“Information Rich” = “Messy Data” … Let’s deal with it!

- Messy data is inevitable
  - My use case: Making sense of Structured Data on the Web
  - Assumptions: Loads of data from heterogeneous providers, no way they all agree on a joint schema, no way they provide consistent information

- Be Tolerant
  - Analysis and Repair are key (our analyses show that lots of otherwise useful data is not readable for consumption [HUH+12]
  - Reasoning is needed:
    - No agreement upfront on schemas, but “linkage” between used vocabularies can be fruitfully exploited, e.g. a la [HHP09],
    - We also need reasoning & a query language that can take confidence/trust into account [HBPS11], [ZLPS12]
    - But: Keep it Simple, e.g. a la OWL LD [GHKP12]

**Conclusion:** Information-rich programming needs to allow for consuming and making sense of inconsistent, partial, and wrong information!
References


