

# *Nouvelles théories de représentation des connaissances: sur les traces de Marc Jamouille*

1. Ontologies
2. Web sémantique
3. Terminological Markup framework



# *Ontologies*

➔ Pourquoi ? De nouveaux besoins

➔ Comment ?

- Structure des ontologies
- Création d'une ontologie
- Types d'ontologies
- Fusion d'ontologies

➔ Exemples, applications



# *Ontologies: pourquoi ?*

- ➔ Formalisation informatique d'une représentation des connaissances (vocabulaire contrôlé, thésaurus, classification...): c'est une nouvelle manière de travailler les systèmes classificatoires.
- ➔ Rendre compte des perceptions différentes d'une même réalité
  - Nécessité de travailler non en fonction des termes que l'on utilise mais des réalités qu'ils désignent.



# *Ontologies: pourquoi ?*

- ➔ Logique forte: permet l'inférence des propriétés
  - Conséquences: cohérence, automatisation (entre classes d'une même ontologie, entre ontologies)
  - NB: Dans la construction des vocabulaires contrôlés « historiques », les règles étaient localisées en fonction des évolutions voulues par les concepteurs.



# Comment ?

- ➔ Objectif : fixer des règles de construction pour pouvoir ensuite en faire une exploitation automatique
  - ➔ Dilemne : si les règles sont trop lâches, l'automatisation devient impossible (trop lent ou impossible), si elles sont trop restrictives, on ne peut pas tout exprimer.
  - ➔ Présentation des familles (classes):
    - Logique, Graphes et Réseaux Sémantiques, Langages Objet
- => Logique de description: langage OWL



# *L'ambiguïté d'un terme pour désigner un objet*

« Portable  
»



# Les définitions concurrentes d'un objet



« vél  
o »

enfant

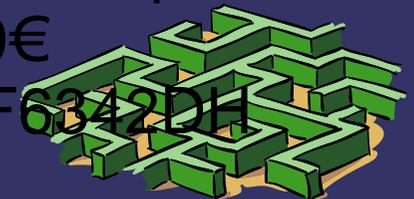
écologiste

catalogue

code route

mono-  
cycliste

- jouet
- moyen de transport
- taille : 1m20
- roues : 2
- guidon : à moustaches
- selle : dure
- véhicule sans permis
- prix : 200€
- ref : WDF6342DH



