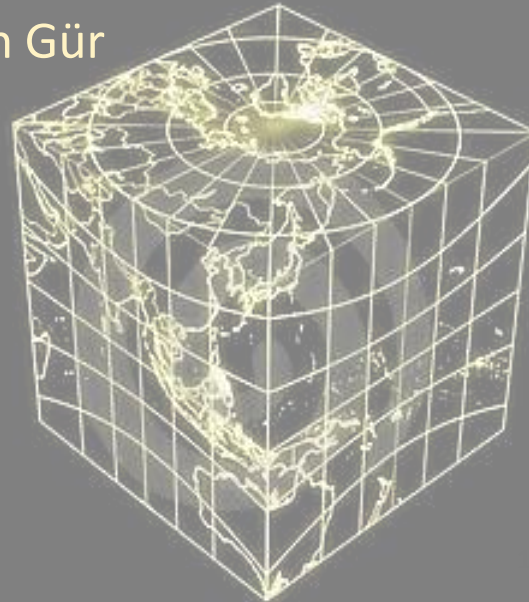
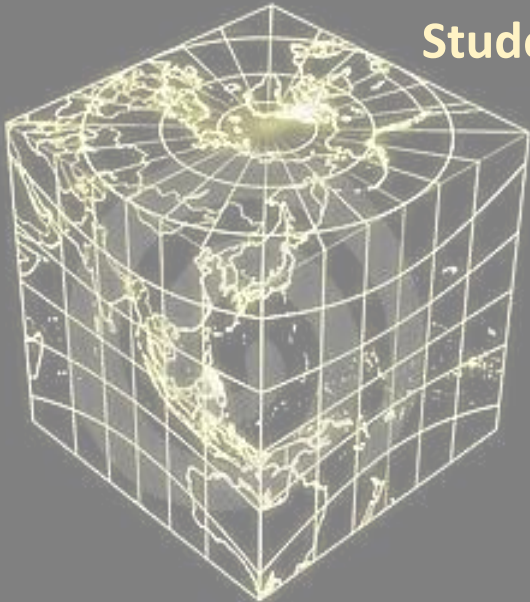


Business Intelligence over Linked Open Spatio-Temporal Data

Student: Nurefşan Gür



Ph.D. Research Progress Report

Supervisor: Torben Bach Pedersen (AAU)

Co-supervisor: Esteban Zimányi(ULB)

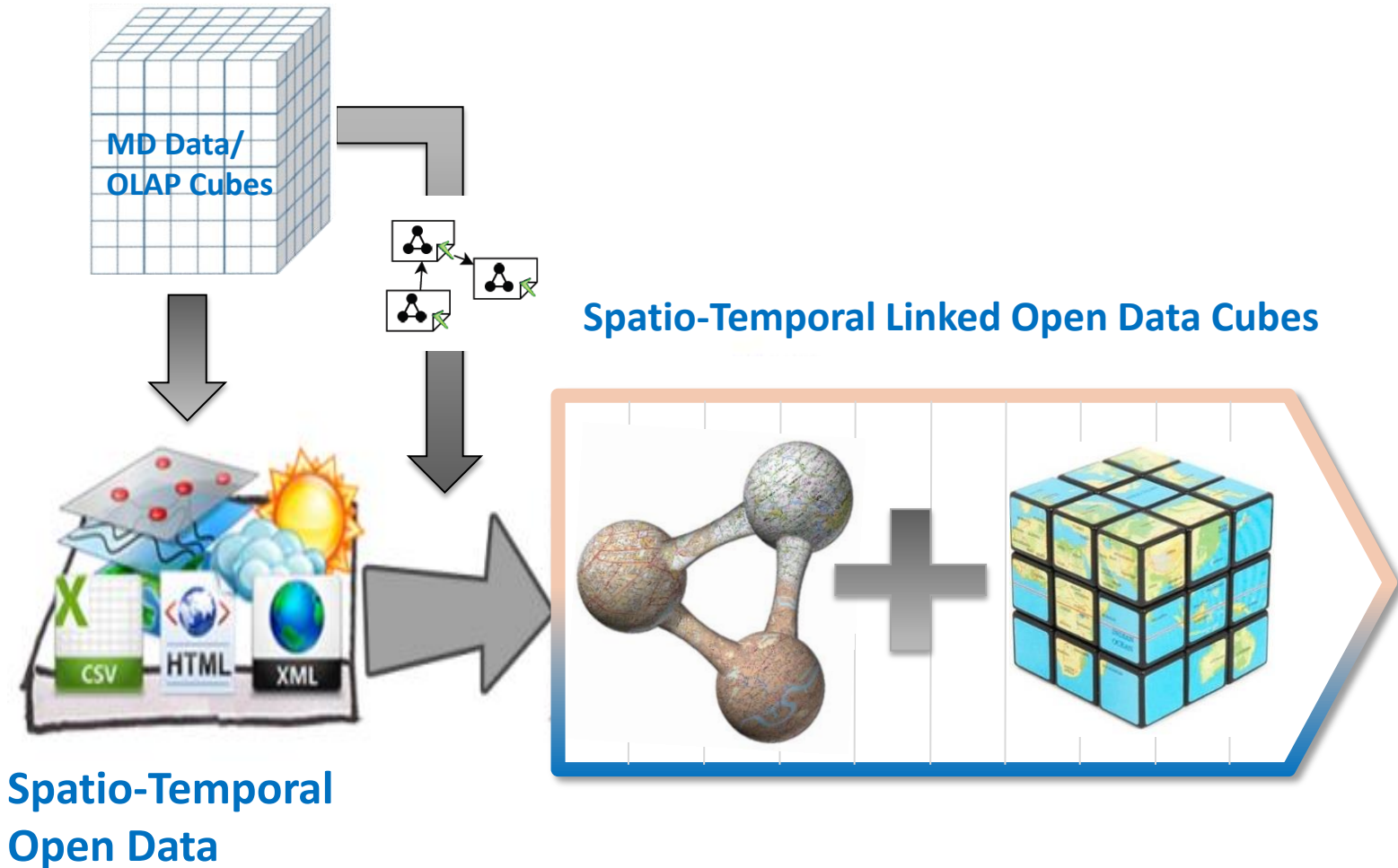
Co-supervisor: Katja Hose (AAU)

CPC Chair: Wolfgang Lehner (DTU)

Time span: August 15, 2013 – November 16, 2016

1. Thesis Summary
2. Background
3. Objectives
4. Research Work
 1. Completed Work
 2. Ongoing Work
 3. Future Work
5. Publications
6. Schedule
7. Ph.D. Courses
8. Discussion

