Data Service Agreements: Toward a Data Supply Chain

Len Seligman, Arnon Rosenthal
The MITRE Corporation
{seligman, arnie}@mitre.org

James Caverlee*
Georgia Institute of Technology
caverlee@cc.gatech.edu

* Work performed while at The MITRE Corporation.
Background

- Organizations seek looser coupling between data providers and consumers
  - E.g., Pub/sub, web services, semantic web, grid
  - More flexible
  - Rapid discovery and use of new information sources
- Challenges
  - Semantic interoperability
  - Semantics management
    - Actively creating new areas of agreement
    - Is needed data actually being collected?
  - Evolution of data supply chains despite decentralized control

See our SIGMOD Record paper, Sept. 2004
A Motivating Example

- How can the customers distinguish between:
  - The data publication is one-time only
  - The data has been published for 3 years, but is about to be discontinued
  - The data will continue to be dependably provided for the next 5 years
Another Example

FUEL_DAILY: Estimated fuel supplies in gallons of depots A and B updated daily

FUEL_WEEKLY: Estimated fuel supplies in liters of depots A and B updated weekly

Who is affected? What are my obligations?
Hold on, we rely on this data!

Typical changes
- Semantics
  - New Definition
  - New Legal Value
- Data quality
- Representation
  - New unit of measure

IIWeb 2004
The Opportunity: Identify and Support Data Obligations

• The current state-of-the-practice is inadequate
  – Either rigid configuration management boards with long approval and change cycles; or
  – Informal, ad hoc agreements with no supporting technical infrastructure

• The problem will only get more challenging
  – Information services (think Web services, grid, p2p) are growing in popularity
  – Organizations increasingly build value-added services on top of data resources they do not control
Our Proposed Solution

• The Data Service Agreement (DSA)
  – Documents data arrangements in formal machine-consumable agreement specifications
  – A DSA identifies:
    • Parties involved – the provider and the customer
    • Data obligations – on both the provider and customer side
    • Exceptional conditions – what to do when obligations are not met
Benefits of the DSA Framework

- Promote an increase in data sharing
  - Ease of adoption
  - Reduced costs through automation
  - Reusable agreement components, etc.
- Increased clarity through explicit obligations
- Technical infrastructure for creating, maintaining, monitoring, enforcing, etc.
- Support tools for:
  - Change impact analysis
  - Violation detection
  - Data flow analysis (e.g. who are critical providers?)
DSAs are Modeled after Contracts

• Contracts typically include:
  – The names and contact information for the participating parties
  – Obligations for all parties, including payment terms and deadlines for delivery
  – Identification of reasonable excuses for performance (due to forces beyond the party’s control); plus allowances to terminate the contract if problems continue
  – Recourse (or damages) for unexcused failure to perform (nonperformance)
What about SLAs?

- Service Level Agreements typically emphasize performance metrics like response time or availability.
- However, we are aware of no prior work on SLAs to address data obligations (e.g., the need to provide data of a certain quality at specified time intervals).
Provider Obligations

- A provider data obligation identifies:
  - What the data to be provided is
  - Constraints on the data (like quality, freshness, etc.)
  - Frequency of the data delivery
  - Preconditions that may hold
  - Restrictions on changes to the data

- Example 1: Data Delivery Obligation
  - The Army Fuel Service is obliged to post fuel depot data every 8 hours for the next 3 months
  - Exception: If the data is not updated every 8 hours, penalty of $X per hour late

- Example 2: Query Obligation
  - The Army Fuel Service will allow the Logistics Analyst to query the table ‘Fuel’ up to 10 times per day
  - Exception: If the Fuel Service changes its schema, it will notify each customer 30 days in advance
Customer Obligations

• A customer data obligation identifies:
  – How the data should be protected
  – How the data can be used
  – Appropriate compensation to the provider

• See the paper for more details…
Supply Chain Analogy

- An automobile manufacturer wants not only access to their suppliers’ stocks of parts; they want assurances that the parts will be available as needed throughout the manufacturing process.
- Similarly, the DSA approach supports ongoing information supply chains.
Our Research Agenda

- Develop a DSA toolkit for creating and monitoring agreements
  - Provide support for DSA monitoring with Agreement Rules
  - Evaluate multiple architectures with respect to level of intrusiveness
  - Identify the technical infrastructure and guidelines that real organizations will need to apply DSAs

- Analysis of DSA graphs
  - To determine what information resources are most critical to the operation of an enterprise

- Org 1
- Org 2
- Org 3

- N
- I
- G
- A
- D
- B
- J
- C
- K
- L
- M
- F
- N