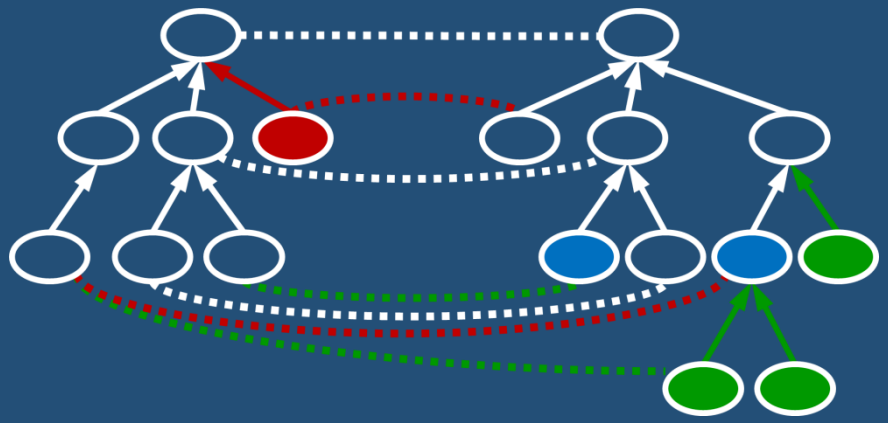


How do Computed Ontology Mappings Evolve?

A Case Study for Life Science Ontologies

Anika Gross, Michael Hartung, Andreas Thor, Erhard Rahm

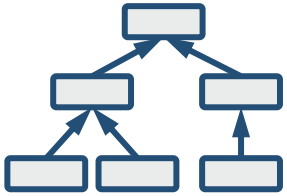
UNIVERSITÄT LEIPZIG

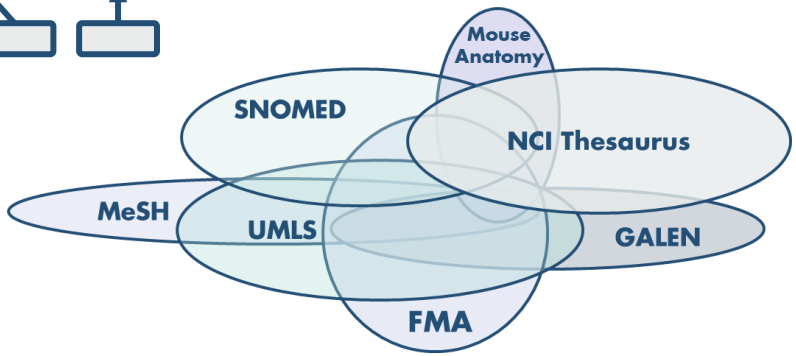
12TH November 2012, EvoDyn Workshop, Boston

