

**ONTOADR A SEMANTIC
RESOURCE DESCRIBING ADVERSE
DRUG REACTIONS TO SUPPORT
SEARCHING, CODING, AND
INFORMATION RETRIEVAL**

Cedric Bousquet^{1,2}

¹ SSPIM, CHU de Saint-Etienne, France

² INSERM U1142/LIMICS, Paris, France

PHARMACOVIGILANCE

- According to the World Health Organization **Pharmacovigilance** is defined as the “science and activities relating to the detection, assessment, understanding and prevention of adverse effects or any other drug-related problem.”
- According to the European Medicines Agency, a safety **signal** is “information on a new or known adverse event that is potentially caused by a medicine and that warrants further investigation [...] The evaluation of safety signals is part of routine pharmacovigilance and is essential to ensuring that regulatory authorities have the most up-to-date information on a medicine’s benefits and risks.”



MEDICAL DICTIONARY FOR REGULATORY ACTIVITIES (MEDDRA)

- Developed In the late 1990s by the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH)
- Highly specific standardised medical terminology to facilitate sharing of regulatory information internationally for medical products used by humans
- Growing use worldwide by regulatory authorities, pharmaceutical companies, clinical research organisations and health care professionals

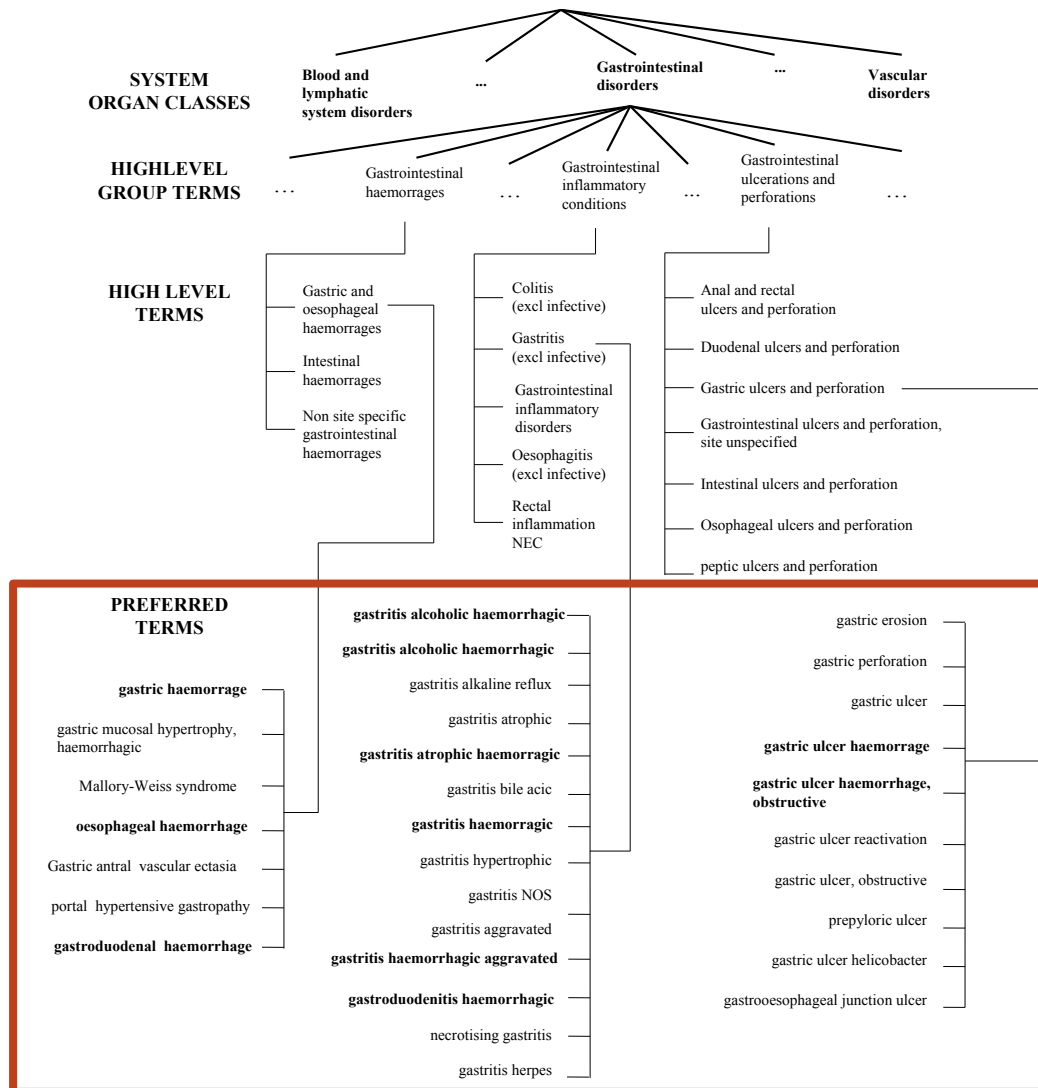


USING MEDDRA GROUPINGS IN SPONTANEOUS REPORTS SYSTEMS TO DETECT SIGNALS

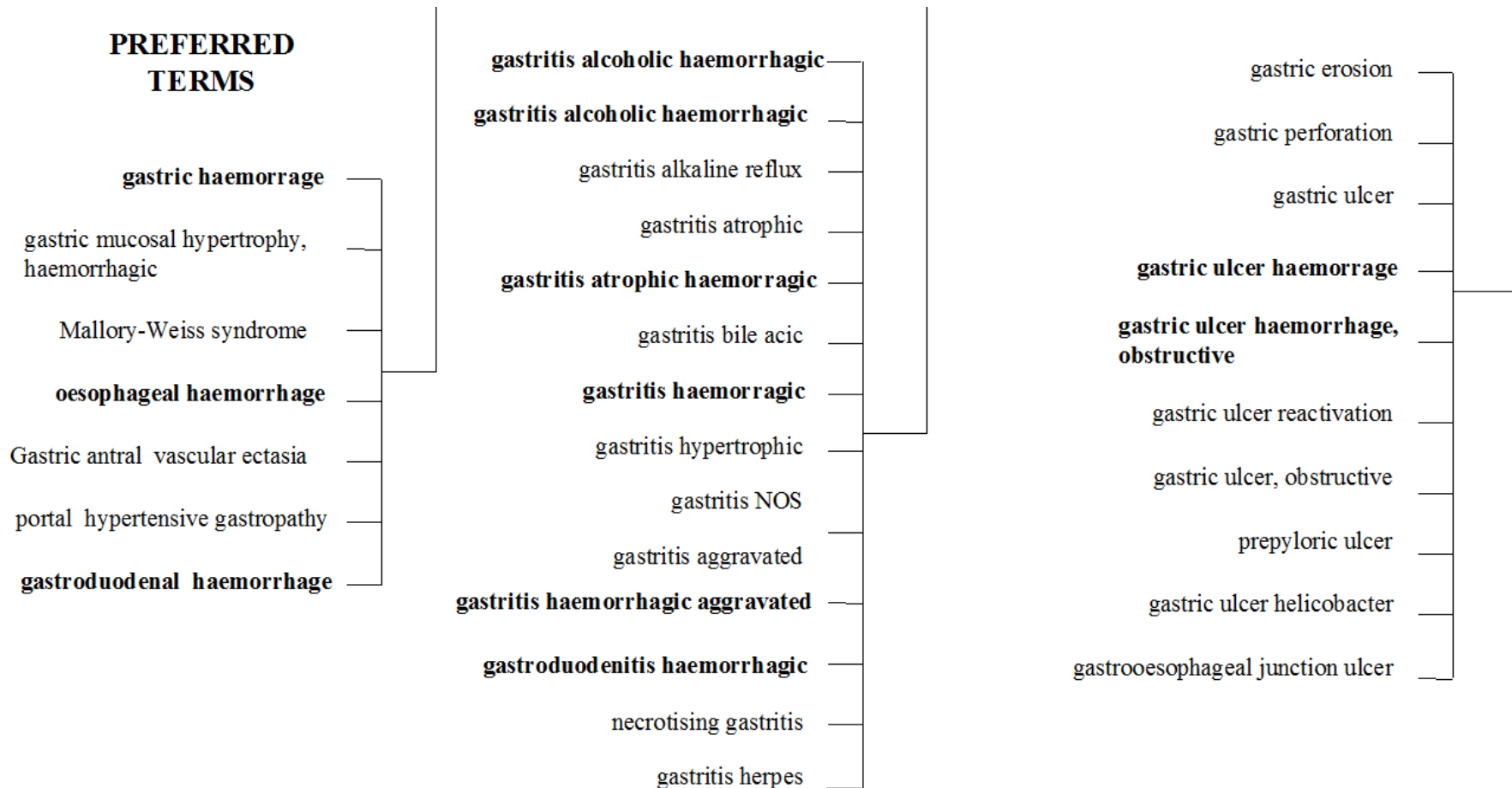
- Spontaneous reporting of ADRs by healthcare professionals to regulatory agencies and/or the Pharma industry
- Manual detection of signals
 - Usually: Review of case reports by an expert
 - Find common patterns between case reports
- Automated signal detection
 - Disproportional number of occurrences of an ADR compared to the expected number
 - Previous experience by LORIA (Yannick Toussaint) in the VigiTermes project
- Classical approach
 - Use the MedDRA terminology by selecting terms belonging to the same branch of the hierarchy



