

Erweiterungen und Realisierungen im Rahmen der Evolutionsanalyse von Datenquellen und Mappings

Michael Hartung

Interdisziplinäres Zentrum für Bioinformatik und Abteilung Datenbanken

Zingst 2008



UNIVERSITÄT LEIPZIG



UNIVERSITÄT LEIPZIG

Abteilung Datenbanken

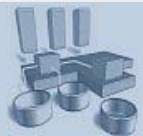
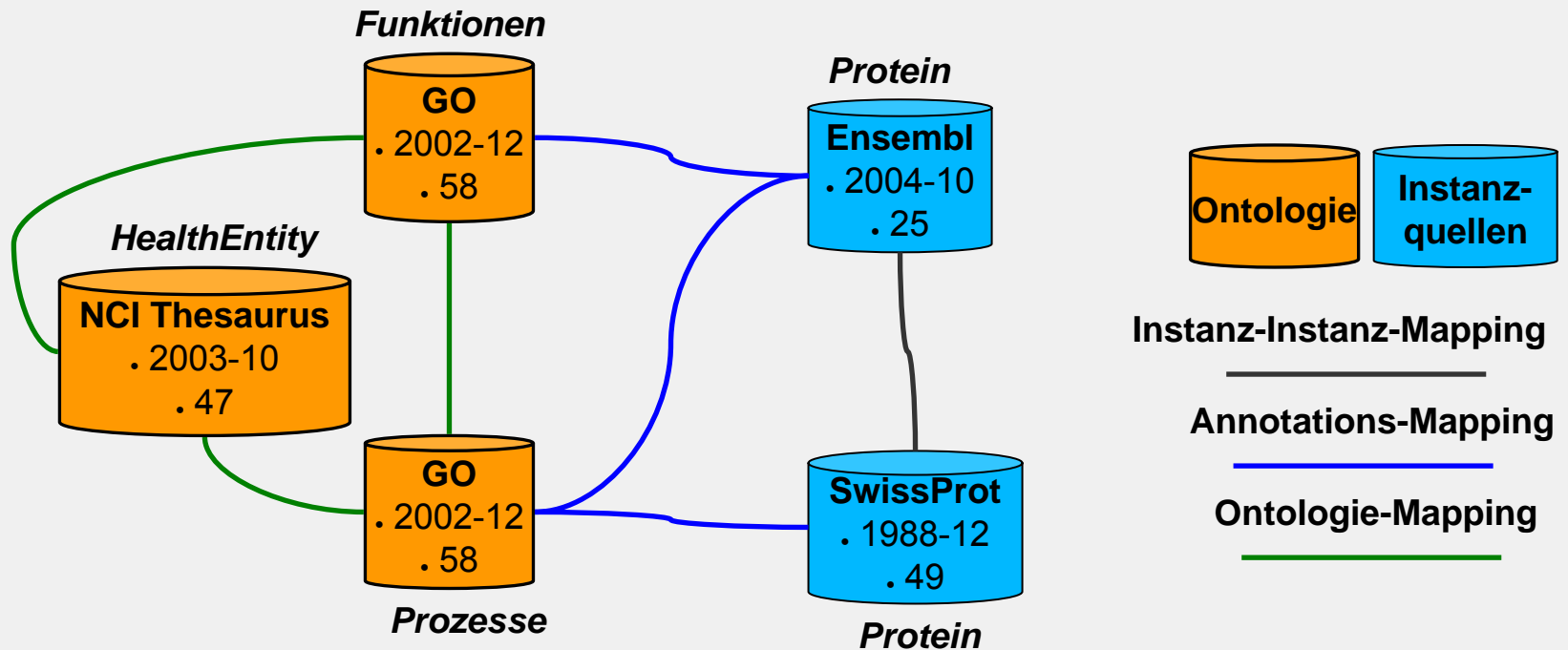
Erweiterungen und Realisierungen im Rahmen der Evolutionsanalyse von Datenquellen und Mappings

Michael Hartung Zingst, 2008

Motivation I

• DILS 2008 – Analyse

- Modelle für Quellen (Instanzquellen, Ontologien) und Mappings (Annotationen, Ontologie-Mappings) + einheitliches Evolutionsmodell + Metriken
- Analyse über iFuice oder direkt über Repository → Import in Excel → Auswertung (Trend Charts, Anwendung der Metriken)
- Fehlende Infrastruktur um noch systematischer Evolution zu analysieren



Motivation II

• Evolutionsmodell für Ontologien

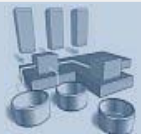
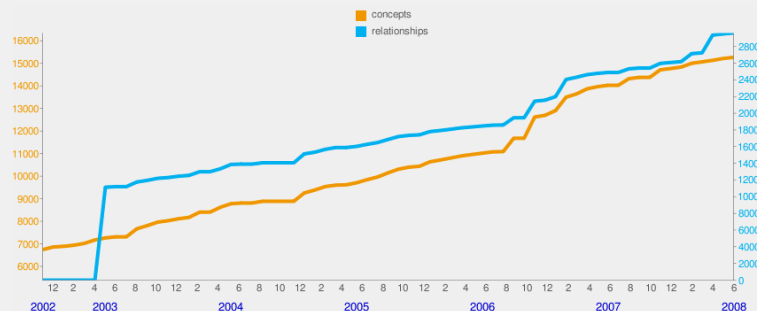
- Neben rein quantitativer Analyse auf Konzeptebene ebenfalls Evolution in Attributen und in der Struktur möglich
- DILS-Analyse zeigte durchaus hohe Löschraten → Frage: Existieren neben Löschungen auch andere Evolutionsoperatoren?
- Notwendigkeit der Einbeziehung weiterer Attribute (Wissen), z.B. Synonyme, Definitionen etc.



| Ontology | Full period (May. 04 - Feb. 08) | | | | | | | Last year (Feb. 07 - Feb. 08) | | |
|----------------------------------|---------------------------------|-----|-----|------|----------|----------|----------|-------------------------------|-----|-----|
| | Add | Del | Obs | obs | add-frac | del-frac | obs-frac | Add | Del | Obs |
| <i>NCI Thesaurus</i> | 627 | 2 | 12 | 42.4 | 1.3% | 0.0% | 0.0% | 416 | 0 | 5 |
| <i>GeneOntology</i> | 200 | 12 | 4 | 12.2 | 0.9% | 0.1% | 0.0% | 222 | 20 | 5 |
| -- <i>Biological Process</i> | 146 | 7 | 2 | 16.2 | 1.2% | 0.1% | 0.0% | 133 | 10 | 2 |
| -- <i>Molecular Function</i> | 36 | 3 | 2 | 6.8 | 0.4% | 0.0% | 0.0% | 69 | 7 | 3 |
| -- <i>Cellular Components</i> | 18 | 2 | 0 | 8.9 | 1.0% | 0.1% | 0.0% | 19 | 3 | 0 |
| <i>ChemicalEntities</i> | 256 | 62 | 0 | 4.1 | 1.8% | 0.5% | 0.0% | 384 | 67 | 0 |
| <i>FlyAnatomy</i> | 5 | 1 | 1 | 3.3 | 0.1% | 0.0% | 0.0% | 6 | 0 | 0 |
| <i>MammalianPhenotype</i> | 65 | 2 | 9 | 6.0 | 1.2% | 0.0% | 0.2% | 74 | 2 | 3 |
| <i>AdultMouseAnatomy</i> | 11 | 0 | 0 | 30.9 | 0.4% | 0.0% | 0.0% | 1 | 0 | 0 |
| <i>ZebrafishAnatomy</i> | 33 | 5 | 1 | 5.5 | 1.8% | 0.3% | 0.1% | 45 | 2 | 1 |
| <i>Sequence</i> | 19 | 3 | 2 | 4.1 | 1.5% | 0.3% | 0.2% | 19 | 0 | 0 |
| <i>ProteinModification</i> | 5 | 2 | 1 | 1.5 | 0.4% | 0.2% | 0.1% | 7 | 0 | 2 |
| <i>CellType</i> | 5 | 1 | 0 | 2.8 | 0.7% | 0.2% | 0.1% | 1 | 0 | 0 |
| <i>PlantStructure</i> | 5 | 0 | 1 | 6.1 | 0.7% | 0.0% | 0.1% | 3 | 0 | 0 |
| <i>ProteinProteinInteraction</i> | 21 | 0 | 0 | 41.7 | 2.7% | 0.0% | 0.2% | 4 | 0 | 0 |
| <i>FlyBaseCV</i> | 1 | 0 | 1 | 2.1 | 0.2% | 0.0% | 0.1% | 0 | 0 | 0 |
| <i>Pathway</i> | 7 | 1 | 0 | 7.9 | 1.3% | 0.2% | 0.0% | 6 | 2 | 0 |