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

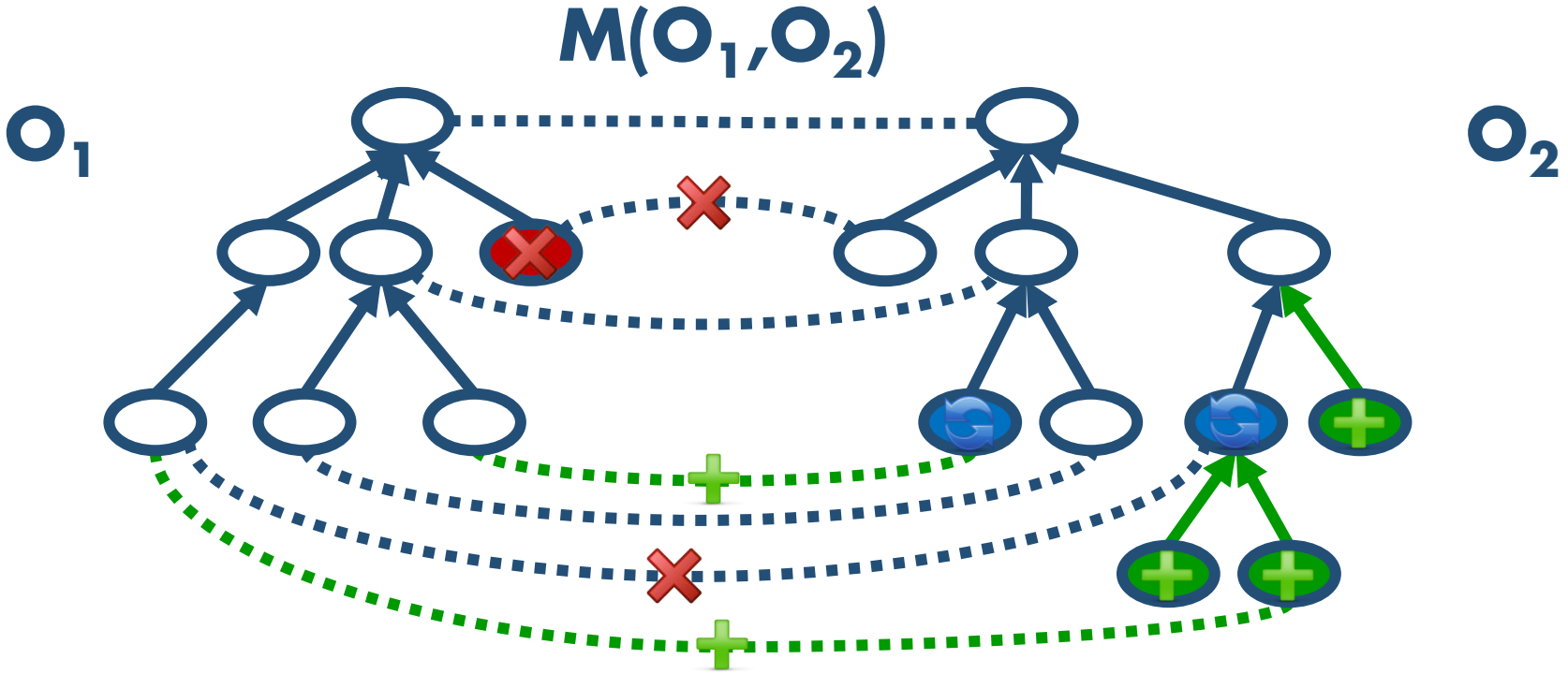
Use of mappings

- 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

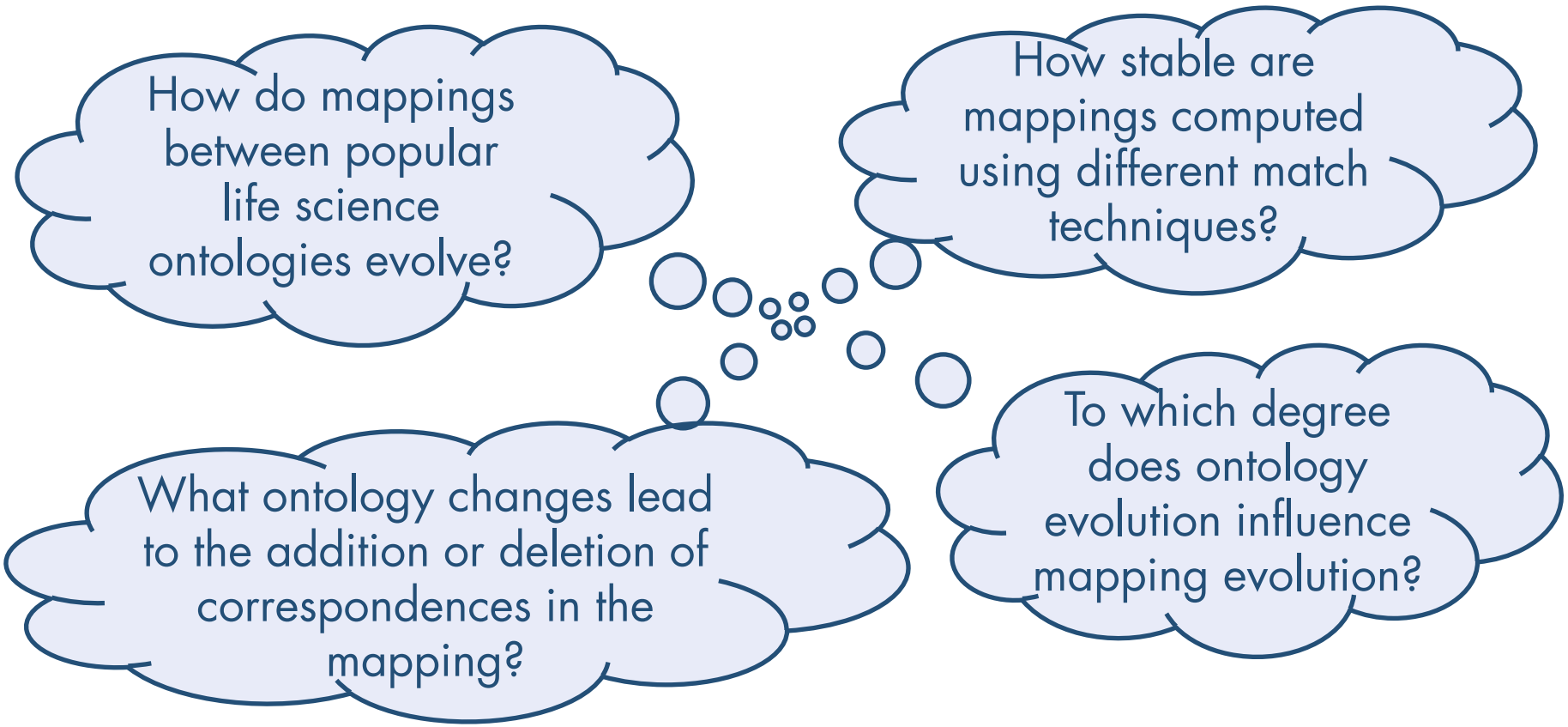
- Ongoing research, new findings → continuous modifications
- Periodical release of new ontology versions
- Ontology changes

