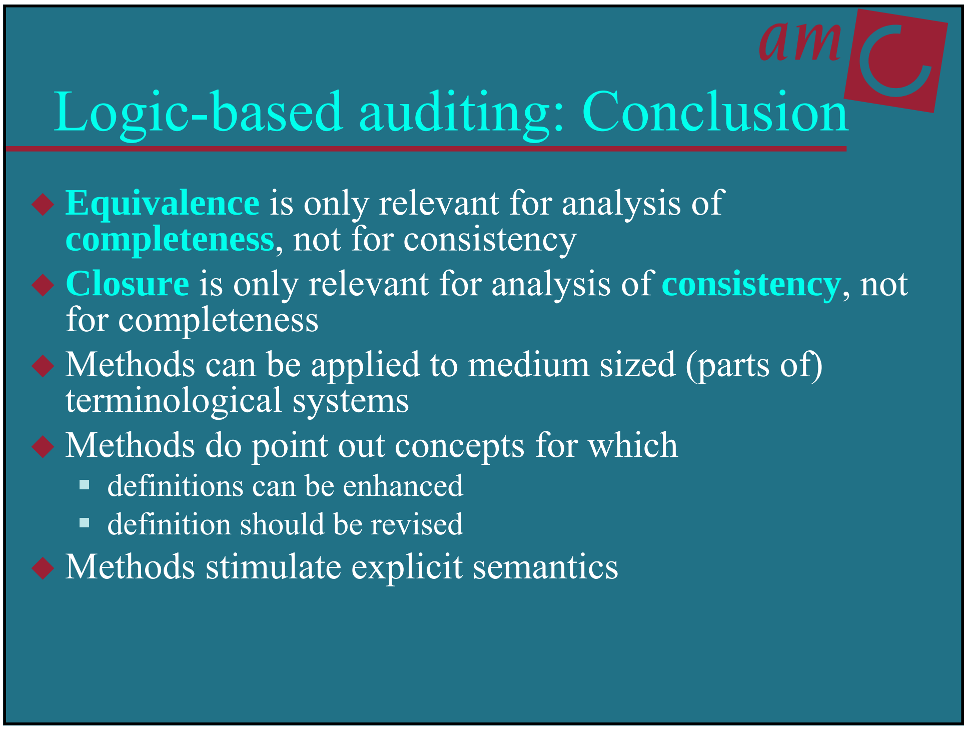
    Staph ⊑ Bact

    Disjoint (Virus, Bact)



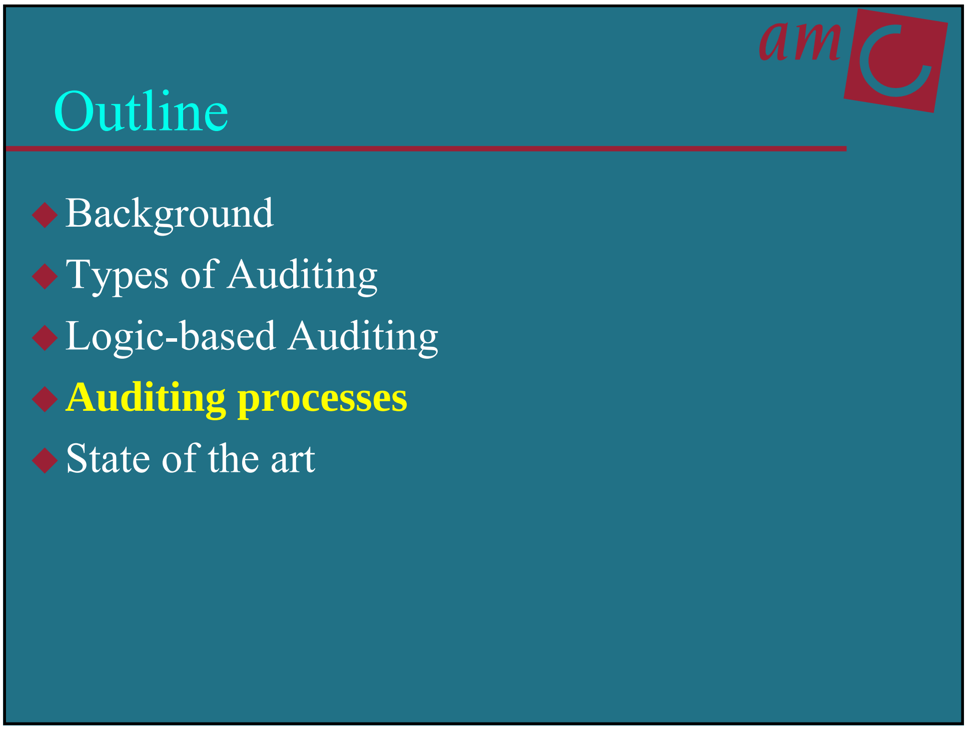
Results: Completeness

- Resulting model is not very complex
- A DL reasoner (RACER, FaCT++) returns sets of equivalent concepts
- Further analysis involves comparing the concepts within each set



