

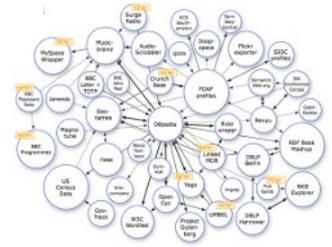
Implementierung von Metriken zur Qualitätsbewertung von RDF-Daten

LOD Cloud

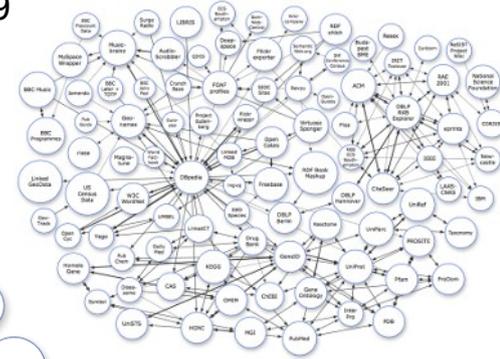
2007



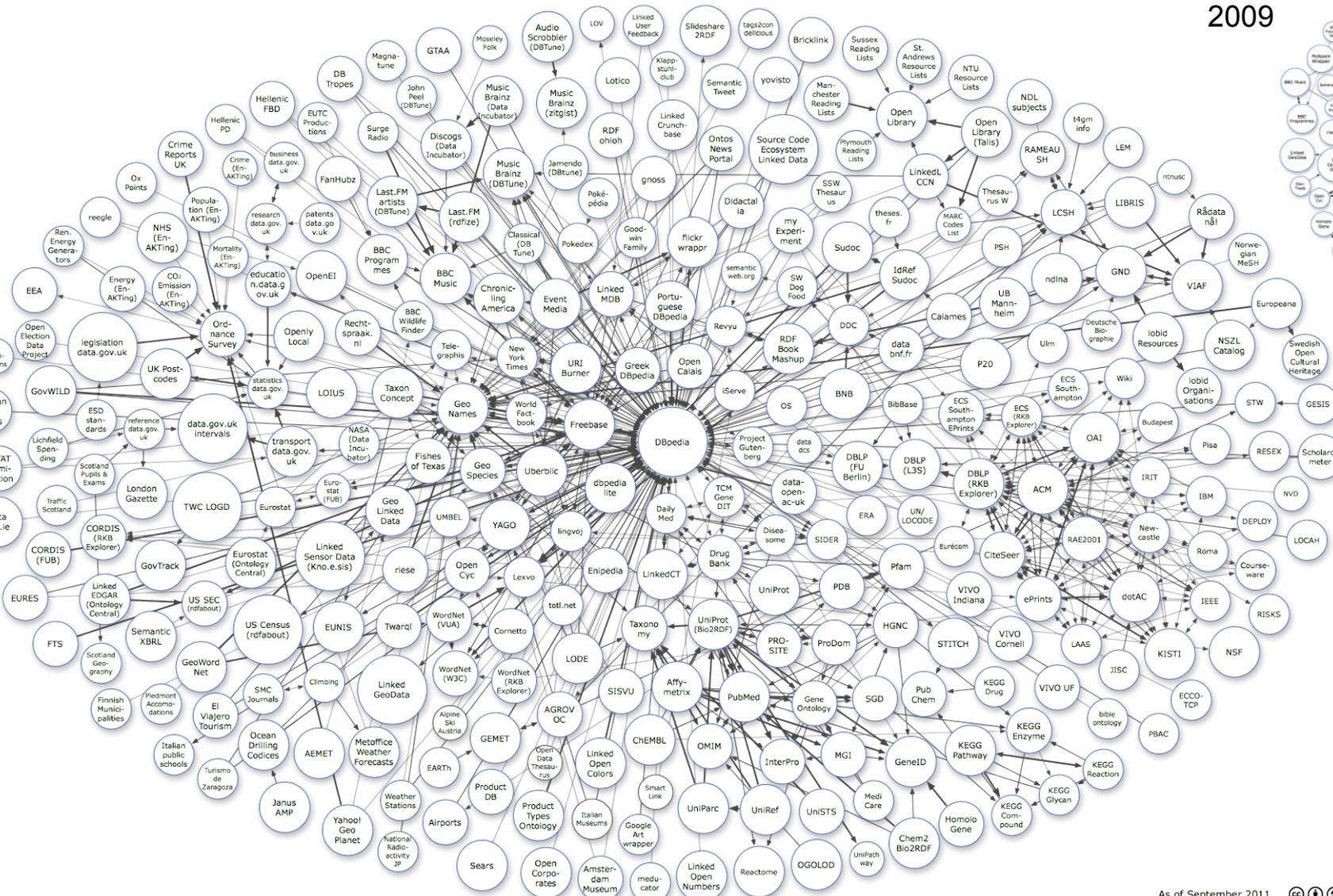
2008



2009

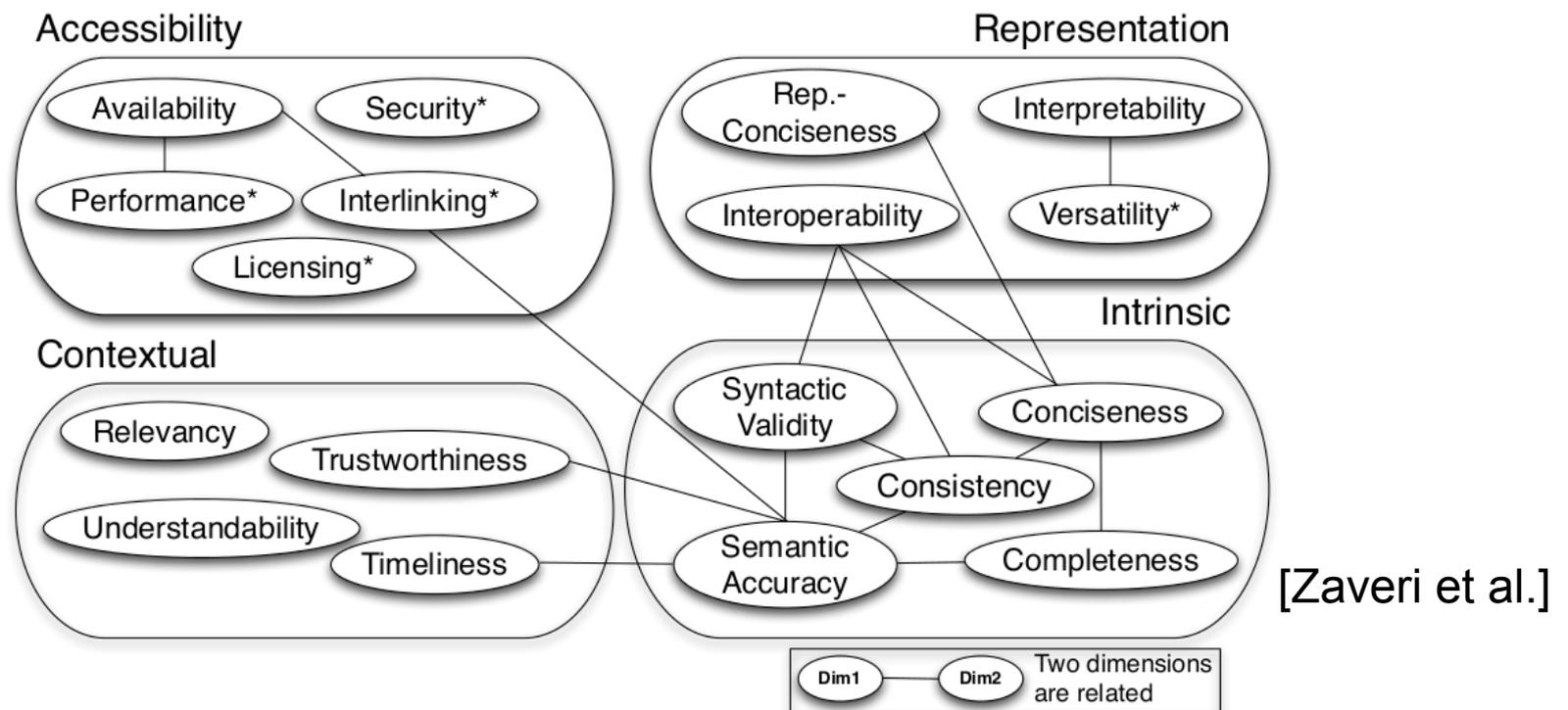


2011



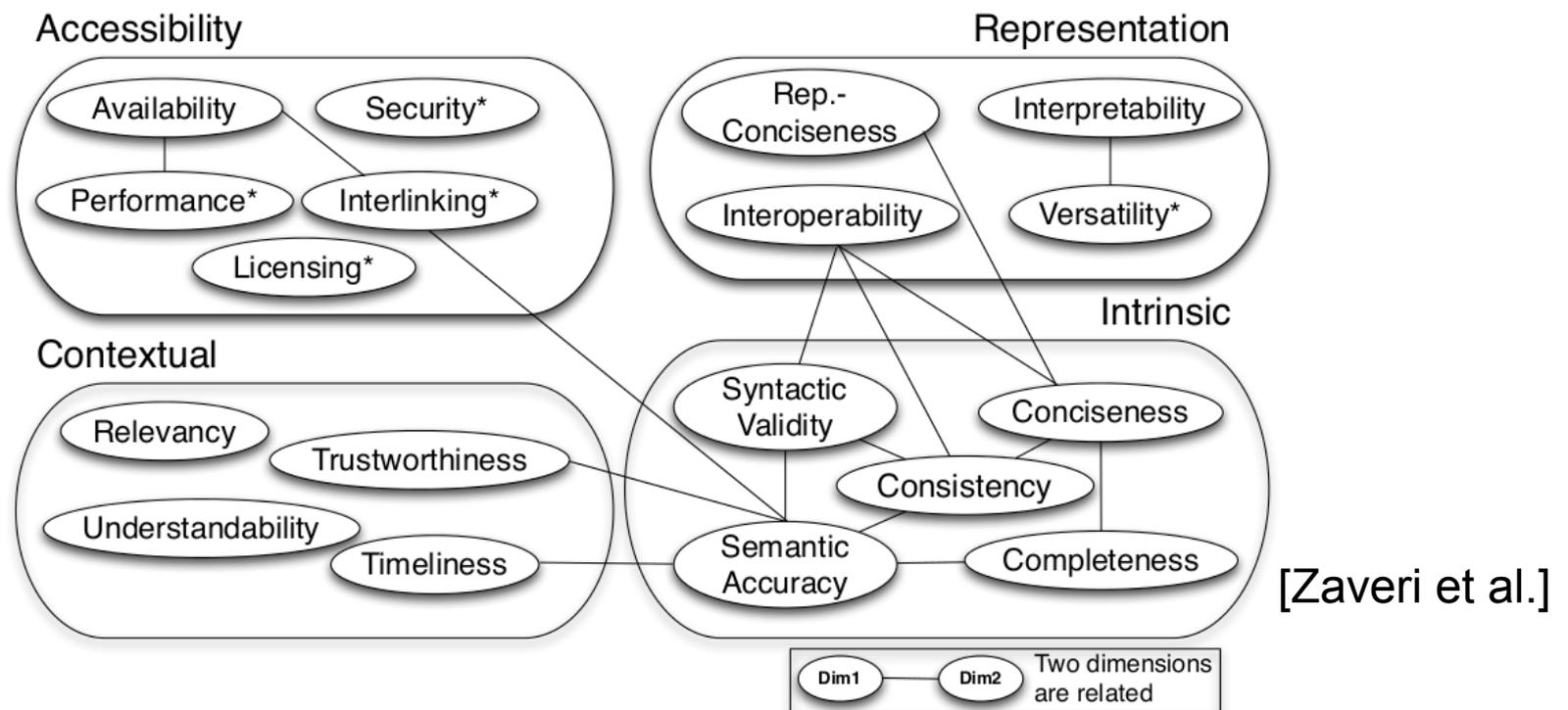
Datenqualität

- Erste tiefere Qualitätsbetrachtungen bereits in 1950ern
- Populäre Veröffentlichungen zu Daten-/Informationsqualität den 1990ern
- Qualitätsbetrachtungen im Web of Data in den letzten Jahren
- Generelle Ansichten:
 - „Fitness for Use“
 - Datenqualität → multidimensional



Datenqualität > Tools

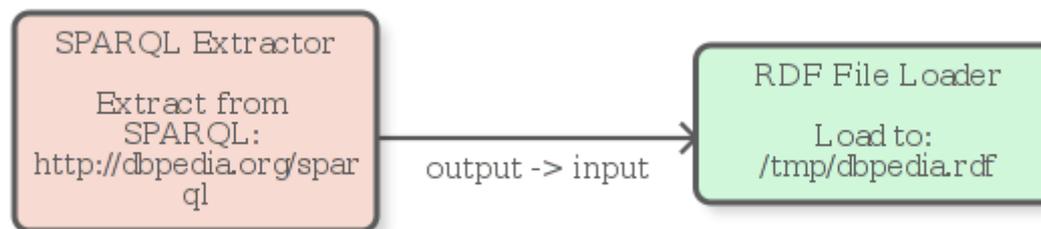