- **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

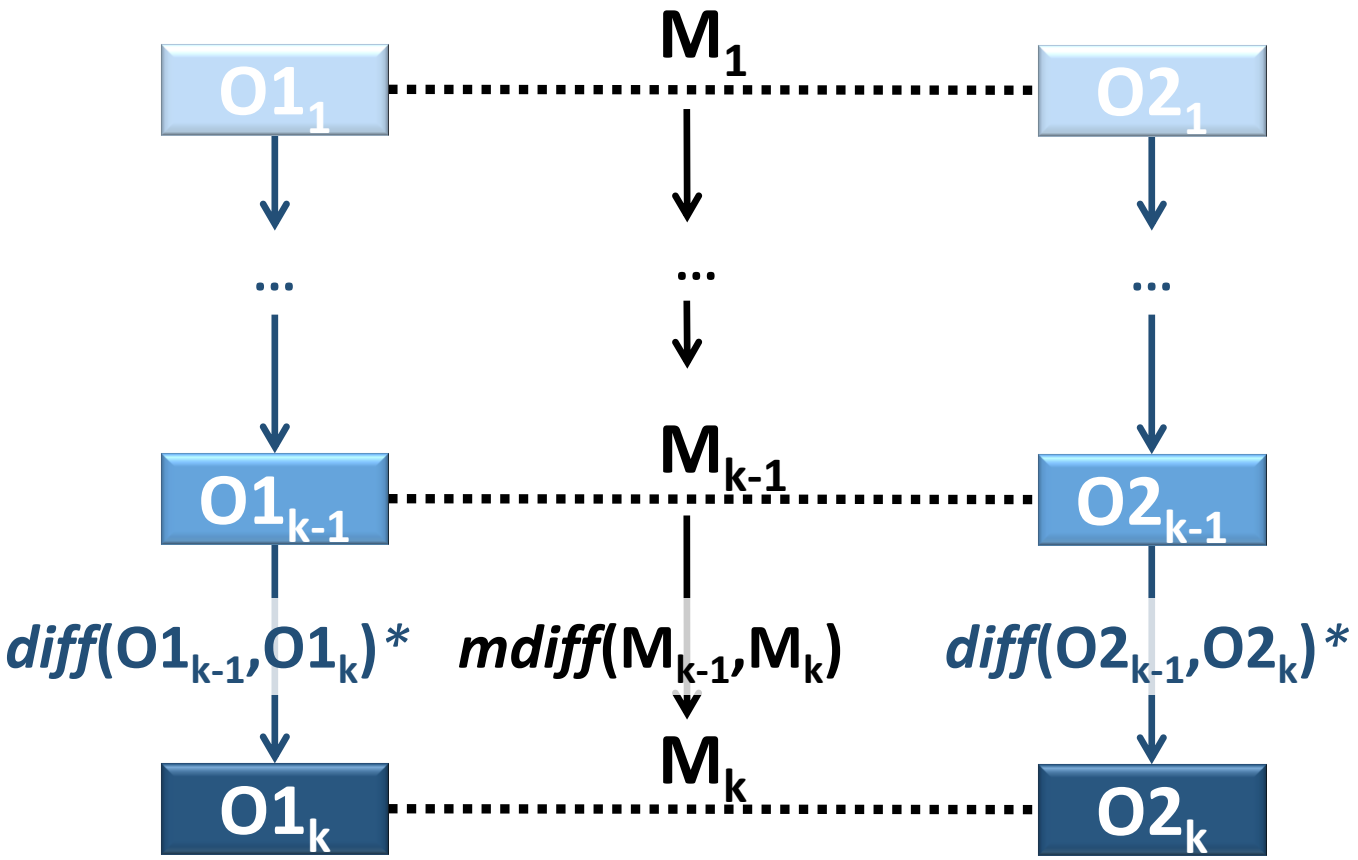
* OAEI = Ontology Alignment Evaluation Initiative

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



General Evolution Scheme



Current version *k*

* 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:

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

Insert new concept, subgraph, relationship, attribute,

…

Reduction set:

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

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

…

Revision set:

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

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

…

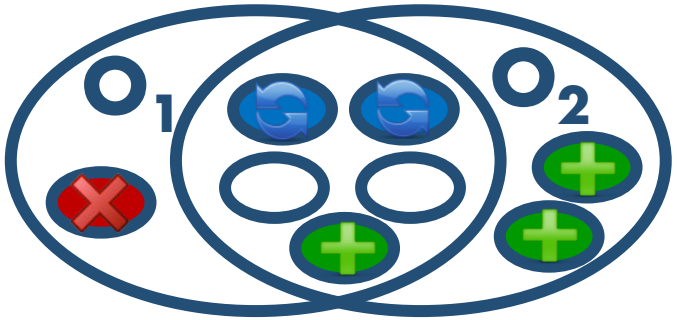
Mapping changes:

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

Measures

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

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



$$\frac{6}{8}$$

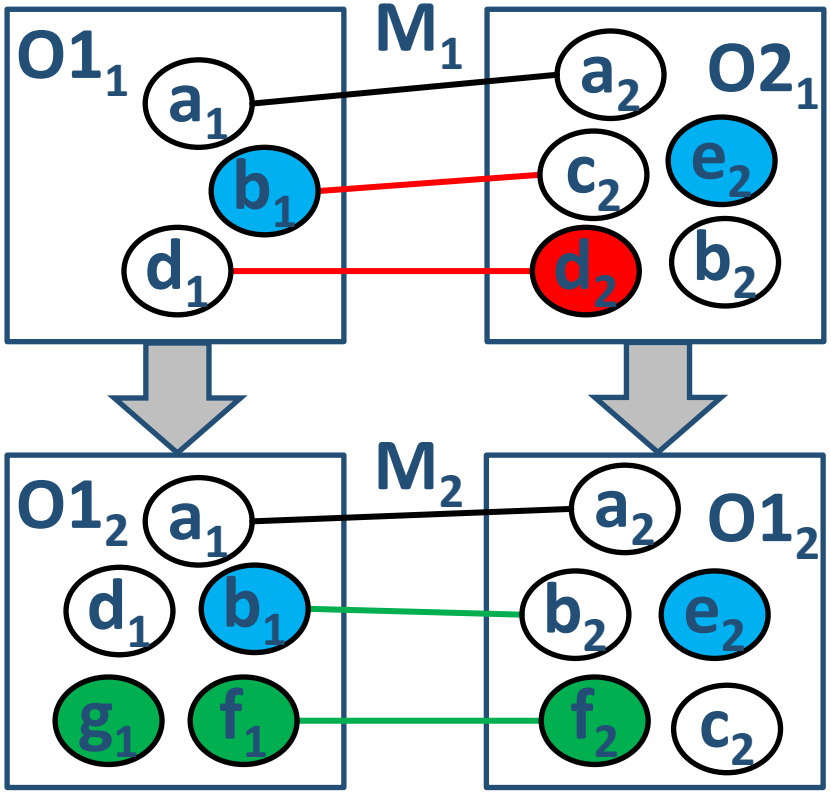
Mapping Change Ratio $MCR(M_{v \to v+1})$

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

Measures

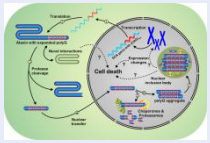
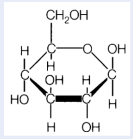
Impact Ratio $IR(O_{Ch}, M_{Ch})$

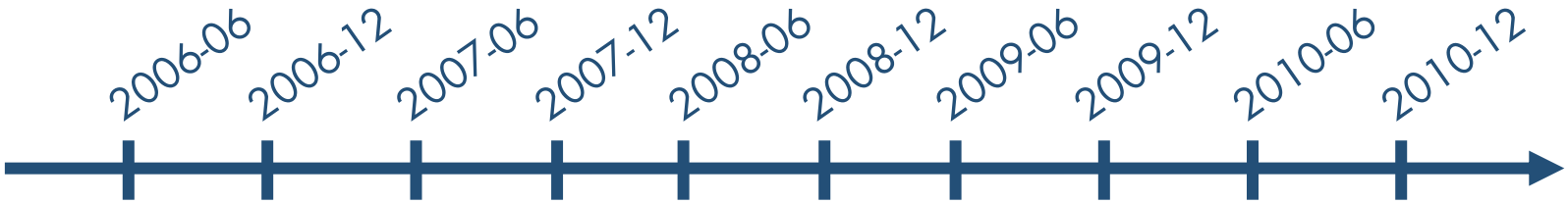
- Share of changed concepts that actually had an impact on changed correspondences
- For instance, fraction of additive ontology changes that led to new correspondences: $IR(Ext, Add)$



