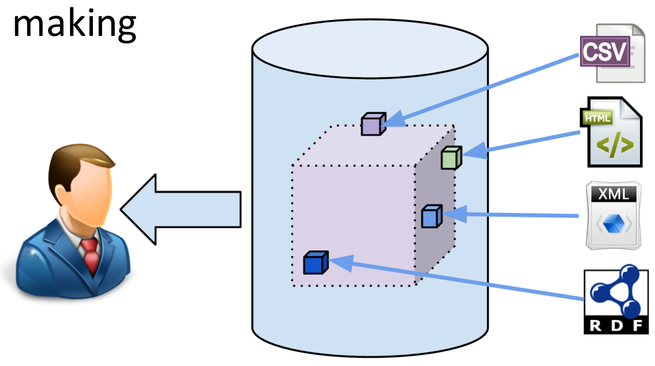
Towards OLAP over Federated RDF Sources

Dilshod Ibragimov (Supervisors Esteban Zimányi, Torben Bach Pedersen)
dibragim@ulb.ac.be, diib@cs.aau.dk
Université Libre de Bruxelles (ULB), Aalborg University (AAU)

1. BI and the Semantic Web

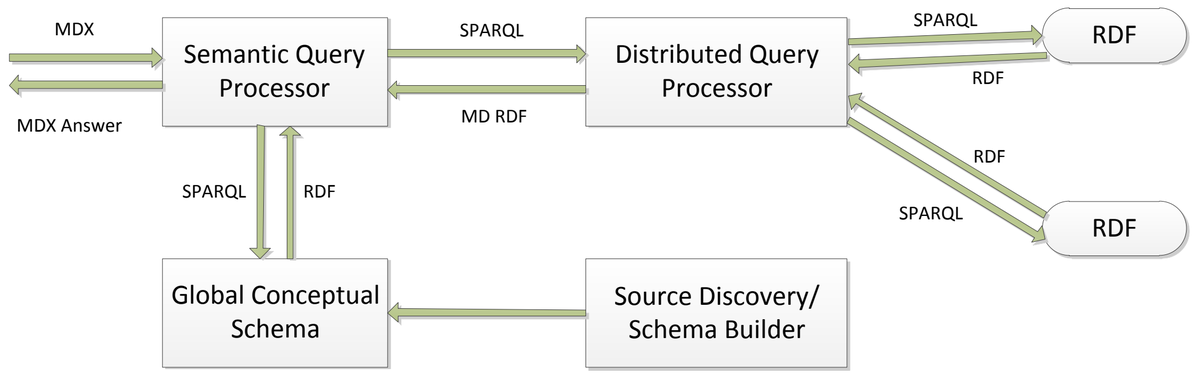
- Business Intelligence tools need to analyze data published on the Web
- OLAP-style analysis of Linked Data may help in better decision making



2. Objective

Design, develop, and evaluate an approach for performing OLAP over federated RDF sources.

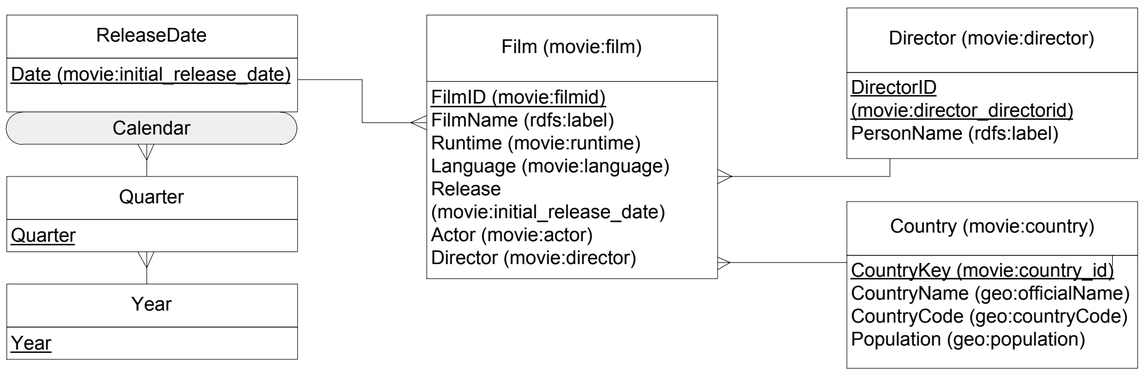
3. System Architecture



- *Global Conceptual Schema* (GSC) – high-level view of the system (expressed in QB4OLAP)
- *Source Discovery/Schema Builder* – discovery of data sources and construction of the GCS
- *Distributed Query Processor* (DQP) – retrieval, in parallel, data from several federated data sources
- *Semantic Query Processor* – conversion of MDX to SPARQL which is sent to the DQP

