Ontology alignment: evaluation and healthcare applications

Ernesto Jiménez-Ruiz

HealthInsight workshop, Oslo, May 20, 2016
Outline

Preliminaries

Ontology Alignment

Healthcare Applications
The presentation in a nutshell

- Introduction to ontologies

- Ontology alignment

- Ontology alignment in practice
  - Semantic annotation and access of clinical letters
  - Semantic enrichment of UK BioBank Cardiac cine-MRI Scans
  - Shared hypothesis testing in the biomedical domain
  - Pistoia alliance mapping project (disease-phenotype domain)
  - Lung Cancer Assistant (LCA)
  - Optique: Scalable End-user Access to Big Data
The presentation in a nutshell

- **Introduction to ontologies**

- **Ontology alignment**

- **Ontology alignment in practice**
  - Semantic annotation and access of clinical letters
  - Semantic enrichment of UK BioBank Cardiac cine-MRI Scans
  - Shared hypothesis testing in the biomedical domain
  - Pistoia alliance mapping project (disease-phenotype domain)
  - Lung Cancer Assistant (LCA)
  - Optique: Scalable End-user Access to Big Data
The presentation in a nutshell

- Introduction to ontologies

- Ontology alignment

- Ontology alignment in practice
  - Semantic annotation and access of clinical letters
  - Semantic enrichment of UK BioBank Cardiac cine-MRI Scans
  - Shared hypothesis testing in the biomedical domain
  - Pistoia alliance mapping project (disease-phenotype domain)
  - Lung Cancer Assistant (LCA)
  - Optique: Scalable End-user Access to Big Data
What is an ontology?

- **Introduces vocabulary** relevant to a domain
  - Anatomy
- **Specifies meaning (semantics) of terms**
  - Heart is a muscular organ that is part of the circulatory system
- **Formalised** using suitable logic
  - Heart SUBCLASSOF MuscularOrgan AND (isPartOf SOME CirculatorySystem)

Borrowed from Ian’s slides: **Ontologies and the Semantic Web: The Story So Far.** Zhejiang University, April 2010
What is an ontology?

- Introduces **vocabulary** relevant to a domain
  - Anatomy
- Specifies meaning (**semantics**) of terms
  - Heart is a muscular organ that is part of the circulatory system
- **Formalised** using suitable logic
  - Heart SUBCLASSOF MuscularOrgan AND (isPartOf SOME CirculatorySystem)

Borrowed from Ian’s slides: **Ontologies and the Semantic Web: The Story So Far.** Zhejiang University, April 2010
What is an ontology?

- Introduces **vocabulary** relevant to a domain
  - Anatomy
- Specifies meaning (**semantics**) of terms
  - Heart is a muscular organ that is part of the circulatory system
- **Formalised** using suitable logic
  - Heart $\text{SUBCLASSOF}$ MuscularOrgan $\text{AND}$ (isPartOf SOME CirculatorySystem)

Borrowed from Ian’s slides: **Ontologies and the Semantic Web: The Story So Far**. Zhejiang University, April 2010
OWL: Web Ontology Language

- World Wide Web Consortium (W3C) standard
- The most widely used ontology modelling language
  - e.g. FMA, NCI, SNOMED CT
- Formal underpinning of OWL is based on formal logic
- Supported by tools and infrastructure

**OWL example axioms**

- `JuvenileArthritis ⊑ JuvenileDisease`
- `PolyArthritis ≡ Arthritis ⊓ ⩾ 5 affects.Joint`
- `Disease ⊓ Joint ⊑ ⊥`
- `JuvenileIdiopathicArthritis @ “Juvenile Rheumatoid Arthritis”`
- `John:Patient ⊓ ∃suffersFrom.JuvenileIdiopathicArthritis`
OWL: Web Ontology Language

- World Wide Web Consortium (W3C) standard
- The most widely used ontology modelling language
  - e.g. FMA, NCI, SNOMED CT
- Formal underpinning of OWL is based on formal logic
- Supported by tools and infrastructure

**OWL example axioms**

- `JuvenileArthritis ⊑ JuvenileDisease`
- `PolyArthritis ≡ Arthritis ⊓ ⩾ 5 affects.Joint`
- `Disease ⊓ Joint ⊑ ⊥`
- `JuvenileIdiopathicArthritis @ “Juvenile Rheumatoid Arthritis”`
- `John:Patient ⊓ ∃suffersFrom.JuvenileIdiopathicArthritis`
What ontologies are good for?

