XLDB · LaSIGE · Faculdade de Ciências da Universidade de Lisboa

FACULDADE DE CIÊNCIAS
UNIVERSIDADE DE LISBOA

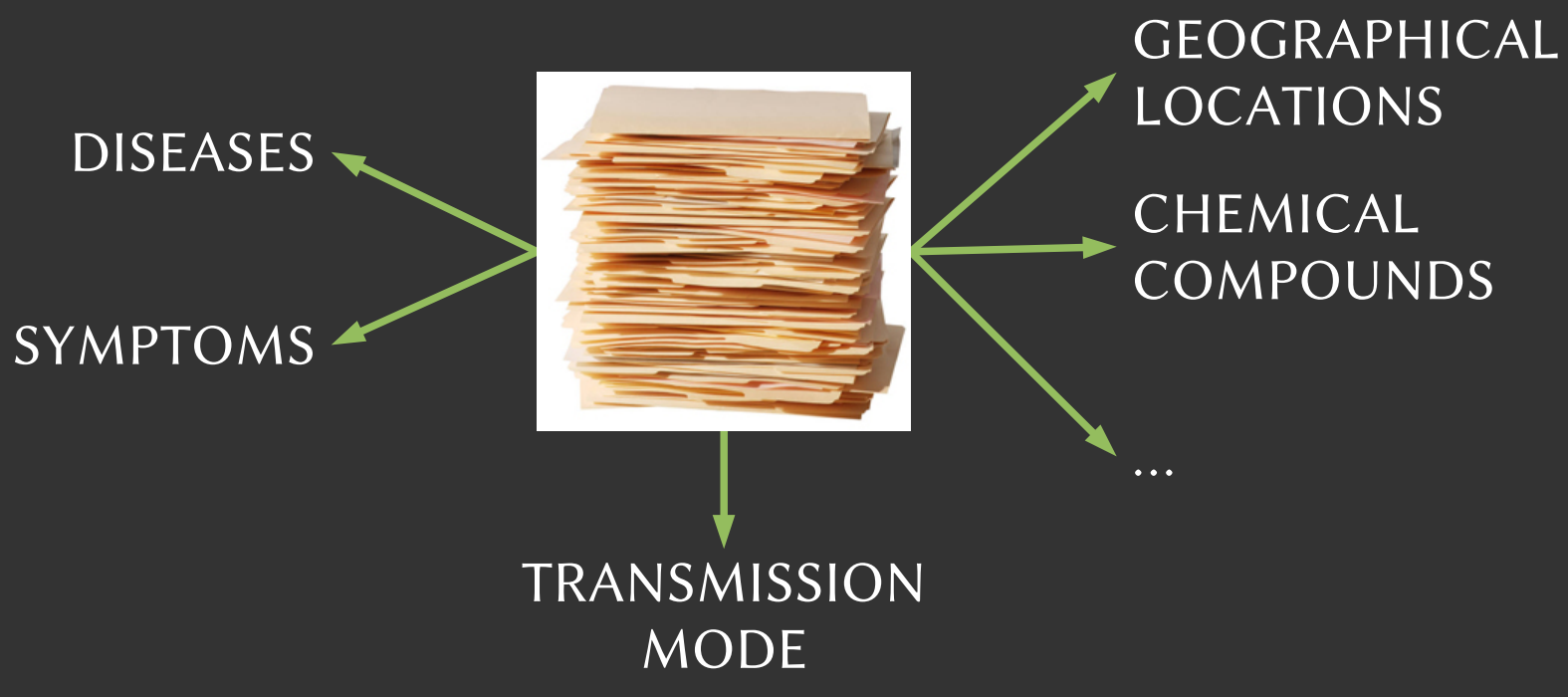
Bringing epidemiology into the Semantic Web

João D Ferreira, Catia Pesquita, Francisco M Couto and Mário J Silva

Graz, July 23th

Introduction
Epidemiological data is heterogeneous

- Data is heterogeneous and from different fields of knowledge



Introduction

Epidemiological data is heterogeneous

LOCATION : France

TIME : from Feb 18th, 2009 to Nov 6th, 2010

PATHOGEN : H1N1 Influenza Virus

HOST : Human

DISEASE : Influenza

Introduction

- Metadata is more *machine-friendly* when taken from controlled vocabularies
- Ontologies bring many advantages:
  - Inference in searching
  - Common vocabulary in shared resources
  - Semantic analysis

