Towards Data Services

Some Preliminary Research Issues

Schahram Dustdar

Hong-Linh Truong

Distributed Systems Group

Institute of Information Systems

Vienna University of Technology

Austria

Joint Work with DB&AI Group: Reinhard Pichler and Vadim Savenkovy

Data Integration

- The Internet of Things and the Internet of Services
  - So much data and so little knowledge about agreements, data quality, and e.g., compliance while using the data

- Data needed for a particular purpose
  - Spans different locations
    - even when data belongs to the same organization
  - Provided by different providers (e.g., companies)
    - Competing or complementary (combinations of) providers
  - Adheres to different policies and laws
    - e.g., the law imposed by a country on data hosted by a company in the country

Some more issues …

- Search and query data: users do *not* want to search all sources, identify relevant sources (e.g., peers), then query and aggregate data

- Data quality/license/policy: users find the data but actually do not know under *which* conditions the data can be used
  - No mechanism to associate quality/license/policy with data sources
  - No automatic data selection based on data concerns

Service-oriented Data Integration

- Combine SOA key concepts and the concept of (peer) data networks with quality considerations in the context of the Internet
  The approach
  - Data is provided under a data as a service (DaaS) concept
    - DaaS (is a Web service)
      - A DaaS might integrate different data sources (including replicated/distributed data)
      - A DaaS might integrate different DaaS (DaaS composition)
  - DaaS is characterized by data concerns
    - Location, metadata, quality of data, licenses, laws, compliance regulations, etc.
  - Data is selected and “composed“ (aggregated) based on data concerns and user request

Data Publishing Services

- What are main concerns when using data?

- How to model data concerns (e.g., QoD - quality of data), e.g., reliability, consistency, completeness, etc. or data-specific aspects of service quality (e.g., recency of data, or query-specific performance), and data service licensing (DSL)
  - QoS focuses on operational aspects of services like performance, reliability, availability, and security

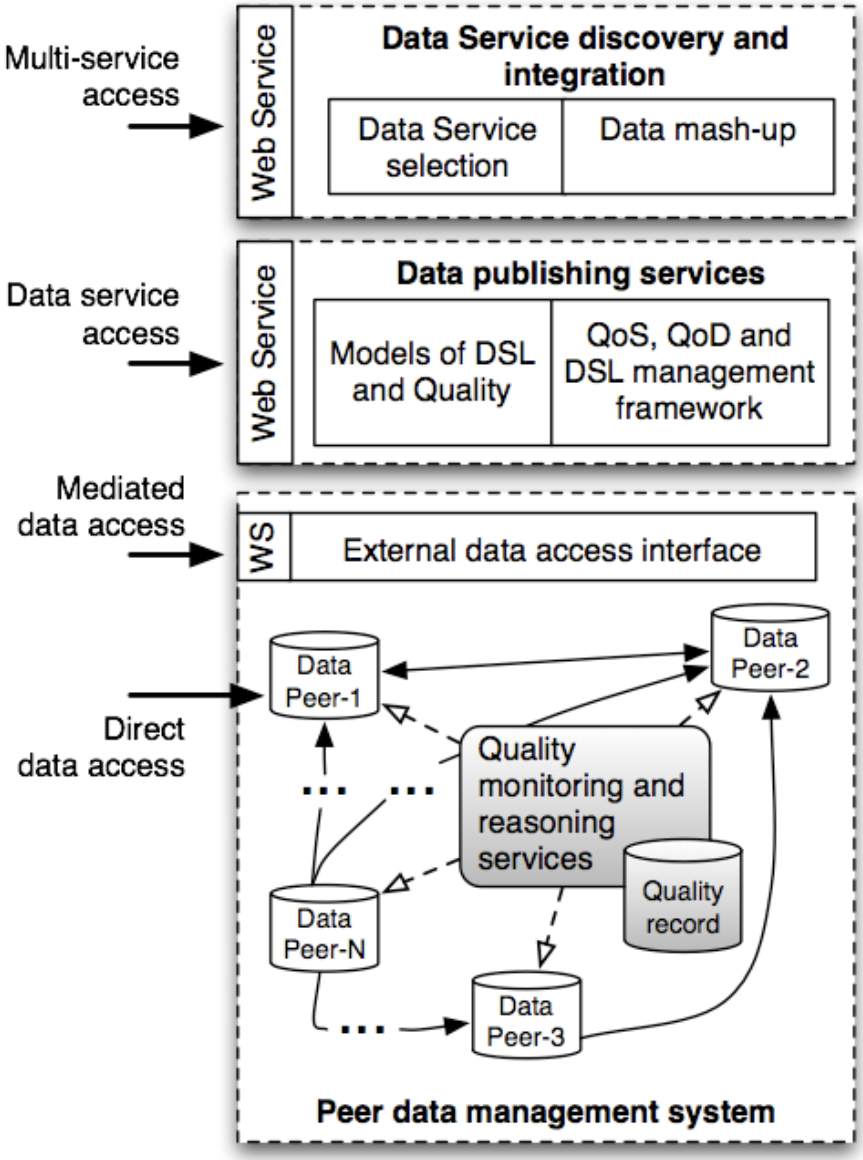
- How to manage DSL, data concerns associated with data?
  - Different models for different services for different purposes
  - Changing DSL/concerns
  - Runtime data selection and usage
  - Metrics

Data service discovery and selection

- How to select DaaS based on user-specific requests and concerns associated with data?
- When can different DaaSs be combined to provide a user request? (mixed data)
- How to compose DaaSs to provide a composite DaaS?
- How to support data mashups of business and e.g., e-science data according to customized needs?
- How to optimize data mashups?

Planned Architecture

- **Peer data management system**
  - Extended traditional P2P data management systems based on quality monitoring and reasoning
- **Data publishing services**
  - Characterize concerns of DaaS
- **Data service discovery and integration**
  - Data selection and mashup



Thanks for your attention!

Schahram Dustdar
Distributed Systems Group
Institute of Information Systems
Vienna University of Technology
Austria

http://www.infosys.tuwien.ac.at/Staff/sd/