- Help identify and resolve disagreements in the domain
  - *Ontology alignment* will play a key role
- Independence of logical/physical schema
- Formulation of queries closer to domain experts
- Incomplete and semi-structured data
- Integration of heterogeneous sources
  - *Ontology alignment* will play a key role
What ontologies are good for?

- Help identify and resolve disagreements in the domain
  - *Ontology alignment* will play a key role
- Independence of logical/physical schema
- Formulation of queries closer to domain experts
- Incomplete and semi-structured data
- Integration of heterogeneous sources
  - *Ontology alignment* will play a key role
Outline

Preliminaries

Ontology Alignment

Healthcare Applications
Ontology alignment: motivation

- An application domain can be modelled with **different points of view and purposes**
- Ontologies with **different naming and modelling conventions** exist for the same domain
- Aligning these ontologies will **enable interoperability** between ontology-based information systems
- **Reusing** vocabulary from domain ontologies is a good practice in ontology engineering
Ontology alignment: definition

- **Ontology alignments or mappings** $\mathcal{M}$ **are sets of tuples** $\langle e_1, e_2, n, \rho \rangle$
  - $e_1, e_2$ are entities in the input ontologies $\mathcal{O}_1$ and $\mathcal{O}_2$
  - $n$ a confidence value between 0 and 1
  - $\rho$ is the semantic relationship between $e_1$ and $e_2$ (e.g. subsumption, equivalence or disjointness)

- **Formalized as OWL 2 axioms**
  - Where the semantic relationship $\rho$ is one of $\{\equiv, \subseteq, \supseteq, \perp\}$
  - Confidence values $n$ are represented as axiom annotations
  - No extra semantics

- **OWL 2 example mappings**
  - $\mathcal{O}_1$:Joint $\equiv \mathcal{O}_2$:Joint
  - $\mathcal{O}_3$:Joint_structure $\equiv \mathcal{O}_2$:Joint
  - $\mathcal{O}_3$:Joint_structure $\equiv \mathcal{O}_2$:Articulation
Ontology alignment: definition

- **Ontology alignments or mappings** $\mathcal{M}$ are sets of tuples $\langle e_1, e_2, n, \rho \rangle$
  - $e_1, e_2$ are entities in the input ontologies $\mathcal{O}_1$ and $\mathcal{O}_2$
  - $n$ a confidence value between 0 and 1
  - $\rho$ is the semantic relationship between $e_1$ and $e_2$ (e.g. subsumption, equivalence or disjointness)

- Formalized as OWL 2 axioms
  - Where the semantic relationship $\rho$ is one of $\{\equiv, \subseteq, \supseteq, \perp\}$
  - Confidence values $n$ are represented as axiom annotations
  - No extra semantics

- **OWL 2 example mappings**
  - $\mathcal{O}_1$:Joint $\equiv \mathcal{O}_2$:Joint
  - $\mathcal{O}_3$:Joint_structure $\equiv \mathcal{O}_2$:Joint
  - $\mathcal{O}_3$:Joint_structure $\equiv \mathcal{O}_2$:Articulation
Ontology alignment: definition

- **Ontology alignments or mappings** $\mathcal{M}$ are sets of tuples $\langle e_1, e_2, n, \rho \rangle$
  - $e_1, e_2$ are entities in the input ontologies $\mathcal{O}_1$ and $\mathcal{O}_2$
  - $n$ a confidence value between 0 and 1
  - $\rho$ is the semantic relationship between $e_1$ and $e_2$ (e.g. subsumption, equivalence or disjointness)

- **Formalized as OWL 2 axioms**
  - Where the semantic relationship $\rho$ is one of $\{\equiv, \subseteq, \supseteq, \bot\}$
  - Confidence values $n$ are represented as axiom annotations
  - No extra semantics

