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."

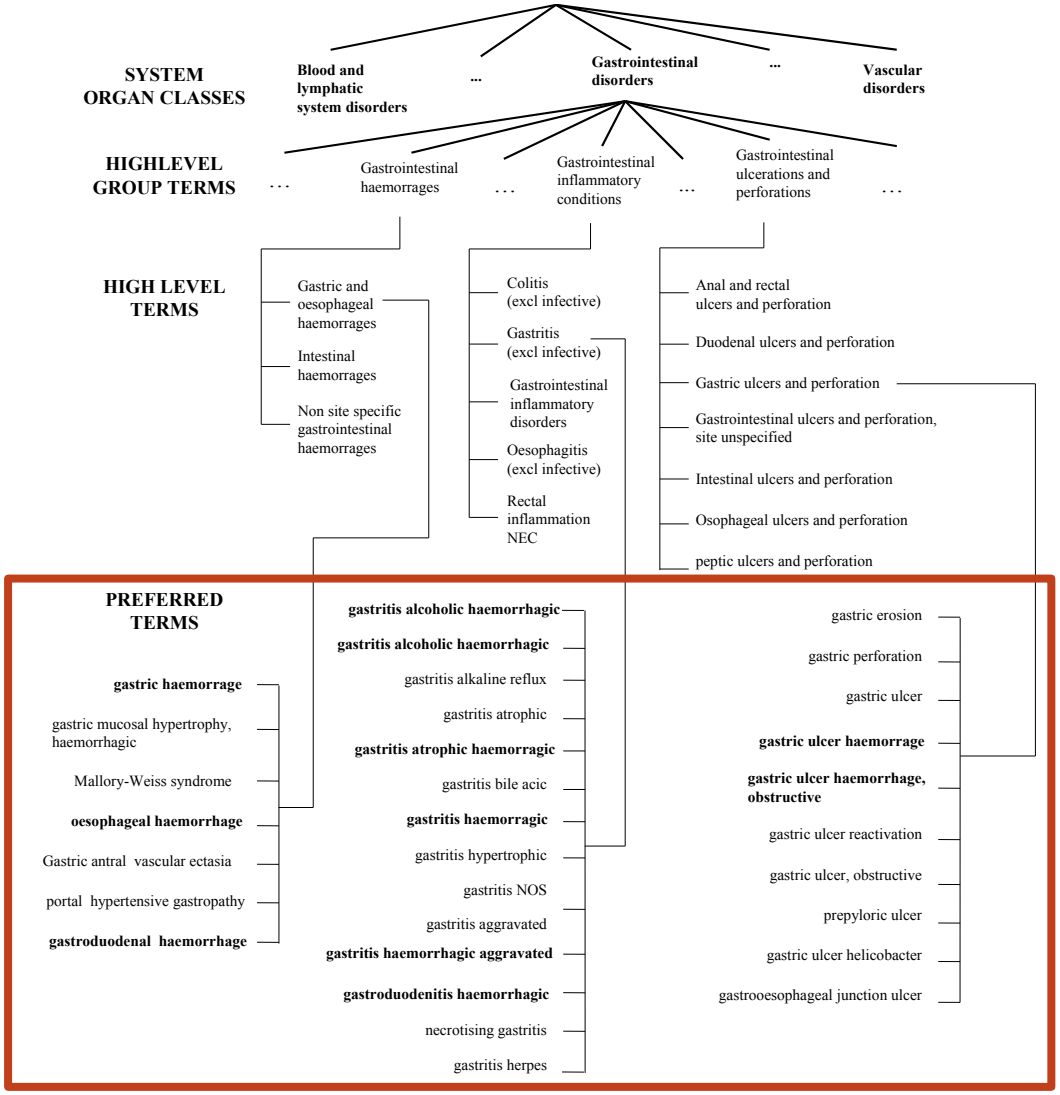
Medical dictionary for Regulatory Activities (MedDRA