Motivation III

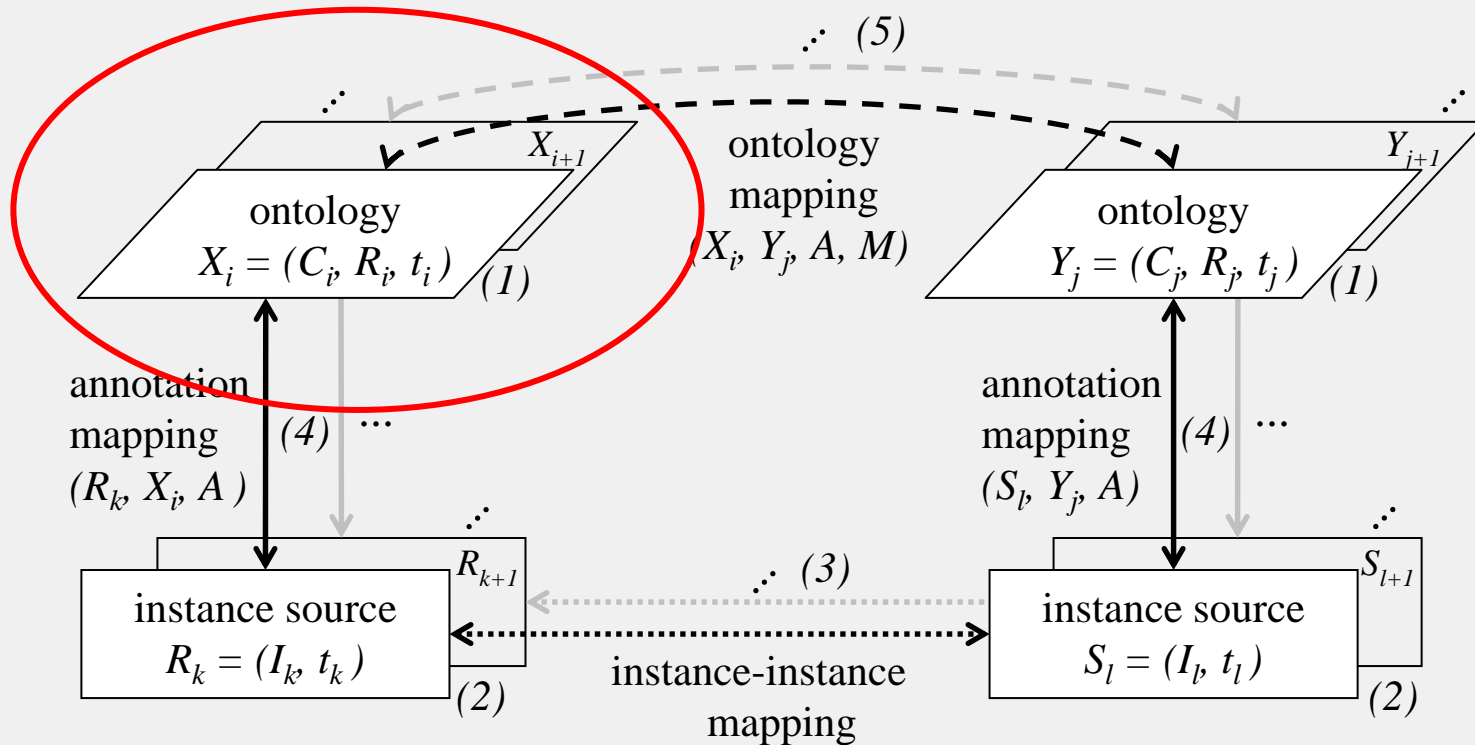
- *Verfügbarkeit der Evolutionsergebnisse*
 - Zitat aus den Reviews: „*It would have been more useful if the authors had developed and released an openly available tool for the use of bioinformatics ontology curators and researchers.*“
 - Bisher: Evolutionsergebnisse lagen mehr oder minder in Excel-Sheets oder sonstigen Dateiformaten vor
 - Anforderungen an ein Tool
 - Online frei verfügbar
 - Leichte Bedienbarkeit und Navigationsmöglichkeiten
 - Generierung übersichtlicher Trend Charts / Tabellen, aber auch Navigation in Detailergebnisse und Suche in allen verfügbaren Ontologien
 - Updates der verfügbaren Ontologien → neueste Ontologieversionen zur Verfügung stellen



- **Motivation**
- **Repository zur Evolutionsanalyse**
 - Ausgangspunkt
 - Modell
 - Import
- **Erweiterte Evolutionsoperatoren**
 - Neue Operatoren auf Konzeptebene
 - Evolution in Attributen
- **OnEX – Ontology Evolution Explorer**
- **Zusammenfassung und Ausblick**

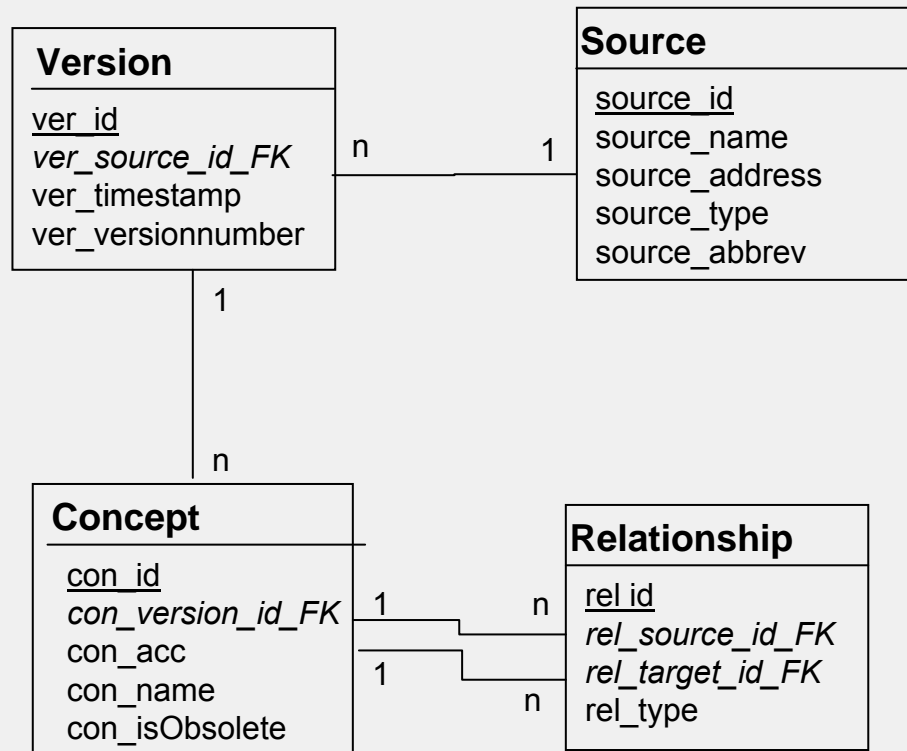


Überblick über zu analysierende Quellen und Mappings



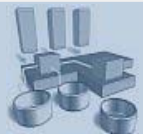
Modell für Ontologieevolutionsanalyse (DILS 2008)