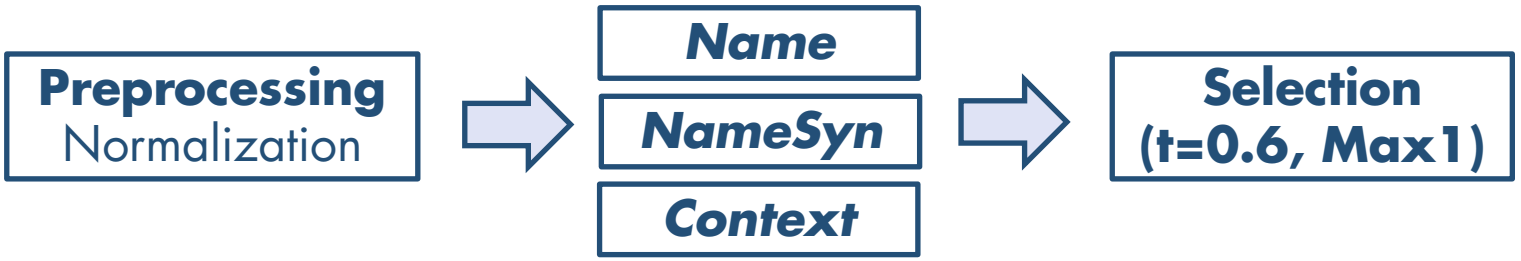
Ontologies	Mapping	ADD: {(b_1,b_2), (f_1,f_2)}	DEL: {(b_1,c_2), d_1,d_2)}
EXT	{f_1, g_1}∪{ f_2}	2/3	0
RED	∅∪{d_2}	0	1
REV	{b_1}∪{ e_2}	1/2	1/2

Evaluation Setup

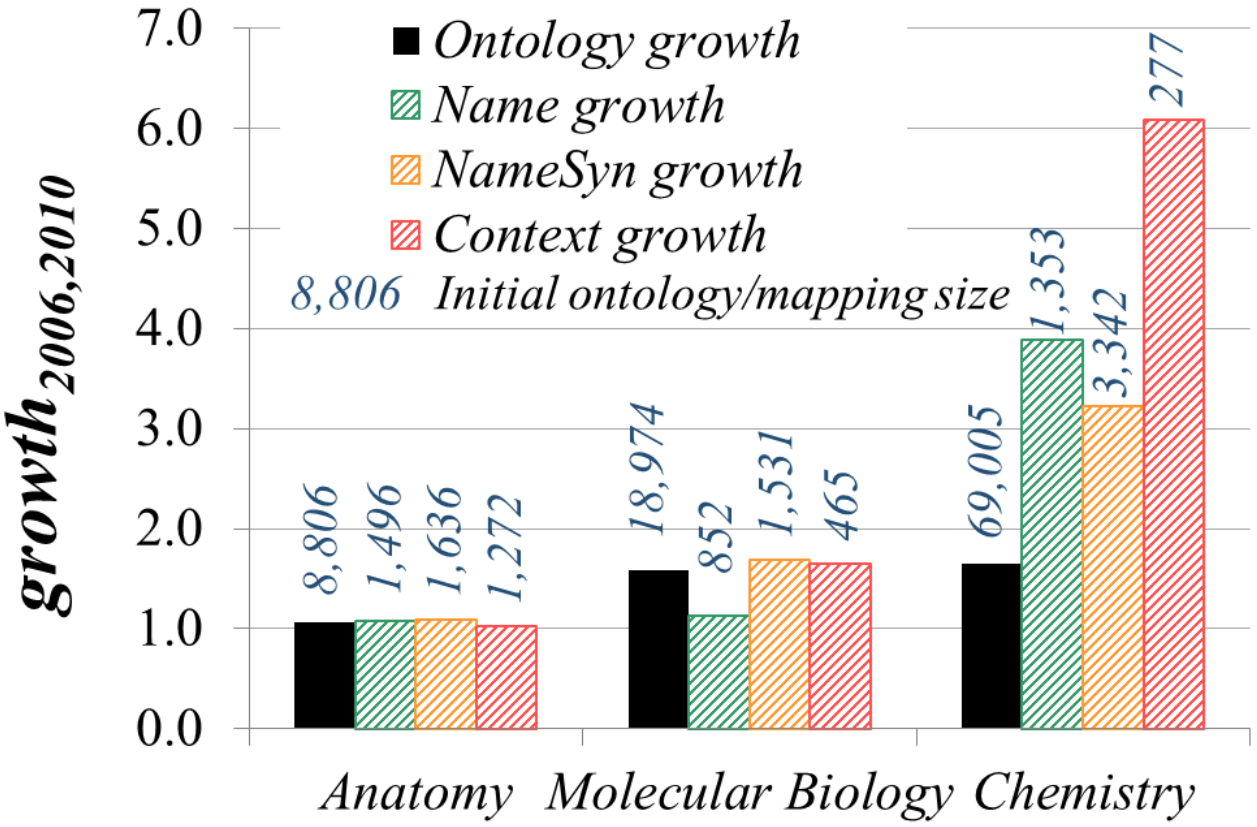
	Anatomy	Adult Mouse Anatomical Dictionary (MA)	NCI Thesaurus Anatomy part (NCITa)
	***Molecular Biology***	Molecular Functions (MF)	Biological Processes (BP)
	***Chemistry***	Chemical Entities of Biological Interest (ChEBI)	NCI Thesaurus (NCIT)