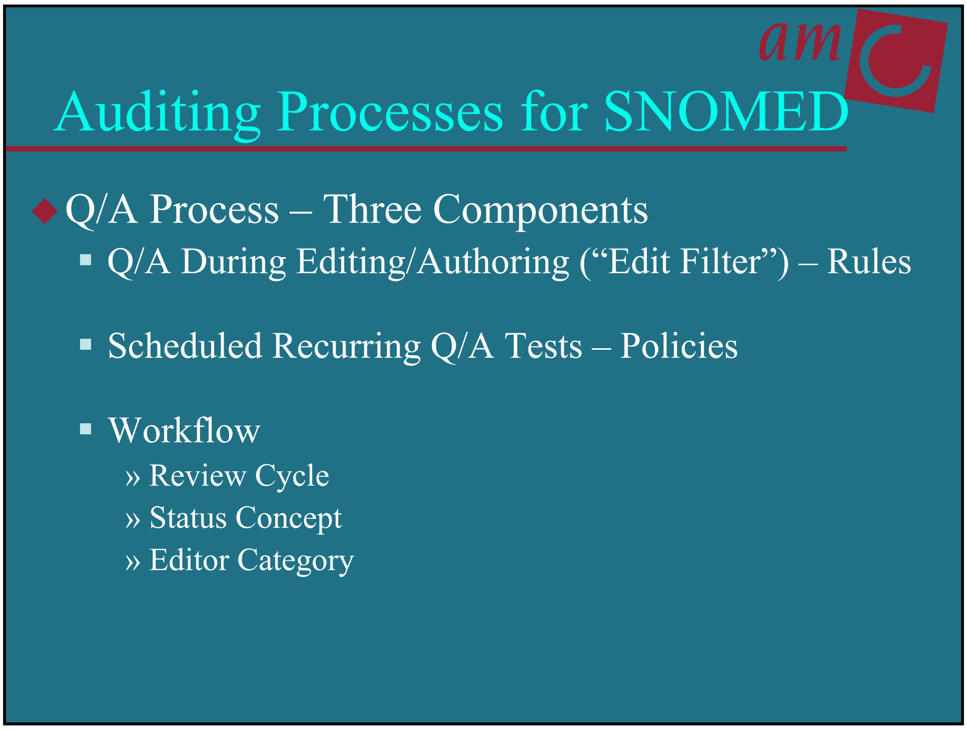
Logic-based auditing: Conclusion

- **Equivalence** is only relevant for analysis of **completeness**, not for consistency
- **Closure** is only relevant for analysis of **consistency**, not for completeness
- Methods can be applied to medium sized (parts of) terminological systems
- Methods do point out concepts for which
  - definitions can be enhanced
  - definition should be revised
- Methods stimulate explicit semantics



