The importance of processes for data quality assessment

Gerhard Navratil

TU Vienna

Problem

- Quality discussion based on technical processes (observation, classification etc.)
- What are cadastral processes?
  - Legal processes?
  - Socio-technical processes? (Ottens, 2004)
- How to deal with quality in such processes?
  E.g. uncertainty reduction and absorption (Bédard, 1986)

Outline

- What do we have?
- 5-tier ontology
- Quality concepts and ontology – how do they connect?
- What is missing?
- Conclusions and future work

Existing Concepts of Quality

- **Data Quality** = Lineage, accuracy, completeness, logical consistency, semantic accuracy, currency (Guptil and Morrison, 1995)
  other categorizations e.g. by Wang and Strong (1996) and Veregin (1999)
- **Uncertainty** = error, vagueness, ambiguity, discord (Fisher, 1999)
- **Fitness for use** (Chrisman, 1984)

5-Tier Ontology

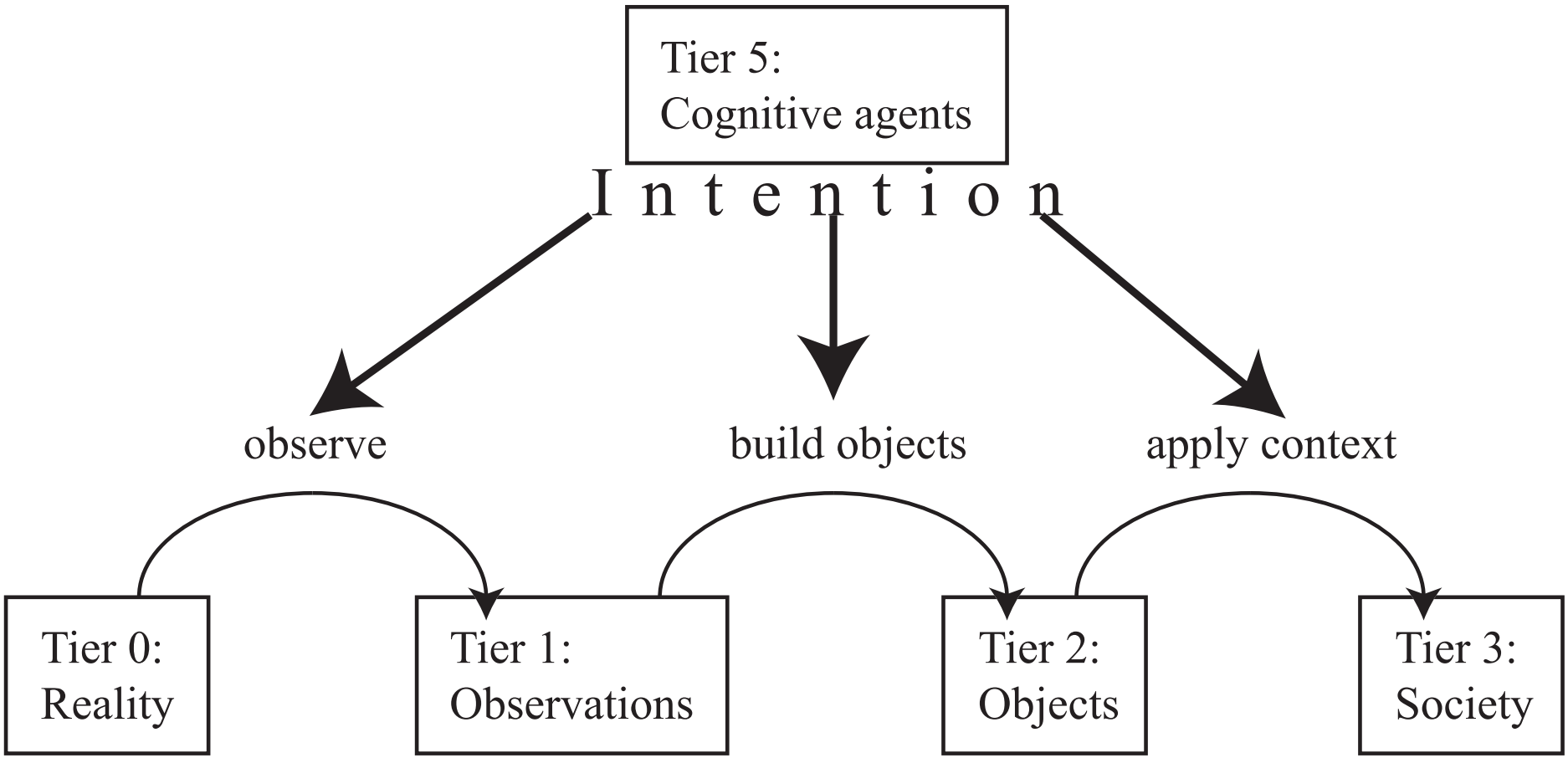
- Tier 0: Physical environment
- Tier 1: Observations of the environment
- Tier 2: World of objects
- Tier 3: Socially constructed reality
- Tier 4: Subjective reality of cognitive agents

Connection to concepts of quality?

Connection Quality/Ontology (1)

- Tier 0 ➔Tier 1: **Observation** process
- Tier 1 ➔Tier 2: **Classification** process (building objects)
- Tier 2 ➔Tier 3: Process of applying **social context**
- Tier 3 ➔Tier 4: Process of **selecting** observations used to obtain knowledge – influenced by intentions, results in a subjective view

Connection Quality/Ontology (2)



Connection Quality/Ontology (3)

- **Observation process**: Statistical methods, described by **data quality** parameters
- **Classification process**: Requires concepts for classification (vagueness), described by **uncertainty** parameters
- **Selection process**: Subjective views on reality, **fitness for use** describes correspondence of producer view with user view

Connection Quality/Ontology (4)

- Tier 0 no quality needed
- Tier 1 uses Data quality
- Tier 2 uses Uncertainty
- Tier 3
- Tier 4 uses Fitness for Use

What is Missing?

- No concept for social context
- No method to capture the quality of legal processes
- No detailed analysis of legal decision making: Judges do not use statistical measures (Twaroch 2005)

Conclusions and Future Work (1)

- Data quality, uncertainty, and fitness for use describe the quality of specific processes
- Concepts do not fully describe the quality of processes in the social context
- Quality measures for processes in the social context are necessary

Conclusions and Future Work (2)

- First step: Investigate processes in the social context – how do they deal with data quality?
- Land readjustment could provide an example where technology, society, and users are influential