Australian National University

COMP3430 / COMP8430
Data wrangling

Lecture 25: Ontology matching
(Lecturer: Peter Christen)

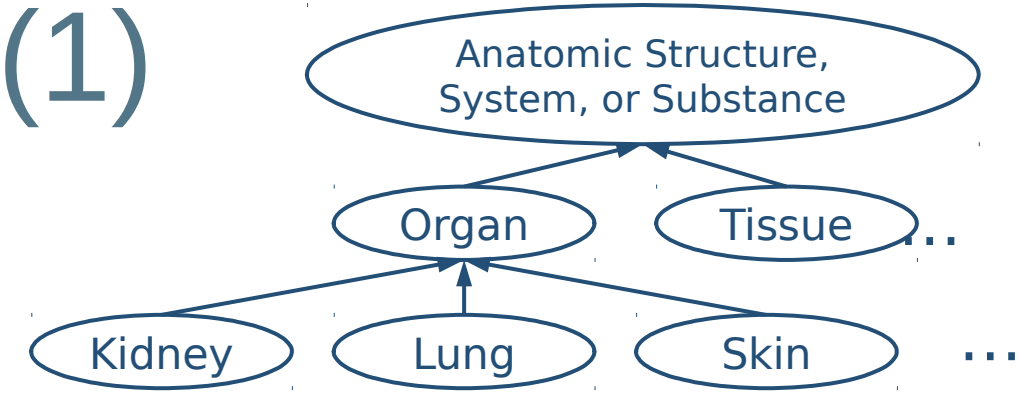
Based on slides by Anika Gross and Michael Hartung
(University of Leipzig)

Lecture outline

- What are ontologies
- Ontology annotations and mappings
- Ontology evolution and trend discovery

What are ontologies? (1)

- Structured representations of knowledge
- Very large ontologies in many domains, for example in the biomedical domain



