



# Publishing Danish Agricultural Government Data as Semantic Web Data

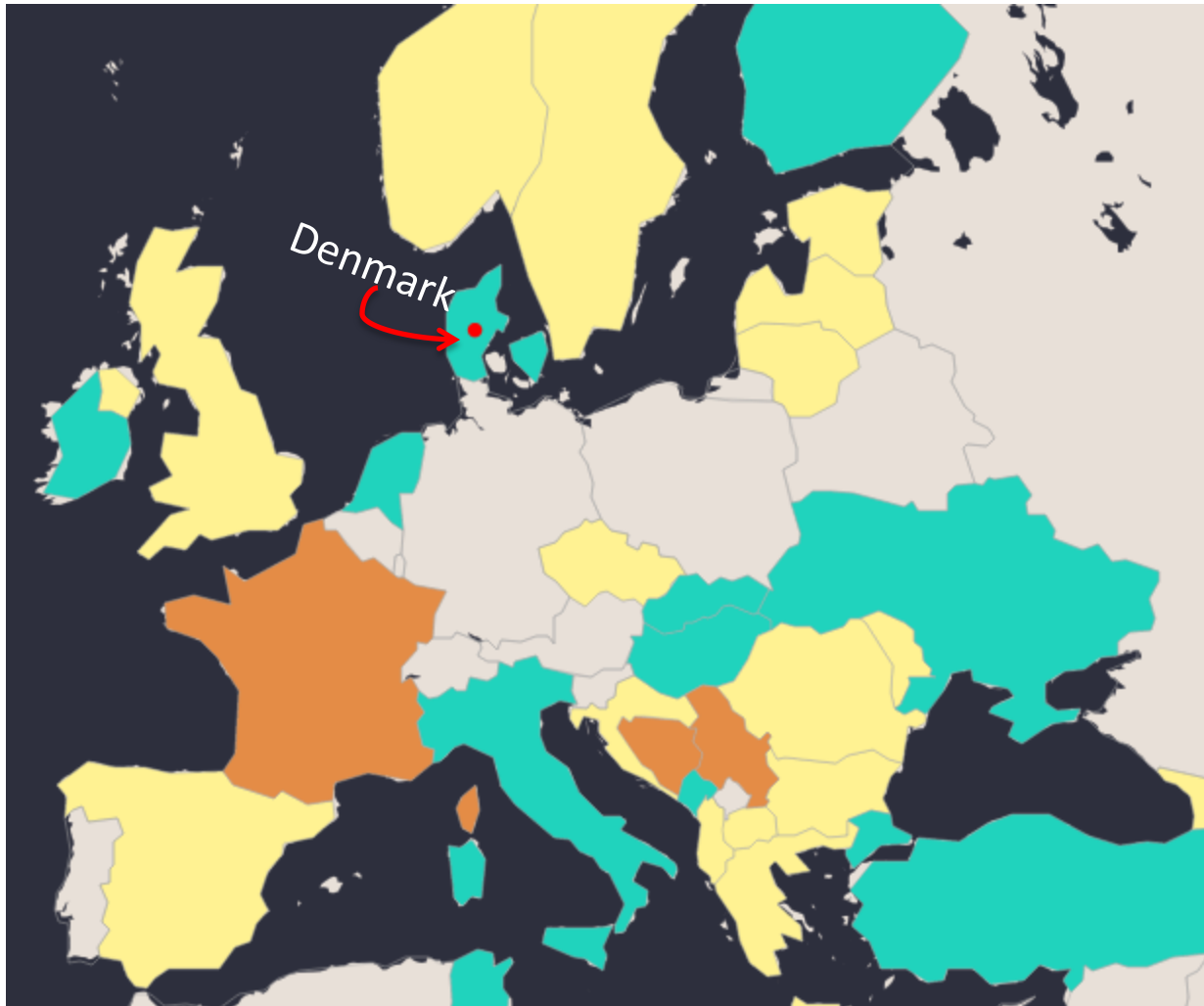
November 10<sup>th</sup>, 2014  
Chiang Mai, Thailand