1. Thesis Summary

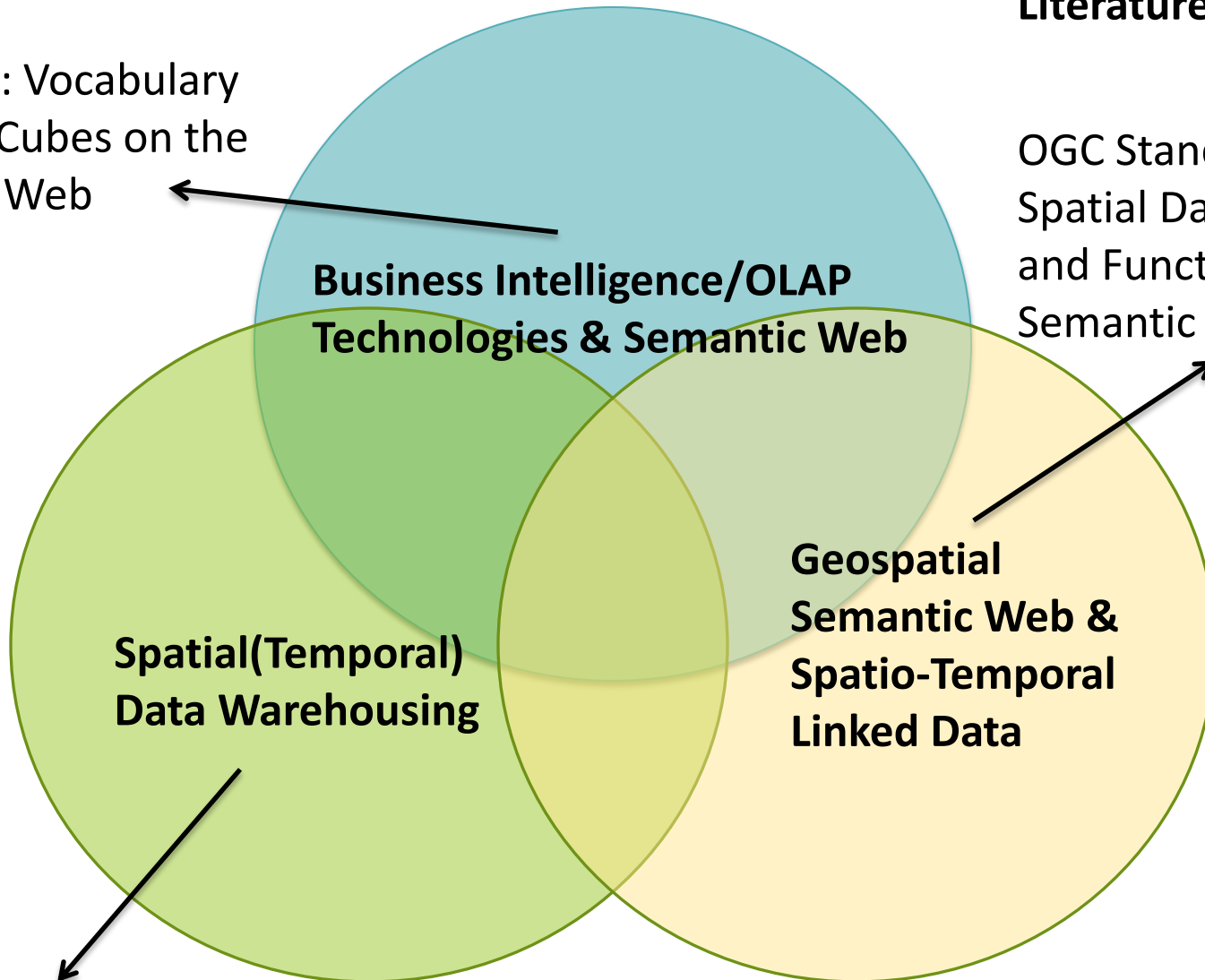


2. Background

Literature Survey

QB4OLAP: Vocabulary for OLAP Cubes on the Semantic Web

OGC Standards for Spatial Data types and Functions on the Semantic Web



MultiDimER Conceptual Model

2. Background

Business Intelligence/OLAP Technologies & Semantic Web

1. “Fusion cubes: Towards self-service business intelligence”, A. Abello, J. Darmont, L. Etcheverry, M. Golfarelli, J.Mazón, F. Naumann and T. B. Pedersen (2011-2013)
2. “QB4OLAP: A New Vocabulary for OLAP Cubes on the Semantic Web”, L. Etcheverry and A. A. Vaisman (2012)
3. “Modeling and querying data warehouses on the semantic web using QB4OLAP”, L. Etcheverry, A. Vaisman, and E. Zimányi. (2014)
4. "Using Semantic Web Technologies for Exploratory OLAP: Survey". A. Abello, O. Romero, T. B. Pedersen, R. Berlanga, V. Nebot, M. J. Aramburu and A. Simitsis. (2014)
5. “OLAP for multidimensional semantic web databases.” A. Matei, K.-M. Chao, and N. Godwin (2015)

2. Background

Spatial(Temporal) Data Warehousing

1. “Advanced Data Warehouse Design: From Conventional to Spatial and Temporal Applications”, E. Zimányi.(2008)
2. “A multidimensional model representing continuous fields in spatial data warehouses”, A. Vaisman and E. Zimányi. (2009)
3. “A Generic Data Model and Query Language for Spatiotemporal OLAP Cube Analysis”
Gómez, Leticia I., Silvia A. Gómez, and Alejandro A. Vaisman. (2012)
4. “Spatial data warehouses. In Data Warehouse Systems: Design and Implementation”,
A. Vaisman and E. Zimányi. (2014)

2. Background

Geospatial Semantic Web & Spatio-Temporal Linked Data

1. "Data models and query languages for linked geospatial data" , Kostis Kyzirakos, Manos Karpathiotakis, Manolis Koubarakis (2011)
2. "SPARQL-ST: Extending SPARQL to Support Spatiotemporal Queries" Matthew Perry, Prateek Jain, Amit P. Sheth (2011)
3. "STRABON : A Semantic Geospatial DBMS" , Kostis Kyzirakos, Manos Karpathiotakis, Manolis Koubarakis (2012)
4. "Representation and querying of valid time of triples in linked geospatial data." Bereta, Konstantina, Panayiotis Smeros, and Manolis Koubarakis. (2013)
5. "Geoknow: Making the web an exploratory place for geospatial knowledge ", S. Athanasiou, D. Hladky, G. Giannopoulos, A. Garcia Rojas, and J. Lehmann. (2014)

3. Objectives

- O1. Data Representation & Integration
- O2. Spatio-Temporal OLAP Query Design and Processing on the Semantic Web
- O3. Business Intelligence over the RDF Warehouse

- **O1. Data Representation & Integration**
 - **Task1.** Initial designation of a semantic model with an ontology, data cleansing/integrating, finally publishing those data sets as linked open data for posing/testing queries which were not available before.
 - **Task2.** Developing and testing a spatially enhanced meta-model for representation of multi-dimensional RDF data on the Semantic Web.

- **O2. Spatio-Temporal OLAP Query Design and Processing on the Semantic Web**
 - **Task3.** Extending the meta-model from T2 with advanced spatial and temporal aspects of MD models. Designing fully formalized SOLAP operators and showing how to translate them into SPARQL.
 - **Task4.** Implementing the designed ST - OLAP queries (in SPARQL within the sufficient support extend) and evaluating the advanced processing of spatio-temporal dimensions and measures of MD linked data through a SPARQL Benchmark.

■ O3. Business Intelligence over the RDF Warehouse

- **Task5.** Implementing the designed MD meta-model and OLAP operators on a real world case study (Gov. – LOD) with extended UDF functions in SPARQL queries.
- **Task6.** Testing and validating these methodologies with a user interface which will be implemented in parallel to the previous objectives.