Anatomy

MGI

Uber Anatomy Ontology

FMA
University of Washington School of Medicine

Medicine

SNOMED CT

NCIthesaurus

Chemistry

ChEBI

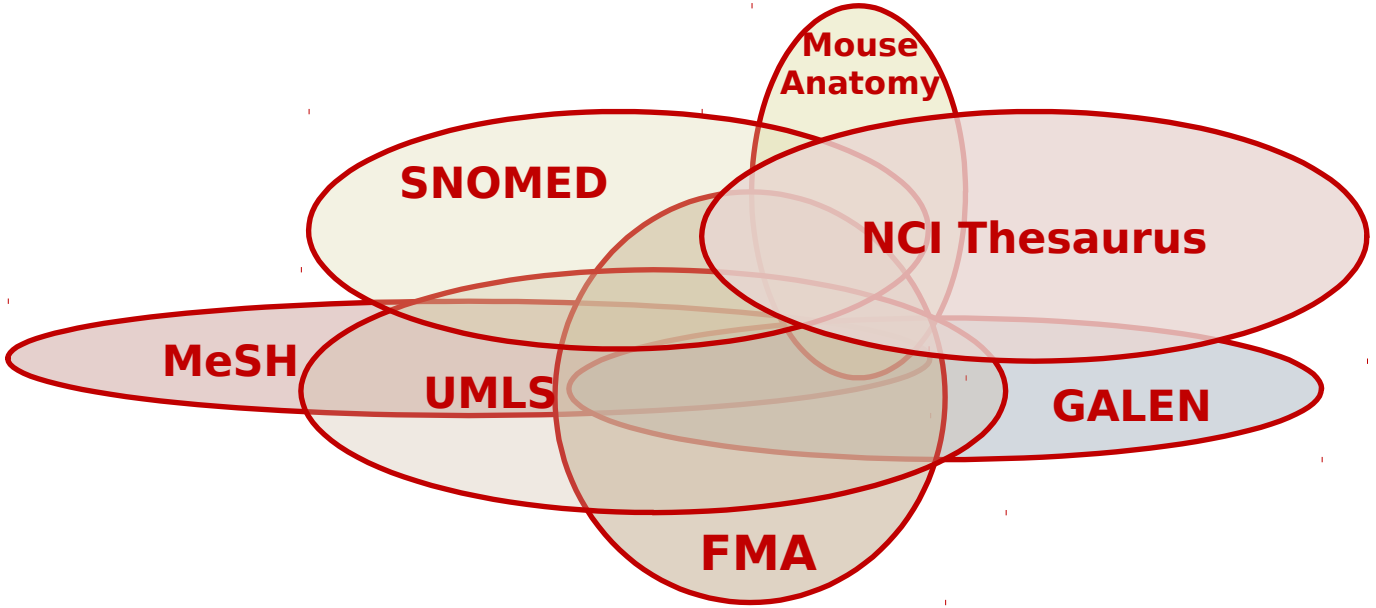
Molecular biology

GO
the Gene Ontology

For examples see: https://en.wikipedia.org/wiki/Ontology_(information_science)#Published_examples

What are ontologies? (2)

- Often multiple interrelated ontologies in a domain (e.g. anatomy)
- We need to identify overlapping information between ontologies
- Create mappings between ontologies



Ontology based annotations

- Standardised semantic descriptions of object properties

Genes, proteins, …

Electronic health records

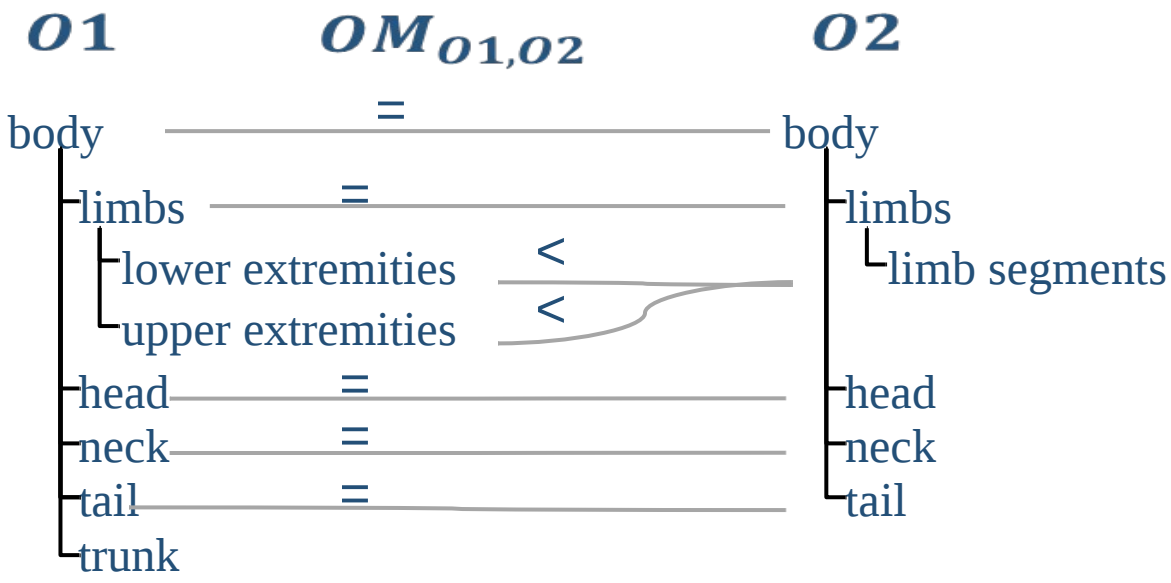
Publications

- Applications:
  - Semantic search, navigation, etc.
  - Functional analysis: Identification of significant characteristics of specific gene/proteins groups

Ensembl ID	GO ID
ENSP00000344151	GO:0015808 (L-alanine transport)
ENSP00000230480	GO:0005615 (extracellular space)
ENSP00000352999	GO:0006915 (apoptosis)