OVERVIEW OF FOUR HIERARCHICAL LEVELS IN MEDDRA



SEARCHING FOR UPPER GASTROINTESTINAL HAEMORRHAGE IN MEDDRA



OBJECTIVE

- The field of pharmacovigilance does not yet benefit from a fully operational ontology to formally represent the MedDRA terms.
- Our objective was to build a semantic resource based on formal description logic to improve MedDRA term retrieval and aid the generation of on-demand custom groupings by appropriately and efficiently selecting terms: **OntoADR**



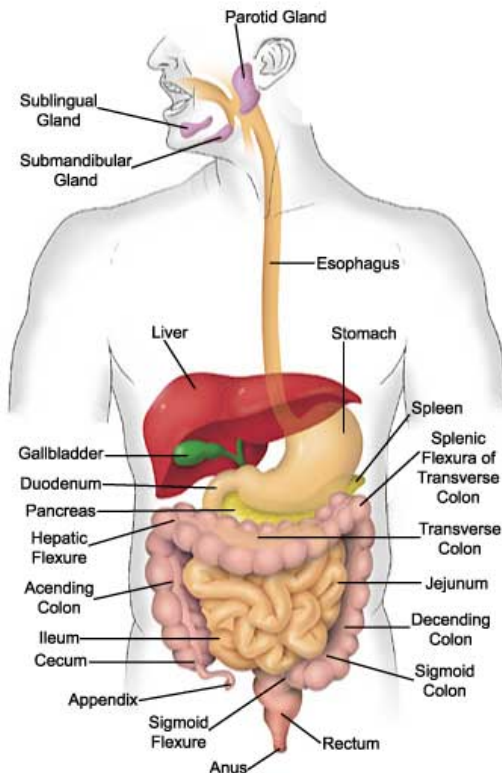
ONTOLOGY OF ADVERSE EVENTS (OAE)

- Takes into account 2300 MedDRA terms
- Does not provide definitions for more than 20,000 MedDRA terms
- OAE only considers ‘pathological bodily process’ descendants (corresponding to both Morphology and Definitional Manifestation in SNOMED-CT).
- In daily routine, some users will use signs, symptoms, or investigations rather than pathological bodily processes.



IMPLICIT VERSUS EXPLICIT KNOWLEDGE REPRESENTATION

- **Implicit knowledge:** the physician knows that gastric ulcer is related to stomach
- **Explicit knowledge:** modelling of relations between MedDRA terms and body structures in the computer program



