

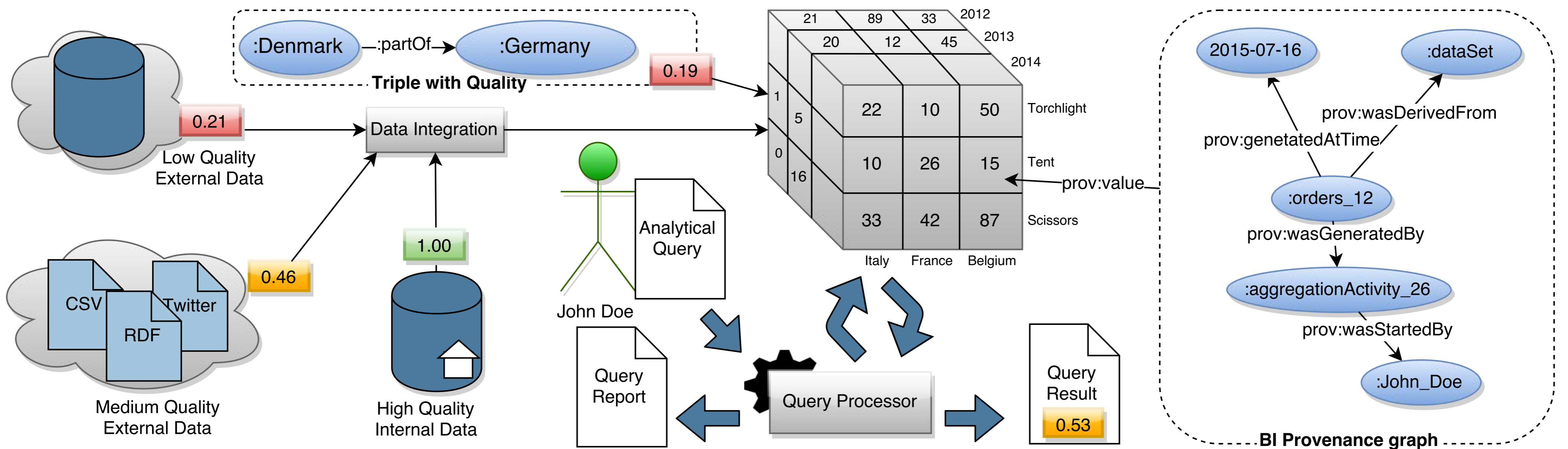
DATA PROVENANCE AND QUALITY IN RDF DATA CUBES



AALBORG UNIVERSITET

Kim Ahlstrøm Jakobsen
Supervisors: Katja Hose, Torben Bach Pedersen
Aalborg University

Project Overview



My Research

- RDF Data Cube Constructions and Optimization [1,2]
 - Denormalization
 - Query Rewriting
- Provenance/Lineage Tracking and Models
- Data Quality Tracking and Models
- Result Explanation and Correctness

Query Rewriting

Snow Pattern Query

```
SELECT ?name sum(?price)
WHERE {
  ?lineitem :extendedprice ?price ;
            :has_order ?order .
  ?order skos:broader ?customer .
  ?customer skos:broader ?nation .
  ?nation :name ?name .
}
GROUP BY ?name
```