Introduction

Ontologies as source of vocabulary

SEARCH

- Searching for data on:
  - Infectious Diseases
  - In Europe
- Inference can help find the relevant resources

Introduction
Ontologies as source of vocabulary

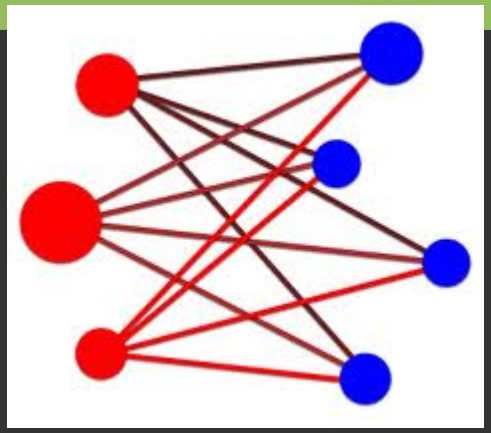
SHARE

- Common vocabulary
- Metadata in RDF can be resolved to known concepts

```
<rdf:Description rdf:about="resource_123">
  <em:disease rdf:resource="&obo;DOID_8469" />
  <em:host rdf:resource="&mesh;D006801" />
  ...
</rdf:Description>
```

Introduction

Ontologies as source of vocabulary



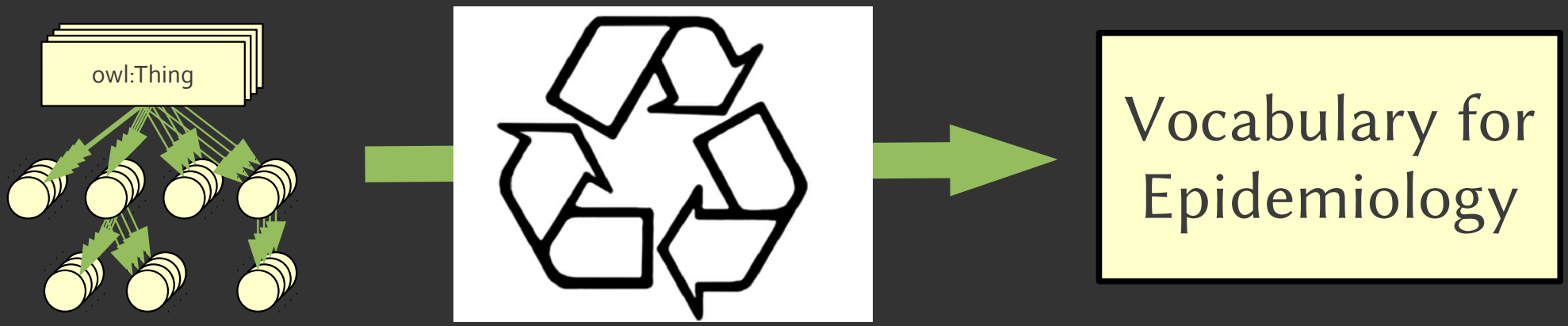
SEMANTIC ANALYSIS

- Exploration of technologies such as:
  - Semantic similarity
  - Ontology matching

