Combining Provenance Management and Schema Evolution

Tanja Auge

University of Rostock
Institute of Computer Science
Data Research Group

10.07.2018

Problem

- Application: Research data management
- Goal: Traceability, reconstructibility and replicability of the path from data collection to publication
- Idea: Saving
  - evaluation query
  - result database
  - original database $\rightarrow$ minimal sub-database
- Question: Which additional information is required to be able to reconstruct a minimal part of the database?
- Constraint: Evaluation of provenance queries with changing data and schemas

Our Aim

- Aim: Combining Provenance Queries, Data and Schema Evolution ⇒ CHASE-based methods
- CHASE:
  - Chasing dependencies into databases or queries
  - Applications: Query Reformulation, **Data Exchange**, **Inverse Schema Mappings**
- Evaluation query: s-t tgds
  ⇒ Chasing s-t tgds into original database
- Calculating minimal sub-database: Inverse mapping of the result database (CHASE-inverse, ...)
- If no inverse exists: Adding Provenance information (Provenance polynomials, ...)
- Expressing (data/schema) evolution: s-t tgds
  ⇒ Chasing s-t tgds into original database

Provenance Management and Schema Evolution