Upper gastrointestinal tract structure

Lower gastrointestinal tract structure

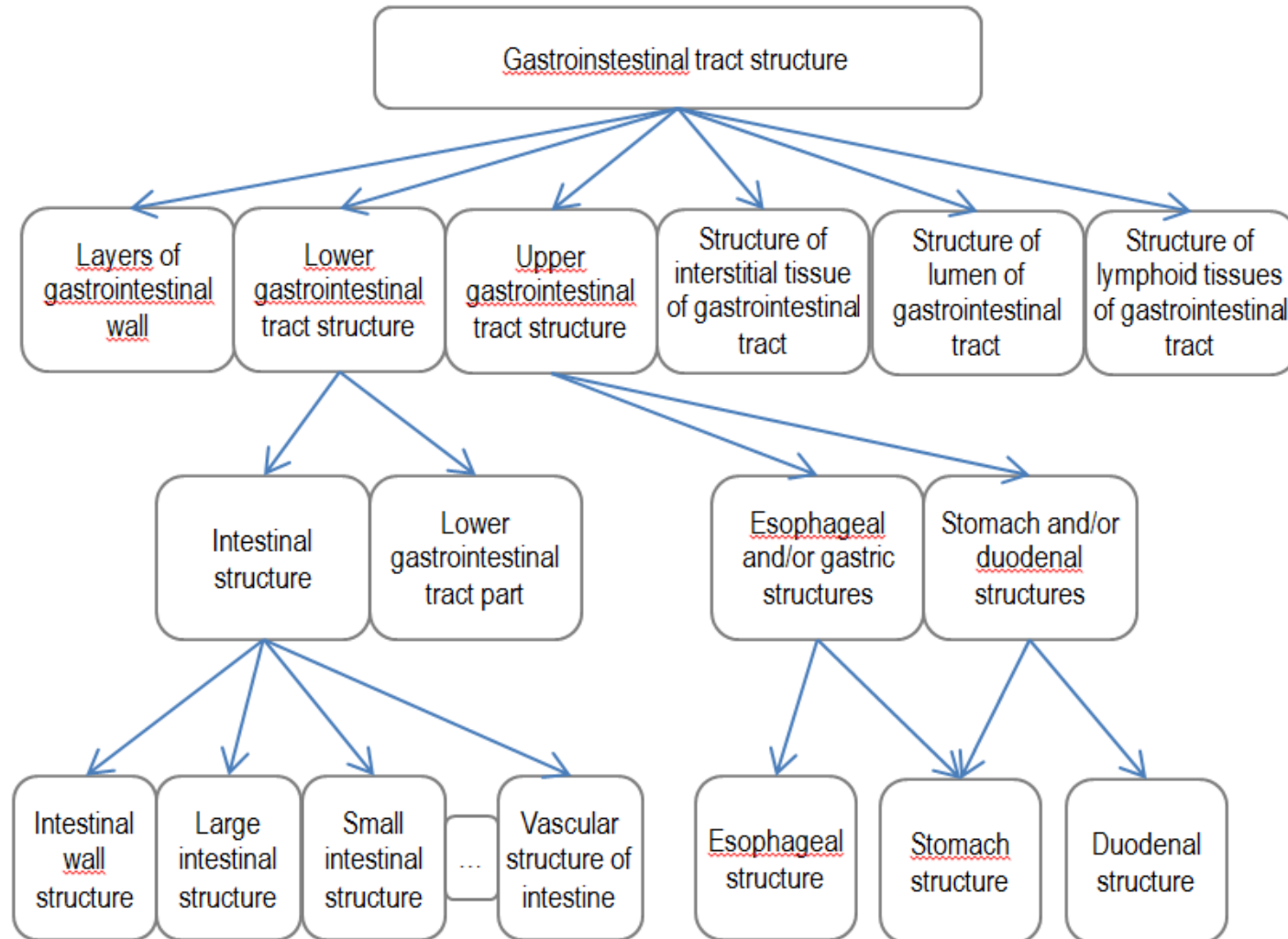


SNOMED CT (SYSTEMATIZED NOMENCLATURE OF MEDICINE -CLINICAL TERMS)