- 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

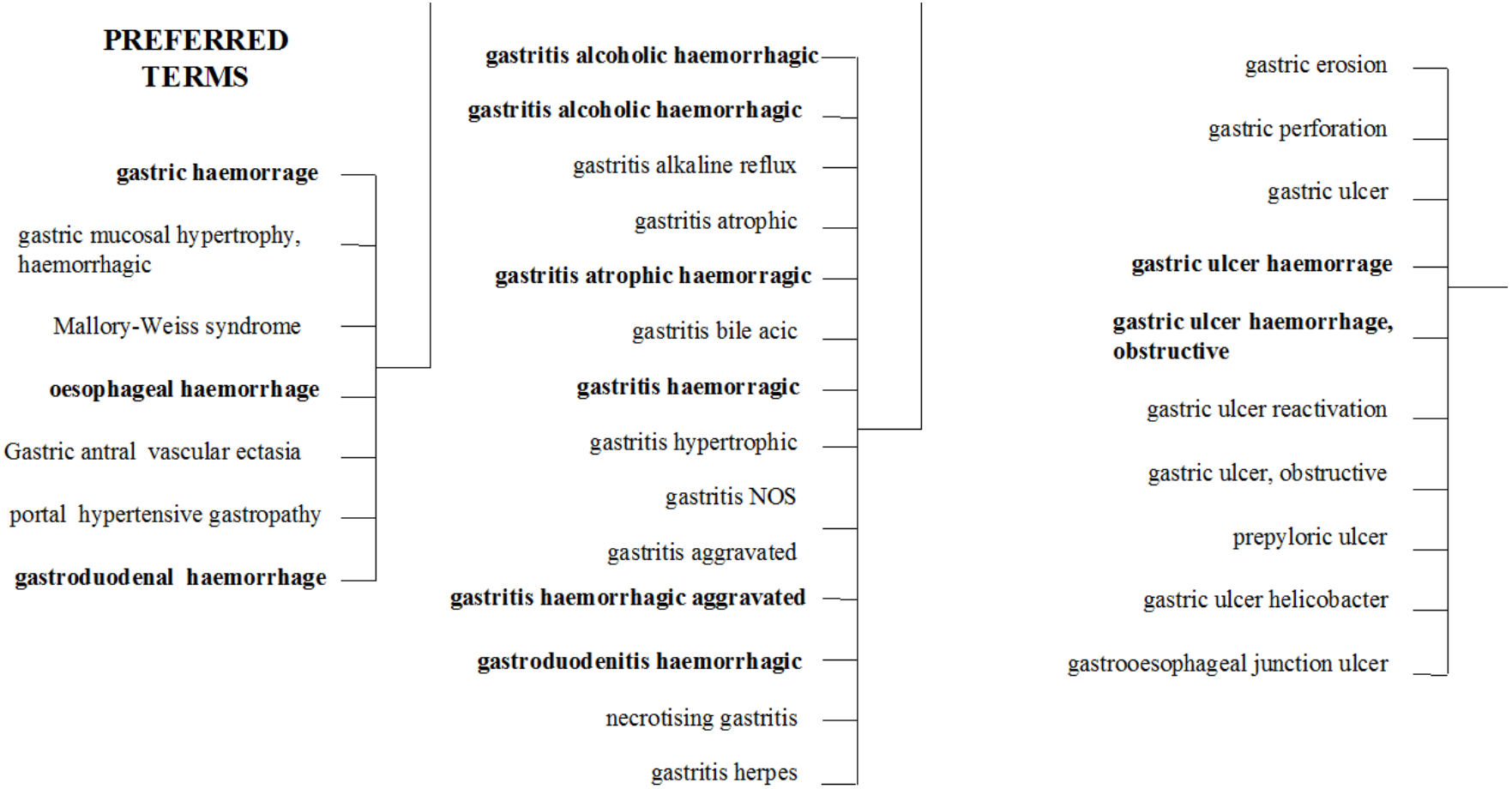
Using MedDRA groupings in spontaneous reports systems to detect signals

