

# HOW DO ONTOLOGY MAPPINGS CHANGE IN THE LIFE SCIENCES?

ANIKA GROSS, MICHAEL HARTUNG,  
ANDREAS THOR, ERHARD RAHM

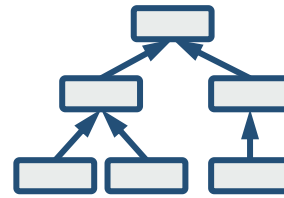
UNIVERSITÄT LEIPZIG

POSTER PRESENTATION, 28<sup>TH</sup> JUNE 2012, DILS, COLLEGE PARK, MARYLAND

# ONTOLOGIES AND ONTOLOGY MAPPINGS

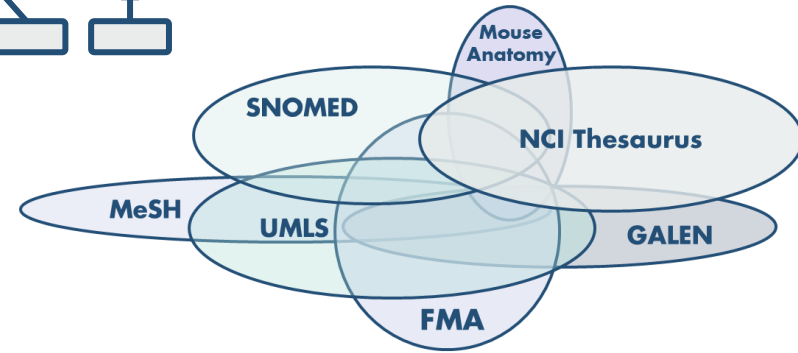
## Ontologies

- Knowledge representation
- Multiple interrelated ontologies in a domain



## Ontology mapping

- Set of semantic correspondences between concepts of different ontologies
- Manual identification or (semi-) automatic matching approaches

