

Evaluating Ontology Matchers on Real-World Financial Services Data Models



Jan Portisch

Data and Web Science Group, University of Mannheim / SAP SE
jan@informatik.uni-mannheim.de / jan.portisch@sap.com

Michael Hladik

SAP SE Product Engineering Financial Services
michael.hladik@sap.com

Heiko Paulheim

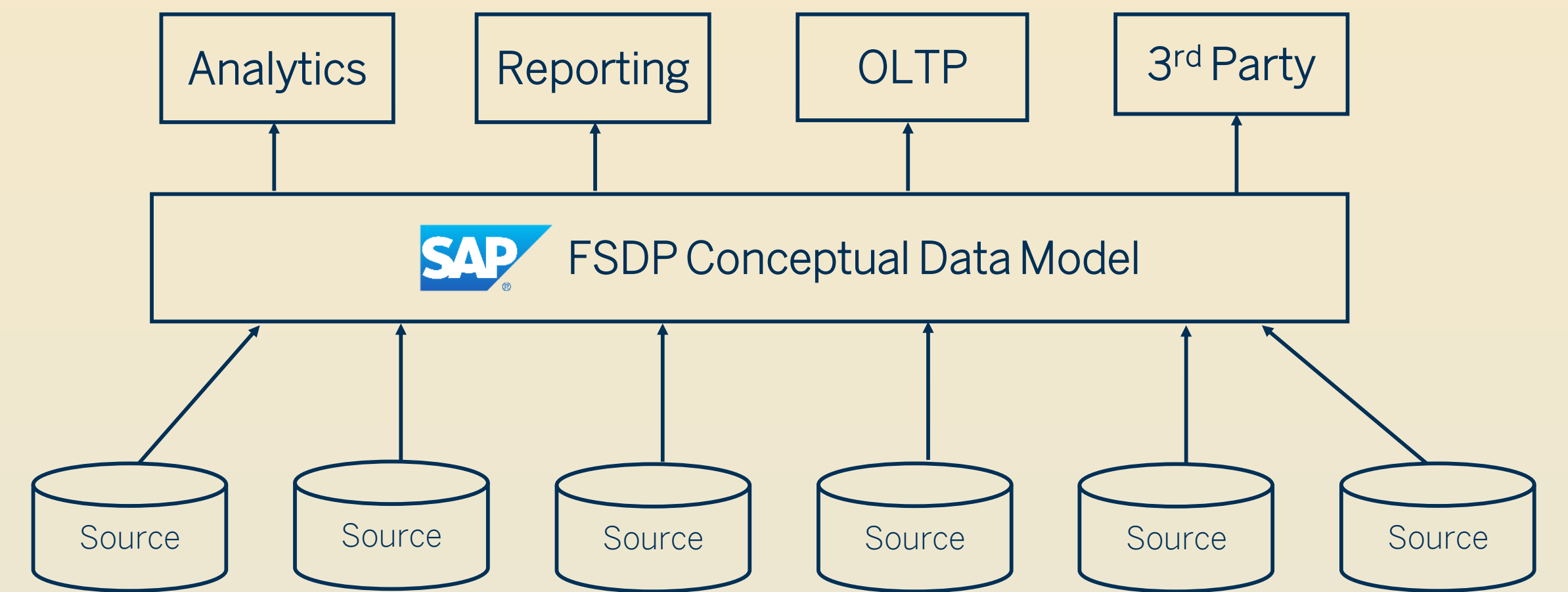
Data and Web Science Group, University of Mannheim
heiko@informatik.uni-mannheim.de



Business Use Case

Research Question (RQ)

“ How do state-of-the-art ontology matchers perform on real-world financial services data models? ”



Problem

- Traditional schema matching problem
- Domain-specific data models
- Heterogenous data structures

Restrictions

- Limited instance data availability
- Mappings shall be explainable

Approach



Translate()