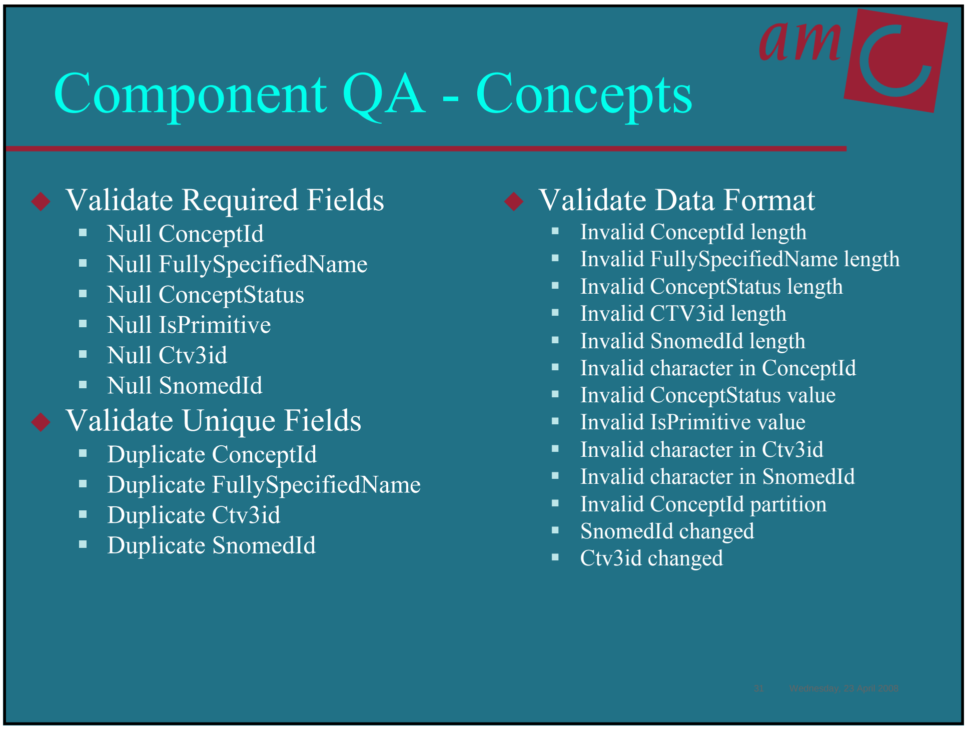
Outline

- Background
- Types of Auditing
- Logic-based Auditing
- **Auditing processes**
- State of the art



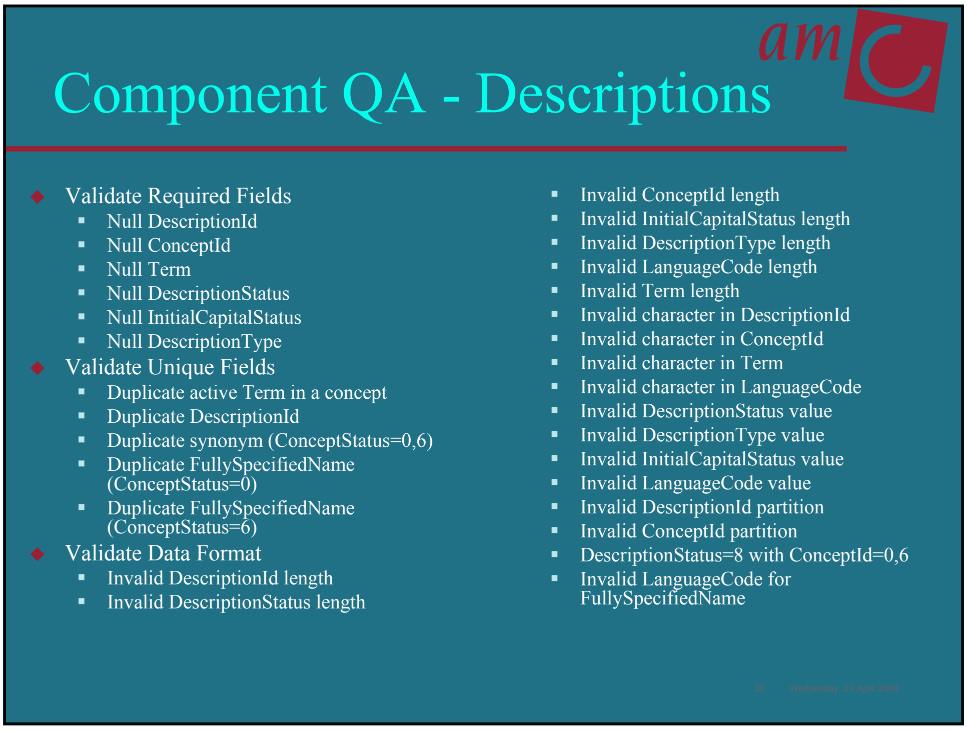
Auditing Processes for SNOMED

- Q/A Process – Three Components
  - Q/A During Editing/Authoring (“Edit Filter”) – Rules
  - Scheduled Recurring Q/A Tests – Policies
  - Workflow
    - » Review Cycle
    - » Status Concept
    - » Editor Category