- *Altes Modell des Repositories für die Evolutionsanalyse von Ontologien*

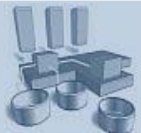


- *Limitierungen*

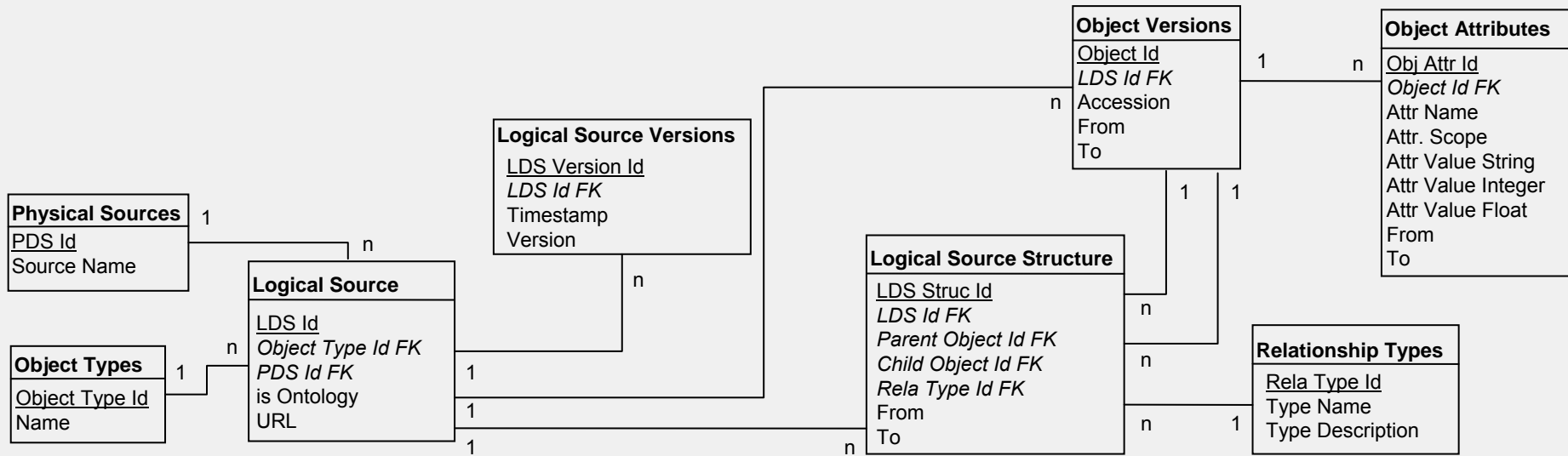
- Fokus auf IDs, Obsolete Status und Relationen
- Versionsorientiert: jede Version wird komplett in das Rep. eingespielt



- *Anforderungen an ein neues Modell*
 - Effizientere Speicherung von Ontologieversionen → Vermeidung von Redundanz
 - Einbeziehung weiterer Attribute (z.B. Synonyme, Definitionen, ...) → Chance weitere Evolutionsoperatoren bzw. Änderungen aufzudecken
 - Attribute sollen ‚Scope‘ unterstützen → auf was beziehen sich Attribute und zugehörige Werte
 - Versionierung auf Konzeptebene, wie auch auf Attribut- und Relationsebene
 - Einfacher Import neuerer Ontologieversionen



Modell im Überblick



| Tmp Objects |
|------------------|
| <u>Accession</u> |

| Tmp Attributes |
|------------------|
| <u>Accession</u> |
| <u>Name</u> |
| <u>Scope</u> |
| <u>Value</u> |

| Tmp Relationships |
|-------------------|
| <u>Parent_acc</u> |
| <u>Child_acc</u> |
| <u>Rel_type</u> |

| Added Objects |
|------------------|
| <u>Accession</u> |

| Added Attributes |
|------------------|
| <u>Accession</u> |
| <u>Name</u> |
| <u>Scope</u> |
| <u>Value</u> |

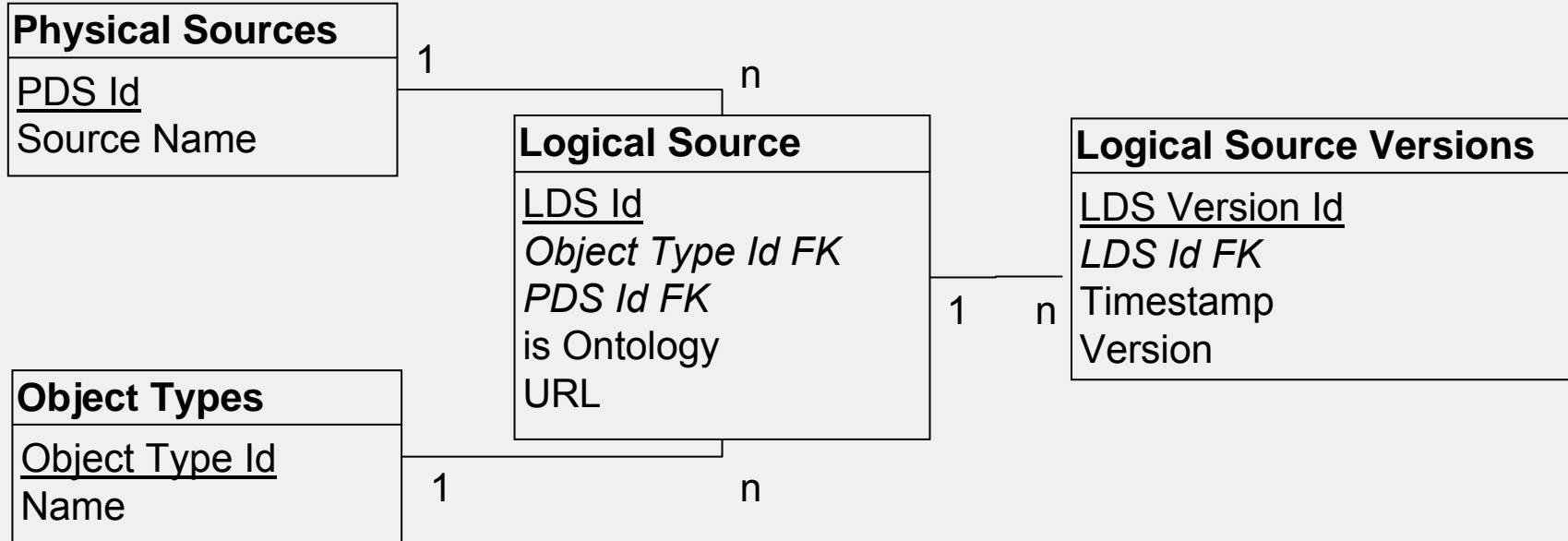
| Added Relationships |
|---------------------|
| <u>Parent_acc</u> |
| <u>Child_acc</u> |
| <u>Rel_type</u> |

| Deleted Objects |
|-----------------|
| <u>Obj Id</u> |

| Deleted Attributes |
|--------------------|
| <u>Obj id</u> |
| <u>Name</u> |
| <u>Scope</u> |
| <u>Value</u> |

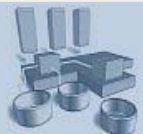
| Deleted Relationships |
|-----------------------|
| <u>Parent obj id</u> |
| <u>Child obj id</u> |
| <u>Rel_type_id</u> |

