



**Hasso
Plattner
Institut**

IT Systems Engineering | Universität Potsdam

Completeness-driven Query Answering in Peer Data Management Systems (PDMS)

Armin Roth, Felix Naumann

DASI '06, Bertinoro

Completeness-driven Query Answering in PDMS

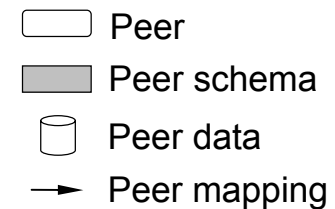
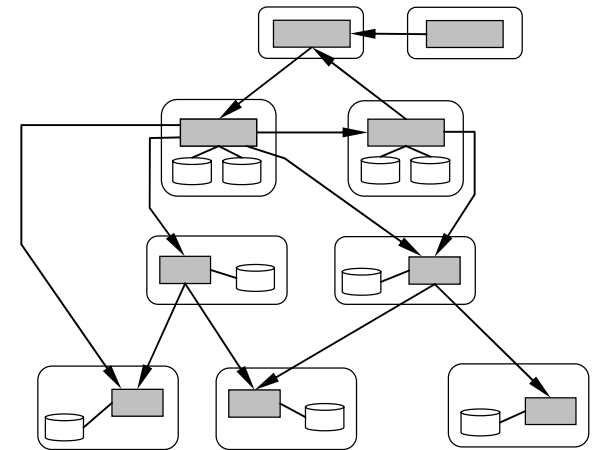
- PDMS

- Network of mediators
- (Directed) schema mappings

[Halevy et al. ICDE 2003,
Franconi et al. DBISP2P 2003,
Calvanese et al. PODS 2004]

- Information loss in mapping paths:

- Selections:
implicit knowledge about peers
- Projections in mappings:
different information about
a real-world entity



Projections in Mappings

- Motivation

- Select a hotel in Milano in expedia.de:
Name and location, but no contact data
- Search for contact data of the hotel on google.de

- Formalization

- Global schema:

`Hotel(name, loc, phone, eMail, prize)`

- **Source 1:** `Expedia.Hotel(name, loc, prize)`

- **Source 2:** `Google.Hotel(name, loc, phone, eMail)`

- (LaV) mappings (including projections)

$\{ (n, l, pr) \mid \text{Expedia.Hotel}(n, l, pr) \} \subseteq \text{Hotel}(n, l, \textit{ph}, m, pr)$

$\{ (n, l, \textit{ph}, m) \mid \text{Google.Hotel}(n, l, p, m) \} \subseteq \text{Hotel}(n, l, \textit{ph}, m, \textit{pr})$

- Query answers (may) contain no NULL values

Observations

- Information loss in mappings:
 - Projections introduce NULL values (i.e., unknown)
 - Selections reduce recall
 - Both accumulate along mapping paths
- Query answering in PDMS:
 - Generally undecidable, in PTIME data complexity for reduced expressivity [[Halevy et al. ICDE 2003](#)]
 - High redundancy in network of mappings
 - Massive scalability problems even with tens of peers
 - Query execution time extremely depends on configuration of the PDMS