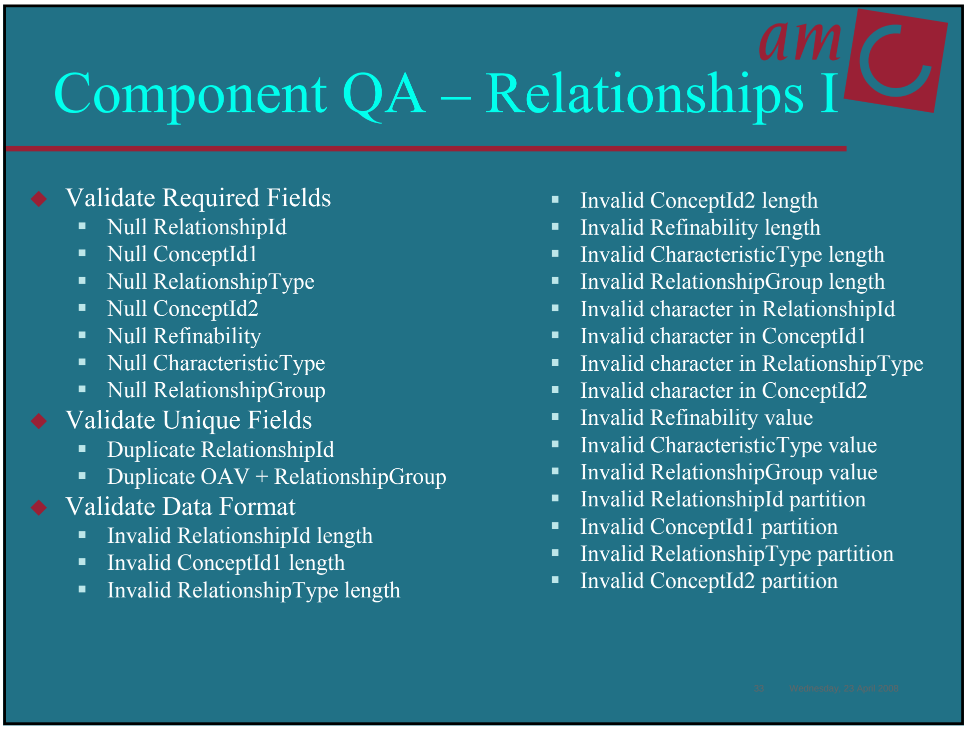
Component QA - Concepts

- Validate Required Fields
  - Null ConceptId
  - Null FullySpecifiedName
  - Null ConceptStatus
  - Null IsPrimitive
  - Null Ctv3id
  - Null SnomedId
- Validate Unique Fields
  - Duplicate ConceptId
  - Duplicate FullySpecifiedName
  - Duplicate Ctv3id
  - Duplicate SnomedId
- Validate Data Format
  - Invalid ConceptId length
  - Invalid FullySpecifiedName length
  - Invalid ConceptStatus length
  - Invalid CTV3id length
  - Invalid SnomedId length
  - Invalid character in ConceptId
  - Invalid ConceptStatus value
  - Invalid IsPrimitive value
  - Invalid character in Ctv3id
  - Invalid character in SnomedId
  - Invalid ConceptId partition
  - SnomedId changed
  - Ctv3id changed