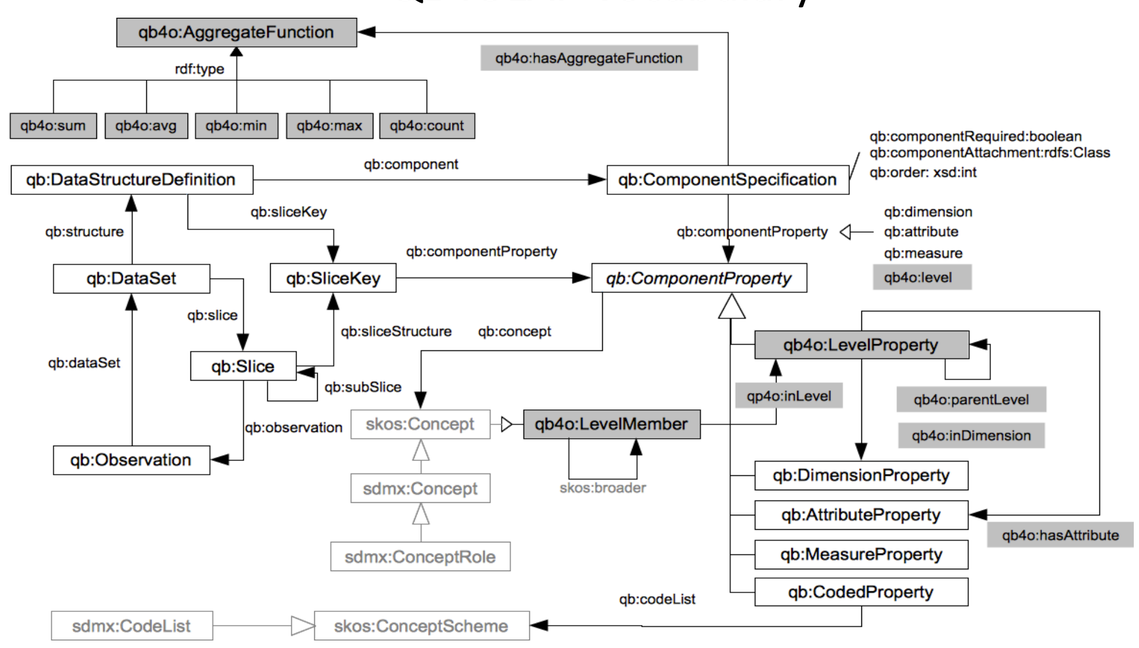
4. QB4OLAP: Global Conceptual Schema and Observations Examples

Global Conceptual Schema




```
## Data structure definition and dimensions
exqb:MoviesCube a qb:DataStructureDefinition ;
    rdfs:label "Movies Cube"@en ;
    ## Dimensions
    qb:component [qb:dimension exqb:ReleaseDateDim ] ;
    qb:component [qb:dimension exqb:DirectorDim ] ;
    qb:component [qb:dimension exqb:CountryDim ] ;
    ## Definition of measures
    qb:component [ qb:measure exqb:Runtime ] ;
    ## Attributes
    qb:component [ qb:attribute exqb:FilmName ] .
exqb:DirectorDim a qb:DimensionProperty ;
    rdfs:isDefinedBy exqb:DirectorID ;
    qb4o:hasAttribute exqb:DirectorID ;
    qb4o:hasAttribute exqb:PersonName .
exqb:ReleaseDateDim a qb:DimensionProperty ;
    rdf:type xsd:dateTime .
exqb:year a qb4o:LevelProperty ;
    qb4o:inDimension exqb:ReleaseDateDim .
exqb:quarter a qb4o:LevelProperty ;
    qb4o:inDimension exqb:ReleaseDateDim ;
    qb4o:parentLevel exqb:year .
## Attributes
exqb:Runtime rdfs:comment "Film Runtime"@en ;
    skos:mappingRelation movie:runtime .
#DataSets
exqb:LinkedMovie a void:Dataset ;
    foaf:homepage <http://data.linkedmdb.org> ;
    dcterms:title "Movies Dataset" ;
    dcterms:description "RDF data about movies" ;
    void:sparqlEndpoint
        <http://data.linkedmdb.org/sparql> ;
    dc:subject exqb:DirectorDim ;
    dc:subject exqb:Runtime .
```

QB4OLAP Vocabulary



Observations

```
http://data.linkedmdb.org/resource/film/930> a qb:Observation;
    qb:dataSet exqb:MoviesDataWarehouse ;
    exqb:Director < http://data.linkedmdb.org/resource/director/1008>;
    exqb:Runtime 158;
    exqb:Country < http://data.linkedmdb.org/resource/country/IN> .
<http://data.linkedmdb.org/resource/director/1008>
    exqb:PersonName "Vivek Agnihotri (Director)" .
```

```
<http://data.linkedmdb.org/resource/country/IN>
    exqb:CountryName "India" ;
    exqb:CountryCode "IN" ;
    exqb:Population "1173108018" .
<http://data.linkedmdb.org/resource/country/DE>
    exqb:CountryName “Germany" ;
    exqb:CountryCode "DE" ;
    exqb:Population "81802257" .
```