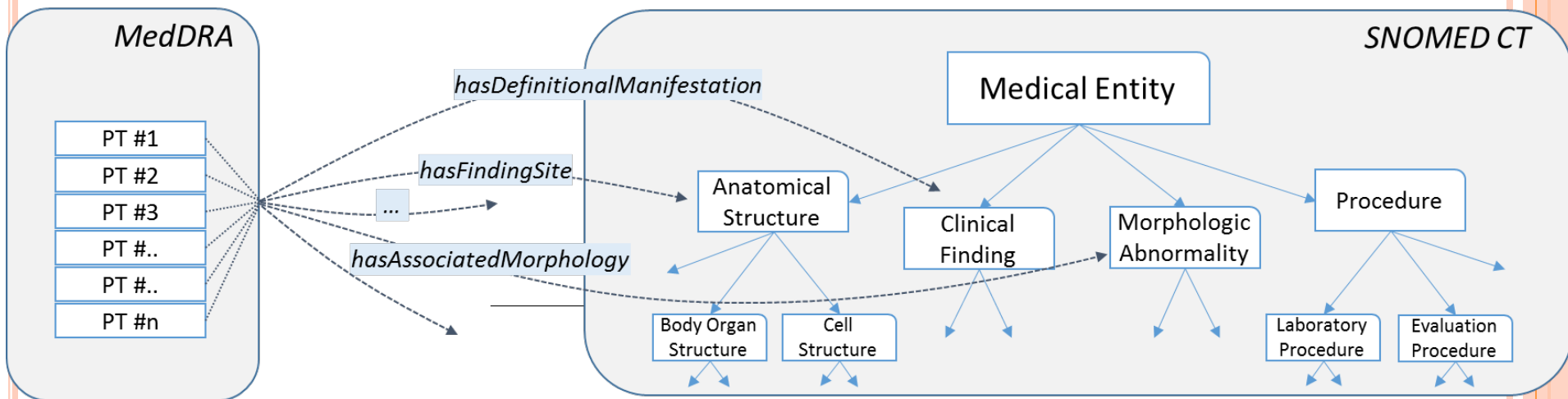
- Most comprehensive and precise clinical health terminology product in the world
- Owned and distributed around the world by The International Health Terminology Standards Development Organisation (IHTSDO)
- Can be characterized as a multilingual thesaurus with an ontological foundation
- Is mapped to other international standards such as MedDRA via the Unified Medical Language System (UMLS)



EXCERPT OF ANATOMICAL LOCATION IN SNOMED-CT



OVERVIEW OF ONTOADR



LIST OF SEMANTIC RELATIONS USED IN ONTOADR

OntoADR findings semantic relations	OntoADR procedures semantic relations
hasFindingSite hasAssociatedMorphology associatedWith ↳ occursAfter ↳ dueTo ↳ hasCausativeAgent hasSeverity hasClinicalCourse hasEpisodicity Interprets hasInterpretation hasPathologicalProcess hasDefinitionalManifestation hasOccurrence	hasFindingMethod hasFindingInformer hasProcedureSite ↳ hasDirectProcedureSite ↳ hasIndirectProcedureSite hasMethod hasComponent hasSpecimen hasFocus hasDirectSubstance hasIntent (interprets)

UNIFIED MEDICAL LANGUAGE SYSTEM (UMLS) METATHESAURUS

- Developed by the NLM (U.S. National Library of Medicine)
- Concepts are organized within a semantic network
- This network maps terms from multiple controlled vocabularies to unique UMLS concepts defined by their concept unique identifier (CUI)
- As UMLS includes both SNOMED-CT and MedDRA concepts, we could extract synonymous mapped concepts (same CUI) from these terminologies



STEP 1: ONE-TO-N MAPPING BETWEEN MEDDRA TERMS AND SNOMED CT CONCEPTS USING MULTIPLE SOURCES

- For each mapping in UMLS, we extracted the definition of the SNOMED-CT concept, and added it to the associated MedDRA term
 - When a MedDRA term was associated with several SNOMED terms (in case of 1-to-n mappings), all properties were extracted and added to the MedDRA term definition
- Also used other mapping resources such as Nadkarni & Darer's propositions of mapping between MedDRA and SNOMED CT