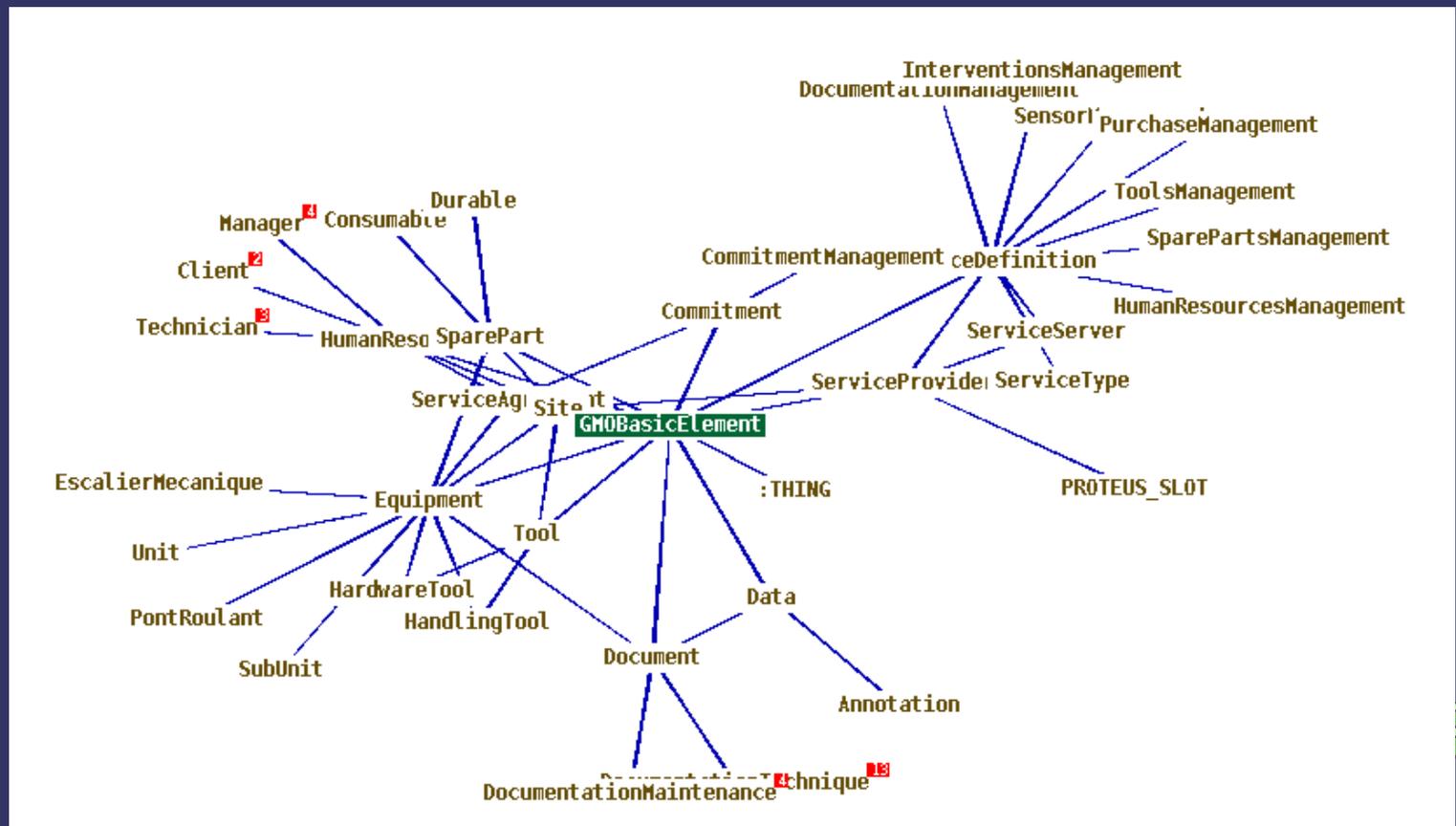
# *La structure d'une ontologie*

- Une ontologie est un cadre formel pour la représentation des connaissances qui repose sur les concepts, définis par leurs attributs dont les relations entre concepts qui la structurent.
- Concepts: ensemble de « choses » ayant des attributs communs (« boîte » dont la dénomination a peu d'importance)
- Attributs (propriétés, les relations font parties des propriétés, ex de relation: « sorte de », « partie de »)
- Les concepts créent des classes et des sous-classes
- Facettes: valeurs autorisées des attributs (cardinalité)
- Instances
- Héritage des propriétés des instances aux classes
- **Toutes les instances sont décrites par une URI!!**



# Un exemple: le projet PROTEUS

- ➔ 136 classes
- ➔ 93 attributs



# *Méthodologies de création d'ontologies*

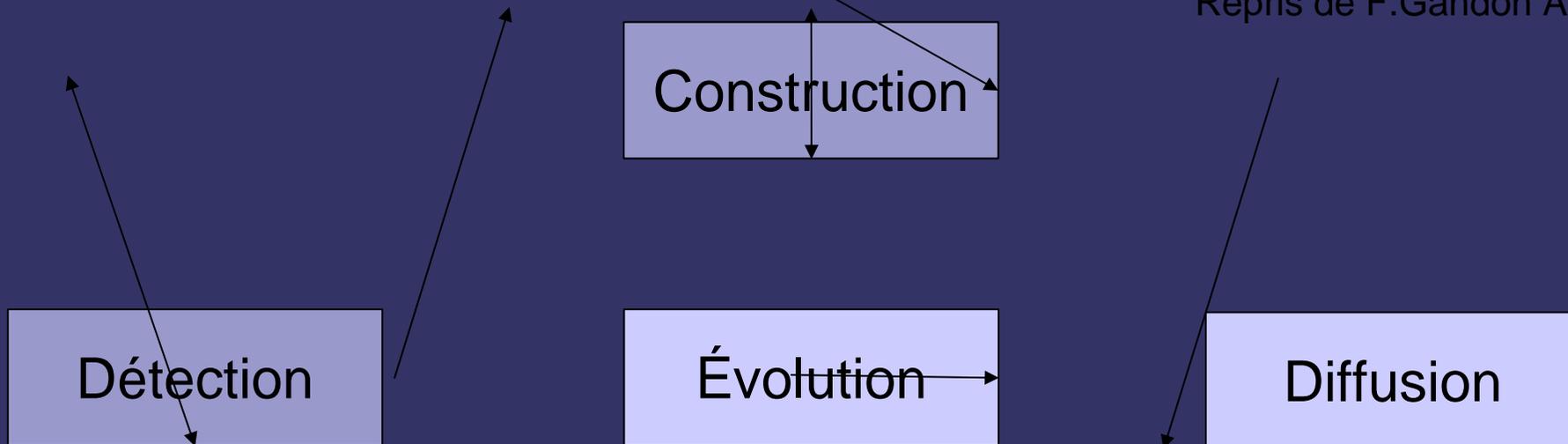
- Cycle de Fernandez, Gomez-Perez & Juristo (1997)
- Développement en V de Robert Stevens (2001)
- *Ontology Development 101 : A Guide to Creating Your First Ontology* - de Noy et McGuinness (2001)

=> permet de définir les concepts (classes), les attributs, les facettes.



# **Cycle** de Fernandez, Gomez-Perez, Juristo (1997)

Repris de F.Gandon ACACI



**Évaluation**

Utilisation



# *En résumé*

Le cadre formel des ontologies est conçu pour permettre:

- ➔ d'éviter les problèmes d'ambiguïté,
- ➔ faciliter la représentation des connaissances concurrentes d'un même objet,
- ➔ raisonner sur les connaissances et aider à la recherche d'information.
- ➔ Faciliter la révision et l'expansion des connaissances sur le domaine étudié