Ontology mappings and alignments

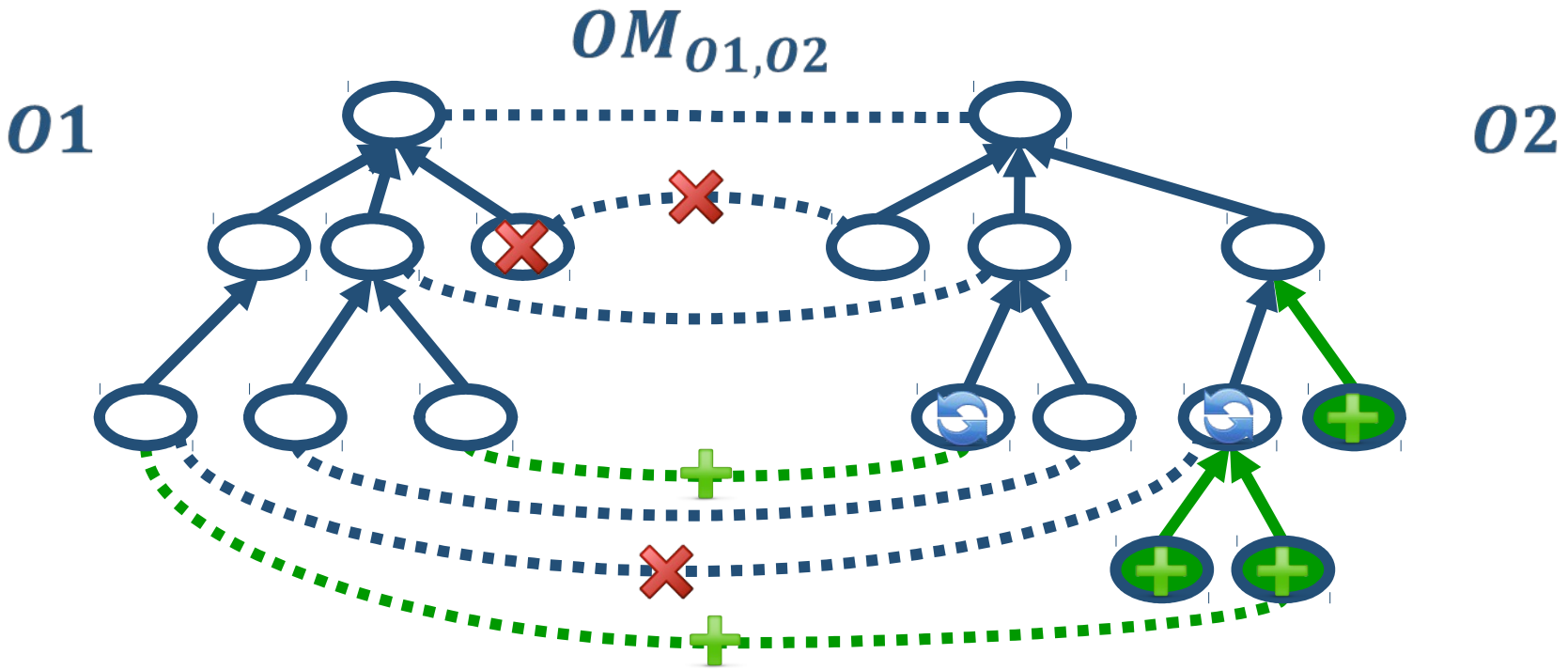
- Overlapping ontologies allow the creation of mappings/alignments
- Useful for data integration and analysis across sources
- **Ontology mapping:** Set of semantic correspondences between concepts of different ontologies
- Manual identification or (semi-) automatic matching approaches
- Use of mappings:
  - Ontology merging (such as creation of an integrated cross-species anatomy ontology)
  - Knowledge transfer (for example experiments for different species)
  - Ontology curation (find missing ontology annotations)



…

Evolution of ontology-based mappings

- Ontologies are not static!
- Research, new knowledge → Continuous changes
- Release of new versions
- Ontology changes:
  - additions
  - deletions
  - updates



