



Trust and Quality for Information Integration: The Data-Metadata-Ontology Continuum

Vipul Kashyap (kashyap@nlm.nih.gov)

National Library of Medicine, NIH

September 1, 2003

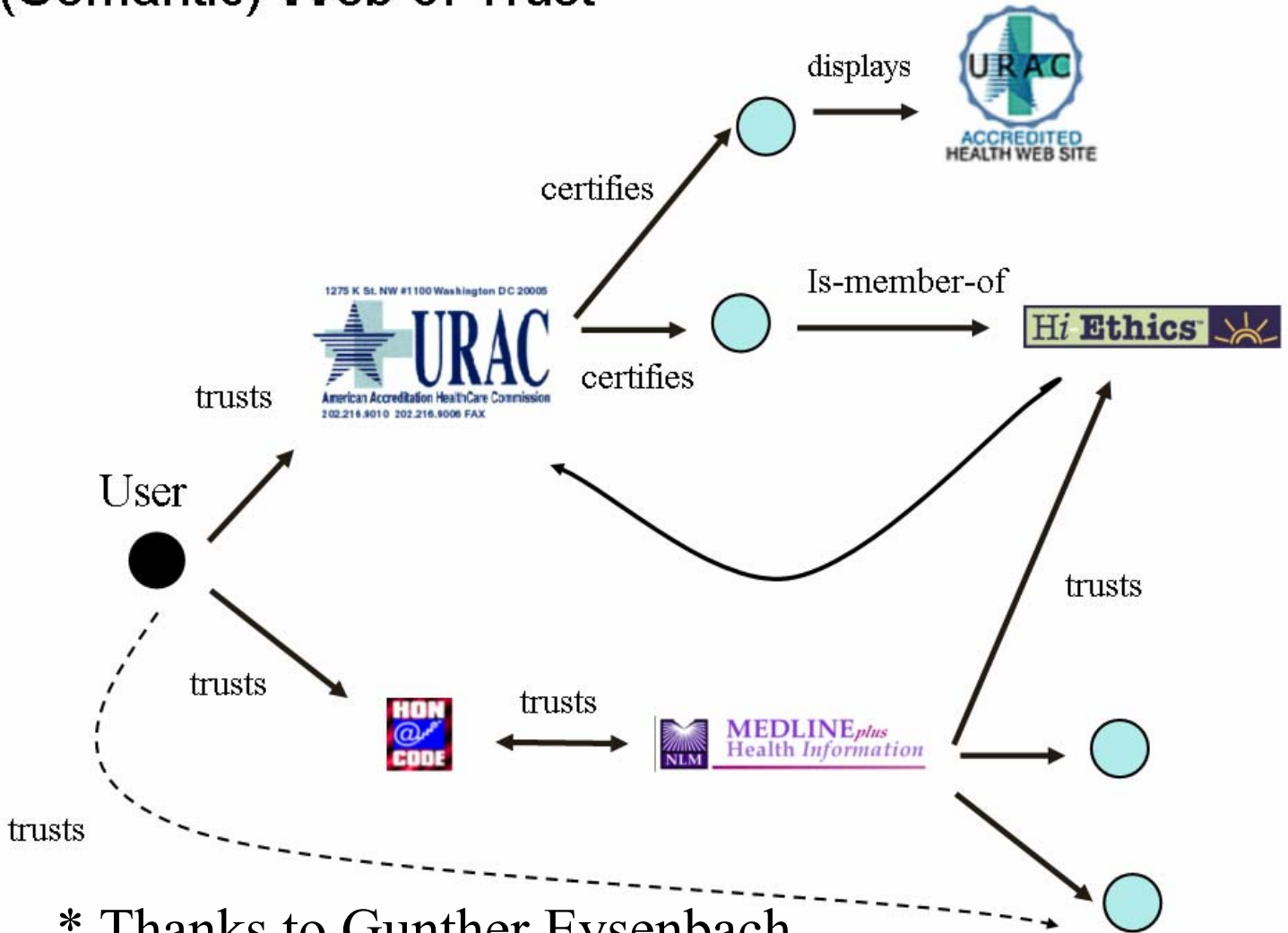
Workshop on Data Quality, Dagstuhl, Germany