Verwaltung der Quellen und Versionen

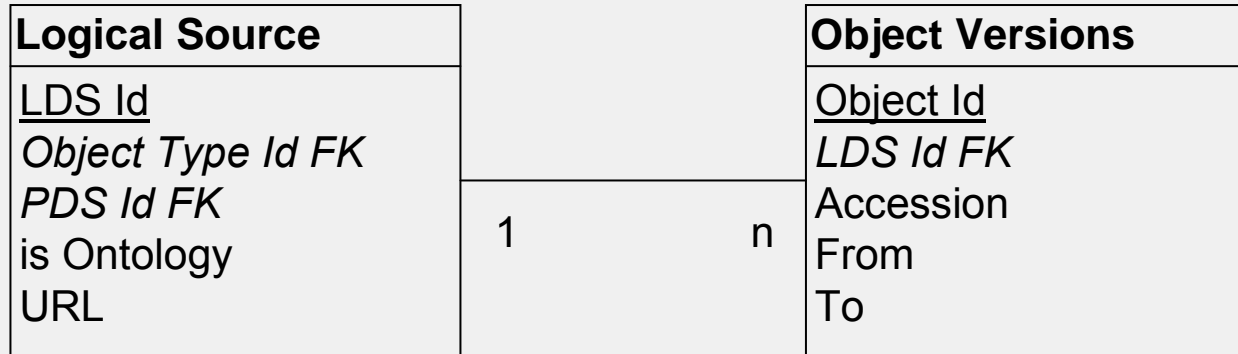


• Quellen und Versionen

- Logische Datenquellen = Physische Datenquelle + Objekttyp
- Jede logische Datenquelle besitzt mindestens eine LDS Version
- LDS-Version: LDS + Timestamp + Versionsnummer
- Unterscheidung von Ontologien und sonstigen Quellen

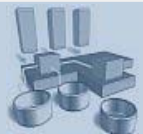


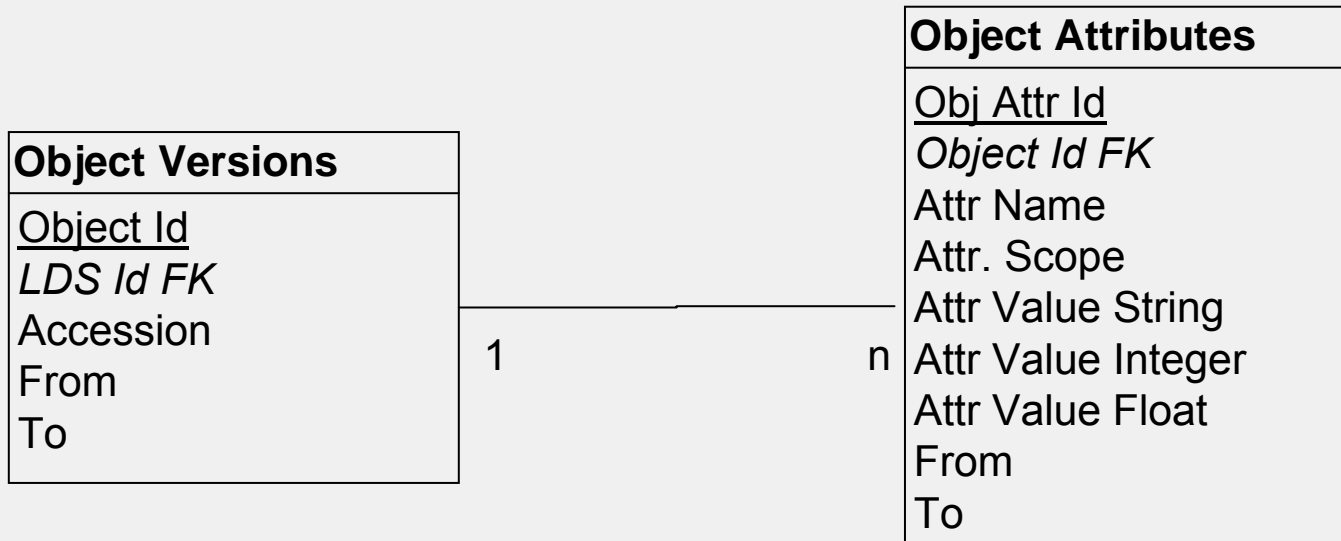
Objekte (Konzepte) einer Quelle



- *Objekte einer Quelle*

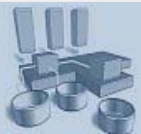
- Jedes Objekt besitzt eine Accession (eindeutige ID)
- Objekt ist ein Gültigkeitszeitraum zugeordnet
 - Z.B. GO:0045308 (protein amino acid binding) ist gültig von (2002-12 bis 2005-01)
 - From-To Notation → Zuordnung zu LDS-Versions möglich
 - Rekonstruktion von Datenquellen zu einem bestimmten Zeitpunkt



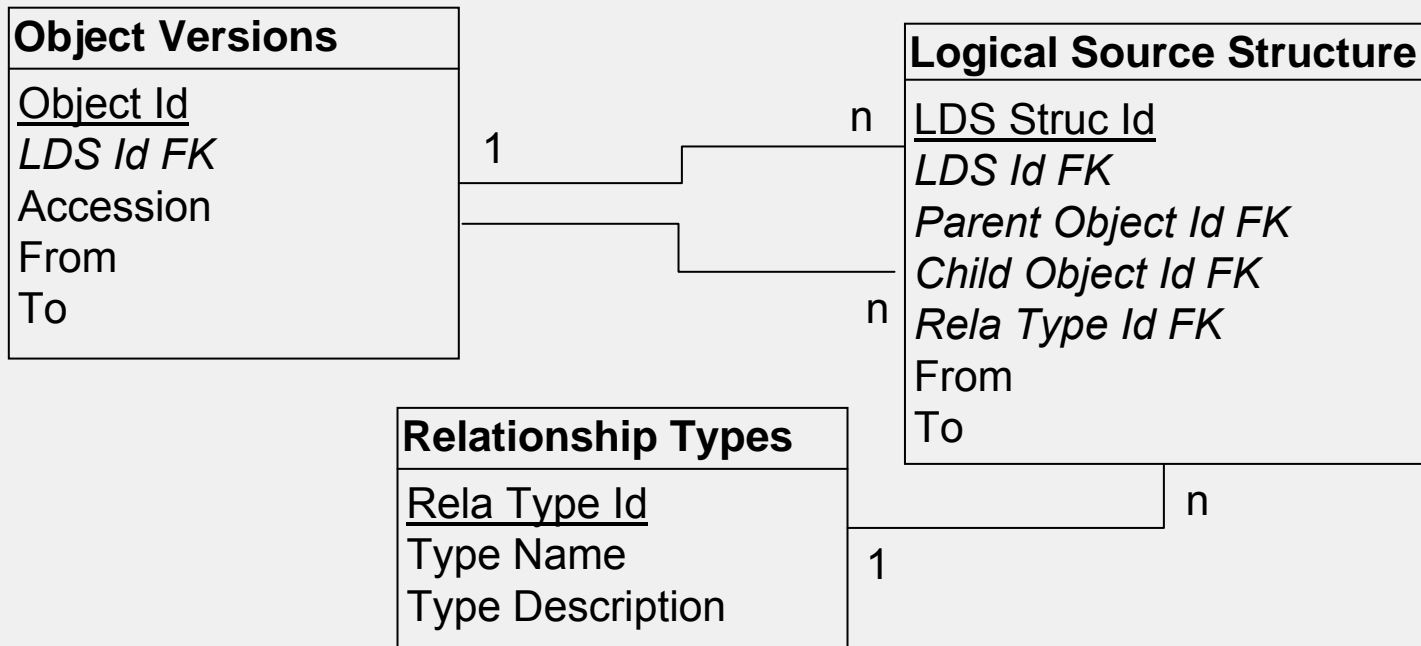


- *Objektattribute*

- Jedes Objekt verfügt über eine unterschiedliche Anzahl von Attributen
- Attribute werden ebenfalls mit Gültigkeitszeiträume annotiert
- Scope eines Attributes im Falle von Synonymen
 - ‚alt_id‘: alternative ID
 - ‚exact‘, ‚narrow‘, ‚broad‘

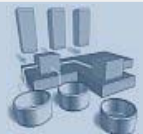


Relationships



• Relationships

- Source – Target Beschreibung der Relationen mit Hilfe der ObjIDs
- Gültigkeitszeiträume für Relationen
- Typen von Relationen, z.B. ‚is_a‘, ‚part_of‘, ...