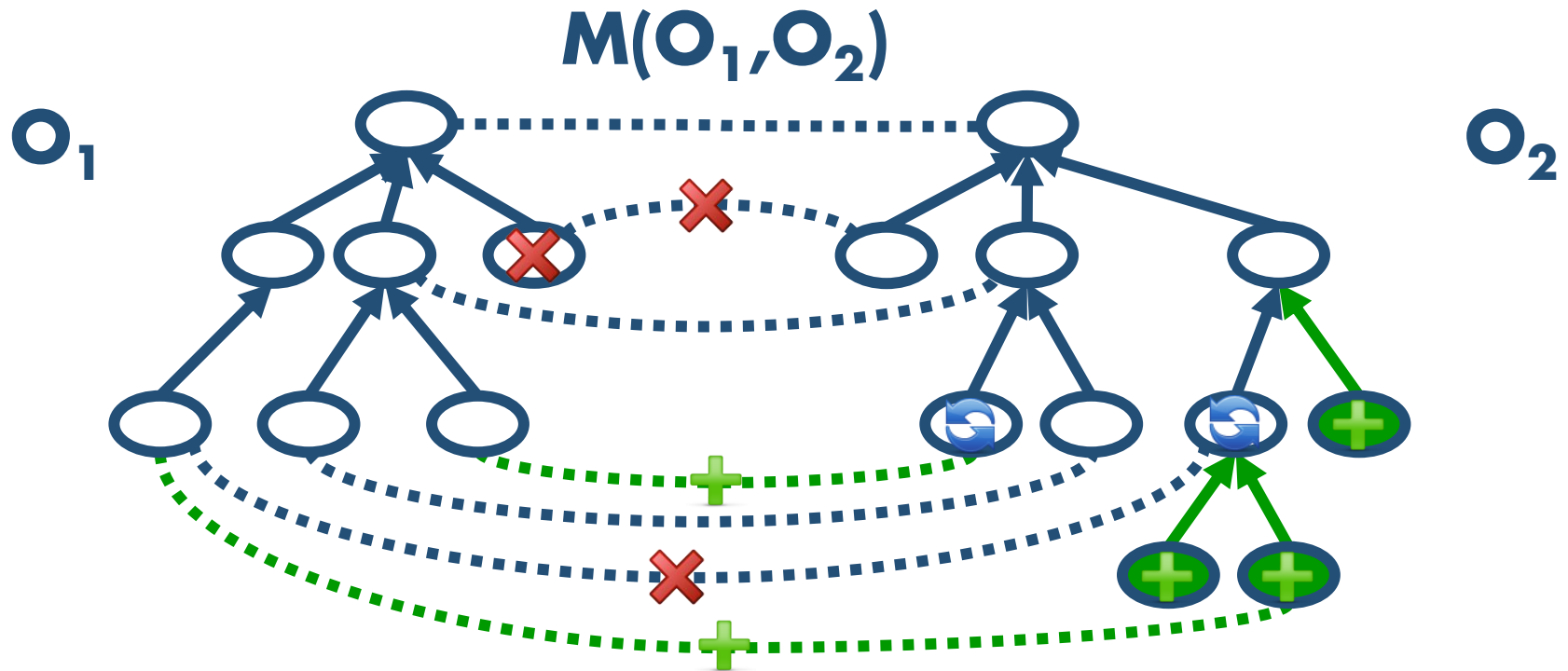


## Use of mappings




- Ontology merging - creation of the integrated cross-species anatomy ontology "Uber ontology"
- Knowledge transfer - experiments for different species
- Ontology curation - find missing ontology annotations
- ...

# ONTOLOGY EVOLUTION

- Ongoing research, new findings → continuous modifications
- Periodical release of new ontology versions
- Ontology changes   
- **Invalidate previously determined ontology mappings?**



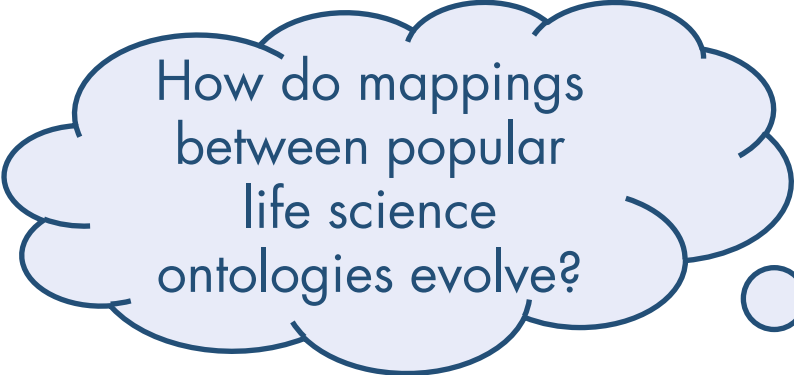
# ONTOLOGY EVOLUTION

- Ongoing research, new findings → continuous modifications
- Periodical release of new ontology versions
- Ontology changes   
- **Invalidate previously determined ontology mappings?**
- Example: Anatomy reference mapping at *OAEI* \*
  - Based on 5 year old versions
  - Quality w.r.t. current ontology versions?
- Re-determination of mappings is an expensive process
  - Manual verification of correspondences
  - Parametrization effort
- Future aim: (semi-)automatic mapping adaptation

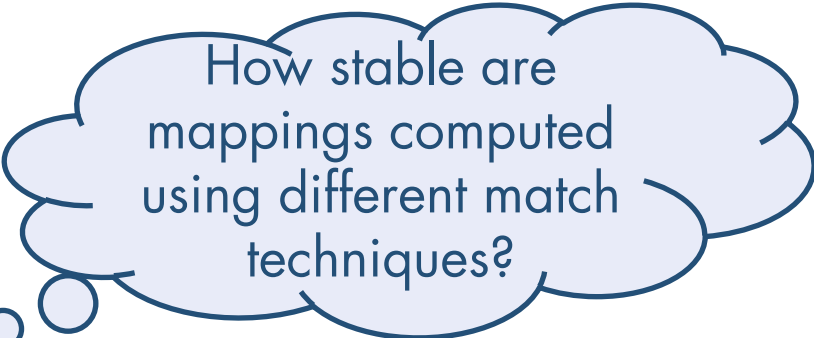
\* OAEI = Ontology Alignment Evaluation Initiative

# CONTRIBUTIONS

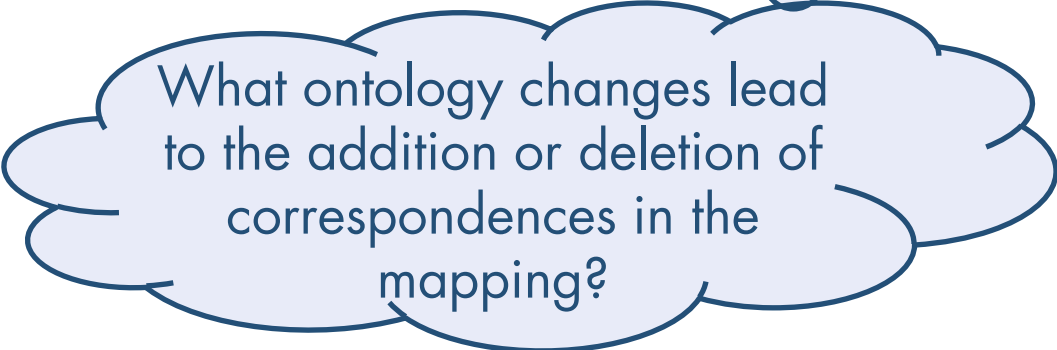
- Investigate evolution of life science ontology mappings
- Generic model for ontology and mapping evolution and their inter-dependencies
- Evaluation for three life science scenarios



