Data Quality in Cooperative Information Systems

Monica Scannapieco

Dipartimento di Informatica e Sistemistica
Università di Roma “La Sapienza”, Italy

Supervisors: Carlo Batini, Tiziana Catarci

Istituto di Analisi dei Sistemi e Informatica
Consiglio Nazionale delle Ricerche, Italy

Supervisor: Paola Bertolazzi
Data Quality: a Multidimensional Concept

(Wang & Strong 1996)

Believability, Accuracy, Objectivity, Reputation, Value-Added, Relevancy, Timeliness, Completeness, ....

(Naumann 2002)

Accuracy, Completeness, Customer Support, Documentation, Availability, Latency, Price, Quality of Service, ....

(Redman 1996)

Content, Scope, Level of Detail, Composition, Accuracy, Completeness, Currency, Appropriateness, Interpretability, ....

(Jarke et al. 1999)

Correctness, Completeness, Minimality, Traceability, Interpretability, Metadata Evolution, Functionality, ....
Data Quality in Cooperative Information Systems

**PITFALL**
Exchanges of low quality data deteriorate the quality in each DB

**OPPORTUNITY**
Same data in different organizations enable quality improvement by comparisons
Data and Data Quality Model and Quality Query Language

- A model according to which cooperating organizations exchange data with associated quality values: Data and Data Quality Model (D2Q)
- A query language according to which cooperating organizations request for data with associated quality values: Quality Query Language (QQL)
Preliminary Results

- Model and Query Language XML-based
- Coupling of Data Graphs and Quality Graphs (one for each dimension)

**D2Q data graph**

- Name
- Surname
- TelephoneNumber
- VIA SALARIA 113
- ROMA
- ITALY
- 00198

**Accuracy D2Q quality graph**

- Accuracy_Name
- Accuracy_Surname
- Accuracy_TelephoneNumber
- Accuracy_City
- Accuracy_Country
- Accuracy_ZIPCode

Links to quality objects/values

Link from a data object/value
An Architecture for Quality Improvement in Cooperative Information Systems

- Quality Factory
- Gateway
- e-S e-Service
- Quality P&S Bulletin Board
- Recording Matcher
- Broker
- Query planning and on-line improvement
- Off-line improvement algorithms
- Rating Service
- Evaluating source reliability
- Interschema knowledge
  Service supporting knowledge
  Historical quality knowledge
- Notification Service
  Quality P&S Bulletin Board
  Enabling quality diffusion, i.e., through P&S paradigm
Strategies for Quality Improvement

The Broker selects the best quality data answering a query and sends it to the requester (query planning based on data quality optimization) and to other providers (On-Line Improvement).

The Record Matcher periodically compares exported data in order to improve their quality.

The notification service multicasts data quality changes.