4. Research Work

4.1 Completed Work (O1 & Partially O2)

4.2 Ongoing Work (O2)

4.3 Future Work (Partially O2 & O3)

4.1 Completed Work


1) Publishing Danish Agricultural Government Data as Semantic Web Data

Agricultural Datasets

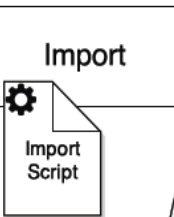


Ministeriet for Fødevarer, Landbrug og Fiskeri 


<http://fvm.dk>




Raw Data

Import 

Import Script



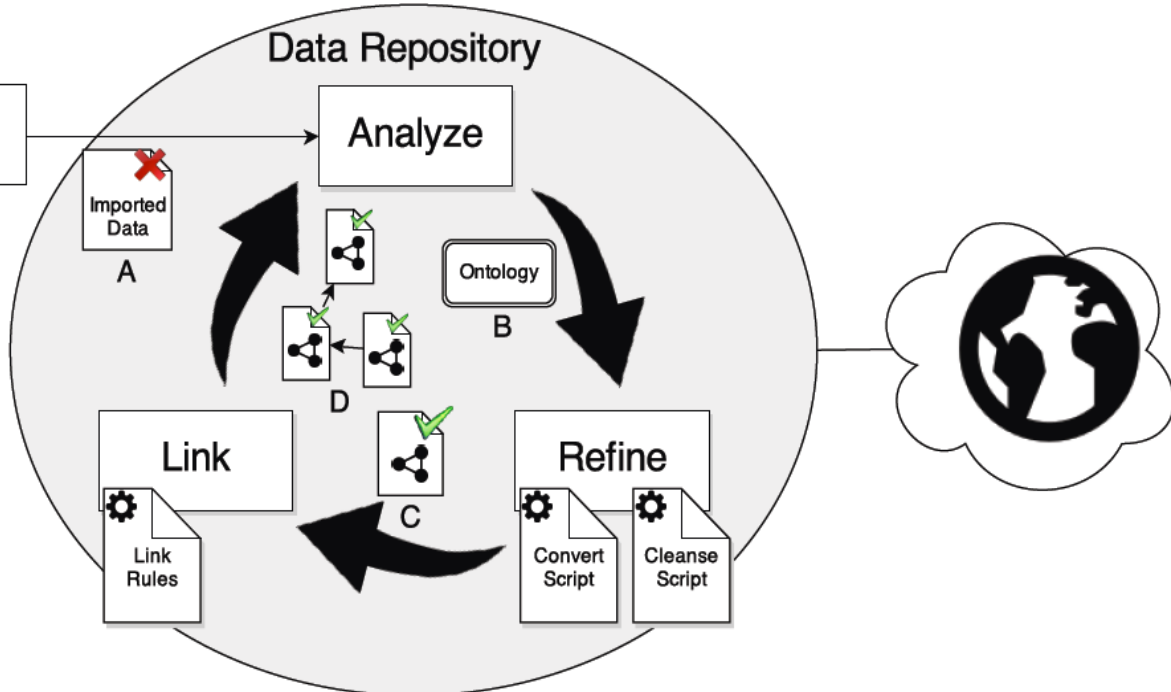
CSV



CVR

<http://cvr.dk>

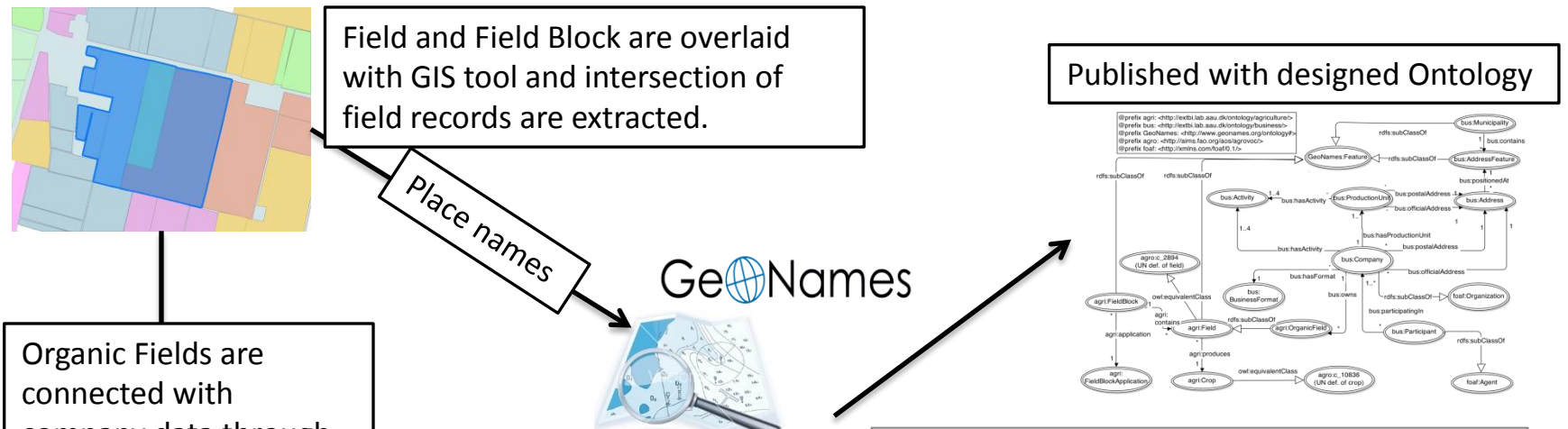
Business Datasets



1) Overall Process Overview

4.1 Completed Work

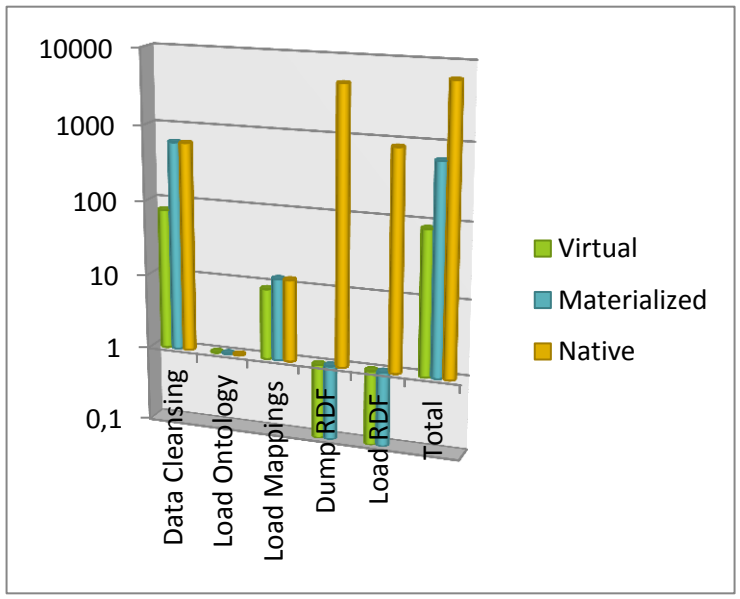
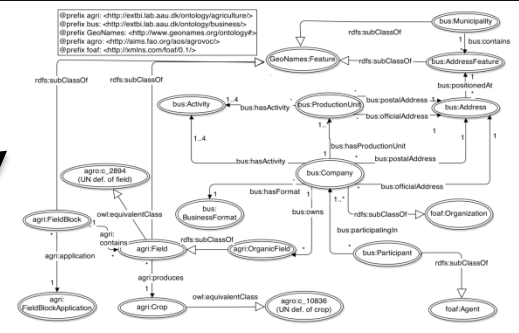
1) Publishing Danish Agricultural Government Data as Semantic Web Data



Organic Fields are connected with company data through their CVR keys

Field and Field Block are overlaid with GIS tool and intersection of field records are extracted.

Published with designed Ontology



2) Analyze/Linking Process

Load times in seconds during the "overall process"

4.1 Completed Work

1) Publishing Danish Agricultural Government Data as Semantic Web Data

```

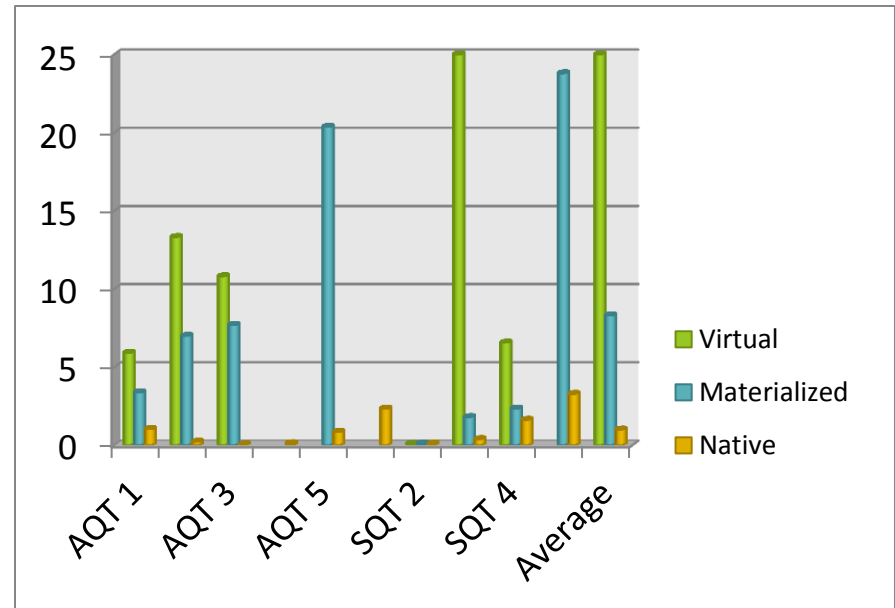
SELECT ?crop COUNT (*) AS ?cnt
FROM <http://extbi.lab.aau.dk/resource/agriculture>
WHERE {
  ? field agri:produces ?crop .
  ? field wgs:long ?long .
  ? field wgs:lat ?lat .
  FILTER (?long > [x - 0.5] && ?long < [x + 0.5]
    && ?lat > [y - 0.5] && ?lat < [y + 0.5]).
} GROUP BY ?crop
  
```

Query 1 : AQT – Counts fields based on the crop they produce

```

SELECT ?name ?address
FROM <http://extbi.cs.aau.dk/resource/agriculture>
FROM <http://extbi.cs.aau.dk/resource/business>
WHERE {
  ?company bus:owns ?organicField.
  ?company bus:name ?name.
  ?company bus:officialAddress ?address
}
GROUP BY ?name ?address
  