- **OWL 2 example mappings**
  - $\mathcal{O}_1$:Joint $\equiv \mathcal{O}_2$:Joint
  - $\mathcal{O}_3$:Joint\textunderscore structure $\equiv \mathcal{O}_2$:Joint
  - $\mathcal{O}_3$:Joint\textunderscore structure $\equiv \mathcal{O}_2$:Articulation
Ontology alignment: systems

- Given two input ontologies $\mathcal{O}_1$ and $\mathcal{O}_2$ generate an alignment $\mathcal{M}$ as output.
- Large number of available ontology alignment systems
- Ontology Alignment Evaluation Initiative (OAEI)

LogMap...
- performs a simple matching but efficient (based on sophisticated lexical and structural indexes),
- can efficiently match semantically rich ontologies containing hundreds of thousands of classes (FMA, NCI and SNOMED),
- incorporates reasoning and repair capabilities
- is one of the top systems in the OAEI evaluation campaign, and
- is the only system participating in all OAEI 2015 tracks.

- Large-scale Interactive Ontology Matching: Algorithms and Implementation. ECAI 2012
- LogMap: Logic-based and Scalable Ontology Matching. ISWC’11
Ontology alignment: systems

- Given two input ontologies $O_1$ and $O_2$ generate an alignment $M$ as output.
- Large number of available ontology alignment systems
- Ontology Alignment Evaluation Initiative (OAEI)

**LogMap** . . .
- performs a simple matching but efficient (based on sophisticated **lexical and structural indexes**),
- can **efficiently match** semantically rich ontologies containing hundreds of thousands of classes (FMA, NCI and SNOMED),
- incorporates **reasoning and repair** capabilities
- is one of the **top systems** in the OAEI evaluation campaign, and
- is the only system participating in **all OAEI 2015 tracks**.

- Large-scale Interactive Ontology Matching: Algorithms and Implementation. ECAI 2012
- LogMap: Logic-based and Scalable Ontology Matching. ISWC’11
Ontology alignment: resources

- **UMLS** (Unified Medical Language System) Metathesaurus
  - Integrates more than one hundred thesauri and ontologies
  - Contains more than 6 million entities

- **BioPortal**
  - Contains more than 500 ontologies
  - Represent a network of ontologies
  - More than 10 million mappings are available
  - Also includes user-submitted alignments
Ontology alignment: resources

- **UMLS (Unified Medical Language System) Metathesaurus**
  - Integrates more than one hundred thesauri and ontologies
  - Contains more than 6 million entities

- **BioPortal**
  - Contains more than 500 ontologies
  - Represent a network of ontologies
  - More than 10 million mappings are available
  - Also includes user-submitted alignments
Ontology alignment: limitations

- **Alignment systems, BioPortal and UMLS techniques...**
  - are mostly based on lexical algorithms,
  - although they also include user-submitted alignments (BioPortal) and expert assessment (UMLS),
  - may lead to undesired cross-references,
  - may lead to logical errors when considering the semantics of the sources being integrated.

<table>
<thead>
<tr>
<th>Aligned ontologies</th>
<th>Generated alignments</th>
</tr>
</thead>
</table>
| FMA ~ NCI          | **FMA:**Joint $\equiv$ **NCI:**Joint  
                     **FMA:**Joint $\equiv$ **NCI:**Articulation  
                     **FMA:**Set_of_joints $\equiv$ **NCI:**Joint  
                     **FMA:**Set_of_joints $\equiv$ **NCI:**Articulation |
| FMA ~ SNOMED CT    | **FMA:**Joint $\equiv$ **SNOMED:**Joint_structure  
                     **FMA:**Set_of_joints $\equiv$ **SNOMED:**Joint_structure |
| SNOMED CT ~ NCI    | **SNOMED:**Joint_structure $\equiv$ **NCI:**Joint  
                     **SNOMED:**Joint_structure $\equiv$ **NCI:**Articulation |
Ontology alignment: limitations

- **Alignment systems, BioPortal and UMLS techniques...**
  - are mostly based on lexical algorithms,
  - although they also include user-submitted alignments (BioPortal) and expert assessment (UMLS),
  - may lead to undesired cross-references,
  - may lead to logical errors when considering the semantics of the sources being integrated.
Ontology alignment: limitations