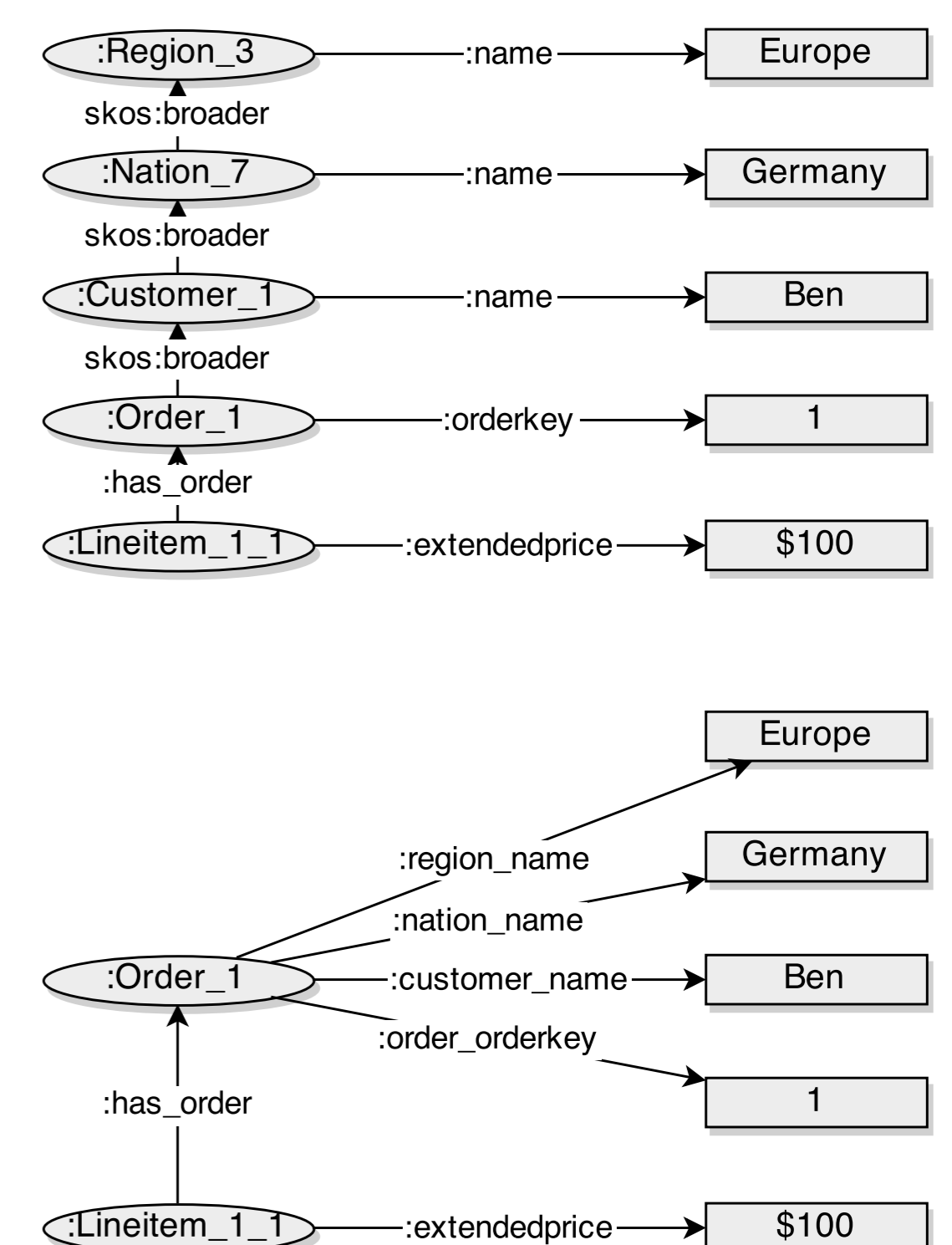
Star Pattern Query

```
SELECT ?name sum(?price)
WHERE {
  ?lineitem :extendedprice ?price ;
            :has_order ?order .
  ?order :nation_name ?name .
}
GROUP BY ?name
```

Merging Level Members

```
1 Function MergeLevel(cube, onto, level, dim) is
2   dimMembers = {};
3   foreach parLevel ∈ onto(level qb4o:parentLevel ?parLevel . ?parLevel qb4o:inDimension dim) do
4     parInsts = MergeLevel(cube, onto, parLevel, dim);
5     foreach parInst ∈ parInsts(?parInst qb4o:inLevel []) do
6       foreach levelMem ∈ cube(?levelMem skos:broader parInst) do
7         dimMembers = dimMembers ∪ {(levelMem, prop, obj) | (prop, obj) ∈ parInsts(parInst ?prop ?obj)};
8       end
9       if cube(?levelMem skos:broader parInst) = {} then
10        dimMembers = dimMembers ∪ {(parInst, prop, obj) ∈ parInsts(parInst ?prop ?obj)};
11      end
12    end
13  end
14  foreach prop ∈ onto(level qb4o:hasAttribute ?prop) do
15    dimMembers = dimMembers ∪ {(levelMem, MergeURIs(level, prop) obj) ∈ cube(?levelMem qb4o:inLevel level. ?levelMem prop ?obj)};
16  end
17  dimMembers = dimMembers ∪ {(levelMem, qb4o:inLevel, level) | levelMem ∈ cube(?levelMem qb4o:inLevel level)};
18  return dimMembers;
19 end
```

Denormalization



Publications

- (In Progress) Kim A. Jakobsen, Alex B. Andersen, Katja Hose, Torben Bach Pedersen: Optimizing RDF Data Cubes for Efficient Processing of Analytical Queries
- Alex B. Andersen, Nurefsan Gür, Katja Hose, Kim A. Jakobsen, Torben Bach Pedersen: Publishing Danish Agricultural Government Data as Semantic Web Data. JIST 2014: 178-186

This PhD project is part of the QWeb project
<http://qweb.cs.aau.dk>