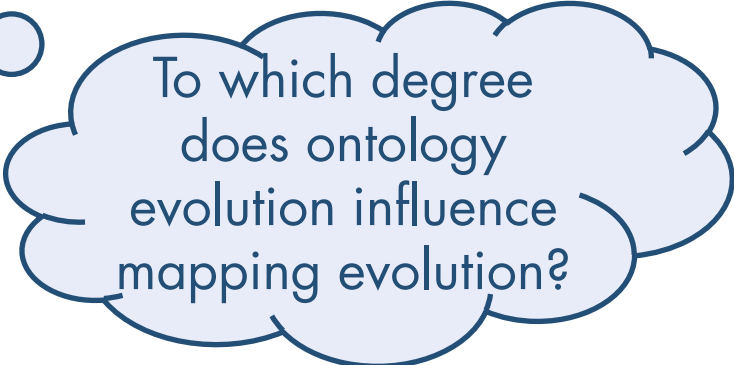
How do mappings between popular life science ontologies evolve?



How stable are mappings computed using different match techniques?

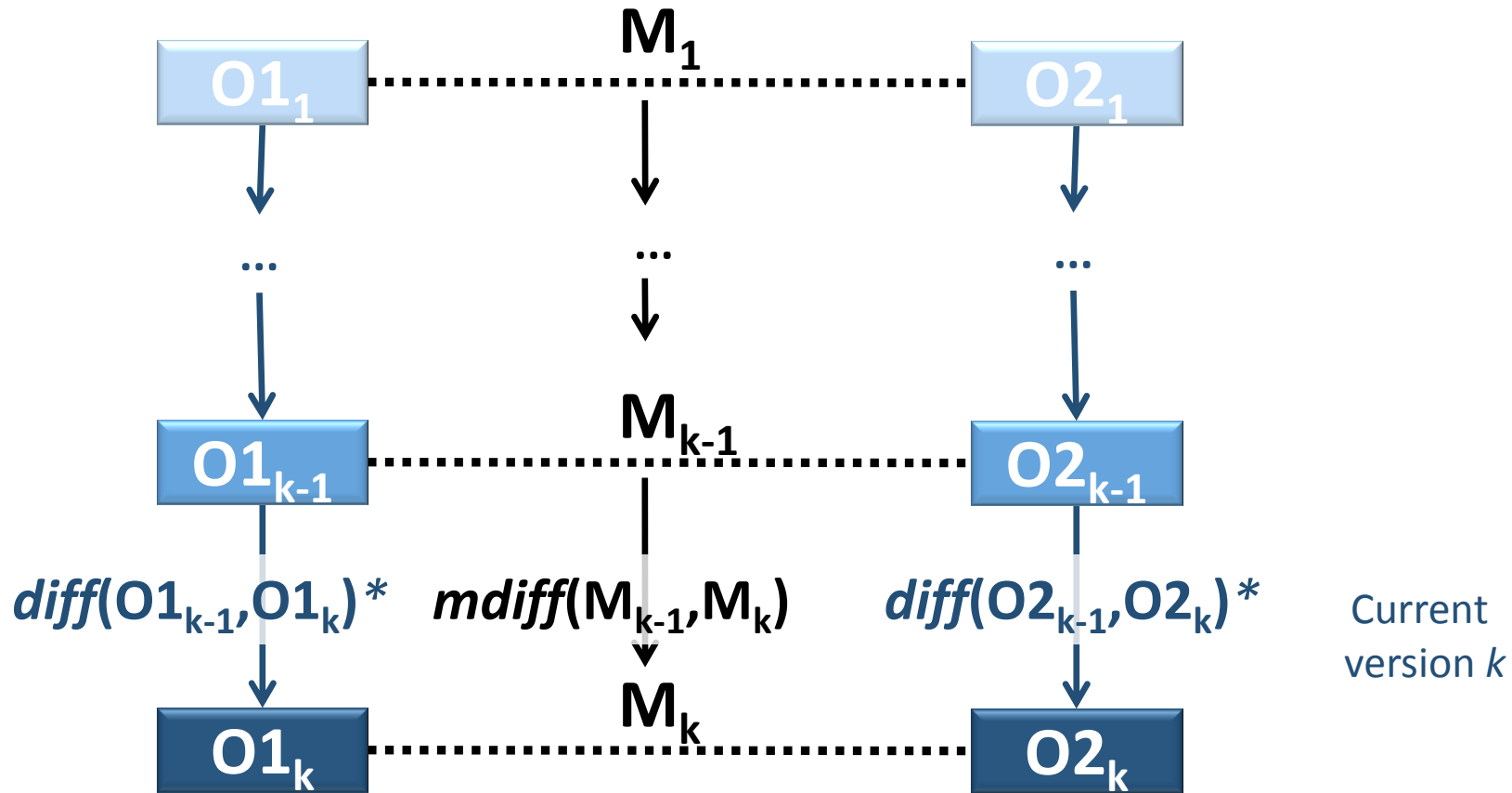


What ontology changes lead to the addition or deletion of correspondences in the mapping?



