

# Bringing epidemiology into the Semantic Web

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

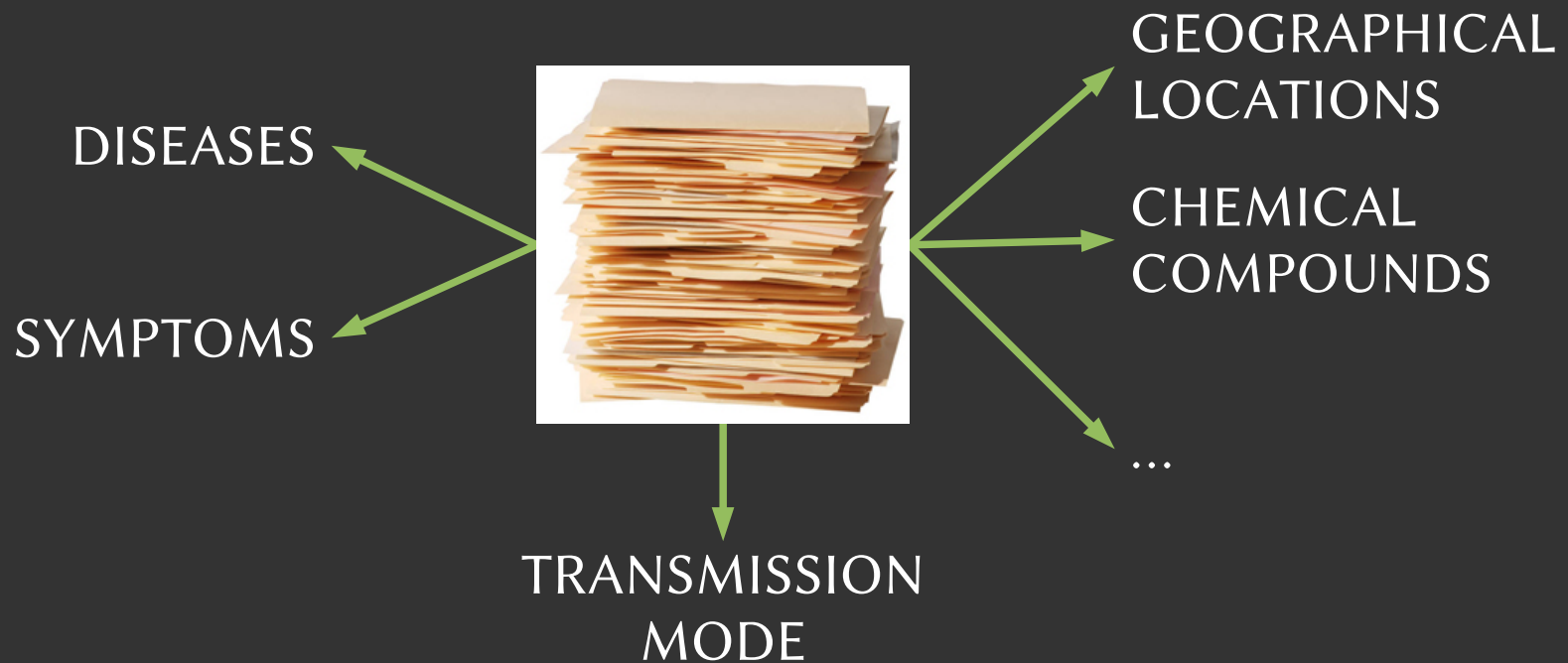
F	7th International Conference on Formal Ontology in Information Systems					July 24 – 27	
O							
I	C	B	O	2	0	1	2
S	3rd International Conference on Biomedical Ontology					July 21 – 25	

Graz, July 23<sup>th</sup>

# Introduction

## Epidemiological data is heterogeneous

- Data is heterogeneous and from different fields of knowledge



# Introduction

Epidemiological data is heterogeneous

## [France Imported cases fraction](#)

Author

2010-06-11

Description

Subject: Epidemiology

Type: Dataset

[See more](#)