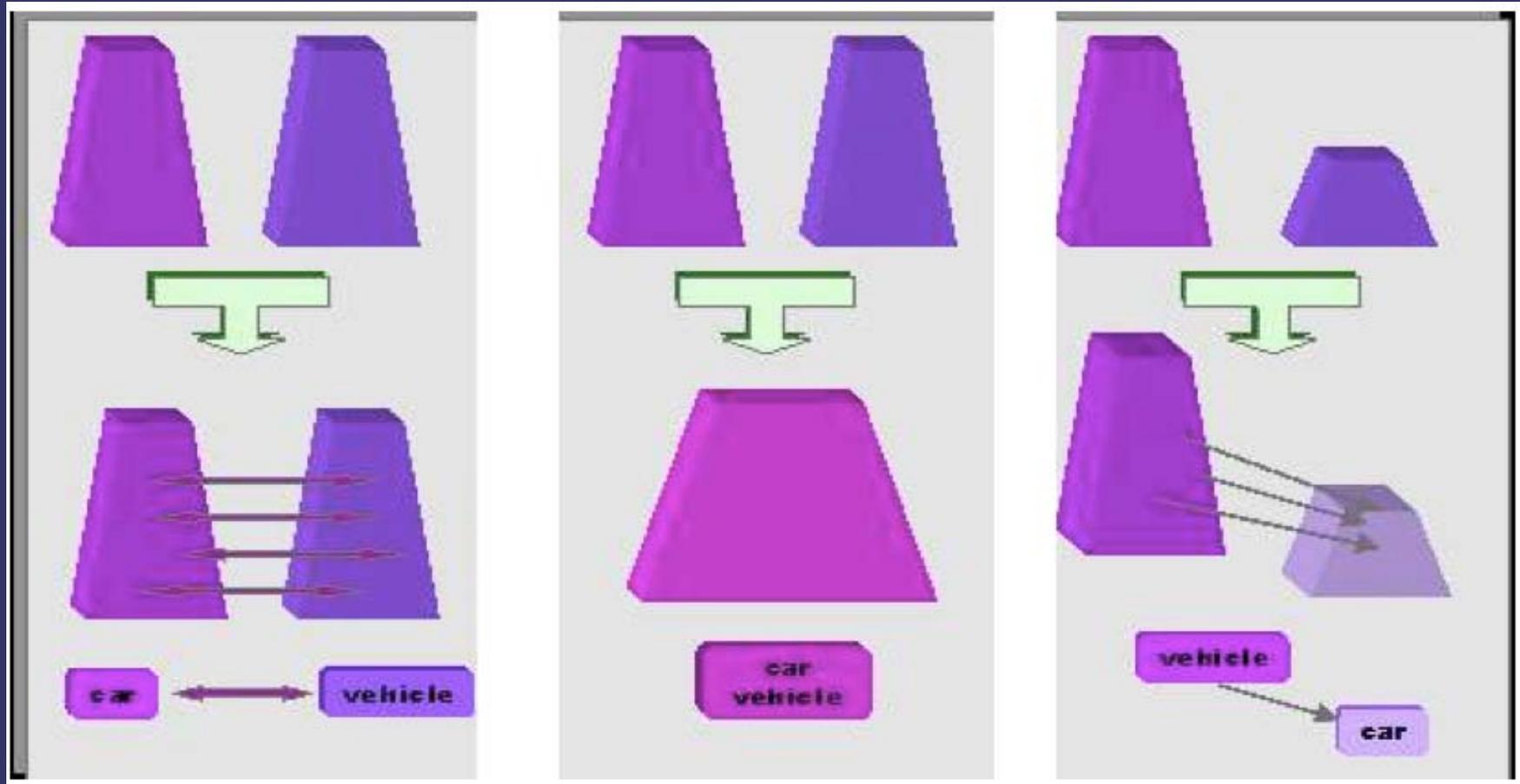
# *Types d'ontologies et leur objectifs*

- ➔ Ontologies applicatives: compréhension commune entre fabricant de logiciels d'un domaine
- ➔ Ontologies de domaines: organiser le savoir, réutiliser le savoir,
- ➔ Meta-ontologies: inter-opérabilité interdisciplinaire, inter-linguistique
- ➔ Autres objectifs des ontologies:
  - Analyse du savoir (distinction du savoir sur un domaine du savoir opérationnel)
  - Expliciter ce qui est implicite (analyse fonctionnelle)



# Fusion d'ontologies

## *“Mapping, Merging, Alignment”*



⊖ Perte d'information!



# *Ontologies médicales*



# ICD-10/11 Ontologie

## ➔ ICD-10

- [https://dkm.fbk.eu/index.php/ICD-10\\_Ontology](https://dkm.fbk.eu/index.php/ICD-10_Ontology)
- Parcourir avec Protégé
- Analyse des exclusions mutuelles

## ➔ ICD-11

- « A Content Model for the ICD-11 Revision », Tu et al., 2009
- Visualisation dans firefox
- <http://informatics.mayo.edu/icd11model/v20090506/index.htm>



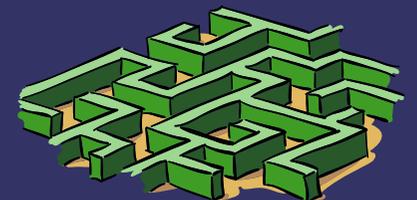
# *Infectious Disease Ontology*

- ➔ « The core Infectious Disease Ontology is an ontology of entities generally relevant to both the biomedical and clinical aspects of infectious diseases, such as 'pathogen', 'host', 'vector', and 'vaccine'. The structure of IDO adheres to the Basic Formal Ontology (BFO). »
- ➔ <http://infectious-disease-ontology.googlecode.com/svn/releases/2010-05-26/ido.owl>



# GeneOntology

- ➔ Pas une réelle ontologie : seulement des relations « is-a » et « part-of »
- ➔ <http://www.geneontology.org/GO.downloads.ontology.shtml>
- Attention : Trop lourde pour être intégrée dans protégé sur une machine normale.