To which degree does ontology evolution influence mapping evolution?

# GENERAL EVOLUTION SCHEME



\* Hartung, M.; Groß, A.; Rahm, E.:

- **COnto-Diff**: Generation of Complex Evolution Mappings for Life Science Ontologies, *Journal of Biomedical Informatics*, 2012.
- **CODEX**: Exploration of semantic changes between ontology versions, *Bioinformatics* 28 (6): 895-896, 2012.

# CHANGE OPERATIONS

## Ontology changes:



### Extension set:

$$Ext(O_{v \rightarrow v+1})$$

Insert new concept,  
subgraph, relationship,  
attribute,

...



### Reduction set:

$$Red(O_{v \rightarrow v+1})$$

Delete existing concept,  
subgraph, relationship,  
set concept to obsolete,

...



### Revision set:

$$Rev(O_{v \rightarrow v+1})$$

Split, merge, substitute,  
move concept,  
change attribute value,

...

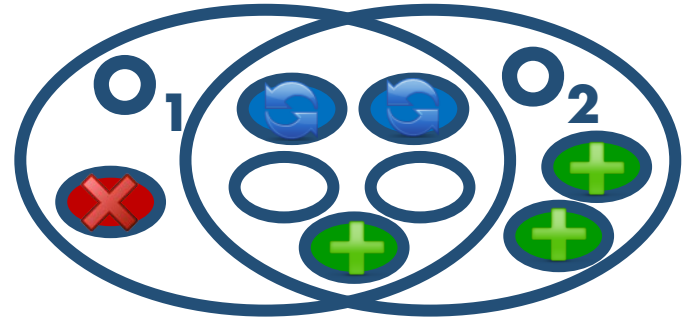
## Mapping changes

- **Addition set:**  $Add(M_{v \rightarrow v+1}) = M_{v+1} \setminus M_v$
- **Deletion set:**  $Del(M_{v \rightarrow v+1}) = M_v \setminus M_{v+1}$

# MEASURES

## Ontology Change Ratio $OCR(O_{v \rightarrow v+1})$

- Degree of ontology changes during evolution from  $O_v$  to  $O_{v+1}$
- Fraction of concepts in **Ext**  $\cup$  **Red**  $\cup$  **Rev** versus all concepts ( $O_v \cup O_{v+1}$ )



6  
—  
8

## Mapping Change Ratio $MCR(M_{v \rightarrow v+1})$

- Degree of mapping changes during evolution from  $M_v$  to  $M_{v+1}$
- Fraction of correspondence in **Add**  $\cup$  **Del** versus all correspondences ( $M_v \cup M_{v+1}$ )

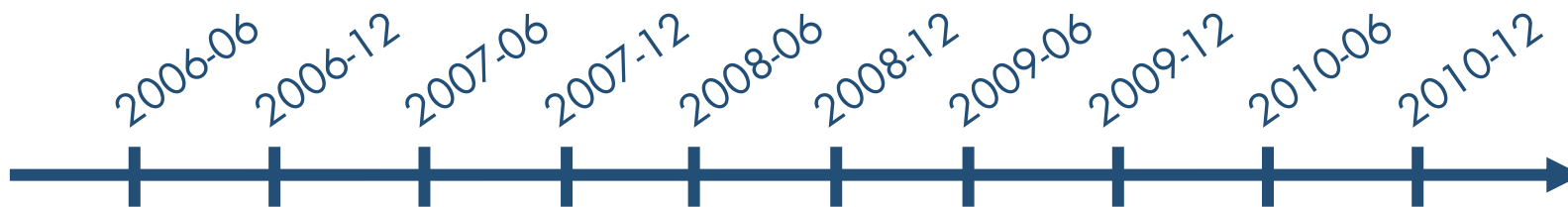
## Impact Ratio $IR(O_{Ch}, M_{Ch})$

- Share of changed concepts that actually had an impact on changed correspondences
- Fraction of additive ontology changes that led to new correspondences:  $IR(\mathbf{Ext}, \mathbf{Add})$