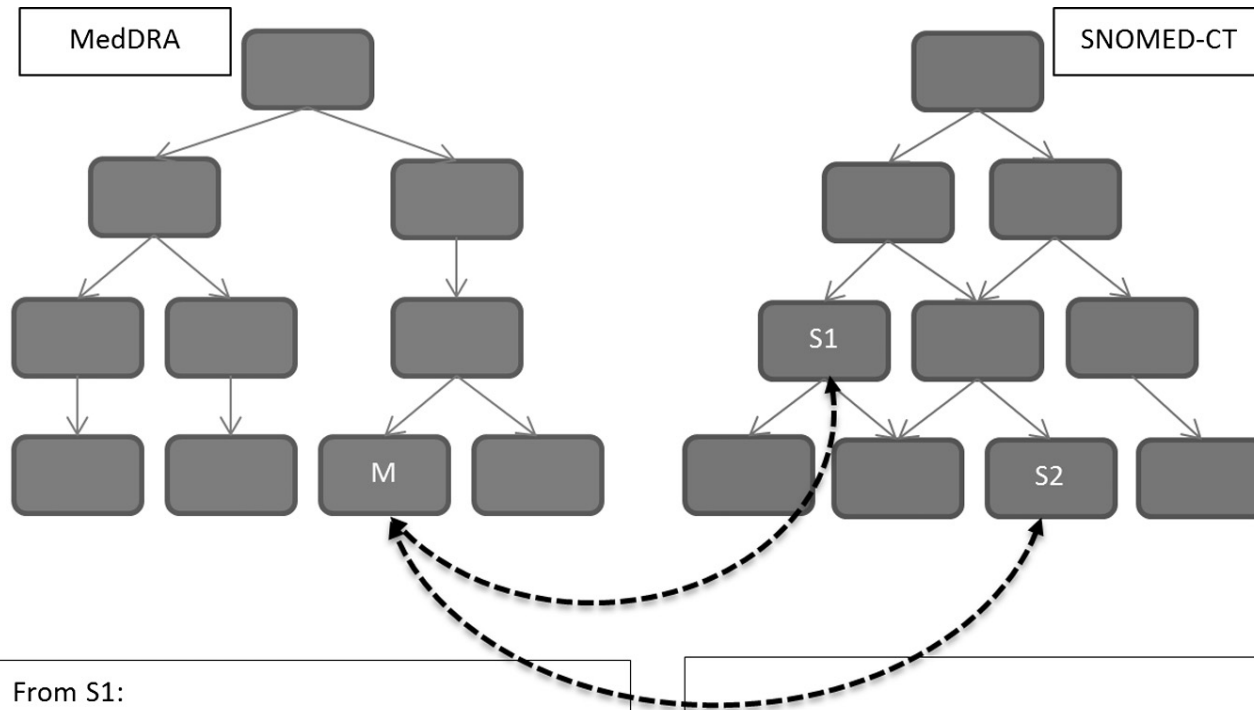


STEP 2: MERGING OF SEMANTIC DEFINITIONS USING SEMI-AUTOMATIC METHODS

- We implemented a simple algorithm for automatic creation of properties from the MedDRA label
 - E.g., when the algorithm detects a given string S_x (hemorrhage, perforation, etc.) in a MedDRA label, it automatically adds a corresponding P_x property (manifestation, morphology, etc.) to the definition
- Combining MedDRA and SNOMED-CT related information introduced unnecessary relations
 - We removed duplicated relationships resulting from multiple mappings between MedDRA and SNOMED-CT, especially inferred relationships



ILLUSTRATION OF FILTERING/INFERRING PROCESS IN ONTOADR.



From S1:
M hasFindingSite 'lower limb structure'
M hasMorphology 'fracture'
~~M dueTo 'traumatic injury'~~

S1 U S2:
M hasFindingSite 'bone structure of tibia'
M hasMorphology 'fracture'
M dueTo 'traumatic injury'



STEP 3: IMPLEMENTATION OF FORMAL DEFINITIONS USING DATABASE REPRESENTATION

- We used EL++ description logic in previous version of OntoADR, but were limited by the impossibility of performing negation and disjunction queries
 - For example, users will remove all congenital diseases or terms related to pathogenic agents when searching for ADRs
- Two kinds of definitions
 - In intension, by a description (specification of a number of predicates) that defines the set, e.g. hasFindingSite 'KidneyStructure'
 - In extension (by naming or designating each individual which is part of it), (e.g. {renal failure, pyelonephritis, etc.})
- We decided to work on the extension (set of MedDRA terms) rather than on the intension (description logic query).
 - Computation with the DL operators, negation and disjunction can then be replaced by set operators, difference, and union