```

Query 2 : SQT – Finds the company addresses that owns organic fields



Query Runtimes in seconds

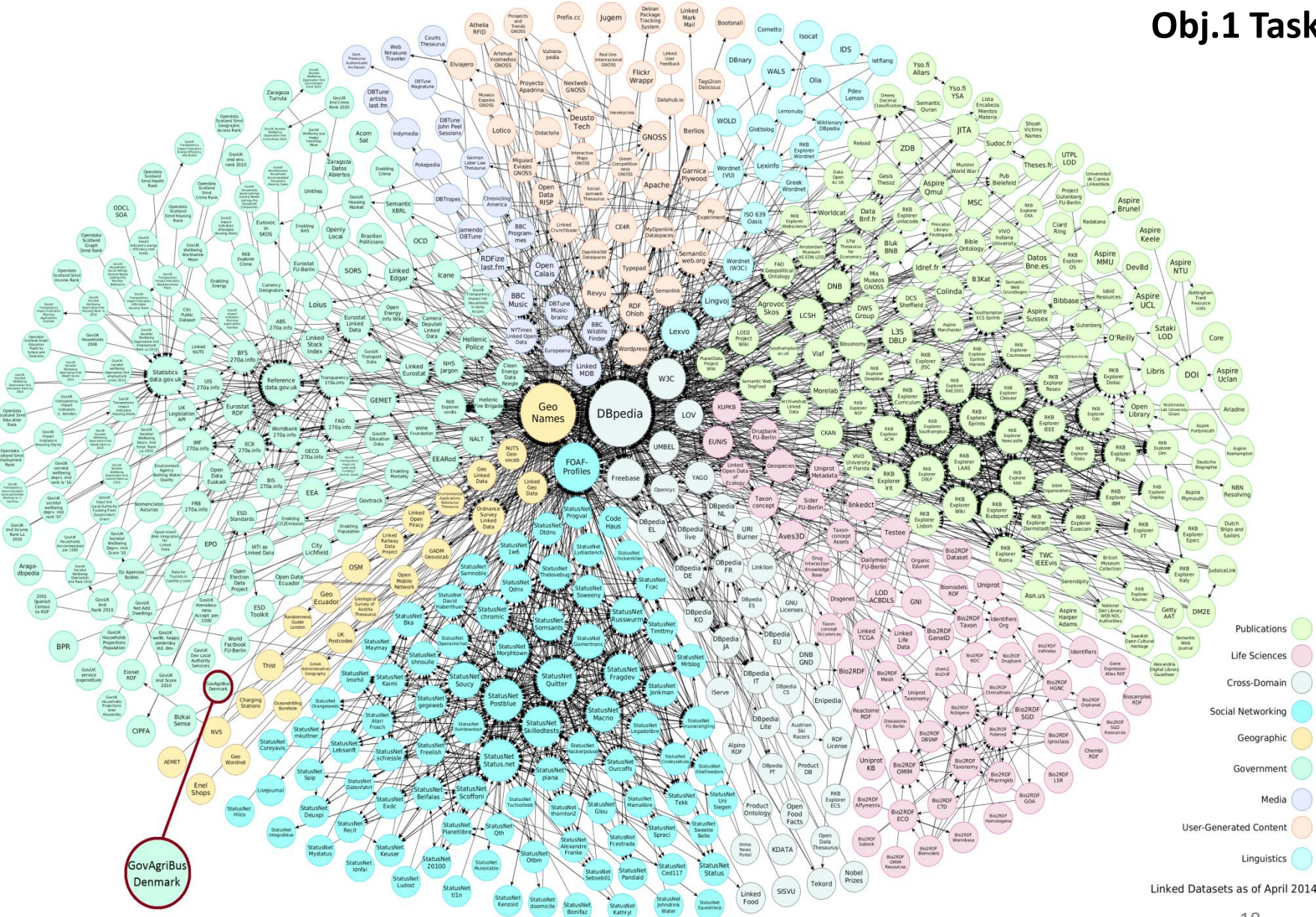
3) Query Evaluation

Project website: <http://extbi.lab.aau.dk/>

Endpoint: <http://extbi.lab.aau.dk/sparql>

4.1 Completed Work

Obj.1 Task1



Linked Datasets as of April 2014

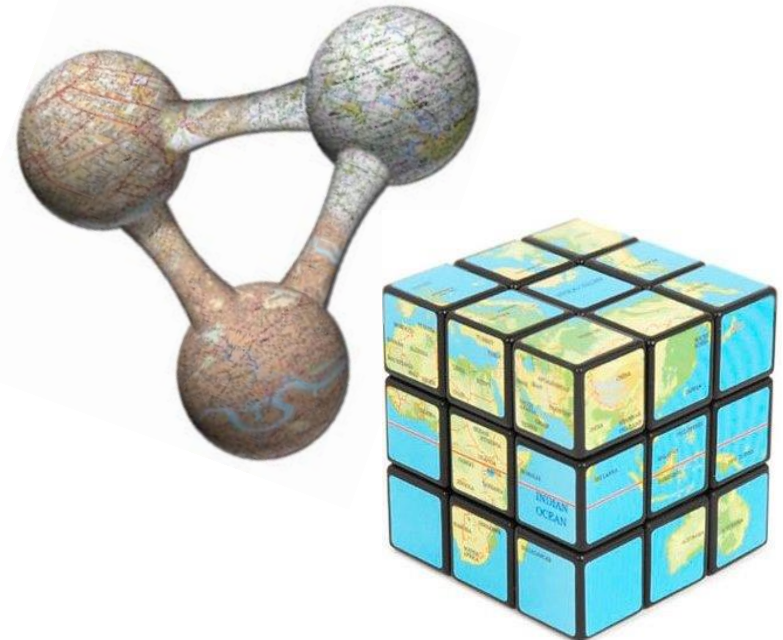
Source : <http://lod-cloud.net/>

4.2 Ongoing Work - *Completed*

2) Modelling and Querying Spatial Data Warehouses on the Semantic Web

Research Context

- **Making it possible ; Geospatial Semantic Web and Semantic OLAP Cubes function together**
- Understanding the spatial data types and operations on them
- Extending the most recent OLAP Cube Vocabulary with spatial concepts
- Describing SOLAP operators formally that can function on Spatial Data Warehouses and the Semantic Web



4.2 Ongoing Work - Completed

2) Modelling and Querying Spatial Data Warehouses on the Semantic Web

Preliminaries: Spatial Data & Operations



"POLYGON((1.500000 3.500000,2.500000 4.500000,1.500000 5.500000,0.500000 4.500000,1.500000 3.500000))"



"LINESTRING(-20.50000 20.50000, 40.50000 25.50000, 60.50000 40.50000, 80.50000 120.50000)"

- "POINT(-99.1332 19.4326)"

Ex: Geometry Primitives in WKT

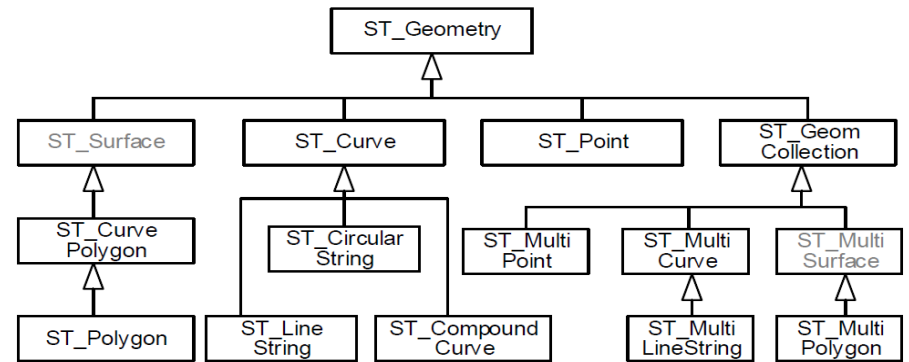


Figure: Hierarchical Spatial Data Schema

Class	Operators
Spatial Aggregation	{Union, Intersection, Buffer, ConvexHull, MBR - Minimum Bounding Rectangle}
Topological Relation	{Intersects, Disjoint, Equals, Overlaps, Contains, Within, Touches, Covers, CoveredBy, Crosses }*
Numeric Operation	{Perimeter, Area, # of Interior Rings, Distance, Haversine Distance, NN – Nearest Neighbor, # of Geometries}

Table : Spatial Operations

- **OGC:** Open Geospatial Consortium
- **RCC8:** Region Connection Calculus
- **DE9-DIM:** Dimensionally Extended Nine-Intersection Model

4.2 Ongoing Work - Completed

QB4OLAP Vocabulary by

L. Etcheverry, A. Vaisman.

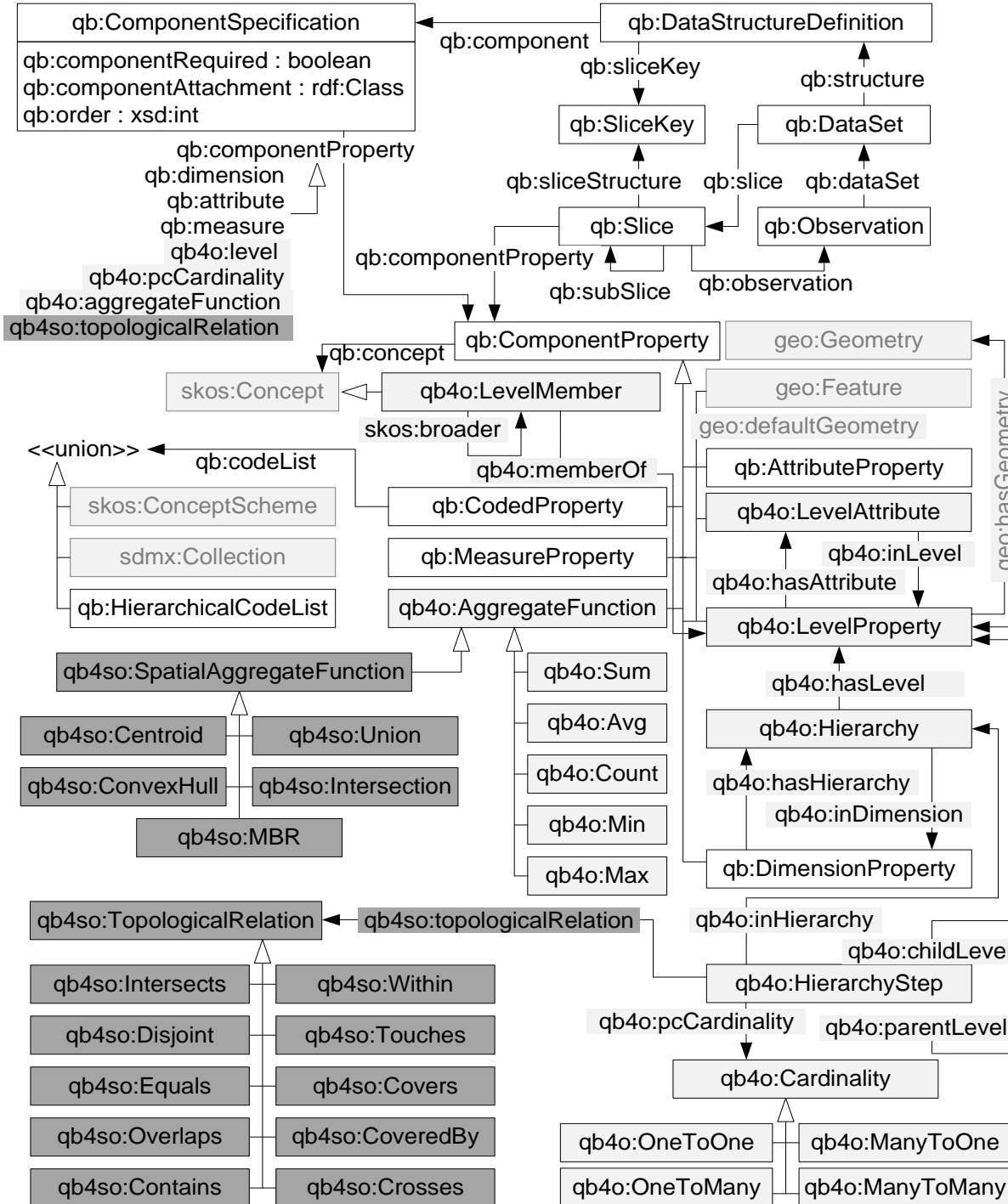
available under:

<http://purl.org/qb4olap/cubes>

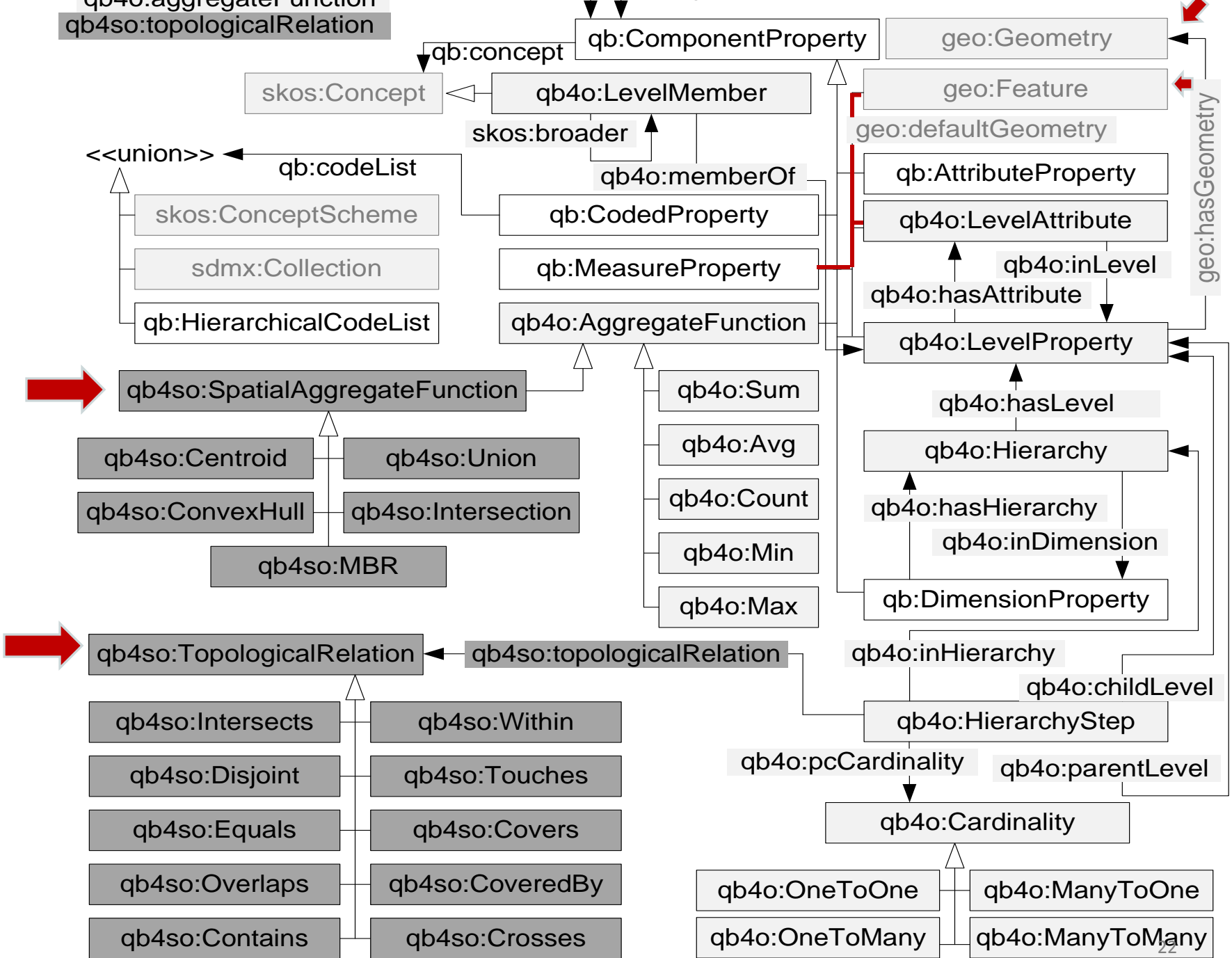
QB4SOLAP Meta-Model

extended to be available under:

<http://purl.org/qb4solap/cubes>



QB4OLAP, SKOS, OGC Properties & Classes	
QB4SOLAP Properties & Classes	
RDF Cube Properties & Classes	



4.2 Ongoing Work - *Completed*

Spatial Data Warehouse Members in QB4SOLAP

triple = (s,p,o) $\in T = (I \cup B) \times I \times (I \cup B \cup Li)$ where I is IRIs, B is blank nodes and Li is a *Literal*.

Spatial Dimensions: Dimensions are identified as spatial if only they have at least a spatial hierarchy.

Spatial Hierarchy Steps: If a hierarchy is spatial, hierarchy steps are extended with topological relations (T_{rel}) which exists for spatial levels in hierarchies.

$T_{HS} \in T$ and $:_t_{shs} \in B$ has rollup relation $R = (L_c, L_p, card, T_{rel})$.

$T_{HS} = (:_t_{shs} \text{ a } qb4o:HierarchyStep; qb4o:childLevel slh_{ci}; qb4o:parentLevel slh_{pi}; qb4o:pcCardinality card; qb4so:topologicalRelation T_{rel})$

where $slh_{ci} \in L_c$ (*child level in hierarchy*), $slh_{pi} \in L_p$ (*parent level in hierarchy*) and T_{rel} is Topological Relations.

Example: R(River, City, ManyToMany, Intersects)

4.2 Ongoing Work - *Completed*

Spatial Data Warehouse Members in QB4SOLAP

Spatial Levels: QB4SOLAP links a geometry class “geo:Geometry” from OGC Schemas Spatial levels use the geometry members ($g \in G$) in two ways:

- Spatial level without the spatial attributes but as a member of geo:Geometry class (i.e. Country level which has a polygon geometry).

$\forall l_s \in L, \exists t_{ls} \in T \ t_{ls} = (l_s \text{ a qb4o:LevelProperty; hasAttribute } a; \text{ geo:hasGeometry } G)$

- Spatial level in which there is a geometry attribute (i.e. Country level which has a polygon geometry and a capital of the country as a geometry attribute)

$t'_{ls} = (l'_s \text{ a qb4o:LevelProperty; qb4o:hasAttribute } a; \text{ geo:hasGeometry } G; \text{ qb4o:hasAttribute } a:\text{dom}_g)$

Spatial Attributes: Each attribute with geometry domain ($a:\text{dom}_g \in G$) is a member of geo:Geometry class. $t_{as} \in T \ t_{as} = (a:\text{dom}_g \text{ a qb:AttributeProperty; rdfs:domain geo:Geometry; geo:defaultGeometry geo:Feature; rdfs:range } g)$

4.2 Ongoing Work - *Completed*

Spatial Data Warehouse Members in QB4SOLAP

Spatial Measures: QB4SOLAP allows spatial measures in the cube, represented by a geometry thus they use a different schema than conventional (numeric) measures which is from OGC schemas. Spatial measures also require the specification of spatial aggregation functions(S_{agg}).

$t_{ms} \in T$ $t_{ms} = (m_s \text{ a qb:MeasureProperty; rdfs:subPropertyOf sdmx-measure:obsValue; geo:defaultyGeometry geo:Feature})$

The class of the measure value is given with the property rdfs:domain and rdfs:range says what values each property can take such as; *point, polygon, line* etc.

Spatial Facts: F_s relates several levels of which two or more are spatial(l_s). Ideally a spatial fact cube has also spatial measures (m_s) mentioned above.