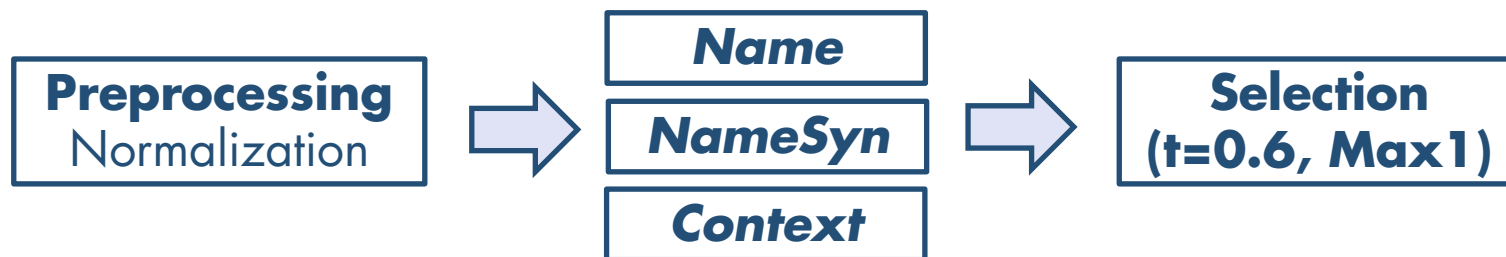


# EVALUATION SETUP

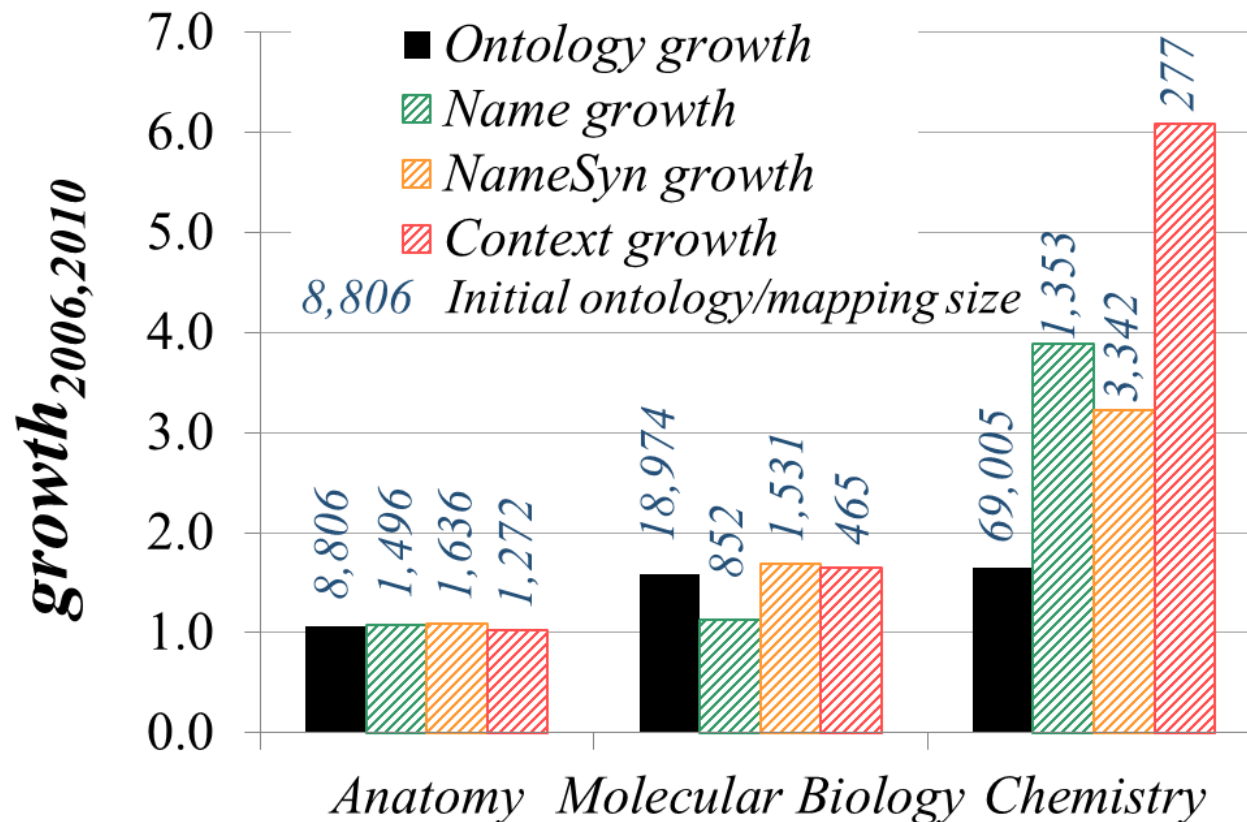
	<h2>Anatomy</h2>		Adult Mouse Anatomical Dictionary (MA)	NCI Thesaurus Anatomy part (NCITa)	
	<h2>Molecular Biology</h2>		Molecular Functions (MF)	Biological Processes (BP)	
	<h2>Chemistry</h2>		Chemical Entities of Biological Interest (ChEBI)	NCI Thesaurus (NCIT)	