  leading to:
  - Pattern recognition
  - Knowledge creation
  - …

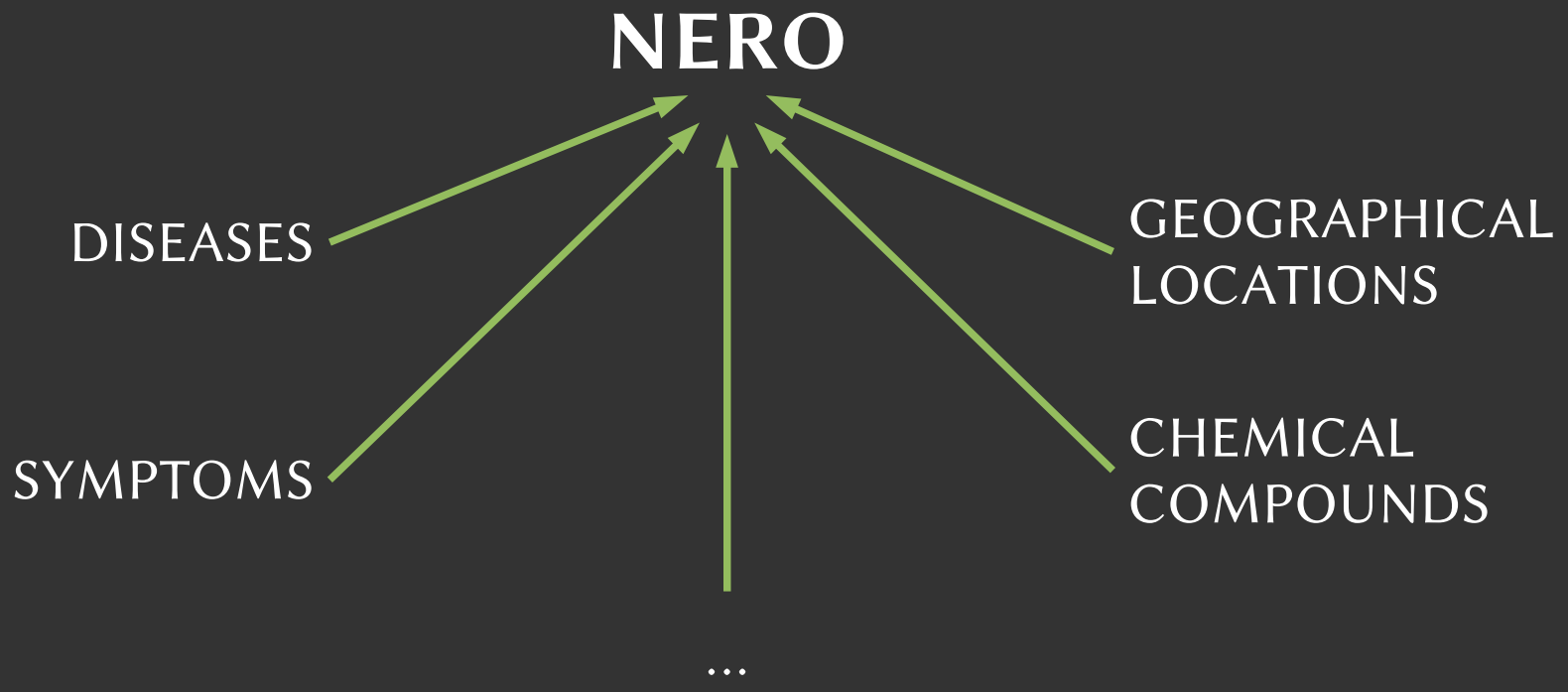
Epidemiology vocabulary

- We need concepts that are relevant in epidemiology
  - Not an ontology from scratch, but a reuse of concepts



NERO

- The *Network of Epidemiology-Related Ontologies*



NERO
Creation

- How to choose the ontologies?

NERO
Creation

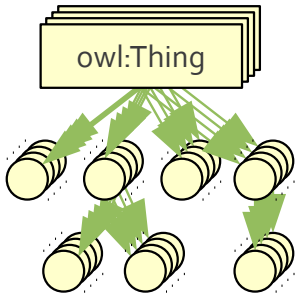
- Inspiration from:
  - General epidemiological needs

NERO
Creation

- Inspiration from:
  - General epidemiological needs
  - The *Epidemic Marketplace* (http://epimarketplace.net)

NERO
Creation

Epidemiology-related ontologies



NERO

applied on

Epidemiological resources

feedback

applied on

EM Epidemic Marketplace

NERO
NERO requirements

- Set of requirements that ensure:
  - interoperability
  - cohesion
- Requirements inspired on:
  - Epidemiological needs
  - W3C
  - OBO Foundry