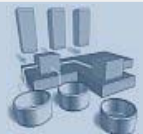


Temporäre Tabellen für Import

| Tmp Objects | Tmp Attributes | Tmp Relationships | Added Objects | Added Attributes | Added Relationships |
|------------------|---|--|------------------------|---|---|
| <u>Accession</u> | <u>Accession</u> <u>Name</u> <u>Scope</u> <u>Value</u> | <u>Parent_Acc</u> <u>Child_Acc</u> <u>Rel_Type</u> | <u>Accession</u> | <u>Accession</u> <u>Name</u> <u>Scope</u> <u>Value</u> | <u>Parent_Acc</u> <u>Child_Acc</u> <u>Rel_Type</u> |
| | | | Deleted Objects | Deleted Attributes | Deleted Relationships |
| | | | <u>Obj_Id</u> | <u>Obj_Id</u> <u>Name</u> <u>Scope</u> <u>Value</u> | <u>Parent_Obj_Id</u> <u>Child_Obj_Id</u> <u>Rel_Type_Id</u> |

- *Temporäre Tabellen*

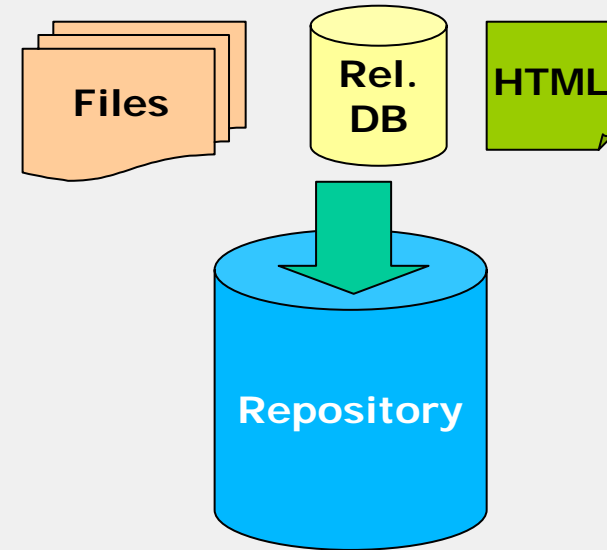
- Erfassung von Objekten, Relationen und Attributen einer neuen Quellversion beim Import
- Zusätzlich Tabellen zum erfassen der gelöschten bzw. hinzugefügten Objekte / Relationen / Attribute



Gestaltung des Imports neuer Quellversionen

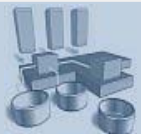
- *Einfache Beschreibung einer Quellversion*

- Objekttyp + physische Quelle, Ontologie: ja/nein
- Timestamp und Versionsnummer einer Quellversion
- Location: JDBC-Adresse, URL, File, ...
- Verweis auf Importerklasse
- Import mehrerer Versionen über ein File möglich



- *Zweistufiger Importansatz*

- Quellenspezifische 1. Stufe
 - Zugriff auf Quellversion mit Hilfe der Informationen über die Location, z.B. Download eines OBO-Files, Scannen einer Website, ...
 - Extraktion von Objekten, Relationen und Attributen → Integration in die temporären Tabellen (tmp_objects, tmp_relationships, tmp_attributes)
- Generische 2. Stufe
 - Abgleich der aktuellen Versionen in den Tmp-Tabellen mit letzter verfügbarer Version im Rep. → Berechnung des Diff für add und delete Tabellen
 - Integration des Diffs in das Repository



Beispiel eines Imports - I

- *Process@GeneOntology vom 2007-06 (letzte Version 2007-05)*

Objekte der neuen Version:

| <i>accession</i> | <i>name</i> |
|------------------|---------------------------|
| GO:0055048 | anastral spindle assembly |
| GO:0055049 | astral spindle assembly |
| GO:0051225 | spindle assembly |
| | ... |

Aktuell gültige Objekte im Rep.:

| <i>accession</i> | <i>name</i> | <i>from-to</i> |
|------------------|---------------------------|----------------|
| GO:0051227 | mitodic spindle assembly | 2005-01 - |
| GO:0051226 | meitodic spindle assembly | 2005-01 - |
| GO:0051225 | spindle assembly | 2005-01 - |
| | ... | ... |

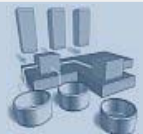


Hinzugefügte Objekte:

| <i>accession</i> | <i>name</i> |
|------------------|---------------------------|
| GO:0055048 | anastral spindle assembly |
| GO:0055049 | astral spindle assembly |
| | ... |

Gelöschte Objekte:

| <i>accession</i> | <i>name</i> |
|------------------|---------------------------|
| GO:0051227 | mitodic spindle assembly |
| GO:0051226 | meitodic spindle assembly |
| | ... |



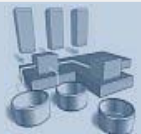
Beispiel eines Imports - II

- *Process@GeneOntology vom 2007-06 (letzte Version 2007-05)*

Nach Import im Rep.:

| <i>accession</i> | <i>name</i> | <i>from-to</i> |
|------------------|---------------------------|-------------------|
| GO:0051227 | mitotic spindle assembly | 2005-01 – 2007-05 |
| GO:0051226 | meitotic spindle assembly | 2005-01 – 2007-05 |
| GO:0055048 | anastral spindle assembly | 2007-06 - |
| GO:0055049 | astral spindle assembly | 2007-06 - |
| GO:0051225 | spindle assembly | 2005-01 - |
| ... | ... | ... |

- Ebenfalls für *Attribute* auf Basis (accession, name, scope, value) und *Relationen* auf Basis (sourceAcc, targetAcc, rel_Type)



Vermeidung von Redundanzen - Objekte

- Vergleich der Anzahl aller Objekte in den Versionen mit den Objekten im Repository

| Objekt-Typ | Quelle | #Objekte (Rep.) | #Objekte (alle Vers.) | #Versionen | Anteil in % |
|--------------------|--------------------------------------|-----------------|-----------------------|------------|-------------|
| BioChemistryEntity | ChemicalEntitiesOfBiomedicalInterest | 21.905 | 435.256 | 31 | 5,0 |
| AnatomicalEntity | CellTypeOntology | 923 | 15.276 | 20 | 6,0 |
| AnatomicalEntity | FlyAnatomyOntology | 6.252 | 91.964 | 15 | 6,8 |
| Concept | FlyBaseControlledVocabulary | 715 | 5.462 | 8 | 13,1 |
| Process | GeneOntology | 15.666 | 634.772 | 59 | 2,5 |
| Component | GeneOntology | 2.309 | 101.250 | 59 | 2,3 |
| Function | GeneOntology | 9.129 | 445.149 | 59 | 2,1 |
| AnatomicalEntity | AdultMouseAnatomyOntology | 2.764 | 47.590 | 18 | 5,8 |
| Experiment | ProteinProteinInteractionOntology | 949 | 17.194 | 23 | 5,5 |
| Protein | ProteinModificationOntology | 1.379 | 17.136 | 16 | 8,0 |
| PhenotypeEntity | MammalianPhenotypeOntology | 6.403 | 106.934 | 19 | 6,0 |
| AnatomicalEntity | PlantStructureOntology | 865 | 18.769 | 24 | 4,6 |
| Process | PathwayOntology | 624 | 13.567 | 25 | 4,6 |
| Concept | SequenceOntology | 1.589 | 35.727 | 29 | 4,4 |
| AnatomicalEntity | ZebraFishAnatomyOntology | 2.375 | 22.506 | 13 | 10,6 |
| HealthEntity | NCIThesaurus | 66.294 | 2.417.508 | 48 | 2,7 |

- Ca. 98 – 85 % Einsparung gegenüber der alten Verfahrensweise
- Abhängigkeit:* Wie viele Versionen einer Quelle liegen vor und wie viele Änderungen treten bei einem Versionswechsel auf?