- 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 occurences 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



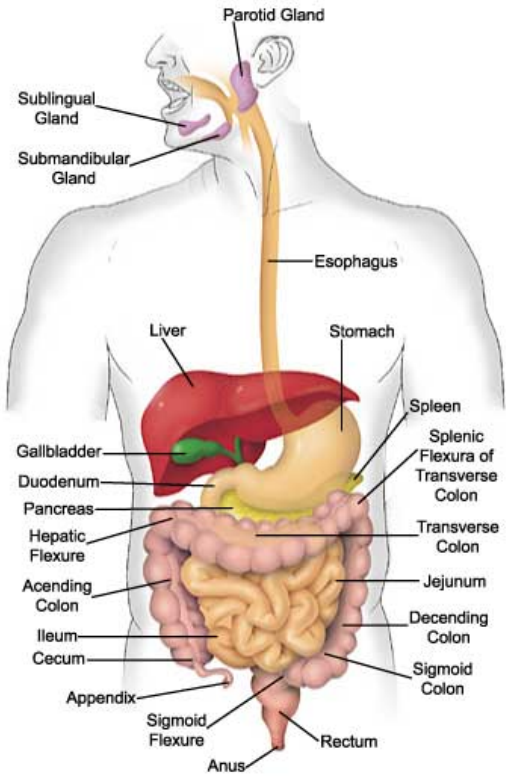
Objective

- 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



Upper gastroinstestinal tract structure