- Meta-data based matchers

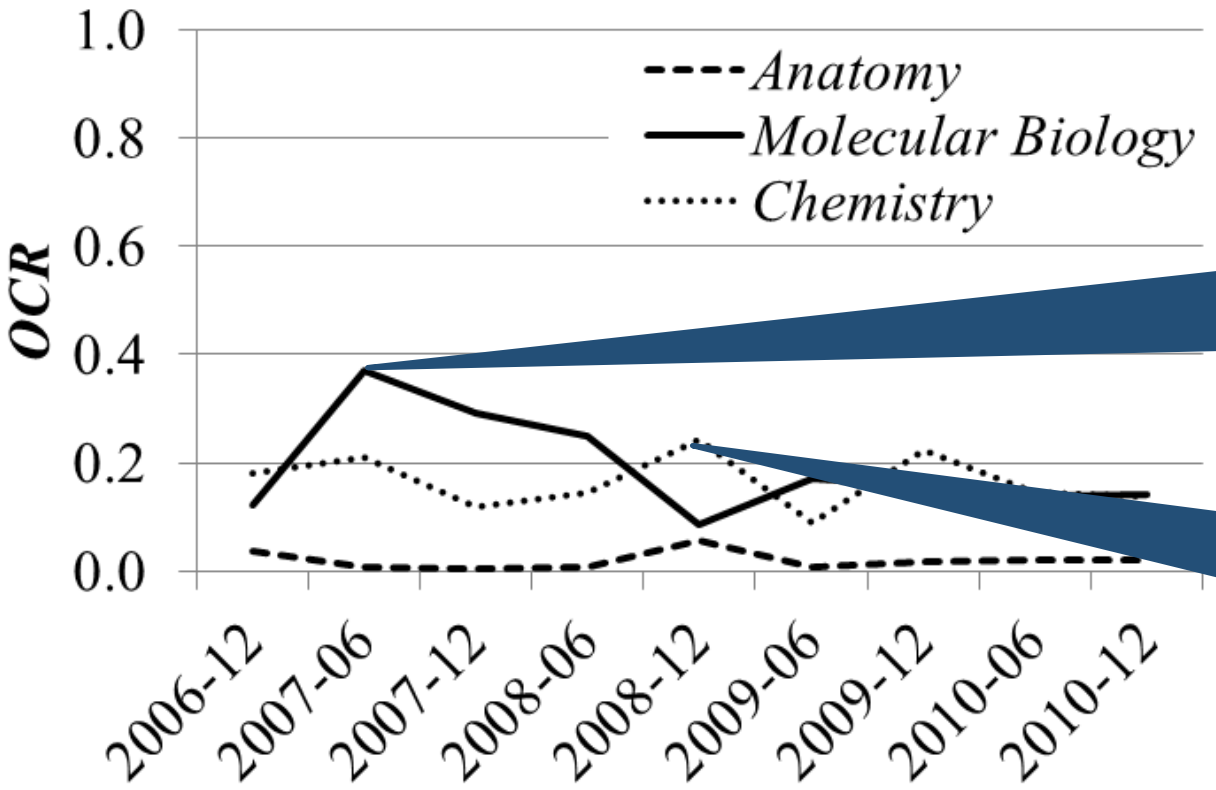


Ontology and Mapping Growth



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