- Meta-data based matchers



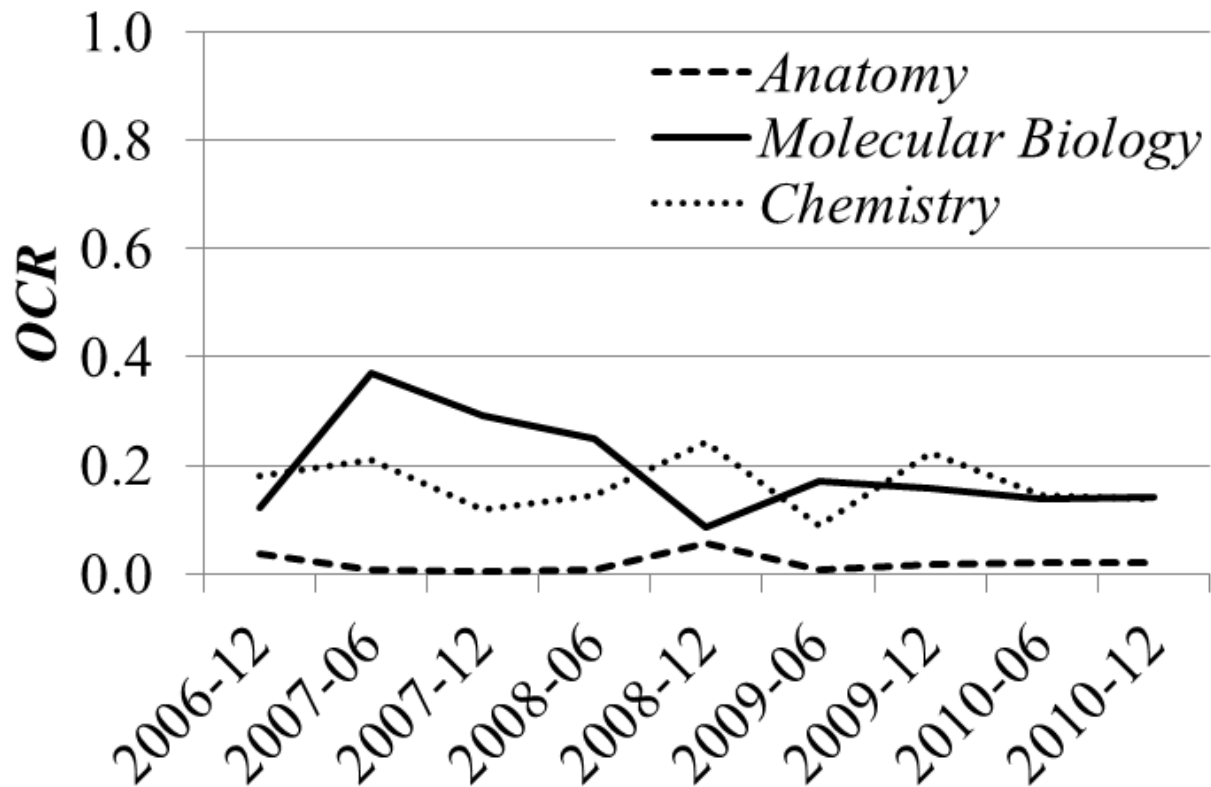
# ONTOLOGY AND MAPPING GROWTH



2006 → 2010:

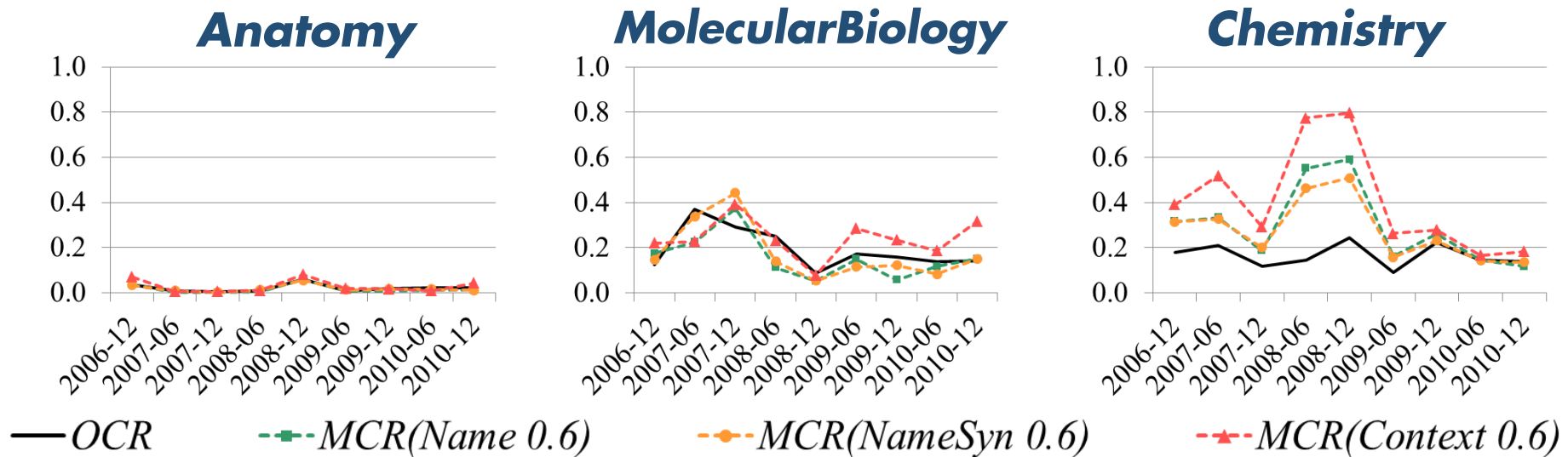
- Slight ontology growth for *Anatomy* (10%)
- 60-70% for *MolecularBiology* and *Chemistry*
- Mapping growth similar to ontology growth, except for chemistry

# ONTOLOGY CHANGE RATIO



- *Anatomy*: only few changes
- *Molecular Biology*: high change rates until 2007 (nearly 40%)
- *Chemistry*: change rates around 20%

# COMPARISON OF CHANGE RATIOS



## Correlation between the ontology and mapping change factors

**Domains** *Anatomy*: few mapping changes, relatively stable

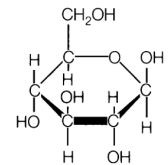
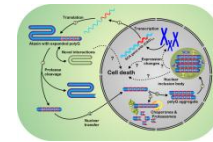
*MolecularBiology, Chemistry*: high degree of mapping changes ,  
between 10 and 80 %

**Matcher** *Name*: relatively stable

*Context*: most heavily influenced by ontology evolution,  
2008, *Chemistry* 80% mapping changes

# CONCLUSIONS & FUTURE WORK

- Study the evolution of ontology mappings
  - General evolution scheme and measures (change factors)
- Evaluation for ontology mappings in three life science domains + comparison of three match strategies
  - Correlation between ontology and mapping change factors
  - Different stability for different match techniques and domains
  - Impact of ontology on mapping changes
    - Most correspondence *Add* / *Del* are caused by ontology *Ext* / *Red*
    - Surprisingly high degree of mapping changes caused by ontology *Rev*



## Future Work

- Use known ontology changes to semi-automatically adapt ontology mappings (without completely new mapping determination)



# How do ONTOLOGY MAPPINGS CHANGE IN THE LIFE SCIENCES?



## POSTER SESSION THIS AFTERNOON

UNIVERSITÄT LEIPZIG

### How do Ontology Mappings Change in the Life Sciences?

Anika Groß<sup>1,2</sup>, Michael Hartung<sup>1,2</sup>, Andreas Thor<sup>1</sup>, Erhard Rahm<sup>1,2</sup>

<sup>1</sup> Department of Computer Science, University of Leipzig  
<sup>2</sup> Interdisciplinary Centre for Bioinformatics, University of Leipzig  
contact: gross@informatik.uni-leipzig.de

#### Motivation

**Ontology Evolution**

- Ongoing research, new findings → continuous modifications
- Periodical release of new ontology versions

**Ontology Mappings**

- Set of semantic correspondences between concepts of different ontologies
- Possible invalidation of previously determined ontology mappings due to ontology evolution

**Aims:**

- Investigate evolution of life science ontology mappings
- Study impact of ontology evolution on mapping evolution
- Future Work: Use known ontology changes to semi-automatically adapt ontology mappings

#### Change Operations

**Ontologies Changes:**

- Extension set**  $Ext(O_{t+1})$ : Add new concept, subgraph, relationship, attribute, ...
- Reduction set**  $Red(O_{t+1})$ : Delete existing concept, subgraph, relationship, ...
- Revision set**  $Rev(O_{t+1})$ : Split, merge, substitute, move concept, ...