Vermeidung von Redundanzen - Relationen

- Vergleich der Anzahl aller Relationen in den Versionen mit den Relationen im Repository

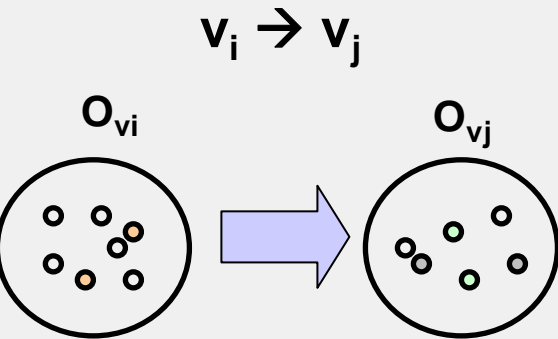
| Objekt-Typ | Quelle | #Relationen (Rep.) | #Relationen (alle Vers.) | #Versionen | Anteil in % |
|--------------------|--------------------------------------|--------------------|--------------------------|------------|-------------|
| BioChemistryEntity | ChemicalEntitiesOfBiomedicalInterest | 46.943 | 624.761 | 31 | 7,5 |
| AnatomicalEntity | CellTypeOntology | 1.842 | 28.214 | 20 | 6,5 |
| AnatomicalEntity | FlyAnatomyOntology | 10.748 | 151.416 | 15 | 7,1 |
| Concept | FlyBaseControlledVocabulary | 755 | 5.340 | 8 | 14,1 |
| Process | GeneOntology | 36.644 | 1.009.647 | 59 | 3,6 |
| Component | GeneOntology | 5.541 | 154.519 | 59 | 3,6 |
| Function | GeneOntology | 13.968 | 478.748 | 59 | 2,9 |
| AnatomicalEntity | AdultMouseAnatomyOntology | 3.680 | 59.085 | 18 | 6,2 |
| Experiment | ProteinProteinInteractionOntology | 1.099 | 17.984 | 23 | 6,1 |
| Protein | ProteinModificationOntology | 2.343 | 25.418 | 16 | 9,2 |
| PhenotypeEntity | MammalianPhenotypeOntology | 8.372 | 119.066 | 19 | 7,0 |
| AnatomicalEntity | PlantStructureOntology | 1.329 | 27.398 | 24 | 4,9 |
| Process | PathwayOntology | 783 | 15.451 | 25 | 5,1 |
| Concept | SequenceOntology | 2.412 | 44.722 | 29 | 5,4 |
| AnatomicalEntity | ZebraFishAnatomyOntology | 9.154 | 72.370 | 13 | 12,6 |
| HealthEntity | NCIThesaurus | 91.979 | 2.758.921 | 48 | 3,3 |

- Ca. 97 – 85 % Einsparung gegenüber der alten Verfahrensweise



Evolutionsooperatoren im bisherigen Modell

Versionswechsel

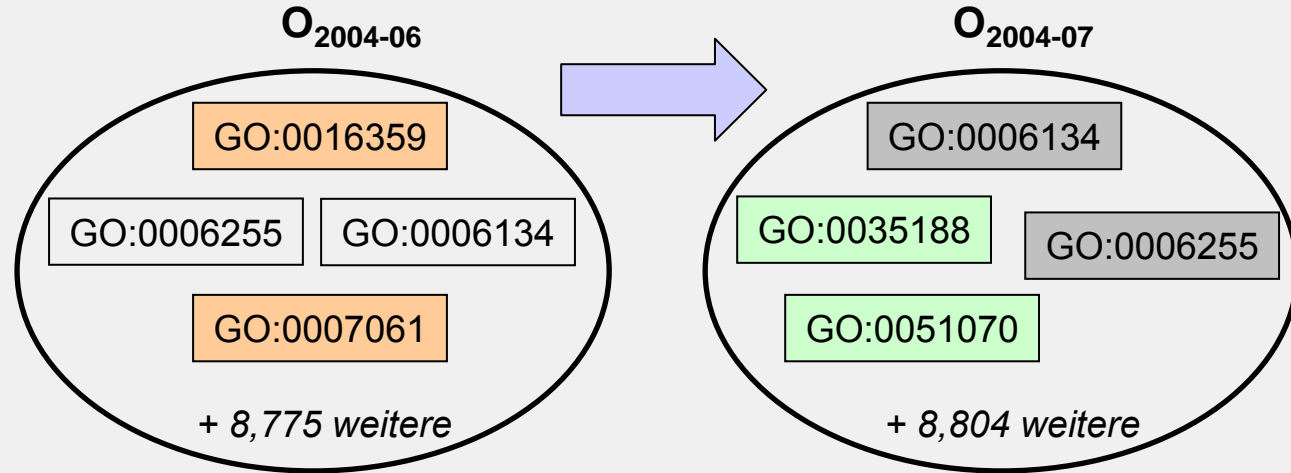


$$add_{v_i, v_j} = O_{v_j} \setminus O_{v_i}$$

$$del_{v_i, v_j} = O_{v_i} \setminus O_{v_j}$$

$$toObs_{v_i, v_j} = O_{v_j, obs} \cap O_{v_i, nonObs}$$

Beispiel: Evolution von Process@GO



Veränderte Objekte:



add_{2004-06,2004-07}



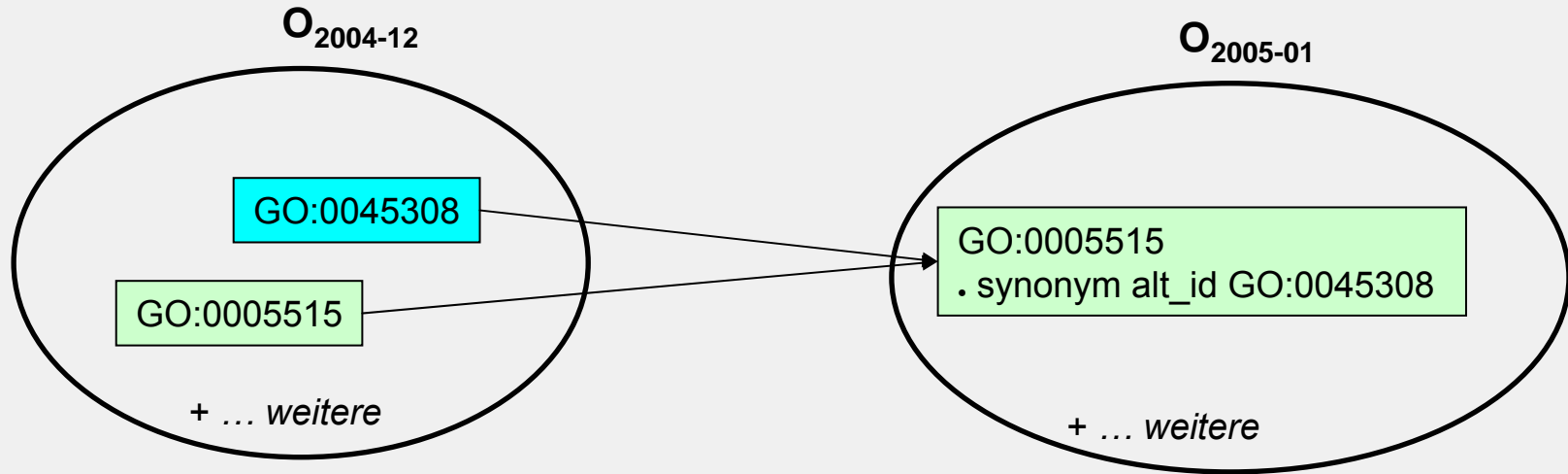
del_{2004-06,2004-07}



toObs_{2004-06,2004-07}

Neuer Evolutionsoperator - Fuse

- Bsp.: „protein amino acid binding“ (GO:0045308) wurde in 2005-01 (bestehend seit 2002-12) zu „protein binding“ fusioniert (GO:0005515)