- Verschiedene Ansätze zur Evaluierung und Verbesserung der Qualität von RDF-Daten
- Tools decken meist nur wenige Dimensionen ab
- Kombination der bestehenden Tools bisher kaum möglich



Datenqualität > Tools

- **UnifiedViews**

- Allgemeines Framework zur Verarbeitung von RDF-Daten
- Verschiedene wiederverwendbare *Data Processing Units (DPUs)*
- Verkettung verschiedener DPUs = *Pipeline*
- Backend zum Abarbeiten der Pipelines
- Web-UI als Frontend



- Führer „ODCleanStore“ (Studentenprojekt an Uni Prag)
- Integration in den LOD2 Stack
- GPL 3.0/LGPL 3.0

Aufgabe

- Beschäftigung mit der Problematik „Datenqualität“ im Semantic Web
- Beschäftigung mit dem UnifiedViews Framework
- Implementierung von 3-5 Metriken
- Evaluierung von 1-2 Datasets bzgl. der implementierten Metriken
- Präsentation der Ergebnisse

Anforderungen

- Java
- Maven
- Vaadin

Quellen und Links

- [Zaveri et al.] Zaveri, Rula, Maurino, Pietrobon, Lehmann, Auer: 'Quality Assessment Methodologies for Linked Open Data',
- UnifiedViews: <https://grips.semantic-web.at/display/UDDOC/Introduction>
- UnifiedViews Source Code: <https://github.com/UnifiedViews>
- DPUs: <https://grips.semantic-web.at/display/UDDOC/Creation+of+Plugins>
- Implementierte DPUs: <https://github.com/mff-uk/DPUs>
- Vaadin: <http://vaadin.com>