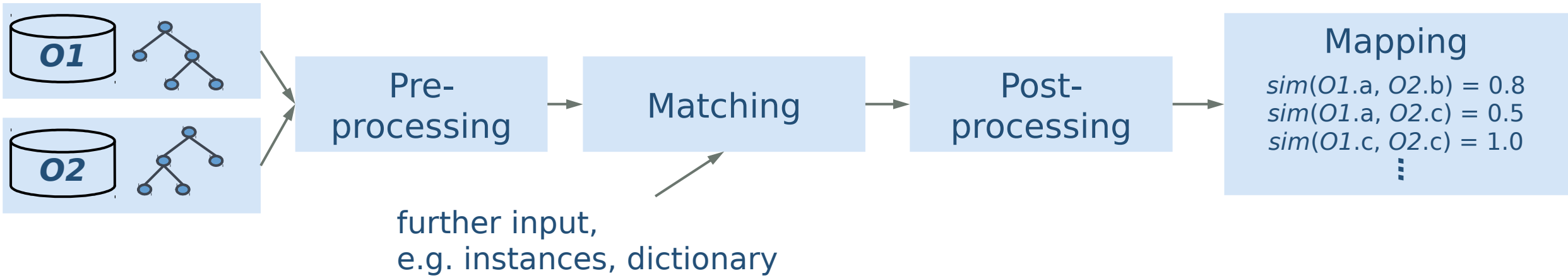
Reuse of existing mappings

- Create new ontology mappings
  "Indirect" matching: combine existing mappings to create new mappings between so far unconnected sources

- Create up-to-date ontology mappings
  Migration of outdated mappings to currently valid ontology versions

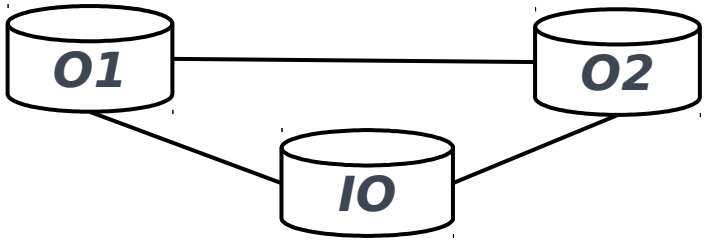
Ontology matching workflow

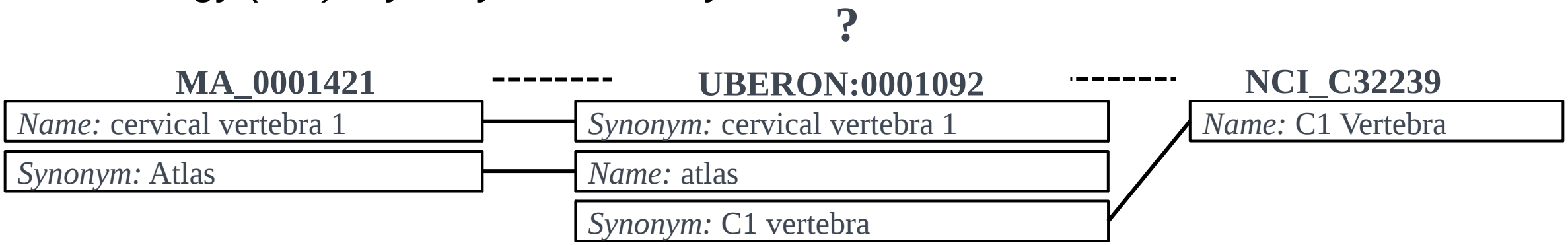
- Manual creation of mappings between very large ontologies is too labor-intensive
- Semi-automatic generation of semantic correspondences: linguistic, structural, instance-based matching techniques
  (see lecture on schema matching, lecture 11)



Mapping composition

- Indirect composition-based matching
- Via intermediate ontology (*IO)* or hub ontology *(HO)*, synonym dictionary, etc.