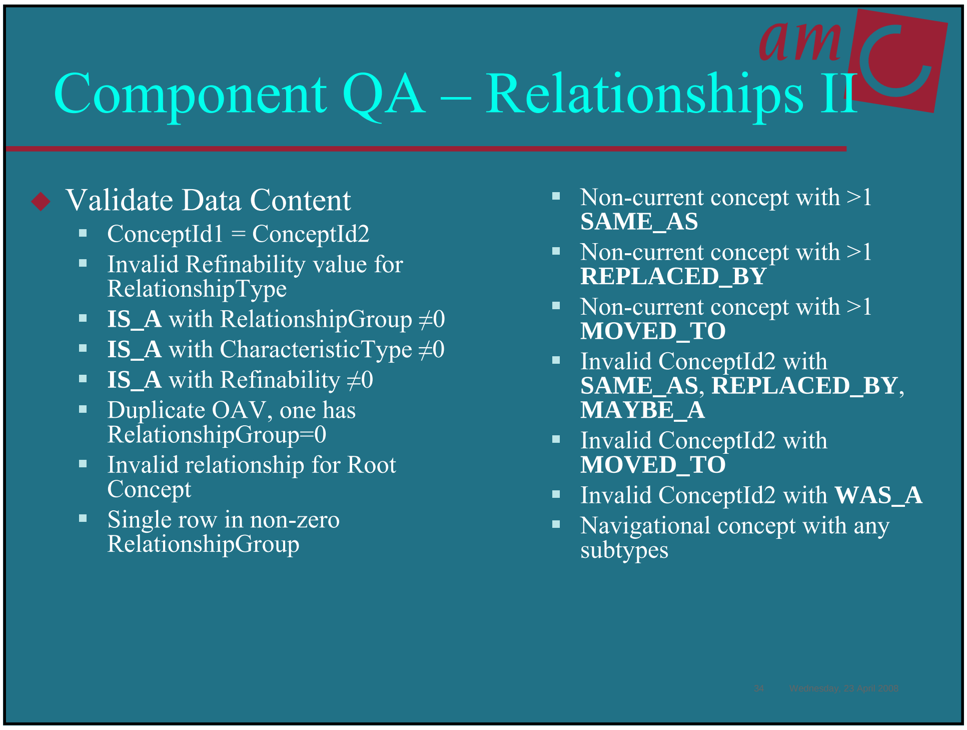
Component QA - Descriptions

- Validate Required Fields
  - Null DescriptionId
  - Null ConceptId
  - Null Term
  - Null DescriptionStatus
  - Null InitialCapitalStatus
  - Null DescriptionType
- Validate Unique Fields
  - Duplicate active Term in a concept
  - Duplicate DescriptionId
  - Duplicate synonym (ConceptStatus=0,6)
  - Duplicate FullySpecifiedName (ConceptStatus=0)
  - Duplicate FullySpecifiedName (ConceptStatus=6)
- Validate Data Format
  - Invalid DescriptionId length
  - Invalid DescriptionStatus length
  - Invalid ConceptId length
  - Invalid InitialCapitalStatus length
  - Invalid DescriptionType length
  - Invalid LanguageCode length
  - Invalid Term length
  - Invalid character in DescriptionId
  - Invalid character in ConceptId
  - Invalid character in Term
  - Invalid character in LanguageCode
  - Invalid DescriptionStatus value
  - Invalid DescriptionType value
  - Invalid InitialCapitalStatus value
  - Invalid LanguageCode value
  - Invalid DescriptionId partition
  - Invalid ConceptId partition
  - DescriptionStatus=8 with ConceptId=0,6
  - Invalid LanguageCode for FullySpecifiedName



Component QA – Relationships I

- Validate Required Fields
  - Null RelationshipId
  - Null ConceptId1
  - Null RelationshipType
  - Null ConceptId2
  - Null Refinability
  - Null CharacteristicType
  - Null RelationshipGroup
- Validate Unique Fields
  - Duplicate RelationshipId
  - Duplicate OAV + RelationshipGroup
- Validate Data Format
  - Invalid RelationshipId length
  - Invalid ConceptId1 length
  - Invalid RelationshipType length
  - Invalid ConceptId2 length
  - Invalid Refinability length
  - Invalid CharacteristicType length
  - Invalid RelationshipGroup length
  - Invalid character in RelationshipId
  - Invalid character in ConceptId1
  - Invalid character in RelationshipType
  - Invalid character in ConceptId2
  - Invalid Refinability value
  - Invalid CharacteristicType value
  - Invalid RelationshipGroup value
  - Invalid RelationshipId partition
  - Invalid ConceptId1 partition
  - Invalid RelationshipType partition
  - Invalid ConceptId2 partition



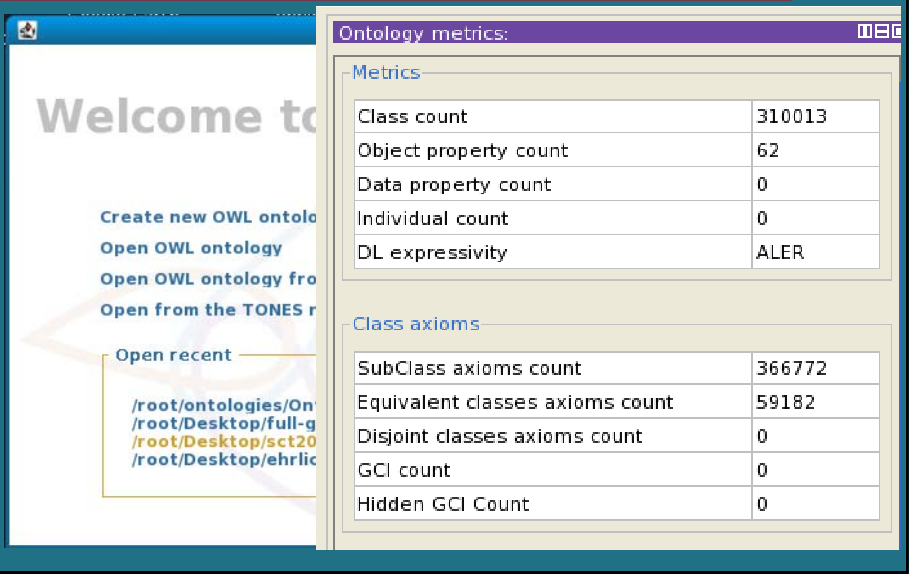
Component QA – Relationships II

- Validate Data Content
  - ConceptId1 = ConceptId2
  - Invalid Refinability value for RelationshipType
  - **IS_A** with RelationshipGroup $\neq 0$
  - **IS_A** with CharacteristicType $\neq 0$
  - **IS_A** with Refinability $\neq 0$
  - Duplicate OAV, one has RelationshipGroup=0
  - Invalid relationship for Root Concept
  - Single row in non-zero RelationshipGroup
  - Non-current concept with >1 **SAME_AS**
  - Non-current concept with >1 **REPLACED_BY**
  - Non-current concept with >1 **MOVED_TO**
  - Invalid ConceptId2 with **SAME_AS**, **REPLACED_BY**, **MAYBE_A**
  - Invalid ConceptId2 with **MOVED_TO**
  - Invalid ConceptId2 with **WAS_A**
  - Navigational concept with any subtypes