NERO
NERO requirements

- Five examples:
  - Relevant domain
  - Textual definitions
  - Synonyms
  - Publicly available
  - Cross-references

NERO
NERO requirements

- Five examples:
  - Relevant domain
  - Textual definitions
  - Synonyms
  - Publicly available
  - Cross-references

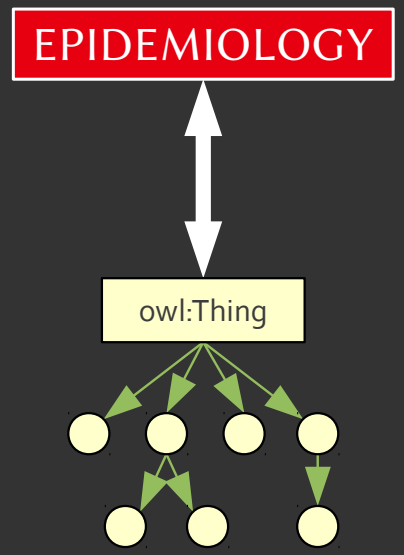
ATTENTION:
These requirements are *guidelines*

NERO
The ontologies

- Three types of ontologies:
  - 1) Ontologies specific to epidemiology
  - 2) Ontologies of generic scope
  - 3) Ontologies focused on a single domain

NERO
The ontologies

- 1) Ontologies specific to epidemiology
  - BioCaster Ontology
  - Epidemiology Ontology
  - Dictionary of Epidemiology

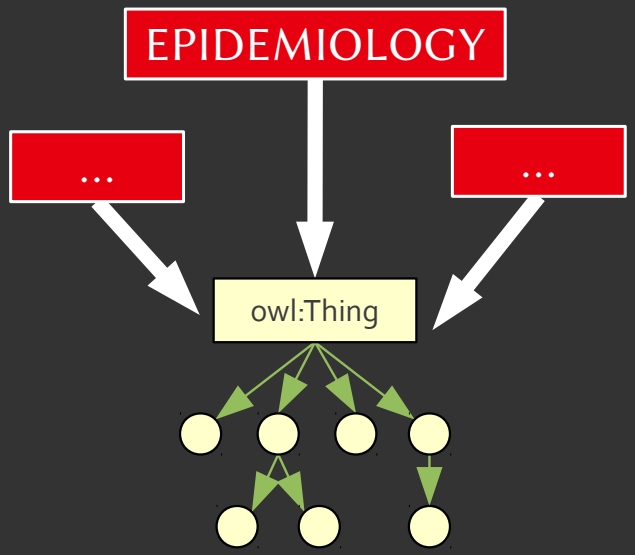


- These terminologies lack:
  - structure, scope, depth, ...

NERO
The ontologies

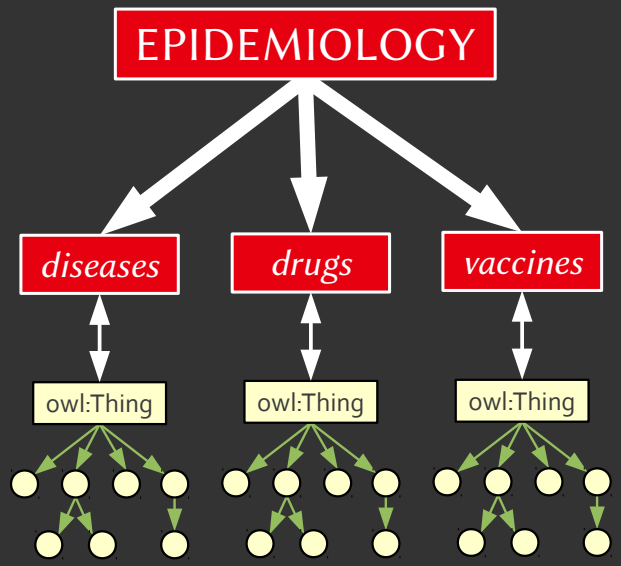
- 2) Ontologies of generic scope
  - UMLS
  - SNOMED-CT
  - MeSH

- Branches are hard to choose



NERO
The ontologies

- 3) Ontologies focused on a single domain
  - Many ontologies in the OBO Foundry web page
    - Diseases
    - Drugs
    - Vaccines
    - …
  - GeoPlanet (from Yahoo!)