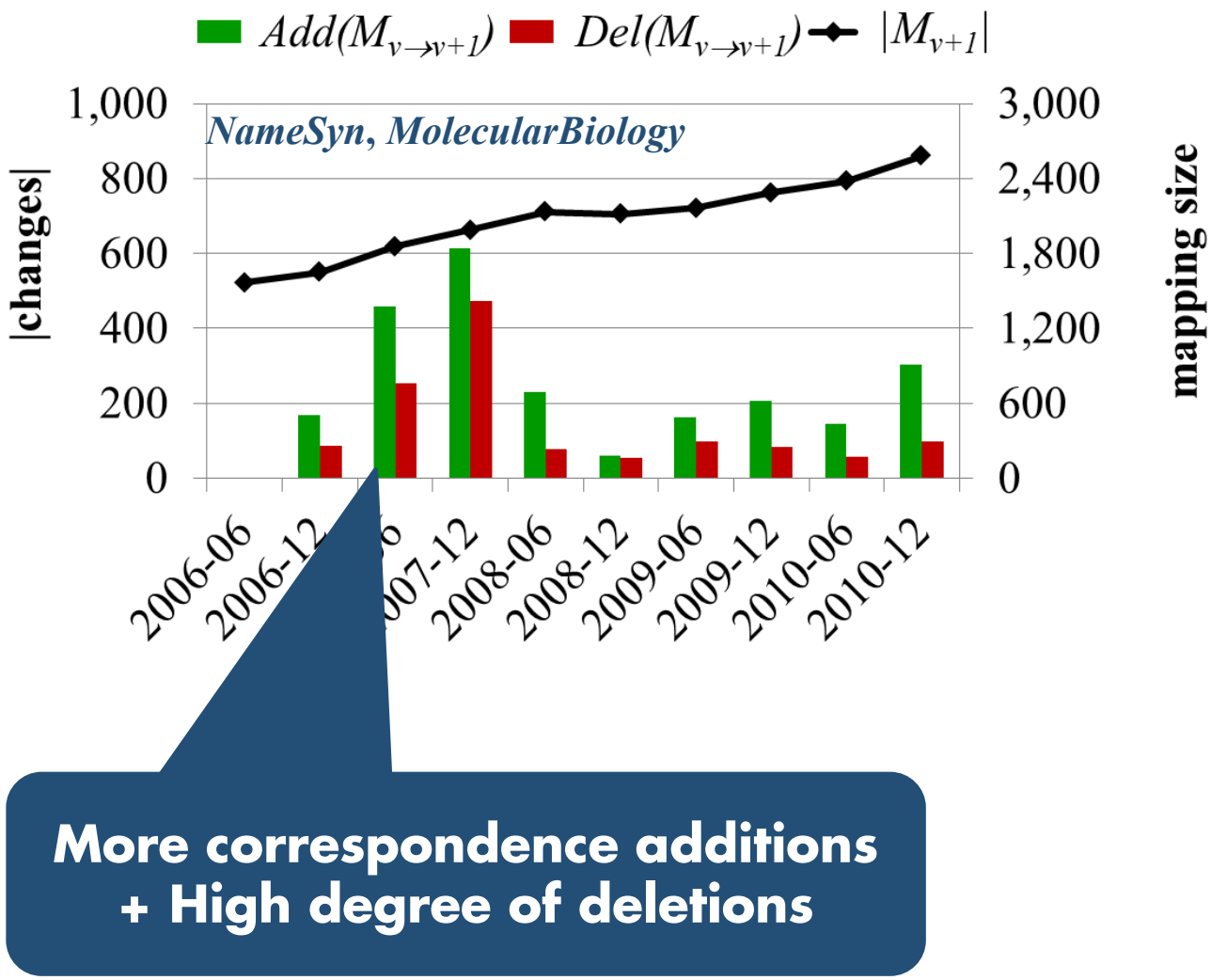
Ontology Change Ratio



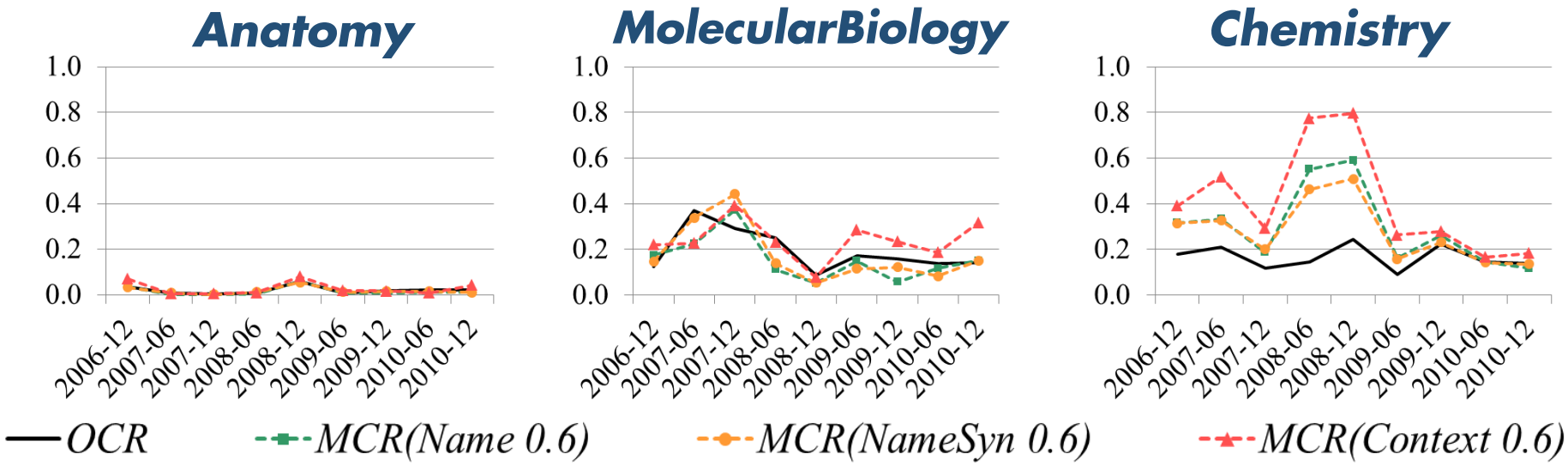
Heavy changes for Molecular Biology (nearly 40%)