# *Exemples*

- ➔ OpenGALEN (informations cliniques) :  
<http://www.opengalen.org/>
- ➔ MeSH (Medical Subject Headings):  
<http://www.nlm.nih.gov/mesh/meshhome.html>
- ➔ UMLS (Unified Medical Language System) :  
[www.nlm.nih.gov/research/umls/](http://www.nlm.nih.gov/research/umls/)



# GALEN

- ➔ The key feature of the GALEN approach is that it provides a model - a set of building blocks and constraints - from which concepts can be composed. By contrast, most traditional classification systems such as ICD9/10, ICPC, or the earlier versions of the Read Codes attempt to provide fully enumerated lists of all of the concepts which might ever be needed. In GALEN, classification of composite concepts is automatic and based on formal logical criteria. In traditional systems, classifications must be performed by hand explicitly for each concept. Traditional systems are usually tuned for use in one primary task, and using them for alternative tasks is difficult. GALEN separates the concept model from the model of use.



# OpenGALEN Browser - og7

TopCategory: MyocardialInfarction  
 hierarchical upwards browser: [dropdown]

MyocardialInfarction [Search] [Top]

- MyocardialInfarction
  - +Infarct
  - +MyocardialDegeneration

There was an error in the callback.

Relationships  Sanctions

## MyocardialInfarction

*Synonyms*  
MyocardialInfarct

Definition (necessary and sufficient)

BodyStructure  
 hasSpecificLocation [Myocardium](#)  
 hasUniqueAssociatedProcess [InfarctingProcess](#)  
 hasPathologicalStatus [pathological](#)

Conventional (necessary) criteria

hasAbnormalityStatus [nonNormal](#)  
 hasAcquisitionMode [AcquisitionMode](#)  
 hasAbsoluteState [acquired](#)  
 hasCountability [discrete](#)

### definition

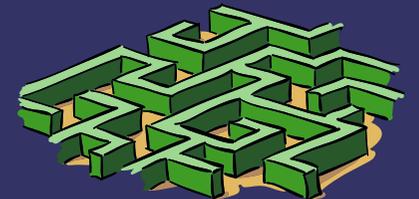
Close

A GALEN concept may either be primitive or composite. A primitive concept has been explicitly added as a child of another concept. In that case its definition is just the name of the concept. A composite concept is defined by its defining criteria (or relationships) and has been placed at its appropriate place in the hierarchy by a classifier.



# *Pratiquement ?*

- ➔ Interface entre les classifications existantes et l'utilisateur



# *Le web sémantique*

- Est l'application principale des théories sur les ontologies.
- Donne leur réel intérêt aux ontologies en permettant la transmission d'information standardisée entre machines.
- L'approche initiale a été celle des méta-données qui se sont progressivement structurées (RDF, RDF-S) et combinées à la théorie sur les ontologies pour former un réel langage de communication entre machines (OWL)

























# *Examples of successes?*



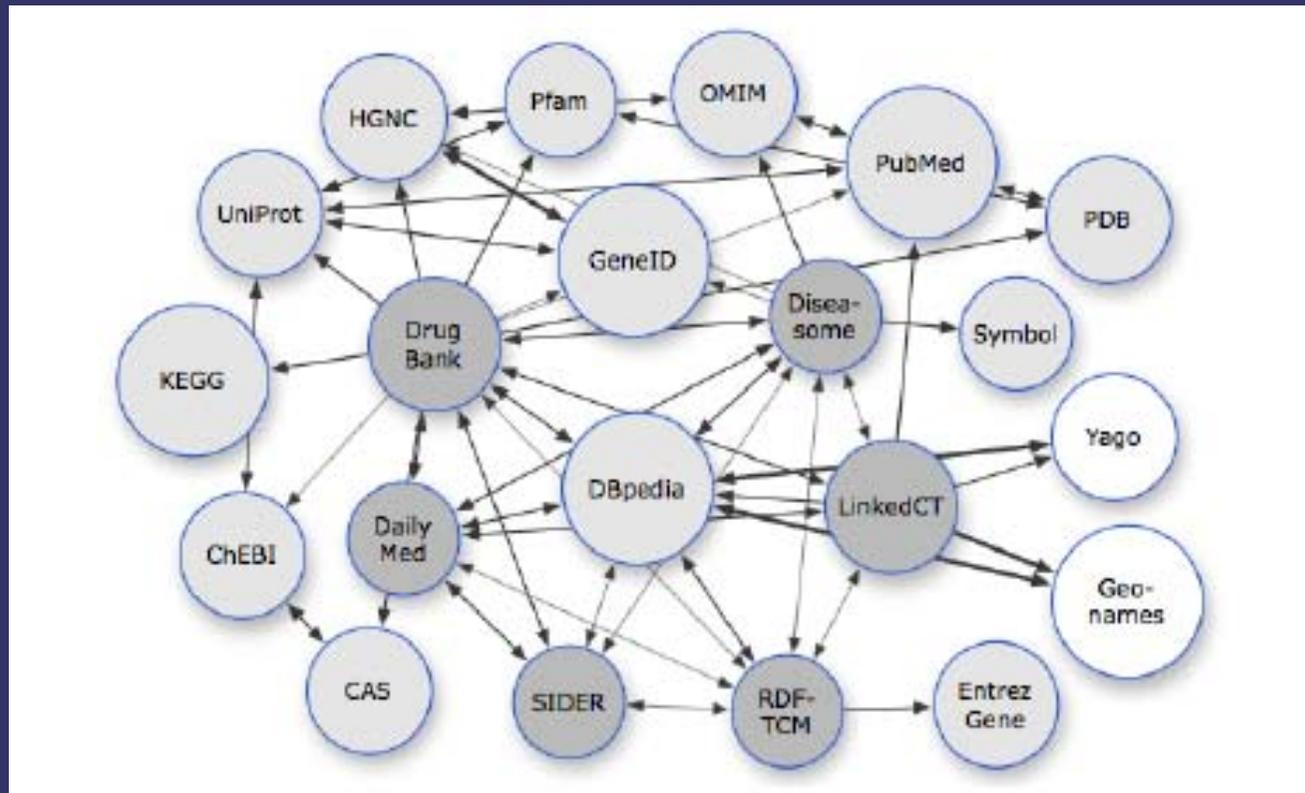
- *« DBpedia is not a commercial product, but it's one of the largest sources of Linked Data on the Web - so we think it deserves to be recognized in our top 10. Essentially, DBpedia extracts structured information from Wikipedia and makes that data available on the Web.*