- **Alignment systems, BioPortal and UMLS techniques...**
  - are mostly based on lexical algorithms,
  - although they also include user-submitted alignments (BioPortal) and expert assessment (UMLS),
  - may lead to undesired cross-references,
  - may lead to logical errors when considering the semantics of the sources being integrated.
Logic-based Assessment of UMLS

- Assessment of the integration of FMA, NCI and SNOMED CT ontologies within UMLS

<table>
<thead>
<tr>
<th>Ontologies</th>
<th>UMLS alignments</th>
<th>Logical errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>FMA-NCI</td>
<td>3,024</td>
<td>655</td>
</tr>
<tr>
<td>FMA-SNOMED</td>
<td>9,072</td>
<td>6,179</td>
</tr>
<tr>
<td>SNOMED-NCI</td>
<td>19,622</td>
<td>20,944</td>
</tr>
</tbody>
</table>

Logic-based Assessment of UMLS

- Assessment of the integration of FMA, NCI and SNOMED CT ontologies within UMLS

<table>
<thead>
<tr>
<th>Ontologies</th>
<th>UMLS alignments</th>
<th>Logical errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>FMA-NCI</td>
<td>3,024</td>
<td>655</td>
</tr>
<tr>
<td>FMA-SNOMED</td>
<td>9,072</td>
<td>6,179</td>
</tr>
<tr>
<td>SNOMED-NCI</td>
<td>19,622</td>
<td>20,944</td>
</tr>
</tbody>
</table>

Logic-based Assessment of BioPortal

- Identification of error-free sets of alignments

<table>
<thead>
<tr>
<th>Ontologies</th>
<th>BioPortal alignments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>#</td>
</tr>
<tr>
<td>BDO-NCIT</td>
<td>1,636</td>
</tr>
<tr>
<td>CCONT-NCIT</td>
<td>2,097</td>
</tr>
<tr>
<td>EFO-NCIT</td>
<td>2,507</td>
</tr>
<tr>
<td>EP-FMA</td>
<td>78,489</td>
</tr>
<tr>
<td>EP-NCIT</td>
<td>2,465</td>
</tr>
<tr>
<td>MA-FMA</td>
<td>961</td>
</tr>
<tr>
<td>OMIM-NCIT</td>
<td>5,178</td>
</tr>
<tr>
<td>SDO-EP</td>
<td>135</td>
</tr>
<tr>
<td>UBERON-FMA</td>
<td>1,932</td>
</tr>
<tr>
<td>ZFA-EFO</td>
<td>427</td>
</tr>
<tr>
<td>ZFA-UBERON</td>
<td>724</td>
</tr>
</tbody>
</table>

Logic-based Assessment of BioPortal

- Identification of error-free sets of alignments

<table>
<thead>
<tr>
<th>Ontologies</th>
<th>BioPortal alignments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>#</td>
</tr>
<tr>
<td>BDO-NCIT</td>
<td>1,636</td>
</tr>
<tr>
<td>CCONT-NCIT</td>
<td>2,097</td>
</tr>
<tr>
<td>EFO-NCIT</td>
<td>2,507</td>
</tr>
<tr>
<td>EP-FMA</td>
<td>78,489</td>
</tr>
<tr>
<td>EP-NCIT</td>
<td>2,465</td>
</tr>
<tr>
<td>MA-FMA</td>
<td>961</td>
</tr>
<tr>
<td>OMIM-NCIT</td>
<td>5,178</td>
</tr>
<tr>
<td>SDO-EP</td>
<td>135</td>
</tr>
<tr>
<td>UBERON-FMA</td>
<td>1,932</td>
</tr>
<tr>
<td>ZFA-EFO</td>
<td>427</td>
</tr>
<tr>
<td>ZFA-UBERON</td>
<td>724</td>
</tr>
</tbody>
</table>

Assessment of alignment systems: OAEI

• **Precision** and **recall** wrt reference alignment or silver/gold standard $|\mathcal{M}_{GS}|$
  
  - Precision (Pre) = $|\mathcal{M} \cap \mathcal{M}_{GS}|/|\mathcal{M}|$
  - Recall (Rec) = $|\mathcal{M} \cap \mathcal{M}_{GS}|/|\mathcal{M}_{GS}|$
  - The F-score $(F) = (2 \times \text{Pre} \times \text{Rec})/(\text{Pre} + \text{Rec})$.  