- Insufficient coverage of the field of knowledge

Conclusions

- Current ontologies are not enough to represent epidemiological domain
- Best approach:

single domain ontologies
+
some branches of
generic ontologies

Conclusions

Terminology	Domain	Ref.	Fulfills requirement #...										In NERO?
			1	2	3	4	5	6	7	8	9	10	
BioCaster	Epidemiology	(Collier *et al.*, 2008)	Y	N	±	Y	Y	Y	N	Y	N	N	No
Epidemiology Ontology	Epidemiology	(HuGE Net, 2007)	Y	Y	N	N	N	Y	N	Y	N	N	Yes
Dictionary of Epidemiology	Epidemiology	(Porta, 2008)	Y	Y	N	N	Y	±	N	Y	N	N	No
UMLS	General	(Lindberg *et al.*, 1993)	Y	Y-	N	Y	Y	Y	Y	±	Y	N	No
MeSH	General	(Lipscomb, 2000)	Y	Y-	N	N	Y	Y	Y	Y	Y	N	Yes
SNOMED-CT®	General	(Stearns *et al.*, 2001)	Y	Y-	Y	Y	Y	Y	Y	N	N	N	No
GeoPlanet™	Geography	(Yahoo!, 2011)	Y	Y	±	Y	±	Y	Y	±	Y	Y	Yes
GeoNames	Geography	(Geonames.org, 2011)	Y	Y	N	Y	N	Y	Y	Y	Y	Y	No
Geo-Net-PT	Geography	(Lopez-Pellicer *et al.*, 2009)	Y	N	Y	Y	±	Y	N	Y	Y	Y	No
OBO ontologies													
ChEBI	Biochemistry	(de Matos *et al.*, 2010)	Y	Y-	Y	Y	Y	Y	Y	Y	Y	Y	Yes
DOID	Diseases	(Osborne *et al.*, 2009)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Yes
ENVO	Environment	(EnvO developers, 2012)	Y	±	Y	±	Y	Y	N	Y	Y	Y	Yes
HP	Symptoms	(Robinson and Mundlos, 2010)	Y	±	Y	Y	Y	Y	Y	Y	Y	Y	Yes
IDO	Diseases	(Cowell and Smith, 2010)	Y	±	Y	N	Y	N	Y	Y	Y	Y	Yes
NCBI Taxonomy	Taxonomy	(Wheeler *et al.*, 2007)	Y	Y-	Y	N	N	N	Y	Y	Y	Y	Yes
NCI Thesaurus	General	(Sioutos *et al.*, 2007)	Y	Y-	Y	Y	Y	Y	Y	Y	Y	Y	Yes
SYMP	Symptoms	(Schriml *et al.*, 2010)	Y	Y-	Y	Y	Y	Y	Y	Y	Y	Y	Yes
TRANS	Disease transmission	(Schriml *et al.*, 2010)	Y	N	Y	±	Y	N	N	Y	Y	Y	Yes
VO	Vaccines	(Yang *et al.*, 2011)	Y	Y	Y	N	Y	N	N	Y	Y	Y	Yes

Contributions

- NERO as a **vocabulary** that can be used to **characterize epidemiological resources**

- Annotated epidemiological resources can be explored in the context of **semantic web**
  - Information Retrieval & Integration

- NERO enables **other technologies**:
  - Ontology Matching, Semantic Similarity, ...

Acknowledgments

XLDB / LaSIGE

FCT
Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

EPIWORK

SEVENTH FRAMEWORK PROGRAMME

XLDB · LaSIGE · Faculdade de Ciências da Universidade de Lisboa

FACULDADE DE CIÊNCIAS
UNIVERSIDADE DE LISBOA

Bringing epidemiology into the Semantic Web

João D Ferreira, Catia Pesquita, Francisco M Couto and Mário J Silva

Find me at: joao.ferreira@lasige.di.fc.ul.pt
http://xldb.di.fc.ul.pt/wiki/Joao_Ferreira

Graz, July 23th