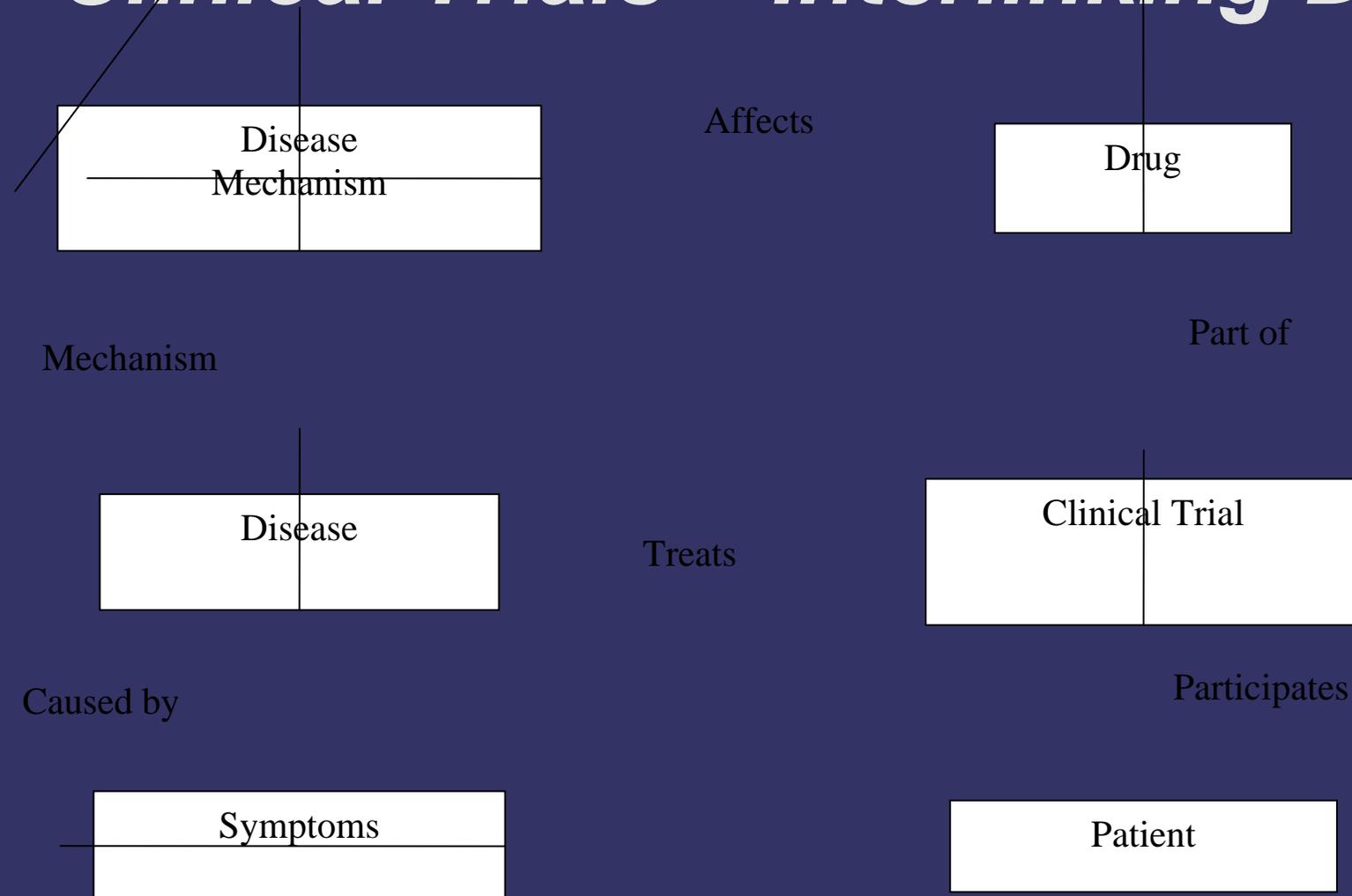
# Semantic Web for Health Care and Life Sciences Interest Group

➔ W3C - <http://esw.w3.org/HCLSIG>

- ➔ [Jen09] Jentzsch, A. and Zhao, J. and Hassanzadeh, O. and Cheung, K. H. and Samwald, M. and Andersson, B., *Linking Open Drug Data*, 2009



