Il in context

Barbara Blaustein
Peter Buneman
Sarah Cohen-Boulakia
Robert Chadduck
Yi Chen
Laura Haas
Tim Finin
James French

Yannis Ioannidis
Subbarao
Kambhampati
Deborah McGuinness
Victor Markowitz
Wang-Chiew Tan
Kenneth Thibodeau
Gary Walter
Usability is the key

Source 1
Source 2
Source n

Integration
underware/knitware

Result

Data Brokers

Data Consumers
Information consumers

• What users want
  – General-purpose integration – provide a big/complete picture of the sources
  – Special purpose – provide a specific query/table view
  – Best effort/precise
  – Static/dynamic
  – Longitudinal integration, synchrony
Information consumers (2)

• What do users want in addition to the “data”?  
  – understanding of quality  
  – explanation of how the data was/were derived  
  – purpose for which it was derived  
  – provenance  
  – explanation  
  – context information (external factors)

• All of these are related
Information Brokers

- Education is the key to everything. There is no silver bullet
- Domain experts need to understand (at least) data models and their semantics
- Computer scientists need to understand the domain (and the semantics of data models)
- Is there a need for a class of “data modellers”? Are they ontologists?
Ontologies

- Do they help? Are they different?
- Is the fact that they are “open world” or “more flexible” important?
- General feeling that there is a continuum between schemas and ontologies, which may work both ways in integration
- The issues of “context” apply to both.
Source selection

• Discovery

• Quality
  – authorship, trust, completeness, performance, cost, staleness, synchrony, cleanliness, availability…
Source understanding

• How do brokers understand the data and create mappings?
  – looking at data more useful than looking at schemas
  – does the richness of the schema/ontology help or hinder?
  – DOCUMENTATION IS ESSENTIAL! (it is a form of annotation that must be included in integration.)
Privacy and Security

• Does integration compromise security and privacy?
  – Goes both ways: we use integration to anonymize and de-anonymize data

• Security may “block” provenance.
Explanation

- Presentation – simple to start with but need to expand
- Granularity of explanation (location of source)
- Annotation/documentation
- Where and how (data and workflow provenance)
- What was the process?
- Where did this data go?

All of this should be carried through in further integration
Usability is the key

Input sources, services sources and curated sources...

- Quality: Trust, completeness, performance, cost, staleness, synchrony
- Sources – data and process
- Explanation / Provenance
  - Documentation/annotation
- Education
  - no silver bullet