

# Schema Mapping Breakout Session

AnHai Doan, Howard Ho, Sergey Melnik,  
**Renée J. Miller**, Dan Suciu, Val Tannen, Ben  
Ashpole, Michael Fitzmaurice, Greg  
Karvounarakis

# Nature of the Problem

- Schema Mapping
  - Representation or specification of the semantics or interpretation of the relationship between models of data
- Problems
  - Creation, maintenance, management, understanding (debugging), interpretation
  - Use (in data exchange, query reformulation, updates, data synchronization, etc.)
  - Specification formalism

# State of the Art (Research)

- Factoring the problem space by mapping management
- Ontology alignment
  - Independent of data transformation
- Mappings for structured data
  - formal semantics for structured relational and XML data
  - GLAV mappings
- Enterprise environments and some work on mappings for peer data sharing

# State of the Art (Research)

- GLAV mappings
  - for relational and nested data, but little work on other models
  - express relationship between queries on two data sources
  - formal representation of structural data transformations and incompleteness
  - solid foundation on:
    - creation, maintenance, (some) debugging
    - use in data exchange, query reformulation, (some) updates
    - operations on mappings (composition, inversion, etc.)

# State of the Practice

- Languages
  - views and queries
    - beginning to see limited use of GLAV
  - procedural mapping languages most common
    - scripting languages, eg., XSLT, general programming languages
    - object-relational mapping
  - many proprietary solutions without formal semantics
    - ETL workflow scripts
    - schema annotation frameworks

# State of the Practice

- Tools
  - visual tools to help in mapping creation, but creation still largely manual
  - debugging manual
  - execution engines for data transformations are pervasive

# Solved Problems, Victories

- Execution engines (optimization, scaling)
- Mapping creation (some impact on products and there is a clear awareness of need)
- Impact on products
  - BEA AquaLogic,
  - MS ADO.NET,
  - IBM WebSphere Information Integrator,
  - IBM Rational Data Architect

# Unsolved Problems

- Formal mapping specification
  - Uncertainty, probabilistic, approximate
  - Do more for other models than relational and XML, eg., unstructured data
  - More general logics, eg., negation, aggregation, recursion, higher-order
  - Formalisms for ETL, EAI (aka web services)
  - Visual specifications
  - Standardization
  - Mappings between Web service specifications
  - Streaming, sensor data

# Unsolved Problems

- Unifying theories of mapping usage (eg., interchange = integration + exchange, integration + update propagation)
- Shared datasets and benchmarks (need to learn from ontology alignment contests)
  - measures of success
- Tools that scale with schema, specification, network complexity
- Usability
  - understanding and debugging mappings
  - design theory for mappings
- Mapping in non-enterprise environments
  - dynamic web or networked environments
- Optimization in execution engines to exploit redundancy
- Runtime environments (synchronization, exception handling)
- Interaction between mappings and privacy

# Challenge domains

- Healthcare
- Life sciences
- Ecology
- Homeland security
- Scientific discovery
- Web communities

# Community challenges and resources

- Incentive systems for creating and sharing mappings and tools
  - Incentives administered by funding agencies
- Standardization
- Educational component, training