# Clinical Trials – Interlinking Data



All autonomous Data sources → Linked Data



# *Interlinking example*

## ➔ For each resource:

- Data source used
- Links used
- SPARQL query
- SPARQL result

## ➔ Resources identified

- Clinical Trials - <http://clinicaltrials.gov/>
- DrugBank - <http://www.drugbank.ca/>
- Diseasome - <http://www.ncbi.nlm.nih.gov/omim/>





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

### INCLUSION CRITERIA

main	Inclusion Category	Constraint
------	--------------------	------------

### EXCLUSION CRITERIA

main	Exclusion Category	Constraint
------	--------------------	------------

#### 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), sulfapyrazone (Anturane) or other uricosuric agents



# ICD-10/11 Ontologie

## ➔ ICD-10

- [https://dkm.fbk.eu/index.php/ICD-10\\_Ontology](https://dkm.fbk.eu/index.php/ICD-10_Ontology)
- Parcourir avec Protégé
- Analyse des exclusions mutuelles

## ➔ ICD-11

- « A Content Model for the ICD-11 Revision », Tu et al., 2009
- Visualisation dans firefox
- <http://informatics.mayo.edu/icd11model/v20090506/index.htm>



# *Terminological Markup Framework*



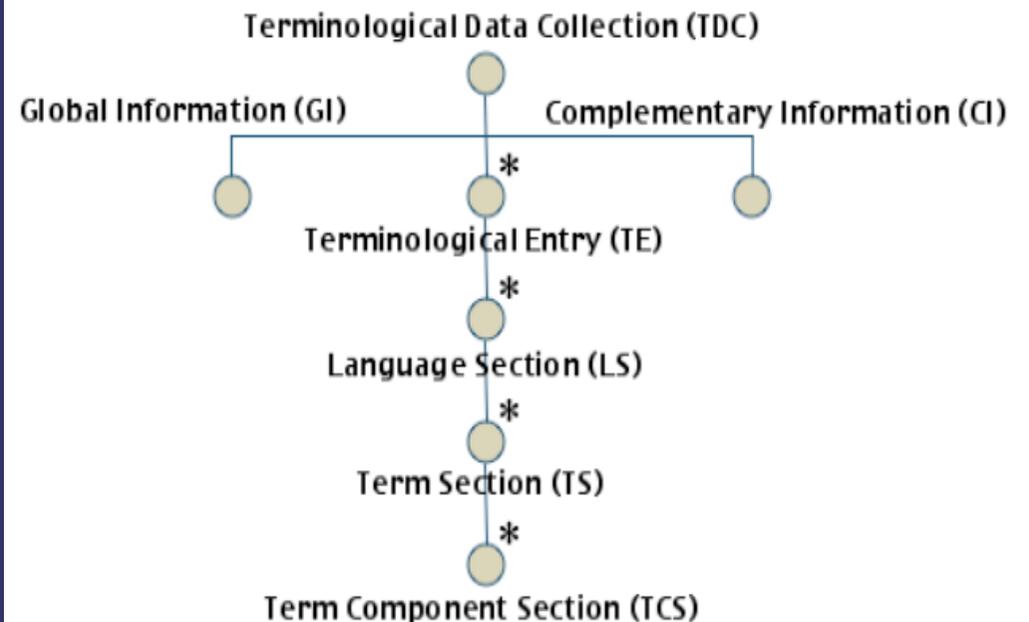
# ISO16642 (TMF) et ISO12620 (DataCat)

## → Terminological Markup Framework

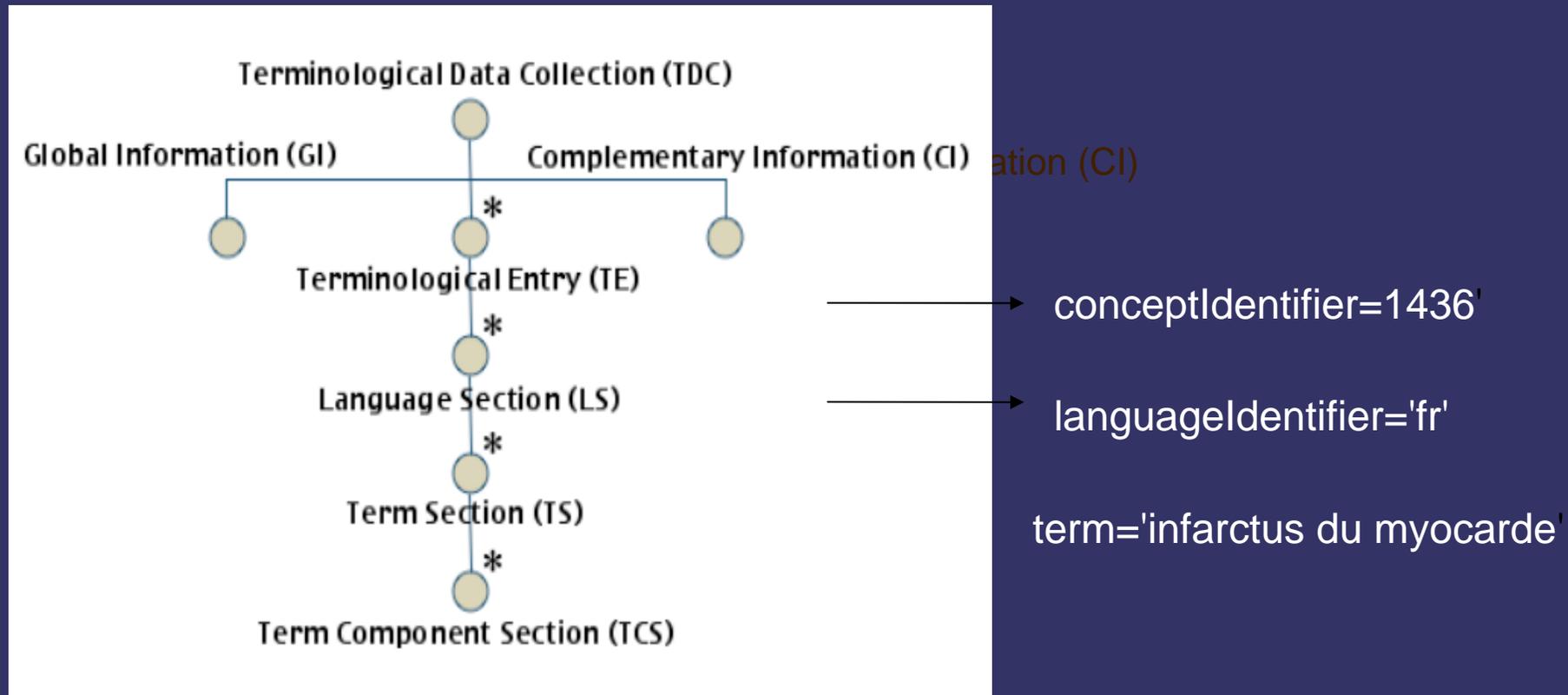
- Méta-modèle terminologique
- ISO 16642