**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



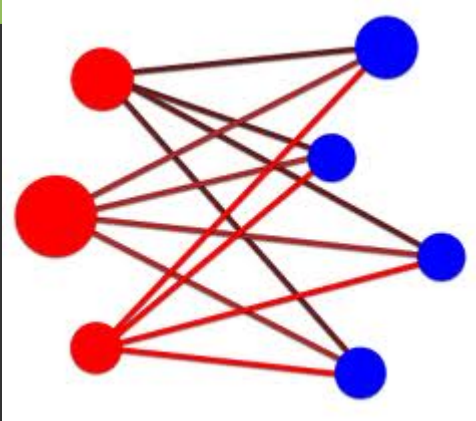
SHARE

- 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

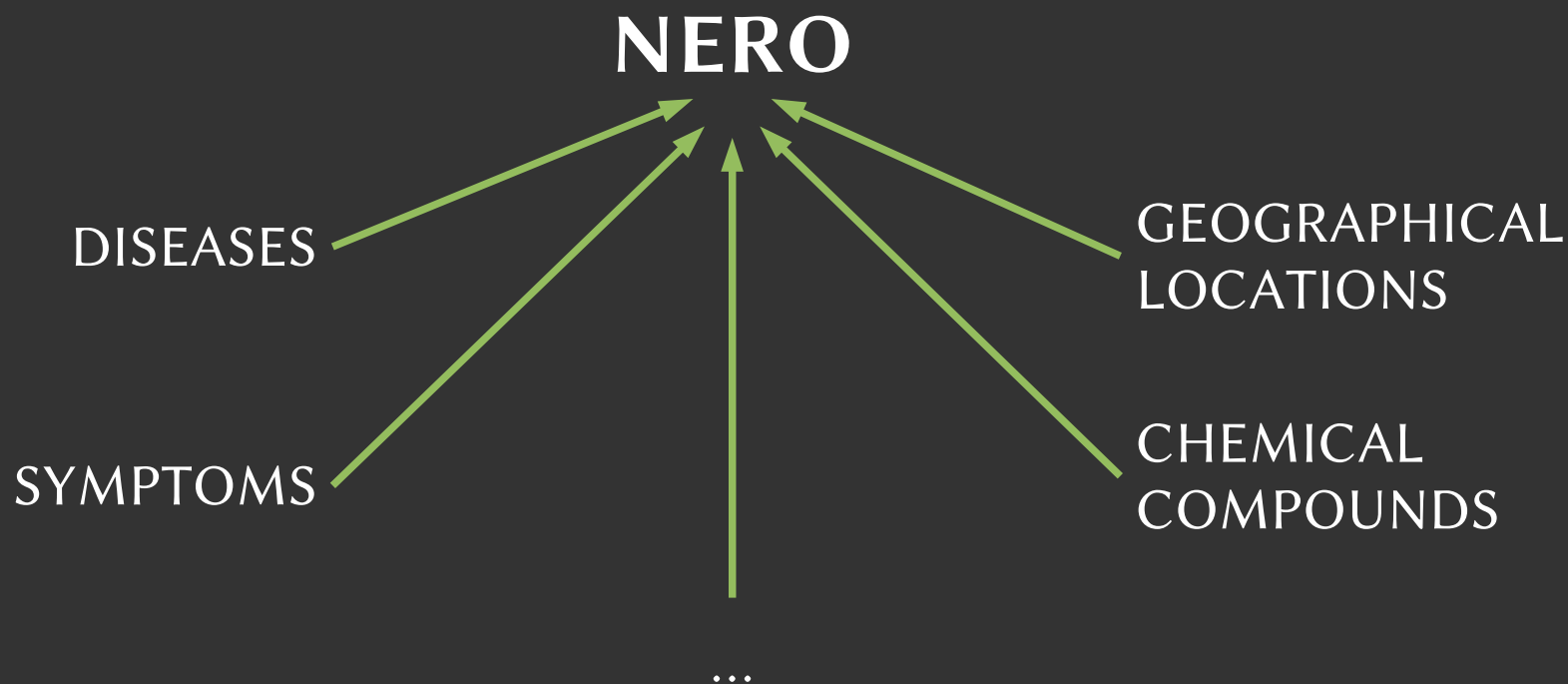
- We need concepts that are relevant in epidemiology
  - Not an ontology from scratch, but a reuse of concepts





# NERO

- *The Network of Epidemiology-Related Ontologies*



# NERO Creation

- How to choose the ontologies?



# NERO Creation

- Inspiration from:
  - General epidemiological needs

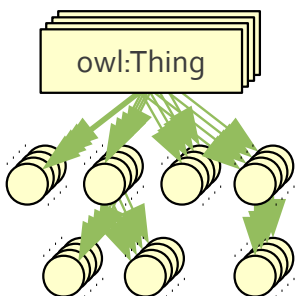
# NERO Creation

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

The screenshot shows the Epidemic Marketplace website. The header is red with the EM logo and tagline 'Epidemic Marketplace ... a platform for integrating and sharing epidemiological data.' There is a search bar and a 'Login' button. Below the header, there are navigation links: // Browse // Upload // Request // Contact // Learn More. The main content area is divided into several sections: a red bar with 'Browse', 'Upload', and 'Request' buttons; a red bar for 'Latest Announcement' with a 'Version 2.2 Online' update; a section for '169 Resources' with a line graph showing an upward trend from Dec 2011 to May 2012; a section for '4170 Visits' with a line graph showing a peak in Jan 2012 and a decline through Apr 2012; and a yellow sticky note with links for 'Recent Requests', 'Popular Requests', 'Recent Uploads', 'Uploaders list', and 'Feed subscription'. There are also links for 'Advanced search', 'Learn more', 'Become a Curator', and 'Developers Corner'.

