

LOD fit for purpose?

Jacco van Ossenbruggen
CWI & VUA

Data quality is relative to the task

- Would you **trust** LOD to control a nuclear power plant?
 - Where are the boundaries?
- How do you make this assessment?
 - Do you really know what your **task** requires?
 - Do the **tools** provide sufficient information?
 - Is there sufficient information in the **data**?

From generic to specific

- Generic:
 - LOD **publishers** like their **data** to be used and reused for as many tasks as possible by as many different users as possible
 - LOD infrastructure **engineers** like their **platforms** to be used and reused for as many tasks as possible by as many different users as possible
- Specific:
 - LOD **users** like to know if the **data & platform** is fit for the purpose of their current task

Trade off: performance for transparency

- FvH: “Semantics = predictable inference”
- But often, we **cannot** predict what a specific tool will do on a specific, but new dataset
- Use case: **vocabulary alignment**
 - performance of alignment tool on new data
 - **Predict** it before running?
 - **Explain** it after?
- The argument for low tech: many use cases for 81% recall (simple & predictable) over 84% recall (high tech, black box, DL, ...)