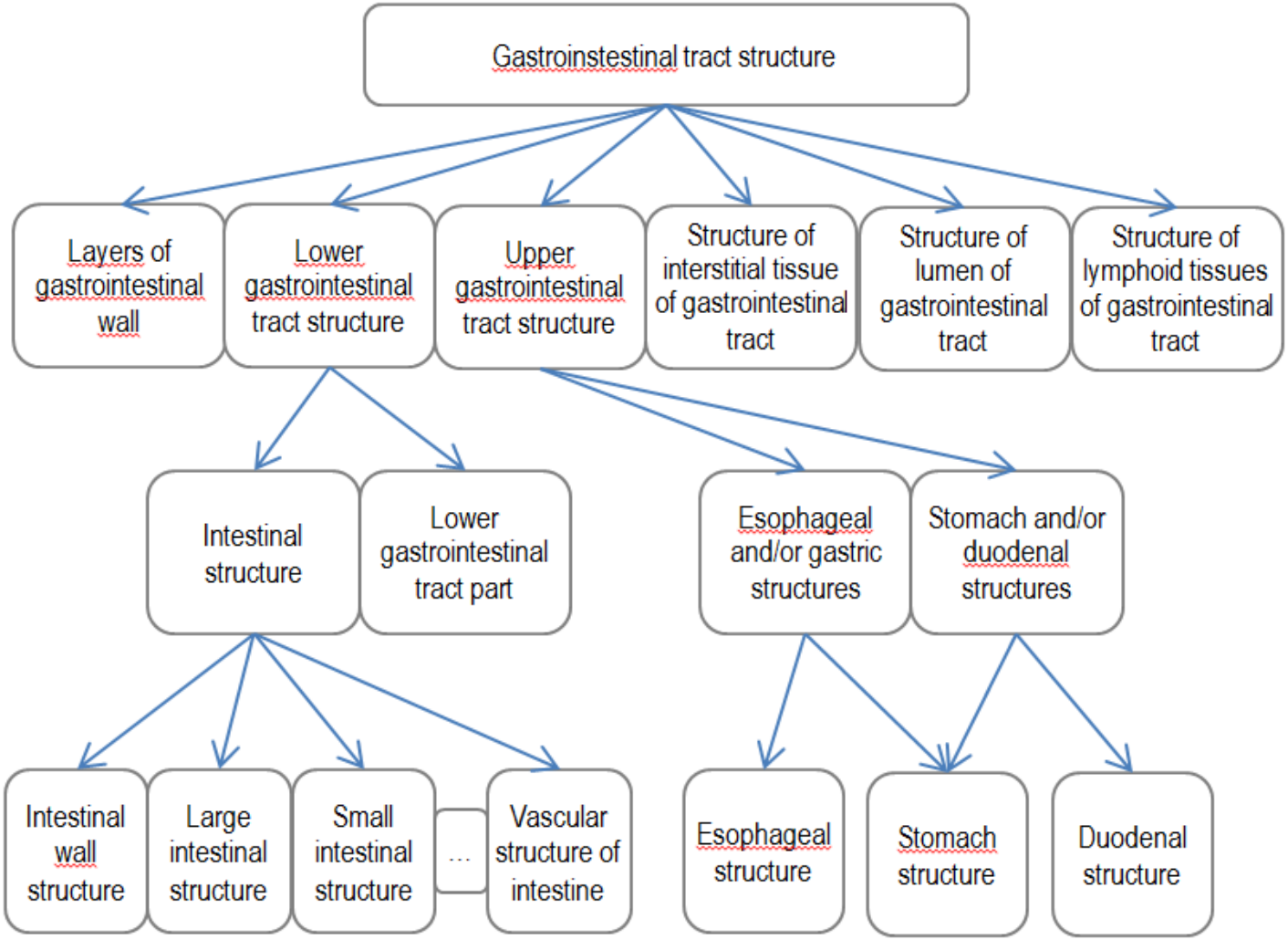
Lower gastroinstestinal tract structure

- **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

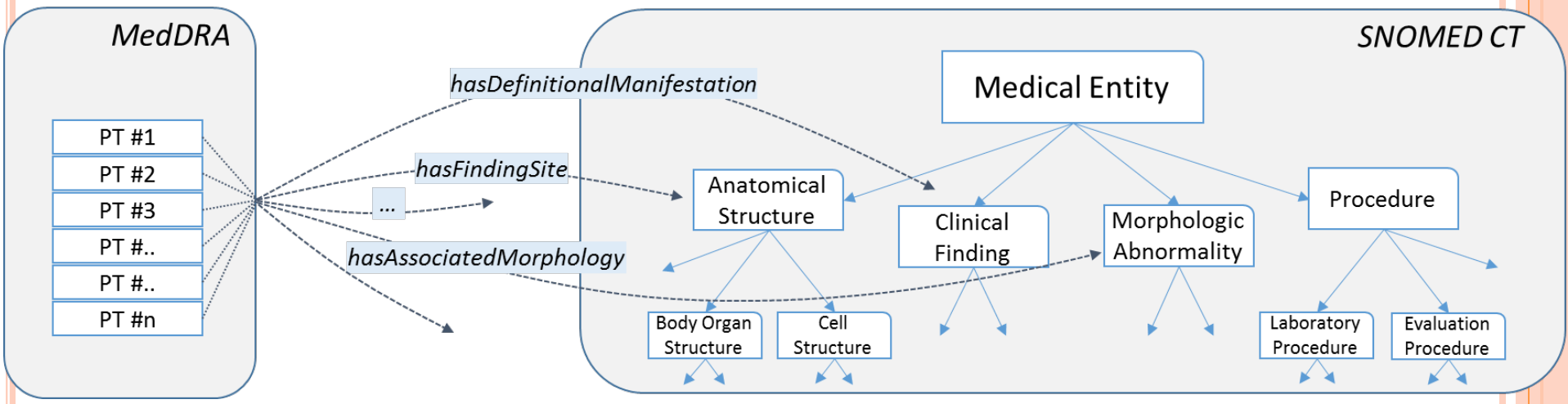
SNOMED CT (Systematized Nomenclature of Medicine -Clinical Terms)