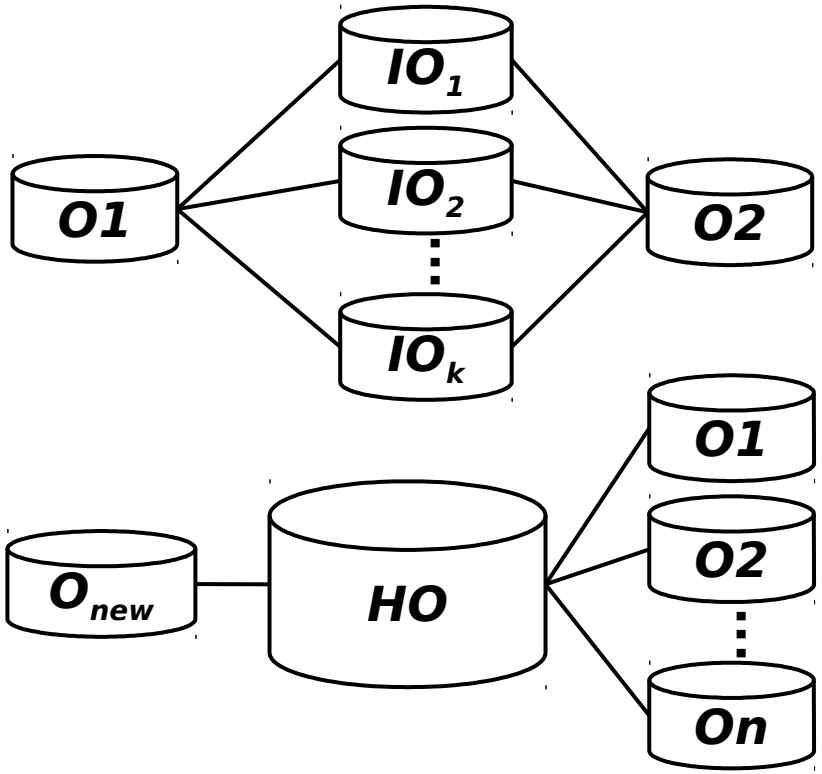




- Find new correspondences via composition
- Reuse existing mappings to increase match quality and save computation time

Indirect matching

- Use mappings to intermediate ontologies IO_1, …, IO_k to indirectly match *O1* and *O2*
- Reduce matching effort by reusing mappings to *IO* → Very fast composition