**Ontology Mapping Changes:**

- Addition set**  $Add(M_{t+1}) = M_{t+1} \setminus M_t$
- Deletion set**  $Del(M_{t+1}) = M_t \setminus M_{t+1}$

#### General evolution scheme

#### Change Ratios

**Ontology Change Ratio (OCR)**

$$OCR(O_{t+1}) = \frac{|Ext(O_{t+1})| + |Red(O_{t+1})| + |Rev(O_{t+1})|}{|O_t \cup O_{t+1}|}$$

**Mapping Change Ratio (MCR)**

$$MCR(M_{t+1}) = \frac{|Add(M_{t+1})| + |Del(M_{t+1})|}{|M_t \cup M_{t+1}|}$$

**Impact ratio (IR)**

$$IR(O_{t+1}, M_{t+1}) = \frac{|(E \in O_{t+1} \setminus O_t) \cap (C \in M_{t+1} \setminus C_t) \cap (C \in C_{t+1})|}{|O_{t+1}|}$$

Fraction of additive ontology changes that lead to new correspondences.

#### Evaluation

**Three Life Science Match Problems**

- Analyze versions between 2006 and 2010

Ontology	Match	Adult Mouse Anatomical Dictionary (MA)	NCI Thesaurus Anatomy part (NCIT)
Anatomy	MA	NCI Thesaurus Anatomy part (NCIT)	
Molecular Biology	GO	Molecular Functions (MF)	Biological Processes (BP)
Chemistry	ChEMBL	Chemical Entities of Biological Interest (ChEBI)	NCI Thesaurus (NCIT)

#### Mapping Changes

More correspondences additions + High degree of deletions

#### 3 Meta-data based Matchers

Preprocessing → Normalization → Selection (I=[0.6, Max])

#### Ontology and Mapping Growth

#### Ontology Change Ratios

#### Correlation between ontology and mapping change features

Different stability for different match techniques  
Contrast → most unstable

#### Impact of Ontology Changes on Mapping Changes

	Ext	Add	Red	Rev	Del	IR
Anatomy	10	18.7%	0.1%	7	0.0%	7.8%
Molecular Biology	2,239	4.6%	0.7%	223	2.4%	8.8%
Chemistry	8,377	11.7%	1.5%	366	3.3%	9.3%

#### Most correspondence additions are caused by ontology extensions

Most correspondence deletions are caused by ontology reductions

Surprisingly high degree of mapping changes caused by ontology revisions

UNIVERSITÄT LEIPZIG

### CODEX

COMPLEX ONTOLOGY DIFF EXPLORER

Michael Hartung, Anika Groß, Erhard Rahm

Department of Computer Science, University of Leipzig  
Interdisciplinary Centre for Bioinformatics, University of Leipzig  
Contact: hartung@izbi.uni-leipzig.de

<http://www.izbi.de/codex>

Use CODEX to explore online semantic changes between different versions of your ontologies

#### Motivation

- Ontologies evolve to meet changed requirements and to incorporate new knowledge
- Diff: set of changes between two ontology versions
- Helpful for mapping adaptation or collaborative ontology curation
- Diff should be compact and human-understandable especially for large life science ontologies
- Small number of semantic changes rather than a long list of "simple" element additions/deletions

Tool to interactively explore a semantically rich diff between ontology versions

#### Demonstration

Diff computation and exploration for different life science ontologies and their versions

- Examples: Sequence Ontology, Adult Mouse Anatomy, Gene Ontology, ...

**I. Different input possibilities:**

- Web URL
- File Upload
- OnEX

**II. Diff Computation**

**III. Multiple exploration facilities:**

- Pie Charts
- Tag Clouds
- Change Navigator
- Tree-based Change Explorer

#### Approach

- Onto-Diff to determine basic and compact diffs using matching and rules
- Supported changes: merge, split, addSubgraph, move, ...

#### Features

- Web GUI and web service interface available
- Multiple input possibilities:
  - Support for OBO and OWL ontologies
  - File upload, copy/paste, URL, OnEX repository
- Several exploration facilities:
  - Customizable overview statistics via pie charts
  - Tag clouds to visualize changes and modified content
  - Tree-based change explorer
  - Impact analysis: Are my categories affected by changes?

#### References

- [1] M. Hartung, A. Groß, E. Rahm: CODEX: Exploration of semantic changes between ontology versions. *Bioinformatics* 28(6), 2012
- [2] M. Hartung, A. Groß, E. Rahm: ONTO-DIFF: Generation of Complex Ontology Mappings for Life Science Ontologies. *Journal of Biomedical Informatics (in Press)*, 2012
- [3] M. Hartung, T. Krüger, A. Groß, E. Rahm: OnEX: Exploring changes in life science ontologies. *BMC Bioinformatics* 10:250, 2009

FUNDING - German Research Foundation

Grant RA497/18-1 - "EVOLUTION OF ONTOLOGIES AND MAPPINGS"