

BIC technical workshop on International Cooperation in Trustworthy ICT

WG1 – Human oriented /citizen trust, privacy and security

Data Provenance

Priscila Solís Barreto

Department of Computer Science

University of Brasilia, Brazil



Introduction

- When we see data on the Web, do we know
 - where it came from?
 - how it got there?
- This information (provenance) is typically lost in the process of copying/ transcribing/ transforming databases
- Provenance is essential to data integrity, currency and reliability.



Introduction and Some Domains

- Motivating Domains
 - Scientific Domains
 - Scientists deal with greater heterogeneity in data and metadata
 - Trust, quality, and copyright of data are significant when using third-party data
 - E-Science
 - Business Domains
 - Virtual organizations, workflows, warehouse environments where lineage information is used to trace the data in the warehouse view back to the source from which it was generated
 - Governamental Domains
 - In Brazil, within the social inclusion policies this is a very important issue. Ex Voting system, taxing system



Applications of Provenance

- Data Quality: use of lineage to estimate data quality and data reliability based on the source data and transformations
- Audit Trail: trace the audit trail of data, determine resource usage and errors in data generation.
- Replication Recipes: allow repetition of data derivation, help maintain its currency and re-do replication
- Attribution: the pedigree can establish the copyright and ownership of data, help to determine liability in case of erroneous data.
- Informational: use of lineage to query metadata for data discovery.



Some examples

- Collecting and modeling provenance from heterogeneous applications and data sources
- Integrating distributed and incomplete provenance information to compose complete provenance models
- Effective management and querying of distributed, semantic provenance repositories for different applications



Some possible actions

- Standardization of provenance models, services, and representations
- Provenance management architectures and techniques
- Analytic provenance and the relationship between provenance and visualization
- Provenance and the semantic web
- Human interpretation of provenance
- Security and privacy implications of provenance
- Provenance and social media
- Provenance implications for trust



International Cooperation

- Application in a specific and visible domain
 - Scientific, Business, Government
 - Multidisciplinary approaches
- Integration of the concept of trust and security in coordinated calls. Example: Smart Cities project, Cloud Computing, Digital Medias