STEP 4: MANUAL CURATION BY A PHARMACOVIGILANCE EXPERT

- Performing curation on the OWL version of OntoADR was difficult because there was no dedicated software for curation and maintenance of ontologies
 - We developed the Ci4SeR tool for this task
- We performed manual curation of approximately 2000 MedDRA terms in 12 months.
 - We focused on high value-added terms for pharmacovigilance selected on the basis of a ranked list of 23 first importance adverse drug events proposed by Trifirò et al.
- We chose not to limit our definitions to necessary and sufficient conditions but rather enlarge them to potentially useful information on the related clinical findings.
 - For example, one may expect to observe increased troponin in most patients presenting myocardial infarct but this remains empirical knowledge



MODELLING IN THE CI4SER TOOL: EXAMPLE OF GASTRODUODENITIS HAEMORRHAGIC

10048712 Gastroduodenitis haemorrhagic	
Semantic definition	
HASFINDINGSITE	Stomach structure [body structure]
	Duodenal structure [body structure]
HASASSOCIATEDMORPHOLOGY	Inflammation [morphologic abnormality]
	Hemorrhage [morphologic abnormality]
ASSOCIATEDWITH	Complication [disorder]
OCCURSAFTER	Gastrointestinal ulcer [disorder]
DUE TO	Portal hypertension [disorder]
	Virus present [finding]
	Helicobacter-associated pyloric ulcer [disorder]
	Non-steroidal anti-inflammatory drug (NSAID)-associated gastropathy [disorder]
INTERPRETS	Endoscopy [procedure] & Abnormal presence of [qualifier value]
	Helicobacter pylori culture [procedure] & Positive [qualifier value]
HASDEFINITIONALMANIFESTATION	Nausea and vomiting [disorder]
	Gastroduodenitis [disorder]
	Vomit: blood present [finding]
	Melena [disorder]
	Epigastric pain [finding]
	Hematemesis [disorder]

USE CASE: UPPER GASTROINTESTINAL BLEEDING (UGIB)

- We developed two queries in order to retrieve candidate terms for UGIB
- **Query 1** was designed to identify any hemorrhage observable in the upper gastrointestinal tract

*hasFindingSite 'Upper gastrointestinal tract structure'
AND hasAssociatedMorphology 'Hemorrhage'*

- **Query 2** is targeting the actual manifestation of an UGIB, taking the risk to be broader than necessary

*Query 1
OR
hasDefinitionalManifestation ('Melena' OR 'Hematemesis')*



RESULTS

$$\text{Query 1 Recall} = \frac{\text{relevant terms} \cap \text{retrieved terms}}{\text{total relevant terms}} = 71.0\%$$

$$\text{Query 1 Precision} = \frac{\text{relevant terms} \cap \text{retrieved terms}}{\text{retrieved terms}} = 81.4\%$$

$$\text{Query 2 Recall} = \frac{\text{relevant terms} \cap \text{retrieved terms}}{\text{total relevant terms}} = 96.7\%$$

$$\text{Query 2 Precision} = \frac{\text{relevant terms} \cap \text{retrieved terms}}{\text{retrieved terms}} = 77.0\%$$



LIMITS

- There may be several ways to represent comparable information due to the high potential compositionality of the SMOMED-CT terms
 - For example, “hasAssociatedManifestation Peripheral Demyelinating Neuropathy” is equivalent to “hasFindingSite Peripheral Nerve Structure” and “hasMorphology Demyelination”
 - Similarly, “hasMorphology Acute Inflammation” is equivalent to “hasMorphology Inflammation” and “hasClinicalCourse Acute”



PERSPECTIVES

- Currently available tools dedicated to MedDRA term selection, such as the MedDRA Browser, only implement string search (searching keywords in the label of terms) and hierarchical browsing
 - A complementary approach that uses a formal semantic representation of MedDRA could substantially decrease the time needed for term selection and improve the precision of the terms used to describe ADRs
- We are developing a graphical user interface that allows the user to query the OntoADR resource using a form-based interface
 - To be adopted by pharmacovigilance staff, the proposed approach should be implemented in a commercial tool



ACKNOWLEDGMENTS

- The research leading to these results was conducted as part of the PROTECT consortium (Pharmacoepidemiological Research on Outcomes of Therapeutics by a European ConsorTium, www.imi-protect.eu) which is a public-private partnership coordinated by the European Medicines Agency.
- The PROTECT project has received support from the Innovative Medicine Initiative Joint Undertaking (www.imi.europa.eu) under Grant Agreement n°115004, resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013) and EFPIA companies' in kind contribution.