The Importance of Quality of Information and Trust

PC AND PIXEL • Thach Bui

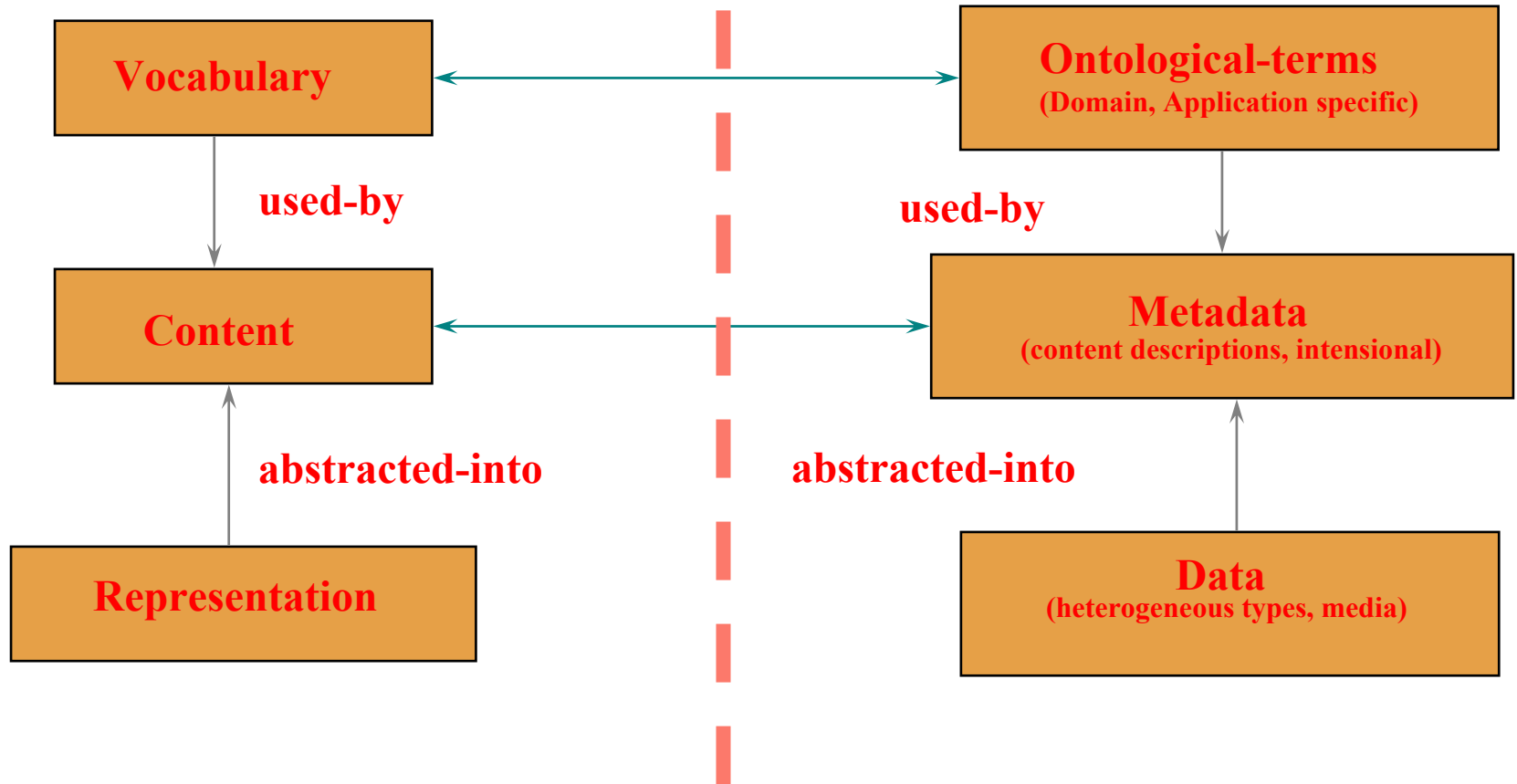


(Semantic) Web of Trust



* Thanks to Gunther Eysenbach

Semantic Web: Information Aspects



Role of Trust and Quality in Information Retrieval and Integration

- Data Quality
 - Is this information source reliable, trustworthy?
 - Does a particular information source have better quality of data?
 - Impacts:
 - Entity matching and identification (Information Retrieval)
 - Record and ID matching (Information Integration)
 - Resolution of conflicting information
- Relationship between DQ and DT
 - Is $DQ = f(DT)$
 - Or $Trust = f(DQ)$?
 - Or is there some notion of fixpoint computation?
- Relationship between the results and data quality
 - Answer = $f(DT, DQ)$?
 - Do these parameters induce a ranking on the set of results?

Role of Trust and Quality in Information Retrieval and Integration

- Metadata Quality (MQ) and Metadata Trust (MT)
 - Trust/Reliability of metadata exported by information sources
 - Trust/Reliability of mappings/morphisms exported by the information source
- Introduction of new dimensions into the equation:
 - $MQ = f(DQ, MT)$?
 - $MT = f(DT, MQ)$?
- MQ and internal consistency
 - Are the mappings internally consistent with each other?
 - Use of category theory based structures?
 - Formalism to provide a mathematical basis for data quality?

Role of Trust and Quality in Information Retrieval and Integration

- Ontology Quality (OQ)
 - Structural Quality of the ontology
 - Notion of semantic richness
 - Notion of internal consistency (no contradictions)
 - Notion of completeness of domain coverage
 - “Atomic Quality” of the ontology: more directly correlated with Ontological Trust (OT)
 - Quality of concepts and relationships
 - Quality of axioms and constraints (for semantically rich ontologies)
 - Notion of “Ontological commitments”
- $OQ = f(\text{Structural Quality})$
 - Employ graph-based comparison approaches
- $OT = f(\text{Atomic Quality})$
 - Investigate “cultural” consensus analysis approaches

Role of Trust and Quality in Information Retrieval and Integration

- The Goal:
 - Identify and formalize interrelationships between the following dimensions
 - DQ/DT
 - MQ/MT
 - OQ/OT
 - Is it generalizable beyond information retrieval and integration?
 - Build on existing information retrieval and integration research
- Evaluation of End-to-End Impact
 - Quality/Trust of answers
 - $f(\text{DQ}, \text{DT}, \text{MQ}, \text{MT}, \text{OQ}, \text{OT}) ?$
 - Ranking of answers