Data Quality Problems

Single-Source Problems

Multi-Source Problems

Schema Level

Instance Level

Schema Level

Instance Level

(Lack of integrity
constraints, poor
schema design)

(Data entry errors)

(Heterogeneous
data models and
schema designs)

(Overlapping,
contradicting and
inconsistent data)

- Uniqueness
- Referential integrity
...

- Misspellings
- Redundancy/duplicates
- Contradictory values
...

- Naming conflicts
- Structural conflicts
...

- Inconsistent aggregating
- Inconsistent timing
...