$t_{fs} = (F_s \text{ a qb:DataStructureDefiniton; qb:component[qb4o:level } l_s \text{ ; rdfs:subPropertyOf sdmx-dimension:refArea; qb4o:cardinality } card \text{ ; qb4so:topologicalRelation } T_{rel} \text{]}; qb:component[qb:measure } m_s \text{ ; qb4o:aggregateFunction } S_{agg} \text{] }).$

4.2 Ongoing Work - Completed

Spatial OLAP Operators

Let $S = (T_{rel} \cup S_{agg} \cup N_{op})$ Spatial Operations given in “Preliminaries”

S-Roll-up: Given a cube C with $d \in C$ and level $l_u \in d$, such that $l \langle a:dom_d \rangle \rightarrow *l_u$ and let roll-up relation R and spatial function S comprise R_s spatial roll-up as:

$$R_s = S(d, l \langle a:dom_d \rangle) \cup R(C, d, l_u) \rightarrow C'$$

S-Drill-down: Inverse of S-Roll-up

S-Slice: Given a cube C with n-dimensions $D = \{d_1, d_2, \dots, d_n\} \in C$, let S_s spatial slice comprised from traditional slice S' (removes a dimension d) and spatial constraint S (fix a single value in one level $L = \{l_1, l_2, \dots, l_n\} \in d$) such that;

$$S_s = S(d, l \langle a:dom_d \rangle) \cup S'(C, d) \rightarrow C' \text{ with n-1 dimensions } D = \{d_1, d_2, \dots, d_{n-1}\} \in C'$$

4.2 Ongoing Work – *Completed*

Spatial OLAP Operators

S-Dice: Dice is analogous to the relational algebra – R selection; $\sigma_{\phi} (R)$. Instead the argument is a cube C thus $\sigma_{\phi} (C)$. In SOLAP dice is not a select operation rather a nested Select and (spatial)Filter operation. S-Dice D_s keeps the cells of a cube C that satisfy a spatial Boolean $S(\Phi)$ condition over spatial dimension levels, attributes and measures.

$D_s = (C, S(\Phi)) \rightarrow C'$ where $S(\Phi) = S(\sigma_{a\phi b}(C)) \vee S(\sigma_{a\phi v}(C))$

and a, b are spatial levels(l_s), attributes($a:dom_g$) or measures(m, m_s) and v is a constant value returns a sub-cube $C' \subset C$.

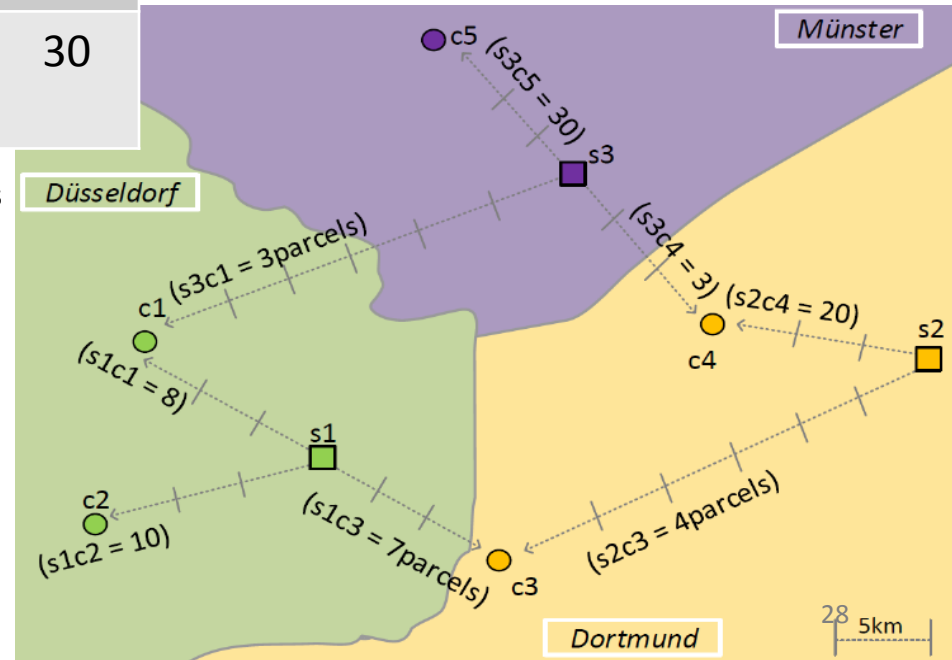
S-Drill-across: In-Progress.

4.2 Ongoing Work - Completed

Example: SOLAP Operator

City	Customer	Supplier Sales (in parcel)			Total Sales
		s1	s2	s3	
Düsseldorf	c1	8	–	3	11
	c2	10	–	–	10
Dortmund	c3	7	4	–	11
	c4	–	20	3	23
Münster	c5	–	–	30	30

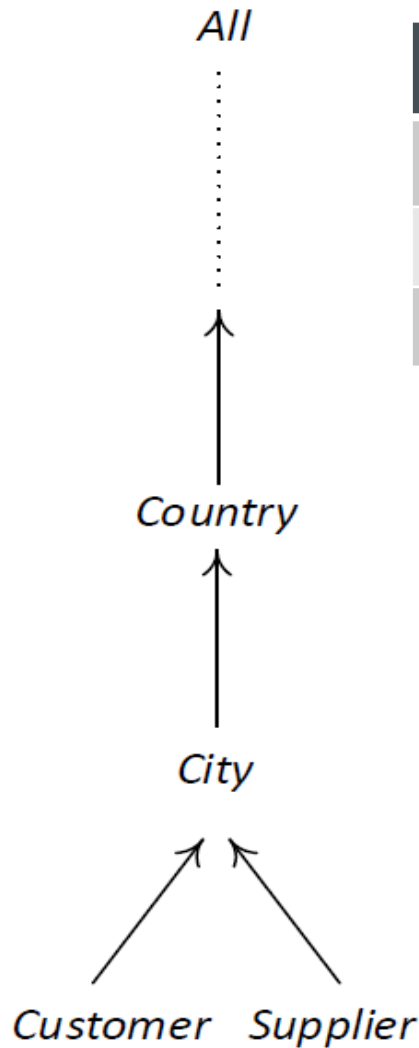
Sales to Customers from Supplier Warehouses in German Cities



4.2 Ongoing Work - Completed

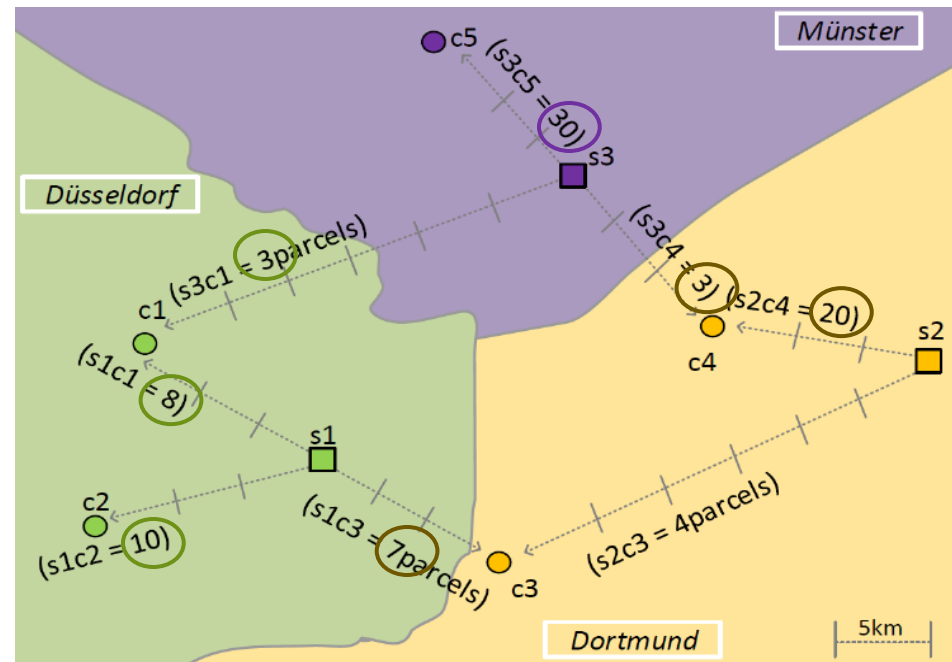
Example: SOLAP Operator

ROLLUP(Sales, (Customer → City), SUM(SalesAmount))