• **Logical errors** of $\mathcal{M}$ wrt $\mathcal{O}_1$ and $\mathcal{O}_2$.

• Computation **times** are also considered.

• **Ontology Alignment Evaluation Initiative (OAEI)**
  
  - An annual campaign
  - 3 biomedical-themed tracks in 2016: Anatomy, LargeBio, Phenotype-Disease tracks

• **Ontology Alignment Evaluation Initiative (OAEI):**
Assessment of alignment systems: OAEI

- **Precision** and **recall** wrt reference alignment or silver/gold standard $|\mathcal{M}_{GS}|$
  - Precision (Pre) = $|\mathcal{M} \cap \mathcal{M}_{GS}|/|\mathcal{M}|$
  - Recall (Rec) = $|\mathcal{M} \cap \mathcal{M}_{GS}|/|\mathcal{M}_{GS}|$
  - The F-score (F) = $(2 \times \text{Pre} \times \text{Rec})/(\text{Pre} + \text{Rec})$.

- **Logical errors** of $\mathcal{M}$ wrt $\mathcal{O}_1$ and $\mathcal{O}_2$.

- Computation **times** are also considered.

- **Ontology Alignment Evaluation Initiative (OAEI)**
  - An annual campaign
  - 3 biomedical-themed tracks in 2016: Anatomy, LargeBio, Phenotype-Disease tracks

**Ontology Alignment Evaluation Initiative (OAEI):**
http://www.ontologymatching.org/
Assessment in the OAEI large biomedical track

- **Ontologies and Reference Alignment**
  - FMA v2.0 (78,989 classes), NCI v.08.05d (66,724 classes) and SNOMED CT v. Jan. 2009 (306,591 classes).
  - Reference alignment based on UMLS

- **Matching problems**
  - FMA-NCI
  - FMA-SNOMED
  - SNOMED-NCI

- **Results**
  - Results are far from perfect when involving SNOMED CT
Assessment in the OAEI large biomedical track

- **Voting of computed alignments**
    - FMA-NCI: 18 contributing (independent) systems
    - FMA-SNOMED: 13 contributing (independent) systems
    - SNOMED-NCI: 9 contributing (independent) systems
  - Alignments voted (i.e. computed) by at least one system
    - Maximum recall
  - Alignments voted (i.e. computed) by all systems
    - Maximum precision
Assessment in the OAEI large biomedical track

**Voting of computed alignments**

  - FMA-NCI: 18 contributing (independent) systems
  - FMA-SNOMED: 13 contributing (independent) systems
  - SNOMED-NCI: 9 contributing (independent) systems
- Alignments voted (i.e. computed) by at least one system
  - **Maximum recall**
- Alignments voted (i.e. computed) by all systems
  - **Maximum precision**
Assessment in the OAEI large biomedical track

- **Voting of computed alignments**
    - FMA-NCI: 18 contributing (independent) systems
    - FMA-SNOMED: 13 contributing (independent) systems
    - SNOMED-NCI: 9 contributing (independent) systems
  - Alignments voted (i.e. computed) by at least one system
    - **Maximum recall**
  - Alignments voted (i.e. computed) by all systems
    - **Maximum precision**
Assessment FMA-NCI: voting

Max recall: 0.96
Max precision: 1.0
Vote \( \geq 5 \): \( P=0.89, R=0.91 \)
Assessment FMA-SNOMED: voting

Max recall: 0.87
Max precision: 0.99
Vote ≥3: P=0.89, R=0.76
Assessment SNOMED-NCI: voting

Max recall: 0.78
Max precision: 0.99
Vote ≥3: P=0.82, R=0.67
User involvement in ontology alignment

- Due to the limits to the quality of automated alignment algorithms.
- Requires tool support
  - “Good” interfaces and services
  - Number of questions must be limited
  - Systems should make the most of user inputs
  - Systems should consider the profile of the user
- Currently evaluated in the OAEI
  - Simulation of users with different error rates