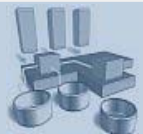


- *Allgemein*
 - Konzept $C1$ in der alten Version v_i wird mit einem Konzept $C2$ der neuen Version v_j fusioniert
 - Erkennung:
 - $C1 \in del_{v_i, v_j}$
 - $C1_{\text{accession}} \in C2_{(\text{synonym alt_id})}$; $C2 \in O_{v_j}$



Nutzung von Fuse in GeneOntology

| From version | To version | Added concepts | Deleted concepts | Fused concepts | ToObsolete concepts |
|--------------|------------|---------------------|--------------------|--------------------|---------------------|
| 2002-12 | 2003-1 | 138 | 20 | 0 | 20 |
| 2003-1 | 2003-2 | 32 | 1 | 0 | 1 |
| 2003-2 | 2003-3 | 50 | 2 | 0 | 3 |
| 2003-3 | 2003-4 | 89 | 6 | 0 | 4 |
| 2003-4 | 2003-5 | 158 | 7 | 0 | 11 |
| 2003-5 | 2003-6 | 94 | 3 | 0 | 11 |
| 2003-6 | 2003-7 | 47 | 1 | 0 | 8 |
| 2003-7 | 2003-9 | 358 | 1 | 0 | 32 |
| 2003-9 | 2003-10 | 142 | 1 | 0 | 0 |
| 2003-10 | 2003-11 | 157 | 2 | 0 | 233 |
| 2003-11 | 2003-12 | 64 | 0 | 0 | 11 |
| 2003-12 | 2004-1 | 87 | 1 | 0 | 2 |
| 2004-1 | 2004-2 | 64 | 4 | 0 | 3 |
| 2004-2 | 2004-3 | 233 | 1 | 0 | 11 |
| 2004-3 | 2004-5 | 228 | 7 | 0 | 18 |
| 2004-5 | 2004-6 | 155 | 1 | 0 | 0 |
| 2004-6 | 2004-7 | 31 | 2 | 0 | 16 |
| 2004-7 | 2004-9 | 116 | 38 | 0 | 13 |
| 2004-9 | 2005-1 | 390 | 1 | 20 | 23 |
| 2005-1 | 2005-2 | 136 | 0 | 7 | 3 |
| 2005-2 | 2005-3 | 155 | 0 | 2 | 4 |
| | | .. | - | - | . |

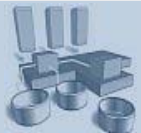


- **Konzeptebene**

- Grundidee: Menge von Objekten in Versionen
- Add, Delete, ToObs, Fuse, ...

- **Attributebene**

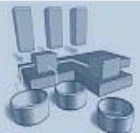
- Grundidee: Menge von Attributen eines Objekts (Konzepts) ebenfalls in Versionen
- *Add*
 - Einfügen eines neuen Attributs, z.B. Konzept bekommt eine Definition
- *Delete*
 - Löschen eines Attributs, z.B. ein nicht benötigtes Synonym wird gelöscht
- *Change*
 - Attributwert wird verändert, z.B. Namensänderung
 - Ersetzungen bei Synonymen → add / delete Kombination



Evolution auf Attributebene – Bsp. „protein binding“

Initial values and changes during evolution: (🟢 initialization in first version, + addition, ✖ deletion, 🔄 change/update)

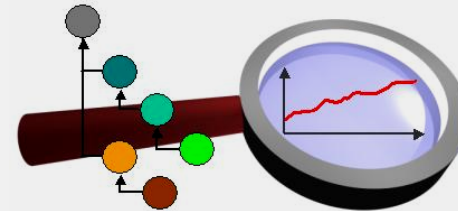
| Date of | | | | |
|---------|---------|---|---------|--|
| 20 | | | | <ul style="list-style-type: none"> ◆ calcium-dependent cell adhesion molecule activity ◆ calcium-independent cell adhesion molecule activity ◆ cell adhesion molecule activity ◆ cell adhesion receptor activity ◆ death receptor adaptor protein activity ◆ death receptor interacting protein activity ◆ death receptor-associated factor activity ◆ glycine receptor-associated protein ◆ nicotinic acetylcholine receptor-associated protein activity ◆ protein degradation tagging activity ◆ protein tagging activity ◆ receptor-associated protein activity ◆ transmembrane receptor protein serine/threonine kinase receptor-associated protein activity |
| 20 | 2005-9 | + | synonym | |
| 20 | 2005-11 | ✖ | synonym | <ul style="list-style-type: none"> ◆ cell adhesion receptor activity |
| 20 | 2006-1 | ✖ | synonym | <ul style="list-style-type: none"> ◆ transmembrane receptor protein serine/threonine kinase receptor-associated protein activity |
| 20 | 2007-10 | ✖ | synonym | <ul style="list-style-type: none"> ◆ calcium-dependent cell adhesion molecule activity ◆ calcium-independent cell adhesion molecule activity ◆ cell adhesion molecule activity ◆ receptor-associated protein activity |
| 20 | 2007-11 | ✖ | synonym | <ul style="list-style-type: none"> ◆ N-methyl-D-aspartate receptor-associated protein activity ◆ death receptor adaptor protein activity ◆ death receptor interacting protein activity ◆ death receptor-associated factor activity ◆ glycine receptor-associated protein ◆ nicotinic acetylcholine receptor-associated protein activity |



Umsetzung der bisherigen Ergebnisse in OnEX

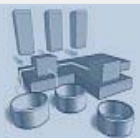
- Ontology Evolution Explorer

- Google Web Toolkit und Erweiterungen → Vortrag Shuangqing He
- 1. Teil: Quantitative Analyse der Evolution von Ontologien
- 2. Teil: Suche in Ontologien und Evolution in Attributen von Konzepten



OnEX
Ontology Evolution Explorer

| Ontology | Trend chart | Evolution in detail | First version date | Latest version date | #concepts (start) | #concepts (latest) | #relationships (start) | #relationships (latest) |
|---|-------------|---------------------|--------------------|---------------------|-------------------|--------------------|------------------------|-------------------------|
| BioChemistryEntity@ChemicalEntitiesOfBiomedicalInterest | | | 2004-10 | 2008-5 | 10236 | 19360 | 11592 | 33889 |
| AnatomicalEntity@CellTypeOntology | | | 2004-6 | 2008-5 | 687 | 857 | 1251 | 1586 |
| AnatomicalEntity@FlyAnatomyOntology | | | 2004-12 | 2007-12 | 6090 | 6222 | 9826 | 10467 |
| Concept@FlyBaseControlledVocabulary | | | 2005-11 | 2008-5 | 658 | 713 | 653 | 698 |
| Process@GeneOntology | | | 2002-12 | 2008-6 | 6741 | 15264 | 0 | 29614 |
| Component@GeneOntology | | | 2002-12 | 2008-6 | 1124 | 2215 | 0 | 4124 |
| Function@GeneOntology | | | 2002-12 | 2008-6 | 5298 | 8850 | 0 | 10255 |
| AnatomicalEntity@AdultMouseAnatomyOntology | | | 2005-8 | 2008-5 | 2416 | 2753 | 2939 | 3463 |
| Experiment@ProteinProteinInteractionOntology | | | 2005-8 | 2008-6 | 194 | 948 | 211 | 989 |
| Protein@ProteinModificationOntology | | | 2006-6 | 2008-4 | 1074 | 1338 | 1568 | 1982 |
| PhenotypeEntity@MammalianPhenotypeOntology | | | 2005-8 | 2008-5 | 4175 | 6338 | 4620 | 7126 |
| AnatomicalEntity@PlantStructureOntology | | | 2005-7 | 2008-6 | 681 | 854 | 980 | 1257 |
| Process@PathwayOntology | | | 2005-11 | 2008-5 | 427 | 600 | 478 | 696 |
| Concept@SequenceOntology | | | 2005-8 | 2008-5 | 981 | 1504 | 1181 | 1935 |
| AnatomicalEntity@ZebraFishAnatomyOntology | | | 2005-11 | 2008-4 | 1389 | 2236 | 4272 | 7603 |
| HealthEntity@NCIThesaurus | | | 2003-10 | 2008-4 | 28740 | 66170 | 33847 | 74872 |