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

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

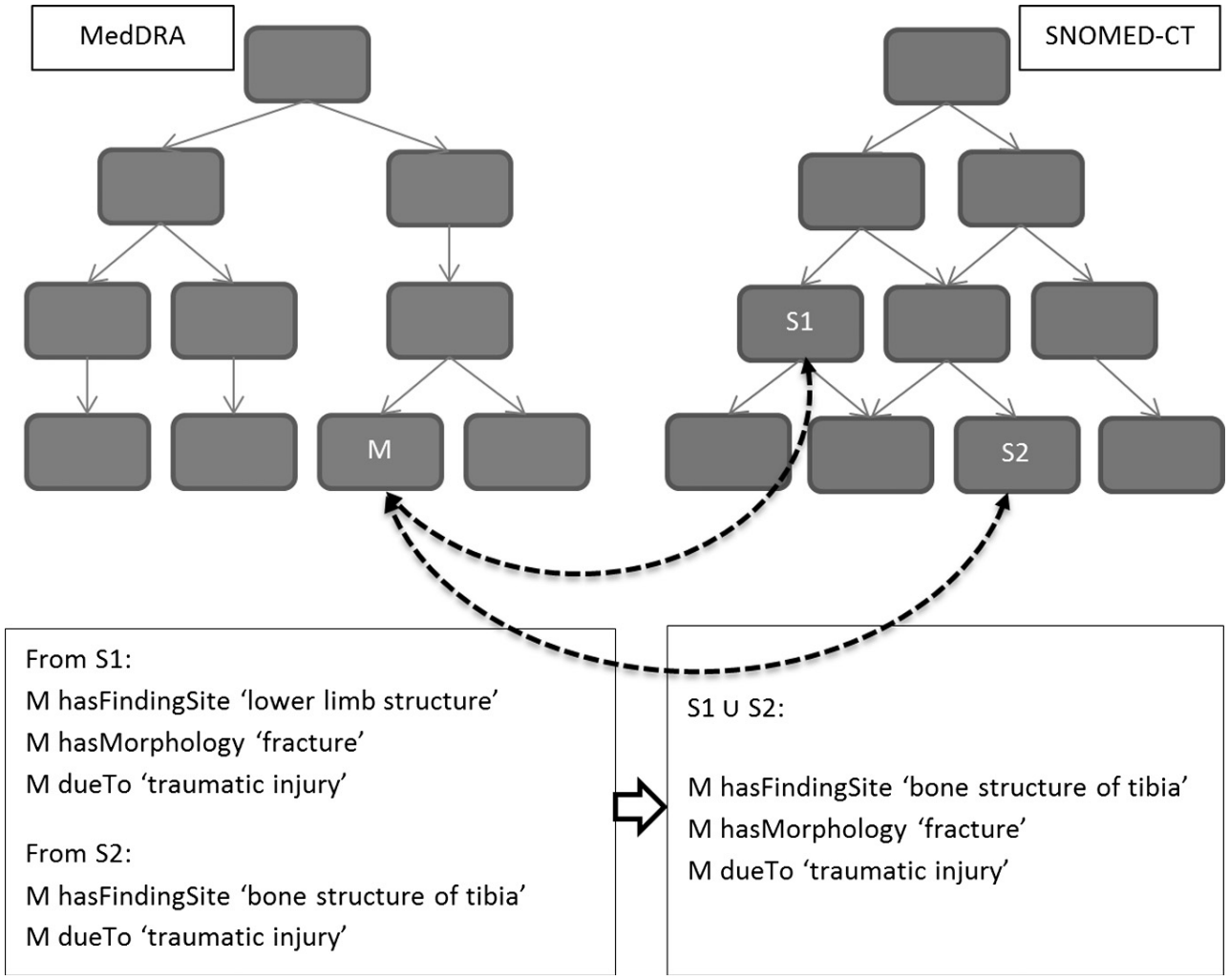
Step 1: One-to-n mapping between MedDRA terms and SNOMED CT concepts using multiple sources

- For each mapping inUMLS, 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 Sx (hemorrhage, perforation, etc.) in a MedDRA label, it automatically adds a corresponding Px 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.