## → Data Categories

- Catégories de données terminologiques
- ISO 12620



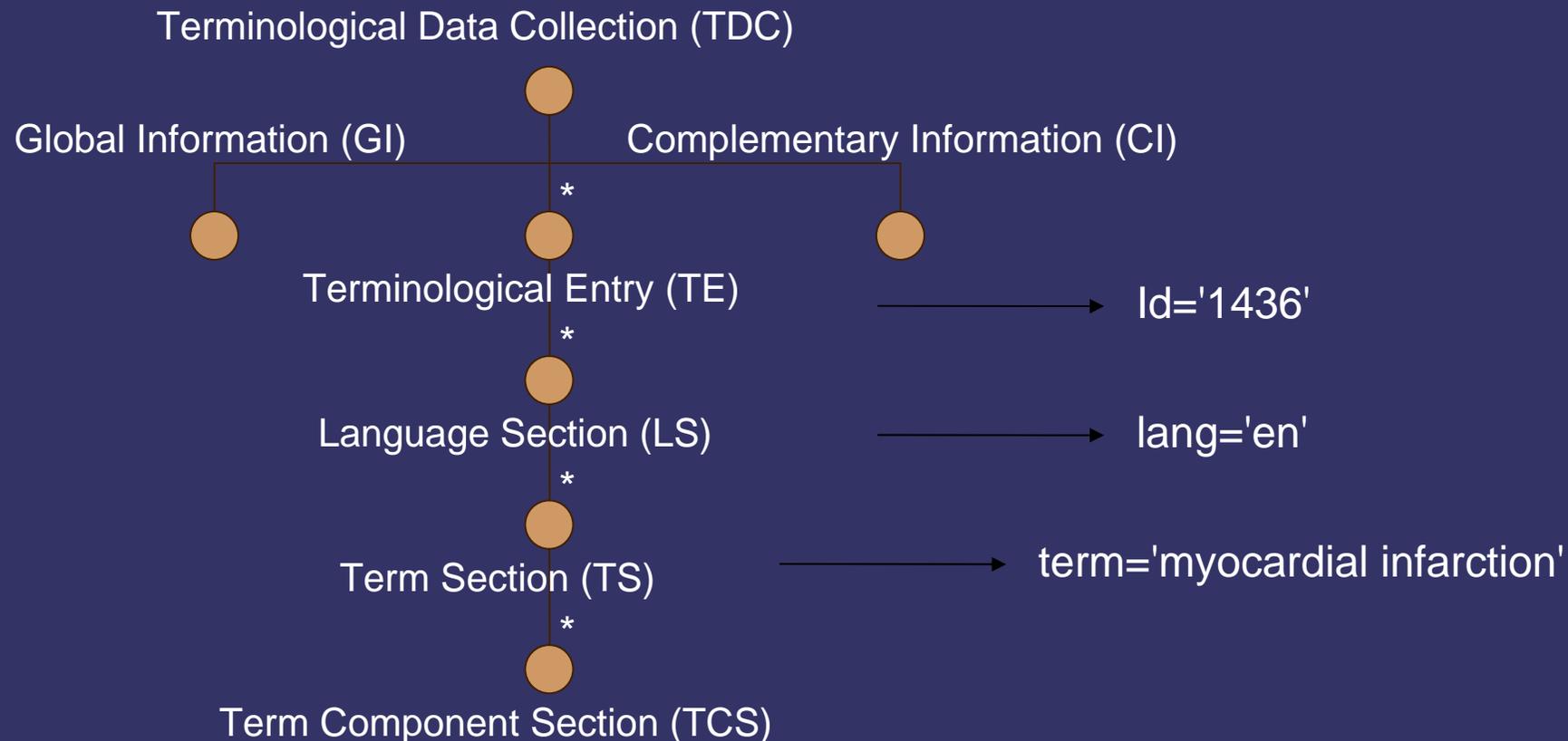
# Terminological Markup Framework



Source: *Terminological Markup Framework skeleton – ISO 16642 (L. Romary)*



# Terminological Markup Framework



Terminological Markup Framework skeleton – ISO 16642 (L. Romary)