→ *IO* should have a significant overlap with *O1* and *O2*

→ IO_1, …, IO_k may complement each other

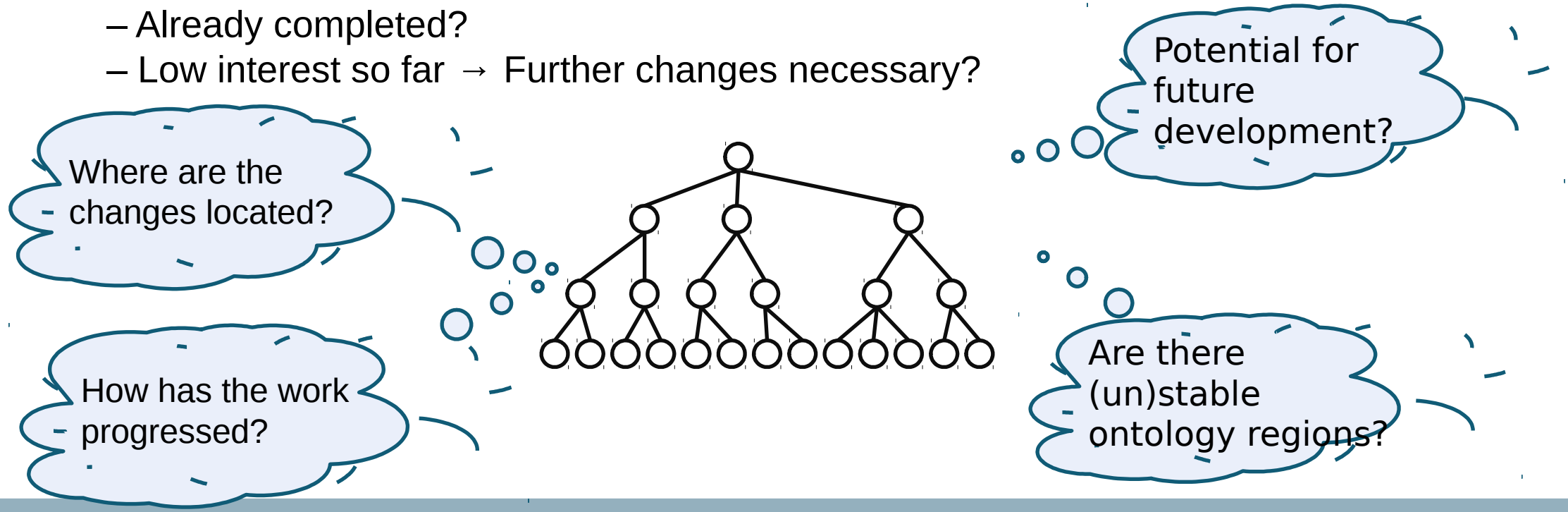
→ Centralized hub *HO*

→ Many mappings to other ontologies

→ O_{new} aligned with any *Oi* via *HO*

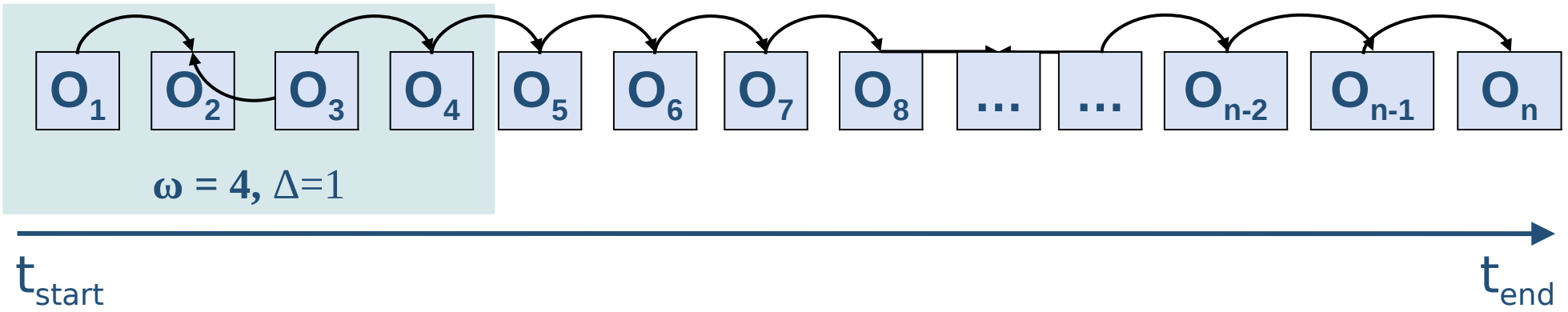
Ontology evolution

- Unstable ontology regions
  - Many modifications → Focus of recent development
  - Impact of changes on ontology-based algorithms or applications → Redo analyses?
- Stable ontology regions
  - Already completed?
  - Low interest so far → Further changes necessary?



Trend discovery

- Trend discovery based on sliding windows
- Monitor region changes over long periods of time
  - Ontology *O*, ontology region of interest *OR*
  - Time interval (t_{start}, t_{end})
  - Sliding window of size ω
  - Step width Δ
- Call region discovery algorithm within ω
  - Collect change intensities for region of interest over time



Outlook and research directions

- Ontologies are becoming increasingly important
  - In the life sciences (for example, conference series *Data Integration in the Life Sciences – DILS*)
  - Knowledge-bases and the semantic Web
  - Internet of Things
- Various research areas
  - Learning to match and map ontologies (semi-) automatically
  - Mapping of dynamic ontologies
  - Parallel algorithms for large-scale ontology matching, mapping and evolution