Chemistry *OCR* around 20%

Mapping Changes



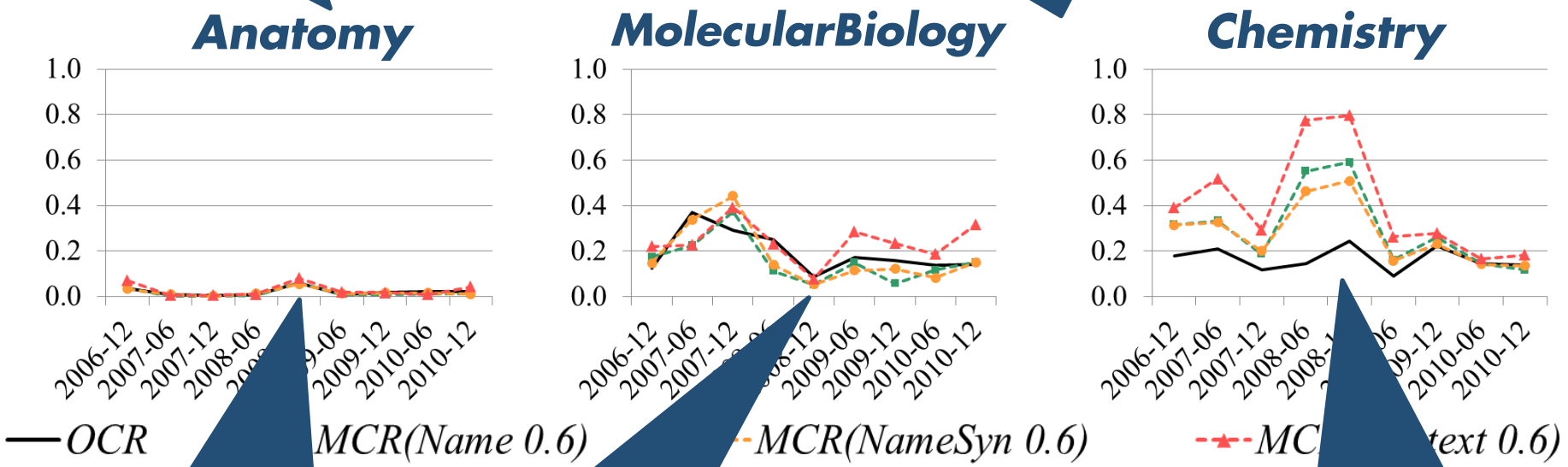
Change Ratios



Change Ratios

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

MolecularBiology, Chemistry: high degree of mapping changes (10 - 80 %)



Correlation between ontology and mapping change factors

Different stability for different matchers
Name: relatively stable
Context: most unstable

Impact of Ontology Changes on Mapping Changes

	|Ext|	IR_{Ext}		|Red|	IR_{Red}		|Rev|	IR_{Rev}	
		→Add	→Del		→Add	→Del		→Add	→Del
Anatomy	95	**18.7%**	0.1%	7	0.0%	**7.8%**	89	**6.8%**	**4.1%**
Molecular Biology	2,359	**4.6%**	0.7%	223	2.4%	**8.8%**	2,209	**3.5%**	**2.1%**
Chemistry	8,377	**11.7%**	1.2%	366	3.5%	**5.3%**	6,441	**8.1%**	**4.0%**

Most correspondence additions are caused by ontology extensions

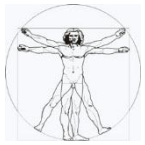
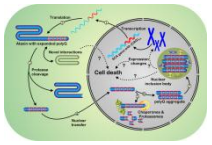
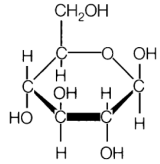
Most correspondence deletions are caused by ontology reductions

Surprisingly high degree of mapping changes caused by ontology revisions

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 Computed Ontology Mappings Evolve? A Case Study for Life Science Ontologies

Funding: German Research Foundation Grant RA497/18-1
"Evolution of Ontologies and Mappings"