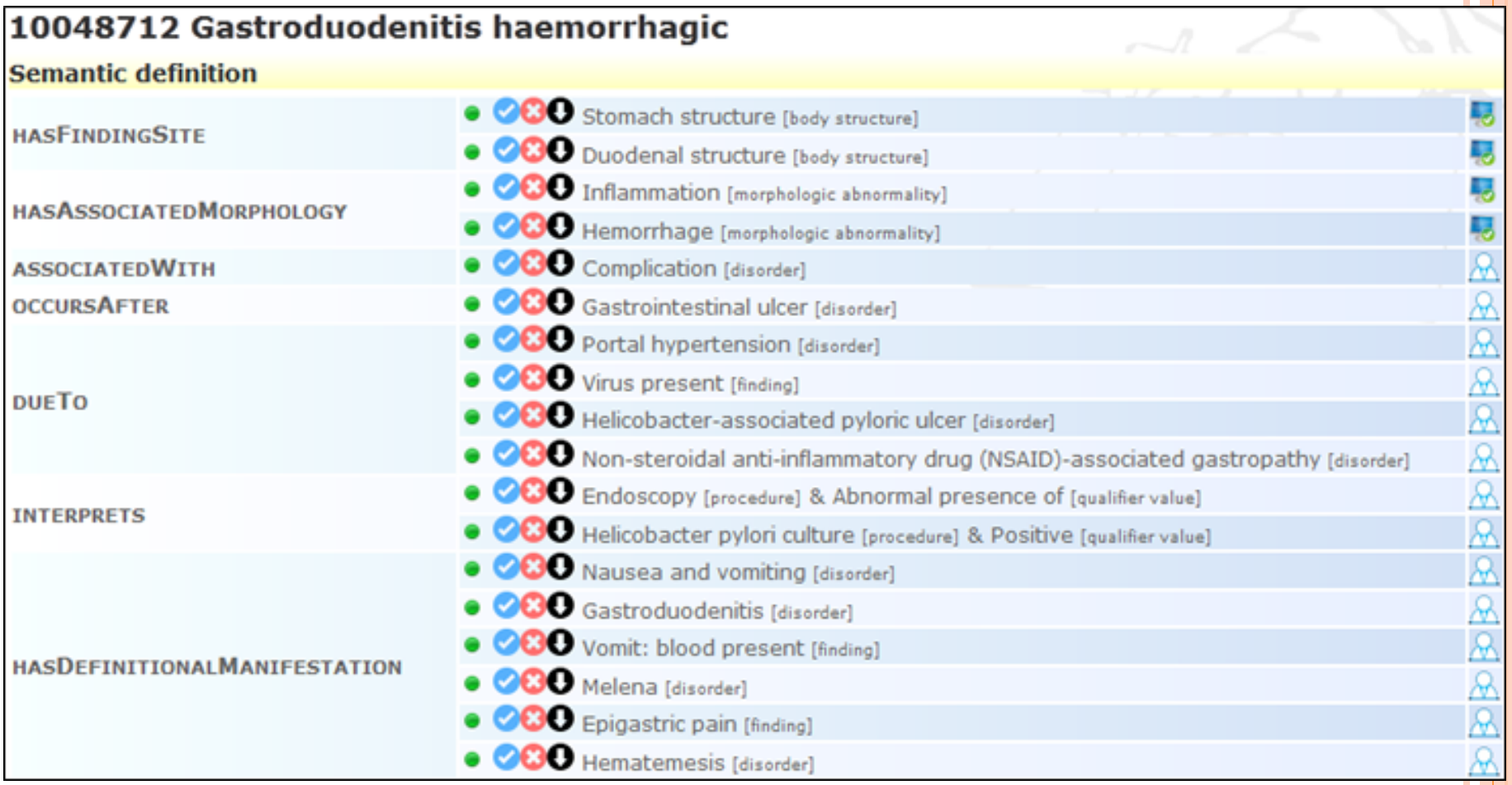
Step 3: Implementation of formal definitions using database representation

- 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



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

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

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

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

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

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”
  - Similarily, “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.