Outline

- Background
- Types of Auditing
- Logic-based Auditing
- Auditing processes
- **State of the art**

State of the art



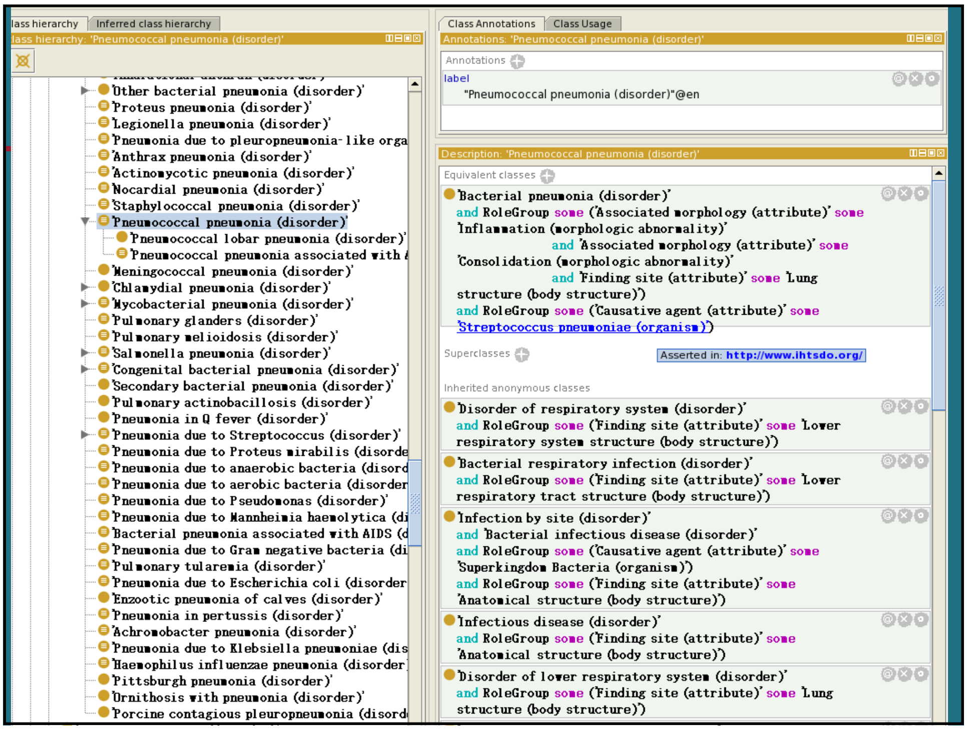
Ontology metrics:

Metrics

Class count	310013
Object property count	62
Data property count	0
Individual count	0
DL expressivity	ALER

Class axioms

SubClass axioms count	366772
Equivalent classes axioms count	59182
Disjoint classes axioms count	0
GCI count	0
Hidden GCI Count	0


lass hierarchy
Inferred class hierarchy
lass hierarchy: 'Pneumococcal pneumonia (disorder)'
Class Annotations
Class Usage
Annotations: 'Pneumococcal pneumonia (disorder)'
Annotations
Description: 'Pneumococcal pneumonia (disorder)'
Equivalent classes
Superclasses
Inherited anonymous classes