City	TSales1
Düsseldorf	21
Dortmund	34
Münster	30

Roll-up : Total sales to customers by city



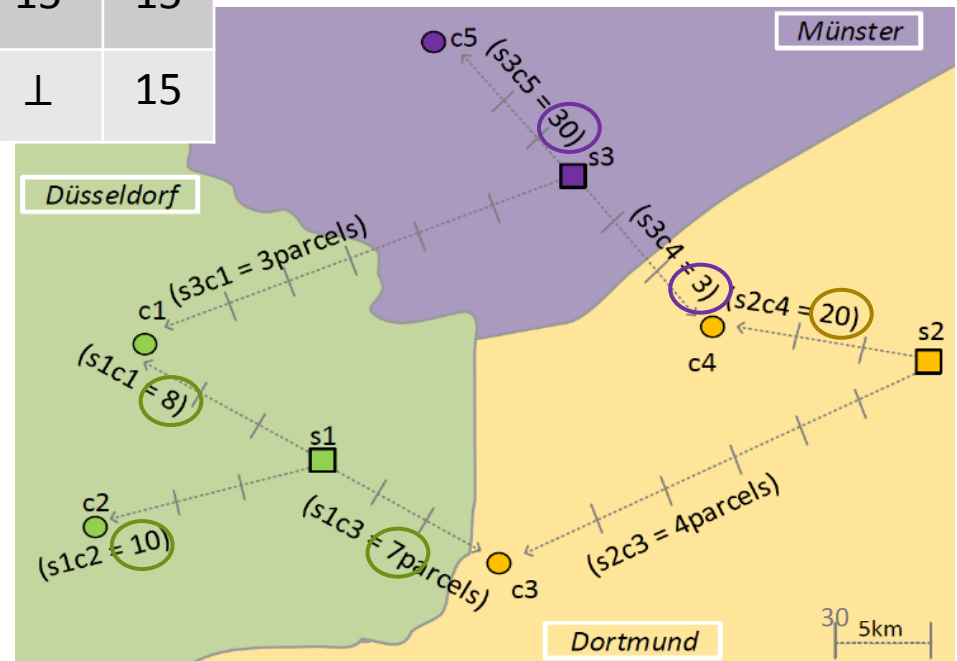
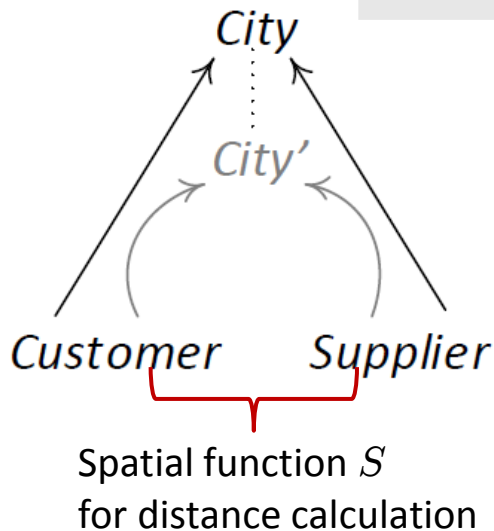
4.2 Ongoing Work - Completed

ROLLUP(Sales, [CLOSEST(Customer,Supplier)] → City'), SUM(SalesAmount))

Customer	To Supplier Distance (in km.s)		
	s1	s2	s3
c1	15	⊥	45
c2	15	⊥	⊥
c3	15	45	⊥
c4	⊥	15	15
c5	⊥	⊥	15

S-Roll-up : Total sales to customers by city of the closest supplier

City	TSales	TSales1
Düsseldorf	25	21
Dortmund	20	34
Münster	33	30



4.2 Ongoing Work - *Completed*

Querying Spatial Data Warehouses on the Semantic Web

Endpoint: <http://164.15.78.30:8890/sparql>

S-Roll-up: Total sales to customers by city of the closest supplier.

```

SELECT ?city ?custName ?sup ?minDistance (SUM(?sales) AS ?totalSales)
WHERE { ?o a qb:Observation ;
        gnw:customerID ?cust ;
        gnw:supplierID ?sup ;
        gnw:cityName ?city
        gnw:salesAmount ?sales .
        ?cust qb4o:inLevel gnw:customer ; skos:broader gnw:city;
        gnw:customerGeo ?custGeo ;
        gnw:customerName ?custName .
        ?sup gnw:supplierGeo ?supGeo .
# Total sales to the closest supplier of the customer
{ SELECT ?cust1 (MIN(?distance) AS ?minDistance)
WHERE { ?o a qb:Observation ;
        gnw:customerID ?cust1 ;
        gnw:supplierID ?sup1 .
        ?sup1 gnw:supplierGeo ?sup1Geo .
        ?cust1 gnw:customerGeo ?cust1Geo .
BIND (bif:st_distance( ?cust1Geo, ?sup1Geo ) AS ?distance)}
GROUP BY ?cust1}
FILTER ( ?cust = ?cust1 && bif:st_distance(?custGeo,?supGeo) =
?minDistance)}
GROUP BY ?city ?custName ?minDistance ?sup

```

4.2 Ongoing Work - *Completed*

Querying Spatial Data Warehouses on the Semantic Web

Endpoint: <http://164.15.78.30:8890/sparql>

S-Dice: Total sales amount to the customers that have orders delivered by suppliers such that their locations are less than 200 km from each other.

```

SELECT ?custName ?distance (SUM(?sales) AS ?totalSales)
WHERE {
    ?o a qb:Observation ;
        gnw:customerID ?cust ;
        gnw:supplierID ?sup ;
        gnw:salesAmount ?sales .
    ?cust gnw:companyName ?custName ;
        gnw:customerGeo ?custGeo .
    ?sup gnw:supplierGeo ?supGeo .
    BIND (bif:st_distance (?custGeo, ?supGeo) AS ?distance)
    FILTER ( ?distance < 200 )
}
GROUP BY ?custName ?distance
ORDER BY ?custName

```

4.2 Ongoing Work - *Completed*

Querying Spatial Data Warehouses on the Semantic Web

Endpoint: <http://164.15.78.30:8890/sparql>

S-Slice: Total sales to customers, located in a city that are within a 100 meter buffer area from a given supplier location.

```

SELECT ?custName (SUM(?sales) AS ?totalSales)
WHERE {
  ?o rdf:type qb:Observation ;
     gnw:customerID ?cust ;
     gnw:salesAmount ?sales ;
     gnw:orderDateID ?time.
  ?cust gnw:customerName ?custName ;
     skos:broader ?city .
  ?city gnw:cityGeo ?cityGeo .
  FILTER bif:st_within (?cityGeo, bif:st_point(-99.1332 , 19.4326),
0.1))
}
GROUP BY ?custName
ORDER BY ?custName

```


4.2 Ongoing Work – *In Progress*

2) Modelling and Querying Spatial Data Warehouses on the Semantic Web

Obj.1, Task 2

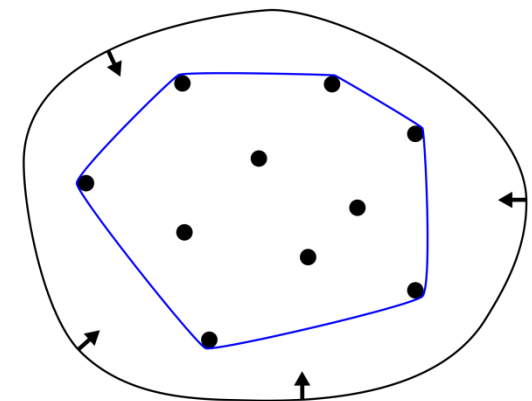
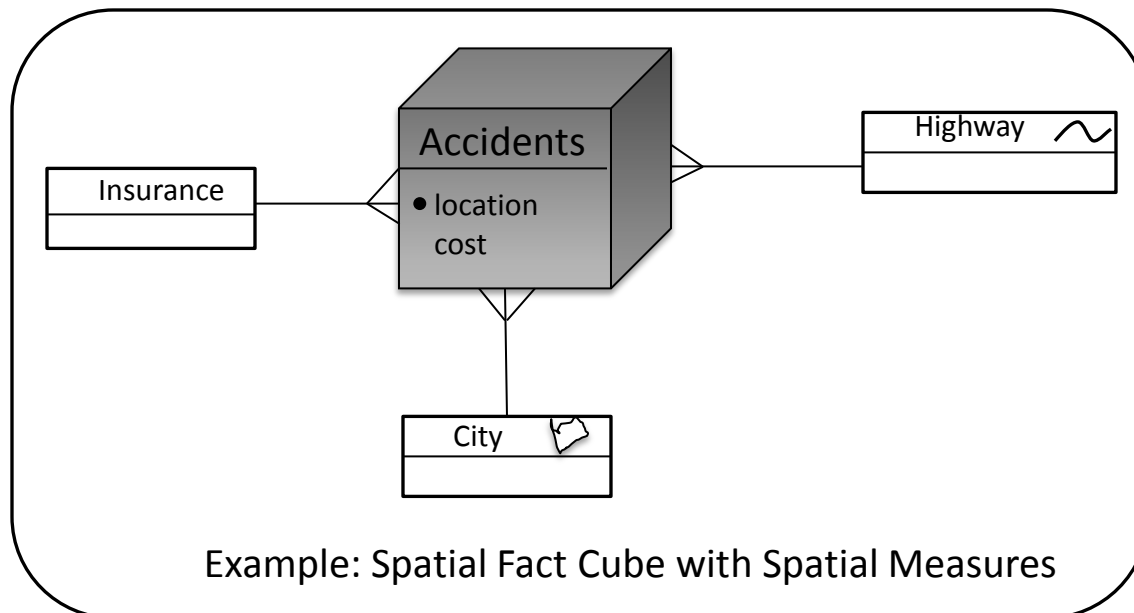
- Complete semantics of SOLAP operators (S-Drill-Across)
- Perform the SPARQL queries of all (possible) defined SOLAP operators on the use case data such as:

S-Drill-Across example: 'Total sales by country compared with those of its bordering countries'

4.2 Ongoing Work – *Research in Progress*

3) A Generic Data Model for Spatio-Temporal OLAP Cubes on the Semantic Web (Extension of 2nd Obj.) Obj.2, Task 3

- Complex spatial data types (Continuous fields; altitude, temperature)
- Spatio-temporal semantics in the conceptual model
- Spatial measures and facts in the use case
- N-ary topological relationships