- **Terminological Entry has several Terms, as a concept has several names**





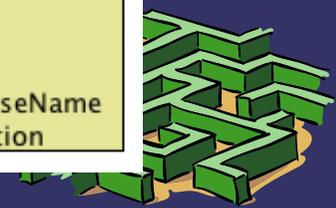
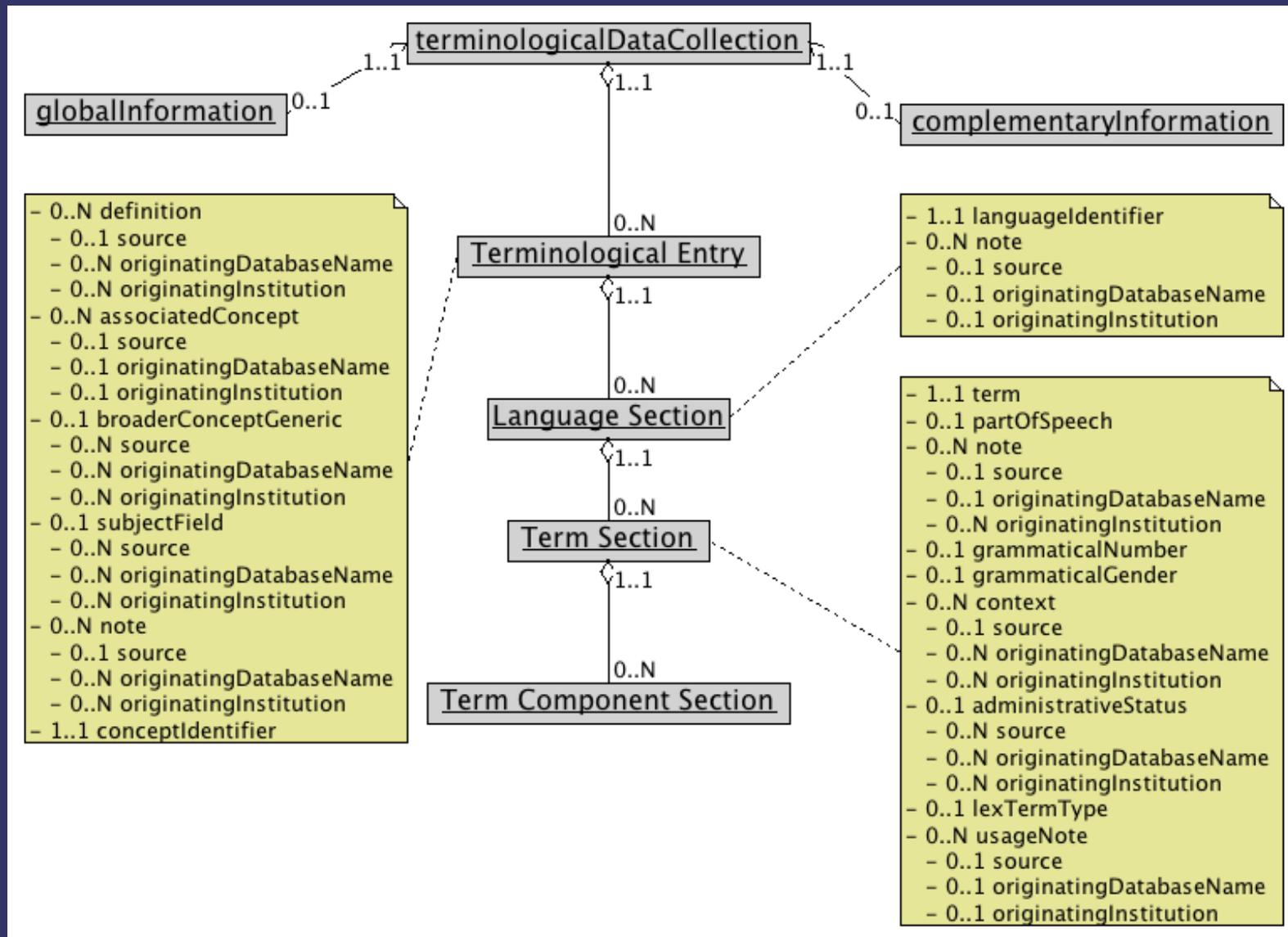
TermSciences  
Portail terminologique scientifique

*termsciences.fr*

- ➔ Portail terminologique (INIST, LORIA, l'ATILF)
- Valoriser, mutualiser les ressources terminologiques (lexiques, dictionnaires, thesaurus)
- Organismes publics de recherche et d'enseignement supérieur
  - constitution d'un référentiel terminologique commun
- ➔ Objectifs :
- Proposer une vision intégrée de la terminologie scientifique



# termsciences.fr modèle



# *Exemple : Infarctus du Myocarde*

- ➔ <http://www.termsciencences.fr/>
- ➔ Infarctus du myocarde



# Conclusion

- ➔ Ces nouvelles techniques permettent de mieux exploiter les masses d'informations disponibles
- ➔ Des possibilités d'échange de données telles qu'on à peine a en réaliser la portée
- ➔ Beaucoup de questions pour la suite:
  - Des outils plus difficiles à s'approprier
  - Demandent plus de moyens humains et financiers