User involvement in ontology alignment

- Due to the limits to the quality of automated alignment algorithms.
- Requires tool support
  - "Good" interfaces and services
  - Number of questions must be limited
  - Systems should make the most of user inputs
  - Systems should consider the profile of the user
- Currently evaluated in the OAEI
  - Simulation of users with different error rates

User involvement in ontology alignment

- Due to the limits to the quality of automated alignment algorithms.
- Requires tool support
  - “Good” interfaces and services
  - Number of questions must be limited
  - Systems should make the most of user inputs
  - Systems should consider the profile of the user
- Currently evaluated in the OAEI
  - Simulation of users with different error rates

Outline

Preliminaries

Ontology Alignment

Healthcare Applications
1. Lung Cancer Assistant (LCA)

- An **ontology-based system** which provides decision support for lung cancer treatment
- LCA exploits the English Lung Cancer Dataset (LUCADA)
- **LUCADA ontology** represents the semantic layer of the LCA,
  - Required **alignment with SNOMED CT**
    - to facilitate interoperability with NHS systems
    - partially done without too support!
  - We offered **LogMap** alignment system to
    - identify the classes in SNOMED CT related to LUCADA
    - extract a lung cancer-specific module of SNOMED CT

1. Lung Cancer Assistant (LCA)

- An **ontology-based system** which provides decision support for lung cancer treatment
- LCA exploits the English Lung Cancer Dataset (LUCADA)
- **LUCADA ontology** represents the semantic layer of the LCA,
- Required **alignment with SNOMED CT**
  - to facilitate interoperability with NHS systems
  - partially done without too support!
- We offered **LogMap** alignment system to
  - identify the classes in SNOMED CT related to LUCADA
  - extract a lung cancer-specific module of SNOMED CT

---

1. Lung Cancer Assistant (LCA)

- An **ontology-based system** which provides decision support for lung cancer treatment
- LCA exploits the English Lung Cancer Dataset (LUCADA)
- **LUCADA ontology** represents the semantic layer of the LCA,
- Required **alignment with SNOMED CT**
  - to facilitate interoperability with NHS systems
  - partially done without too much support!
- We offered **LogMap** alignment system to
  - identify the classes in SNOMED CT related to LUCADA
  - extract a lung cancer-specific module of SNOMED CT

---

2. EU Project Optique

- Scalable End-user Access to Big Data
- Exposing relational data through ontologies

3. UK BioBank Cardiac cine-MRI Scans

- Semantic enrichment of **free-text annotation** of image quality for **UK BioBank** cardiac cine-MRI scans
- Free text annotations added to a **spread-sheet**
- There is a high lexical and semantic **variability** in the annotations
- Ontologies will enhance the processing of the free-text annotations
- Required **ontology alignment**
  - to facilitate interoperability with NHS systems (SNOMED CT) and other UK BioBank parts
3. UK BioBank Cardiac cine-MRI Scans

- Semantic enrichment of **free-text annotation** of image quality for **UK BioBank** cardiac cine-MRI scans
- Free text annotations added to a **spread-sheet**
- There is a high lexical and semantic **variability** in the annotations
- Ontologies will enhance the processing of the free-text annotations
- Required **ontology alignment**
  - to facilitate interoperability with NHS systems (SNOMED CT) and other UK BioBank parts
3. UK BioBank Cardiac cine-MRI Scans

- Semantic enrichment of free-text annotation of image quality for UK BioBank cardiac cine-MRI scans
- Free text annotations added to a spread-sheet
- There is a high lexical and semantic variability in the annotations
- Ontologies will enhance the processing of the free-text annotations
- Required ontology alignment
  - to facilitate interoperability with NHS systems (SNOMED CT) and other UK BioBank parts

- Ferreira et al. Cardiovascular magnetic resonance artefacts. Journal of Cardiovascular Magnetic Resonance 2013
4. Shared hypothesis testing

- Some diseases may be evidenced across **multiple biological scales** (e.g., cellular, molecular, organic, behavioral)
  - Tests (e.g., cell viability, MRI analysis, gait analysis) may evidence a factor
- **Factors** are linked among each other via a **causal relationship**
- A set of factors and the causality relationship constitute a **hypothesis** of the progression of the disease
- **Different specialists** may work on different subparts of the hypothesis
- Specialists may rely on different **domain ontologies** and use different **modelling/naming conventions**
  - Use **ontology alignment techniques**