- 'Other bacterial pneumonia (disorder)'
- 'Proteus pneumonia (disorder)'
- 'Legionella pneumonia (disorder)'
- 'Pneumonia due to pleuropneumonia-like orga
- 'Anthrax pneumonia (disorder)'
- 'Actinomycotic pneumonia (disorder)'
- 'Nocardial pneumonia (disorder)'
- 'Staphylococcal pneumonia (disorder)'
- 'Pneumococcal pneumonia (disorder)'
  - 'Pneumococcal lobar pneumonia (disorder)'
  - 'Pneumococcal pneumonia associated with
- 'Meningococcal pneumonia (disorder)'
- 'Chlamydial pneumonia (disorder)'
- 'Mycobacterial pneumonia (disorder)'
- 'Pulmonary glanders (disorder)'
- 'Pulmonary melioidosis (disorder)'
- 'Salmonella pneumonia (disorder)'
- 'Congenital bacterial pneumonia (disorder)'
- 'Secondary bacterial pneumonia (disorder)'
- 'Pulmonary actinobacillosis (disorder)'
- 'Pneumonia in Q fever (disorder)'
- 'Pneumonia due to Streptococcus (disorder)'
- 'Pneumonia due to Proteus mirabilis (disorde
- 'Pneumonia due to anaerobic bacteria (disord
- 'Pneumonia due to aerobic bacteria (disorder
- 'Pneumonia due to Pseudomonas (disorder)'
- 'Pneumonia due to Mannheimia haemolytica (di
- 'Bacterial pneumonia associated with AIDS (d
- 'Pneumonia due to Gram negative bacteria (di
- 'Pulmonary tularemia (disorder)'
- 'Pneumonia due to Escherichia coli (disorder
- 'Enzootic pneumonia of calves (disorder)'
- 'Pneumonia in pertussis (disorder)'
- 'Achromobacter pneumonia (disorder)'
- 'Pneumonia due to Klebsiella pneumoniae (dis
- 'Haemophilus influenzae pneumonia (disorder
- 'Pittsburgh pneumonia (disorder)'
- 'Ornithosis with pneumonia (disorder)'
- 'Porcine contagious pleuropneumonia (disord

label
"Pneumococcal pneumonia (disorder)"@en

Equivalent classes

```
'Bacterial pneumonia (disorder)'
and RoleGroup some ('Associated morphology (attribute)' some
'Inflammation (morphologic abnormality)'
            and 'Associated morphology (attribute)' some
'Consolidation (morphologic abnormality)'
            and 'Finding site (attribute)' some 'Lung
structure (body structure)')
and RoleGroup some ('Causative agent (attribute)' some
'Streptococcus pneumoniae (organism)')
```

Superclasses

Asserted in: http://www.ihtsdo.org/

Inherited anonymous classes

```
'Disorder of respiratory system (disorder)'
and RoleGroup some ('Finding site (attribute)' some 'Lower
respiratory system structure (body structure)')

'Bacterial respiratory infection (disorder)'
and RoleGroup some ('Finding site (attribute)' some 'Lower
respiratory tract structure (body structure)')

'Infection by site (disorder)'
and 'Bacterial infectious disease (disorder)'
and RoleGroup some ('Causative agent (attribute)' some
'Superkingdom Bacteria (organism)')
and RoleGroup some ('Finding site (attribute)' some
'Anatomical structure (body structure)')

'Infectious disease (disorder)'
and RoleGroup some ('Finding site (attribute)' some
'Anatomical structure (body structure)')

'Disorder of lower respiratory system (disorder)'
and RoleGroup some ('Finding site (attribute)' some 'Lung
structure (body structure)')
```

Contents lists available at ScienceDirect

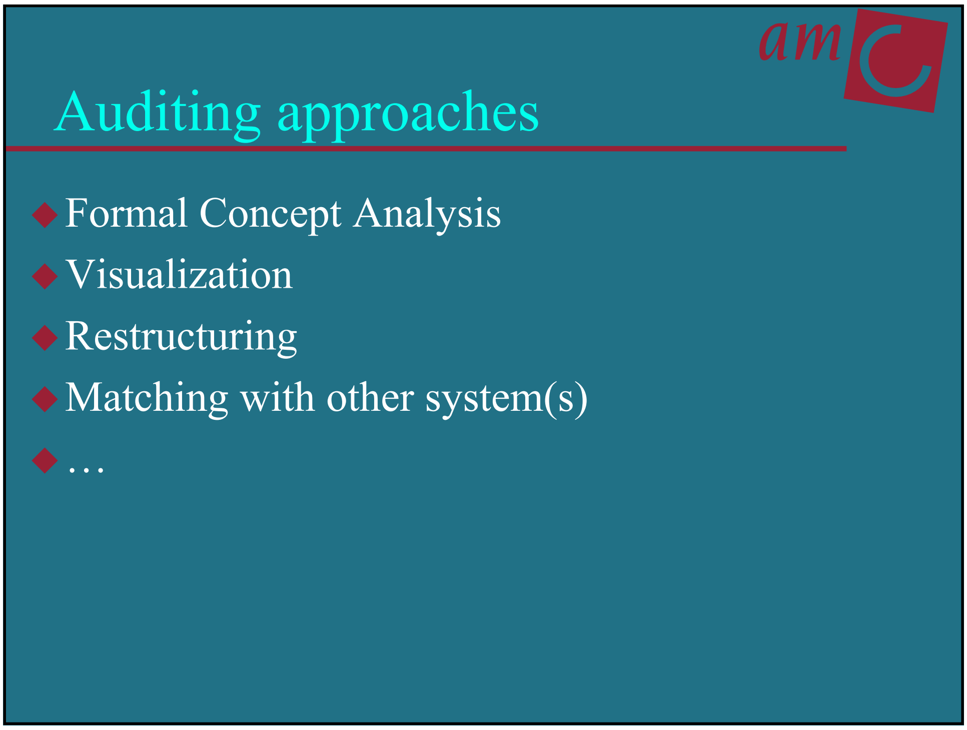
Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Volume 42, Number 3, June 2009