The Translation Process in a Nutshell

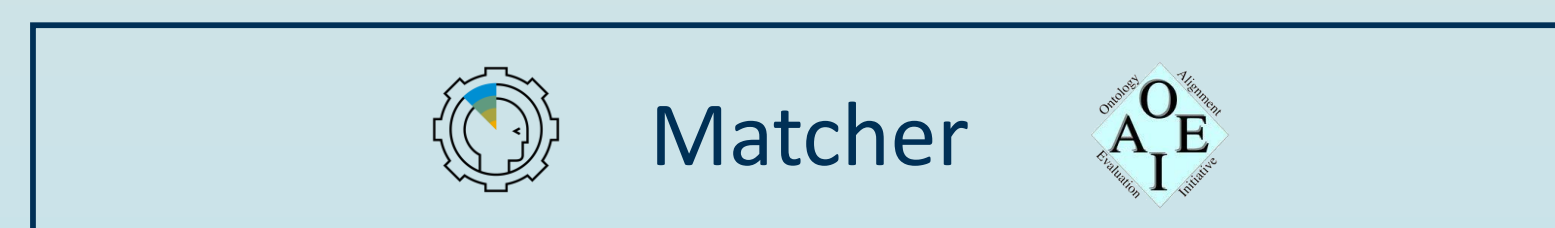
From **Conceptual Data Model** to **Ontology**

- **Entities** → OWL classes
- **Attributes** → `DataType` properties with `MaxCardinality 1`
- **Inheritances** → `rdfs:subClassOf`
- **Mandatory Attributes** → `MinCardinality 1`
- **Keys** → `owl:HasKey`
- **Relationships** → Object properties with cardinalities

Resulting Complexity: $ALN(D)$

The Data (provided by SAP SE, growing over time)

Data Set	Source			Target			# of Corr.	Arity
	C	P _D	P _O	C	P _D	P _O		
D ₁	760	3373	687	438	6878	0	4645	n:n
D ₂	760	4355	687	1	70	0	251	n:n
D ₃	760	4355	687	11	100	0	131	n:n
D ₄	760	4355	687	12	43	0	60	n:n
D ₅	760	4355	687	6	19	0	31	n:n



Alignments

Evaluation of the Alignments using the MELT Framework



Results and Future Work

The Results

		Alod2Vec	AML	LogMap	LogMap Lt	Kepler	Baseline
D ₁	Precision	0.3596	0.6016	0.9628	0.3432	0.6950	0.7210
	Recall	0.7991	0.6129	0.0893	0.7929	0.5681	0.5414
	F ₁	0.4960	0.6072	0.1635	0.4790	0.6252	0.6185
D ₂	Precision	0.5555	-	-	0.4000	0.7143	0.6667
	Recall	0.0199	-	-	0.0239	0.0398	0.0159
	F ₁	0.0385	-	-	0.0451	0.0754	0.0311
D ₃	Precision	0.2333	-	-	0.0769	0.2714	0.2667
	Recall	0.0534	-	-	0.1603	0.1450	0.0612
	F ₁	0.0870	-	-	0.1040	0.1891	0.0994
D ₄	Precision	0.8571	-	-	0.8571	0.8889	0.8571
	Recall	0.1000	-	-	0.1000	0.1333	0.1000
	F ₁	0.1791	-	-	0.1791	0.2319	0.1791
D ₅	Precision	0.0909	-	1.0000	0.1176	0.5000	0.1000
	Recall	0.0323	-	0.0323	0.0645	0.1290	0.0322
	F ₁	0.0476	-	0.0625	0.0833	0.2051	0.0488

Why is it so hard?

- N:N arity
- Domain-specific
- Missing background knowledge

Example

FinRep Equity Method and Minority Interests Flag
→ `DatatypeProperty`, Domain: Partner

“Identification of shares in subsidiaries, associated companies, and joint ventures that are consolidated according to the equity method in the consolidated report (forms B 6.1, B 6.2, B 6.6, and B 6.7) for FINREP [...]”

- Equity Stake Percentage → `DatatypeProperty`,
- Consolidation System → `DatatypeProperty`,
- Method of Consolidation → `DatatypeProperty`,
- Business Partner Relation Type → `DatatypeProperty`,

Challenges

- Current matchers struggle to match real world industry data schemas
- Likely explanations
 - Shallow, weakly structured ontologies
 - Potential overfitting to public data sets
 - Focus on 1:1 arity

Future Work

- Growing the data sets and improving their quality
- Blind alignment track at OAEI
- Evaluating suitable sources and strategies for including background knowledge in the process