UNIVERSITÄT LEIPZIG

Abteilung Datenbanken

Erweiterungen und Realisierungen im Rahmen der
Evolutionanalyse von Datenquellen und Mappings

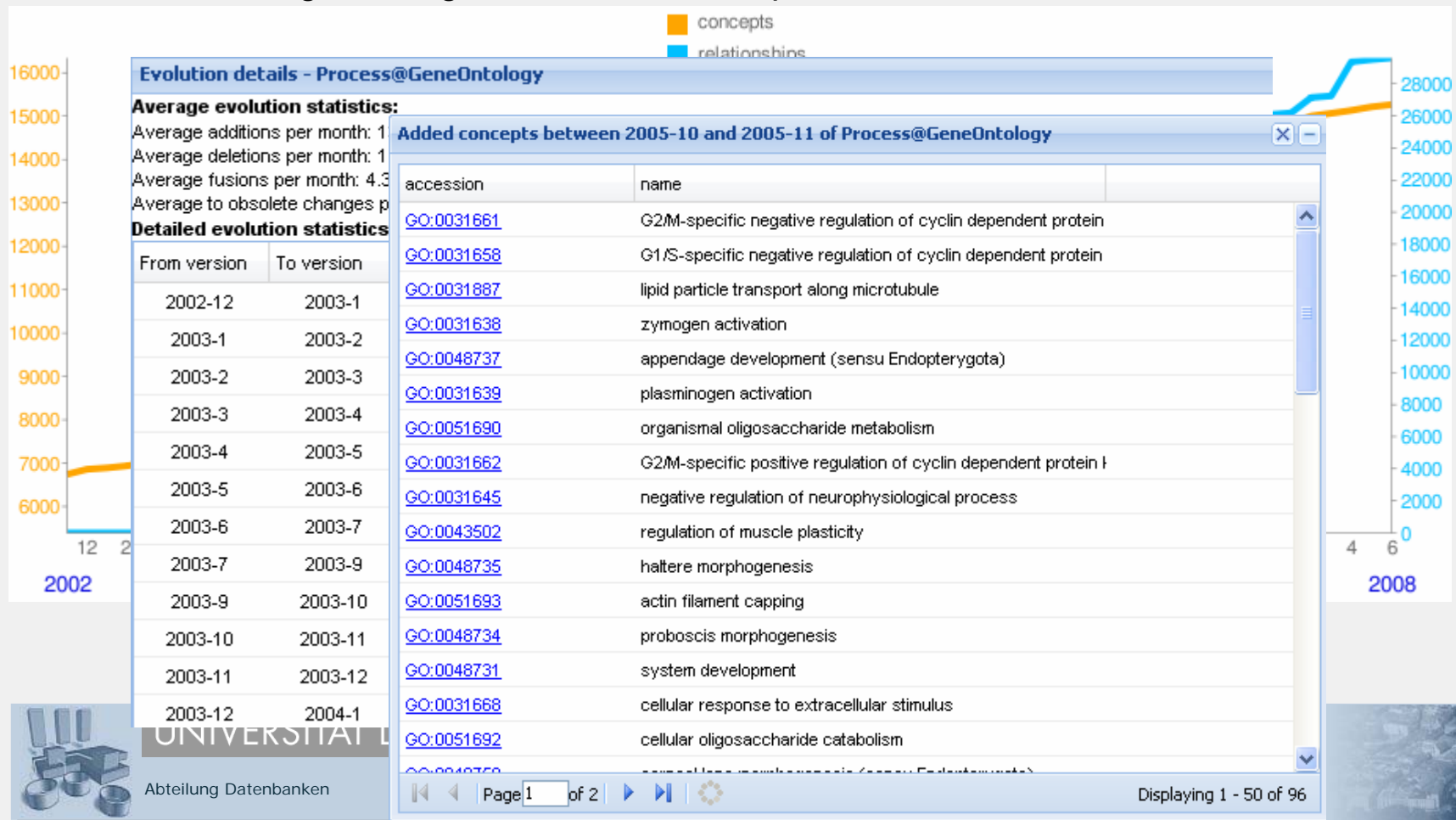
Michael Hartung Zingst, 2008



OnEX – Szenarien I

- *Szenarien der quantitativen Evolutionsanalyse*

- Generierung von Trend Charts für einen ersten Überblick über die Evolution
- Navigieren zu Detailinformationen über Versionsübergänge (add, del, toObs, fuse)
- Nochmaliges Navigieren zu den Konzepten, welche verändert worden



OnEX – Szenarien II

- Szenarien der konzept-basierten Evolutionsanalyse
 - Suche mittels Keywords oder direkt über Accession in verfügbaren Ontologien
 - Präsentation von Suchergebnissen inkl. Kurzinfo über Evolution
 - Details zur Evolution eines einzelnen Konzepts und dessen Attributen

Search in concept names or enter a concept ID directly

Search results - 206 matches in 8 ontologies
(concept ar Evolution of GO:0005515 in Function@GeneOntology

Latest version and basic information:

GO:0005515 as of 2008-6

| | |
|-------------------|--|
| Created: | 2002-12 |
| Related concepts: | GO:0045308 |
| name: | protein binding |
| isObsolete: | false |
| definition: | Interacting selectively with any protein or protein complex (a complex of two or more proteins that may include other nonprotein molecules). |
| synonym(s): | <ul style="list-style-type: none"> ◆ GO:0045308 ◆ protein amino acid binding ◆ alpha-2 macroglobulin receptor-associated protein activity ◆ protein degradation tagging activity ◆ protein tagging activity |

Initial values and changes during evolution: (initialization in first version, addition, deletion, change/update)

| Date of change | Type of change | Changed attribute | New value(s) | Old value(s) |
|----------------|----------------|-------------------|--|---|
| 2002-12 | | isObsolete | ◆ false | |
| 2002-12 | | name | ◆ protein binding | |
| 2003-2 | | definition | ◆ Interacting selectively with any protein. definition_ | |
| 2003-3 | | definition | ◆ Interacting selectively with any protein, or protein complex (a complex of two or more proteins that may include other non-protein molecules). | ◆ Interacting selectively with any protein. definition_ |
| 2003-4 | | name | ◆ protein binding activity | ◆ protein binding |
| | | | ◆ Interacting selectively with any protein, or protein | ◆ Interacting selectively with any protein, or protein |



• **Infrastruktur zur Analyse von Evolution in Datenquellen**

- Effiziente Verwaltung vers. Versionen von Datenquellen in einem Repository
- Einführung weiterer Evolutionsoperatoren auf Basis des generischen Modells der DILS-Analyse
- Zugriff auf die Evolutionsergebnisse über das OnEX System
 - Ergebnisse der quantitativen Analyse
 - Suche nach Konzepten und Details über die Evolution von Konzepten

• **Ausblick**

- Definition, Integration und Erkennung weiterer Evolutionsoperatoren
 - Split eines Konzepts
 - Strukturelle Veränderungen mit Fokus auf den Relationships
- Anwendung des Evolutionskonzepts in anderen Domänen → z.B. bibliographische Domäne: Entwicklung von Forschungsfeldern