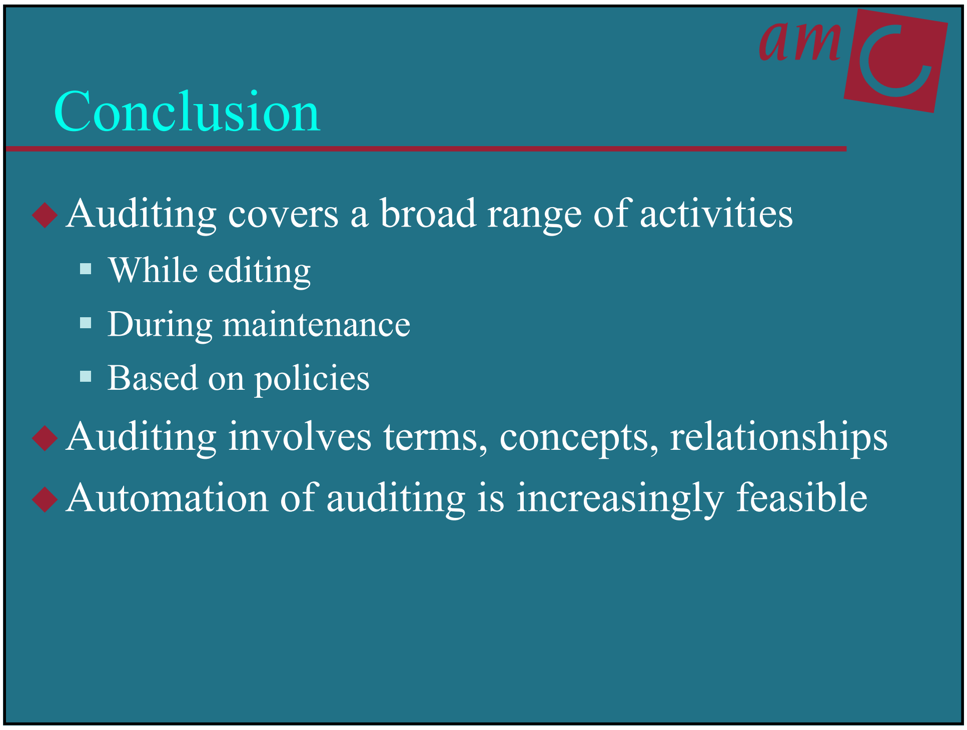
Special Issue: Auditing of Terminologies

Guest Editors: J. Geller, Y. Perl, M. Halper and R. Cornet



Auditing approaches

- Formal Concept Analysis
- Visualization
- Restructuring
- Matching with other system(s)
- …

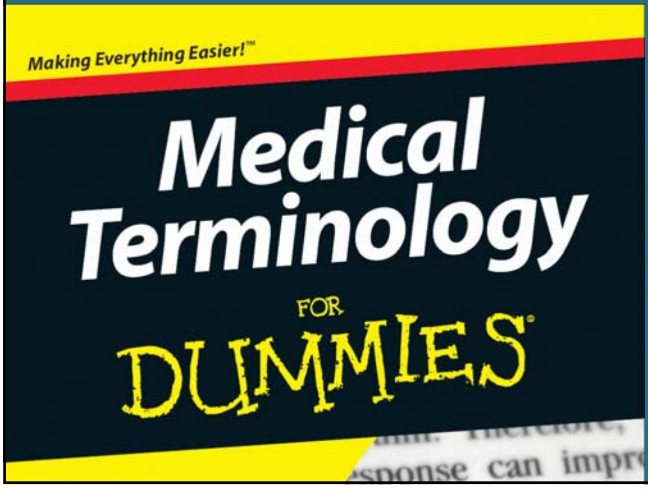


Conclusion

- Auditing covers a broad range of activities
  - While editing
  - During maintenance
  - Based on policies
- Auditing involves terms, concepts, relationships
- Automation of auditing is increasingly feasible

Auditing

- No longer for nerds (only)



Making Everything Easier!™

Medical Terminology FOR DUMMIES®

aim. Therefore,
esponse can impr