# NERO Creation

Epidemiology-  
related ontologies



**NERO**

Epidemiological  
resources



*applied on*

*feedback*

*applied on*



# 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

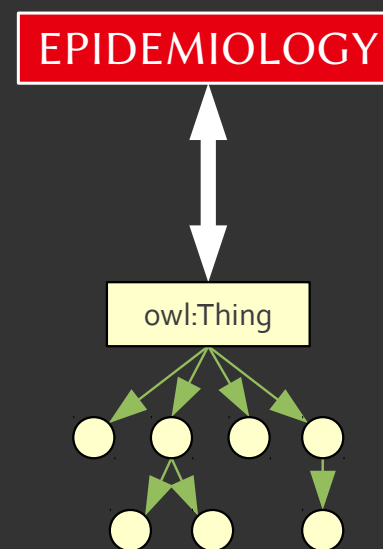
## The ontologies

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

# NERO

## The ontologies

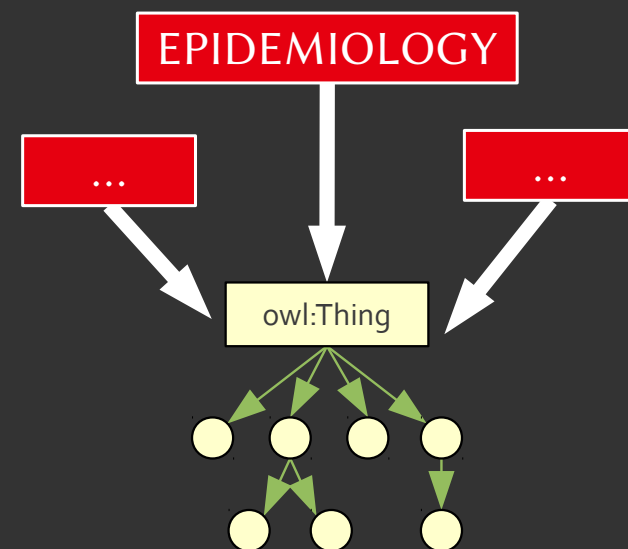
- 1) Ontologies specific to epidemiology
  - BioCaster Ontology
  - Epidemiology Ontology
  - Dictionary of Epidemiology
- These terminologies lack:
  - structure, scope, depth, ...



# NERO

## The ontologies

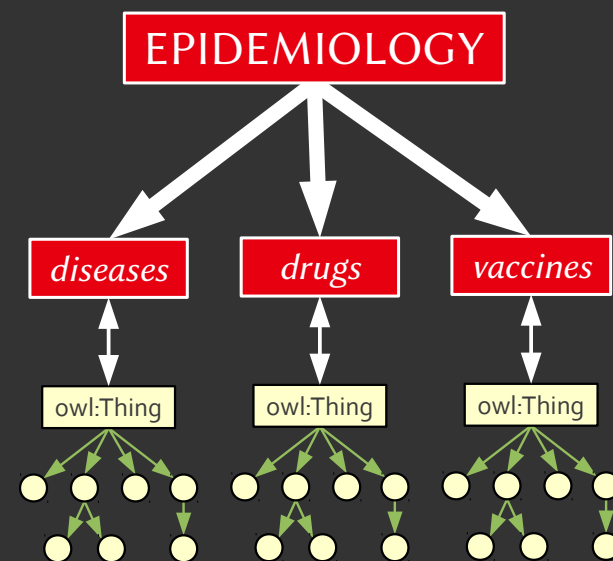
- 2) Ontologies of generic scope
  - UMLS
  - SNOMED-CT
  - MeSH
- Branches are hard to choose



# 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 <i>et al.</i> , 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 <i>et al.</i> , 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 <i>et al.</i> , 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 <i>et al.</i> , 2009)	Y	N	Y	Y	±	Y	N	Y	Y	Y	No
OBO ontologies													
ChEBI	Biochemistry	(de Matos <i>et al.</i> , 2010)	Y	Y-	Y	Y	Y	Y	Y	Y	Y	Y	Yes
DOID	Diseases	(Osborne <i>et al.</i> , 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 <i>et al.</i> , 2007)	Y	Y-	Y	N	N	N	Y	Y	Y	Y	Yes
NCI Thesaurus	General	(Sioutos <i>et al.</i> , 2007)	Y	Y-	Y	Y	Y	Y	Y	Y	Y	Y	Yes
SYMP	Symptoms	(Schriml <i>et al.</i> , 2010)	Y	Y-	Y	Y	Y	Y	Y	Y	Y	Y	Yes
TRANS	Disease transmission	(Schriml <i>et al.</i> , 2010)	Y	N	Y	±	Y	N	N	Y	Y	Y	Yes
VO	Vaccines	(Yang <i>et al.</i> , 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





# 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](mailto:joao.ferreira@lasige.di.fc.ul.pt)  
[http://xldb.di.fc.ul.pt/wiki/Joao\\_Ferreira](http://xldb.di.fc.ul.pt/wiki/Joao_Ferreira)

F	7th International Conference on Formal Ontology in Information Systems					July 24 – 27	
O							
I	C	B	O	2	0	1	2
S	3rd International Conference on Biomedical Ontology					July 21 – 25	

Graz, July 23<sup>th</sup>