5. Source Discovery

Querying large knowledge bases (DBpedia, Yago, Freebase)

Show id, label, and count of items where label contains “Film”, sorted by number of items

```
SELECT ?s ?l COUNT(?s) as ?count
WHERE {
    ?someobj ?p ?s .
    ?s rdfs:label ?l .
    FILTER(CONTAINS(?l, "Film") && (lang(?l) = 'en').
    FILTER (!isLiteral(?someobj)).
} ORDER BY DESC(?count) LIMIT 20
```

S	L	COUNT
http://rdf.freebase.com/ns/m.02nsjl9	Film character	2001832
http://rdf.freebase.com/ns/film.film_character	Film character	1384754
http://rdf.freebase.com/ns/film.actor	Film actor	874840
http://rdf.basekb.com/ns/m.0jsg30	Film performance	673398
http://rdf.freebase.com/ns/film.film	Film	557505

Show instances of the type ns:film.film

```
PREFIX ns: <http://rdf.freebase.com/ns/>
SELECT ?s ?p ?o
WHERE {
    ?s ?p ?o .
    ?s ns:type.object.type ns:film.film .
    FILTER (lang(?o) = 'en').
}
```

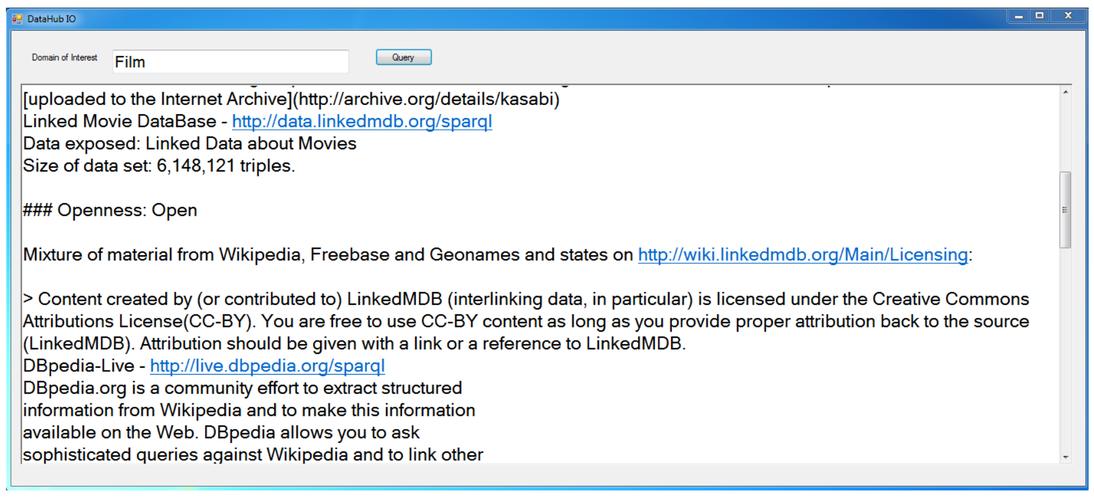
S	P	O
http://rdf.freebase.com/ns/m.0pj5t	rdfs:label	Falling Down
http://rdf.freebase.com/ns/m.0swhj	rdfs:label	Brown Christmas
http://rdf.freebase.com/ns/m.0m2kd	rdfs:label	Stand by Me
http://rdf.freebase.com/ns/m.07cz2	rdfs:label	The Matrix
http://rdf.freebase.com/ns/m.0prk8	rdfs:label	Hamlet

Querying Semantic Web Search Engines (Sindice.com)

http://api.sindice.com/v3/search?q=Film&format=json&fq=format%3ARDF&page=6&facet.field=domain


```
- facet_counts: {
    facet_queries: { },
    facet_ranges: { },
    facet_dates: { },
  - facet_fields: {
      - domain_facet: [
            "elbo.ws",
            161,
            "waptrick.com",
            161,
            "www4.wiwiss.fu-berlin.de",
            156,
            "data.linkedmdb.org",
            148,
            "era.rkbexplorer.com",
            147,
            "bleeper.de",
            133,
            "kisti.rkbexplorer.com",
            131,
            "bulk.resource.org",
            129,
            "dbpedia.openlinksw.com",
            122,
            "semanticweb.org",
            121
        ]
    }
}
```


Querying Data Management Platforms (Datahub.io)



6. Publications

Submitted:

D. Ibragimov, K. Hose, T. B. Pedersen, E. Zimányi. Towards Exploratory OLAP over Linked Open Data – A Case Study. BIRTE 2014

In Progress:

D. Ibragimov, K. Hose, T. B. Pedersen, E. Zimányi. Executing Aggregate SPARQL Queries over Federated Endpoints