---

- EU FP7 MultiScaleHuman project
4. Shared hypothesis testing

- Some diseases may be evidenced across **multiple biological scales** (e.g., cellular, molecular, organic, behavioral)
  - Tests (e.g., cell viability, MRI analysis, gait analysis) may evidence a factor
- **Factors** are linked among each other via a **causal relationship**
- A set of factors and the causality relationship constitute a **hypothesis** of the progression of the disease
- **Different specialists** may work on different subparts of the hypothesis
- Specialists may rely on **different domain ontologies** and use **different modelling/naming conventions**
  - **Use ontology alignment techniques**

---

- EU FP7 MultiScaleHuman project
5. Pistoia alliance ontologies mapping project

- Currently looking for **suitable tool support**
- **Disease and phenotype** domain:
  - Human Phenotype Ontology (HPO) $\leftrightarrow$ Mammalian Phenotype Ontology (MP)
  - Human Disease Ontology (DOID) $\leftrightarrow$ Orphanet and Rare Diseases Ontology (ORDO)
- Motivation: matching human inherited diseases with laboratory studies
- The Pistoia Alliance Ontologies Mapping project co-organises the **OAEI phenotype track**

Pistoia Alliance

6. Semantic annotation of clinical letters (i)

- The output of a visit is known as **clinical letter**, which serves to
  - document the patient’s progress
  - communicate findings among specialists
- Clinical letters are a **key source of knowledge** (specially for rare diseases)
- The letter is typically tagged manually by a specialist responsible for reading and annotating the interesting terms.
  - Common practice in the NHS: Genomics England, OCMR
6. Semantic annotation of clinical letters (i)

- The output of a visit is known as **clinical letter**, which serves to
  - document the patient’s progress
  - communicate findings among specialists
- Clinical letters are a **key source of knowledge** (specially for rare diseases)
- The letter is typically tagged manually by a specialist responsible for reading and annotating the interesting terms.
  - Common practice in the NHS: Genomics England, OCMR
6. Semantic annotation of clinical letters (ii)

- **Manual annotation of clinical letters**
  - Requires exhaustive input from a qualified medical professional
  - Time consuming
  - Leads to the relatively limited use of clinical letters
  - Loss of potentially important information
  - Suitable software support is required to assist the clinician

---

6. Semantic annotation of clinical letters (ii)

- **Manual annotation of clinical letters**
  - Requires exhaustive input from a qualified medical professional
  - Time consuming
  - Leads to the relatively limited use of clinical letters
  - Loss of potentially important information
  - Suitable software support is required to assist the clinician

---

6. Semantic annotation of clinical letters (ii)

- **Manual annotation of clinical letters**
  - Requires exhaustive input from a qualified medical professional
  - Time consuming
  - Leads to the relatively limited use of clinical letters
  - Loss of potentially important information
  - Suitable software support is required to assist the clinician

---

6. Semantic annotation of clinical letters (iii)

- **Special interest on HPO terms**, however...
  - Text annotators may rely on different vocabularies
  - Some domains may require specific vocabulary as well (e.g., head injuries vocabulary)
  - UMLS vocabulary is richer than HPO vocabulary

- **Coordinated vocabularies:**
  - Exploit (and validate) HPO cross-references to UMLS
  - Exploit references of HPO to other (e.g., BioPortal) ontologies
  - Use ontology alignment techniques
6. Semantic annotation of clinical letters (iii)

- **Special interest on HPO terms**, however...
  - Text annotators may rely on different vocabularies
  - Some domains may require specific vocabulary as well (head injuries vocabulary)
  - UMLS vocabulary is richer than HPO vocabulary

- **Coordinated vocabularies:**
  - Exploit (and validate) HPO cross-references to UMLS
  - Exploit references of HPO to other (BioPortal) ontologies
  - **Use ontology alignment techniques**
6. Semantic annotation of clinical letters (iv)

- Coordinated vocabularies:
Questions?

Thank you for your attention

- ernesto.jimenez.ruiz@gmail.com
- ernesto@cs.ox.ac.uk

- Acknowledgements:
  - EU project Optique and the EPSRC project DBOnto.