S-Agg : Convex Hull area of accident locations

4) Query Processing for Spatio-Temporal Linked Data in RDF Warehouses

(Implementation of 3rd task to an extent + Performance Evaluation)

➤ Create a Spatio-Temporal SPARQL Benchmark

From: Available open benchmark data

i.e. BerlinMOD (Berlin Moving Object Database)

➤ Use latest version of QB4SOLAP meta-model

Criteria:

➤ Comparison of STDBMS with ST SPARQL Benchmark

➤ Evaluation of QB4SOLAP queries in SPARQL

4.3. Future Work

Obj.3 Task 5 & 6

5) Interacting with Governmental Linked Open Data via Spatio-Temporal OLAP Operations (Extension to 1st Case Study, full implementation of 3rd & 4th)

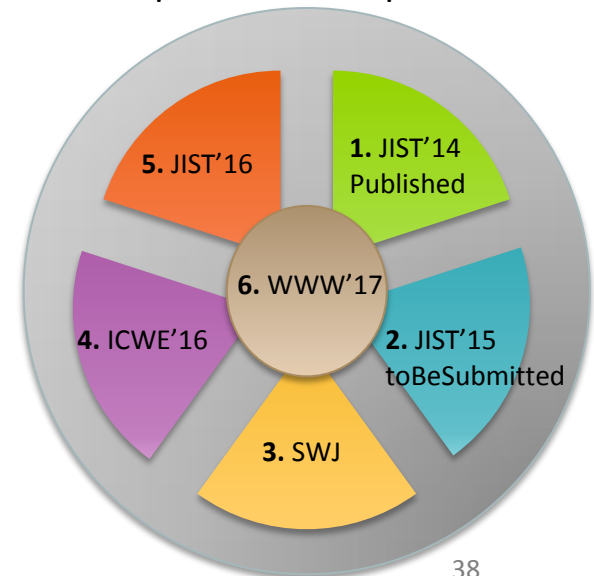
- Elicit a real-world use case from Open Data sources
 - Statistics Denmark <http://statbank.dk/>
 - Open Data Lab <http://www.odlaa.dk/>
- Cleanse, link and publish as RDF with QB4SOLAP
- Extend SPARQL with a (UDF) spatial aggregation function
- Test SOLAP queries

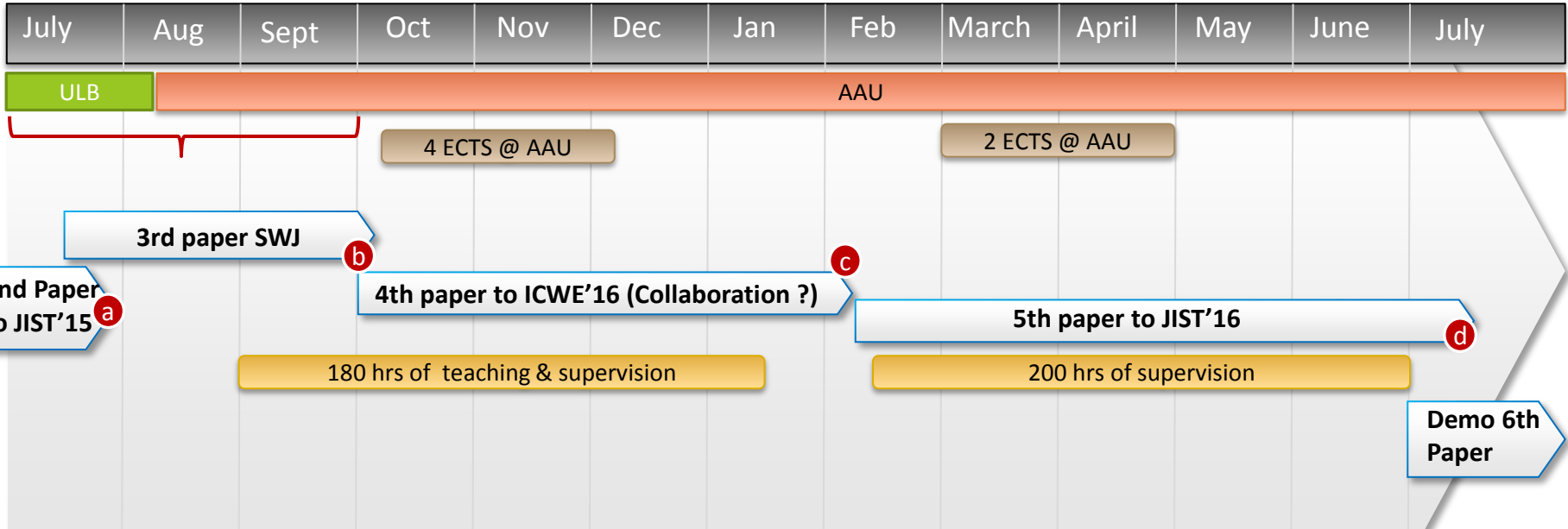
6) Demo LOST-BI: Linked Open Spatio-temporal Data to Business Intelligence (Proof of concept overall objectives & visualization of 5th)

- Implement Queries on a running system with a GUI
- Deployment outside the research group

5. Publications

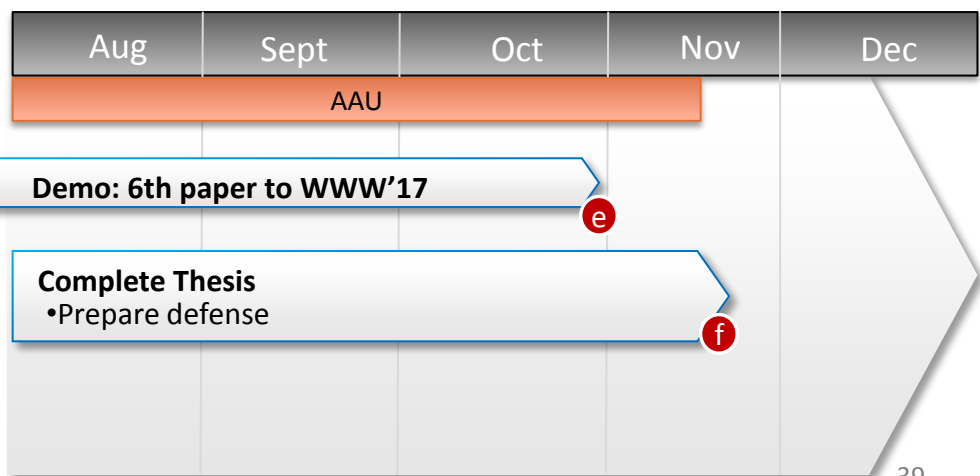
- 1. Publishing Danish Agricultural government Data as Semantic Web Data
- 2. Modeling and Querying Spatial Data Warehouses on the Semantic Web
- 3. A Generic Data Model for Spatio-Temporal OLAP Cubes on the Semantic Web
- 4. *Query Processing for Spatio-Temporal Linked Data in RDF Warehouses*
- 5. Interacting with Governmental Linked Open Data via Spatio-Temporal OLAP Operations
- 6. *Demo LOST-BI: Agricultural Linked Open Spatio-Temporal Data to Business Intelligence*





milestones

- a) Submit 2nd paper (20th of July)
- b) Submit 3rd paper (1st of October)
- c) Submit 4th paper (Beginning of Feb.)
- d) Submit 5th paper (End of July)
- e) Submit 6th Demo paper (Nov.)
- f) Submit the Thesis (Mid. Nov.)



7. P.hD.Courses

Courses	Place	ECTS	G/P	Time	Status
Introduction to the PhD Study	AAU	1.0	G	Fall '13	Finished
PBL in Eng. and Science - Development of Supervisor Skills	AAU	2.0	G	Fall '13	Finished
Semantic Web Warehousing	AAU	2.0	P	Spring'14	Finished
Writing and reviewing Scientific Papers	AAU	3.75	G	Spring'14	Finished
IT4BI Summer school – eBISS	Berlin	2.0	P	Summer'14	Finished
Academic writing in English	AAU	3.0	G	Fall'14	Finished
Library Information Management	AAU	1.0	G	Fall'14	Finished
Col.Oriented DB Man. Systems on Modern Hardware	AAU	2.0	P	Fall'14	Finished
Conference attendance	JIST	1.0	P	Fall'14	Finished
French Language Course	ULB	1.0	G	Spring'15	Finished
Intellectual property, copyright, and knowledge transfer	ULB	2.0	G	Spring' 15	Finished
IT4BI-DC Doctoral Consortium	Barcelona	2.0	P	Summer' 15	Enrolled
Conference Attendance	JIST	1.0	P	Fall'15	Planned
Data Management and Integration on the Web	AAU	2.0	P	Fall'15	Planned
Big Data Management Systems: Concepts, Architectures and Implementation (or Study Group)	AAU	2.0	P	Fall'15	Planned
Project Management and interpersonal skills	AAU	2.0	G	Spring'16	t.b.d.
Conference attendance	t.b.d.	1.0	P	Spring'16	t.b.d.
	Total	30,75	P: 15 + G: 15,75 ECTS		

8. Discussion

Thank you for your attention

nurefsan@cs.aau.dk
nurefsan.gur@ulb.ac.be