Alex B. Andersen, **Nurefşan Gür**, Katja Hose,  
Kim A. Jakobsen and Torben Bach Pedersen

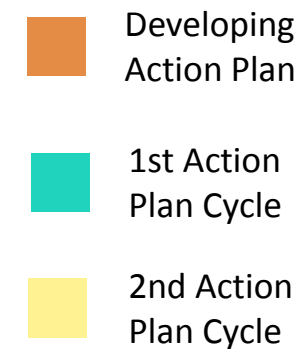
# Outline

1. Motivation
2. Process (Implementation)
3. Use Case
4. Process (Iteration)
5. Evaluation
6. Conclusion

# 1. Motivation: Open Governmental Data



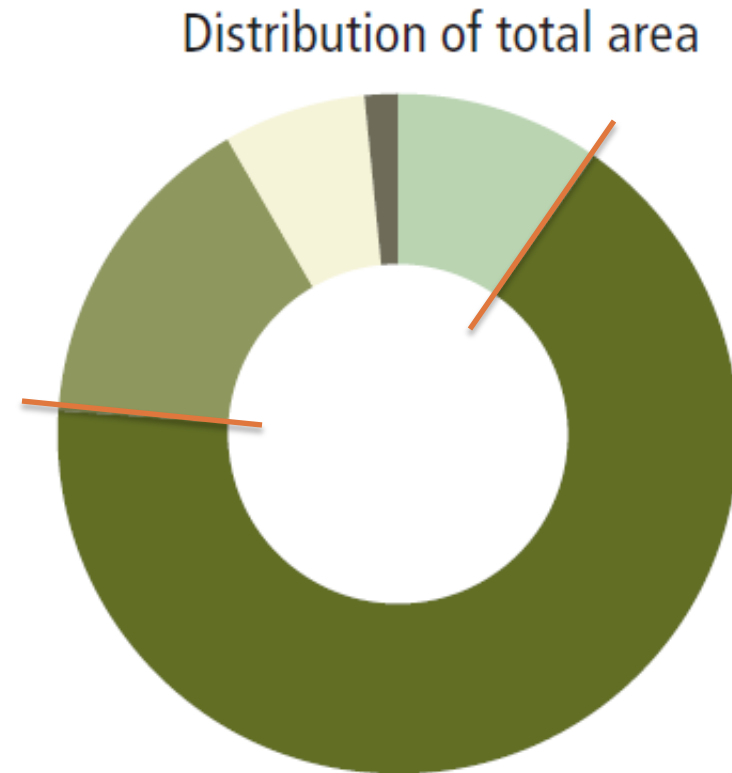
- Denmark :**
- Joined in 2012
  - Commitments on E-Government &
  - Open Data



Open Government Participation

# 1.Motivation: Danish Agriculture

- Artificial surfaces, 9.8 per cent
- ➔ ■ Agricultural areas, 66.3 per cent
- Forests and semi-natural areas, 15.6 per cent
- Wetlands, 6.8 per cent
- Unclassified, 1.5 per cent



# 1.Motivation: Example

Customer needs:



Specific (organic) crops from



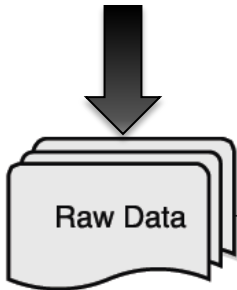
supplier companies



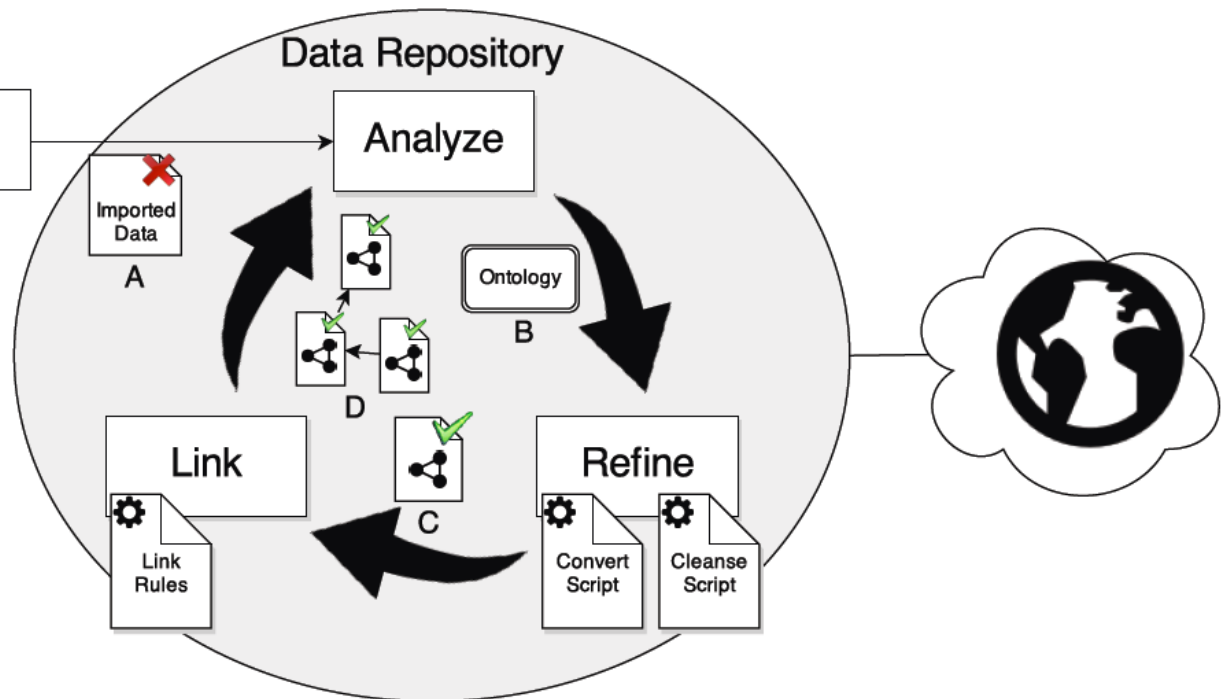
in Nordjylland

# 2.Process (Implementation)

Agricultural Datasets



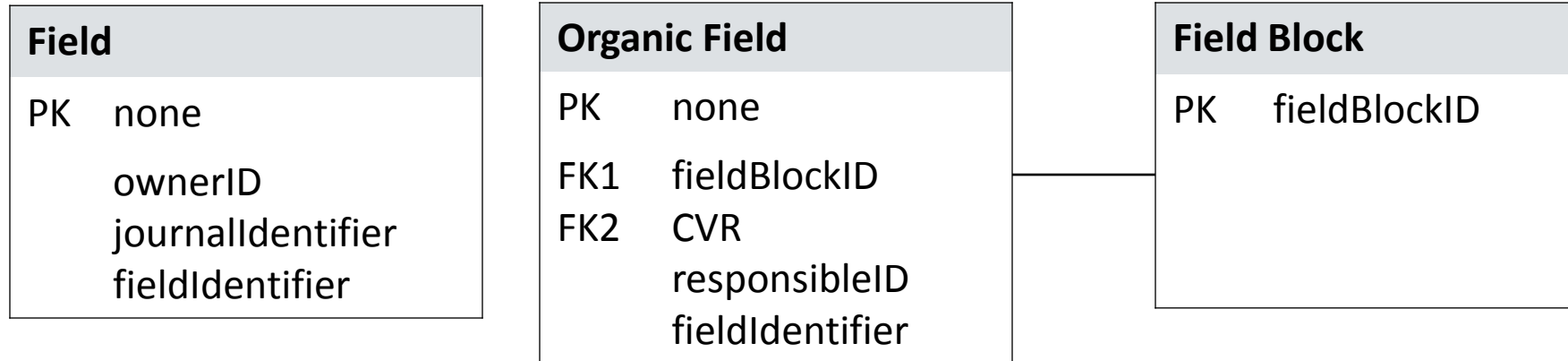
Business Datasets



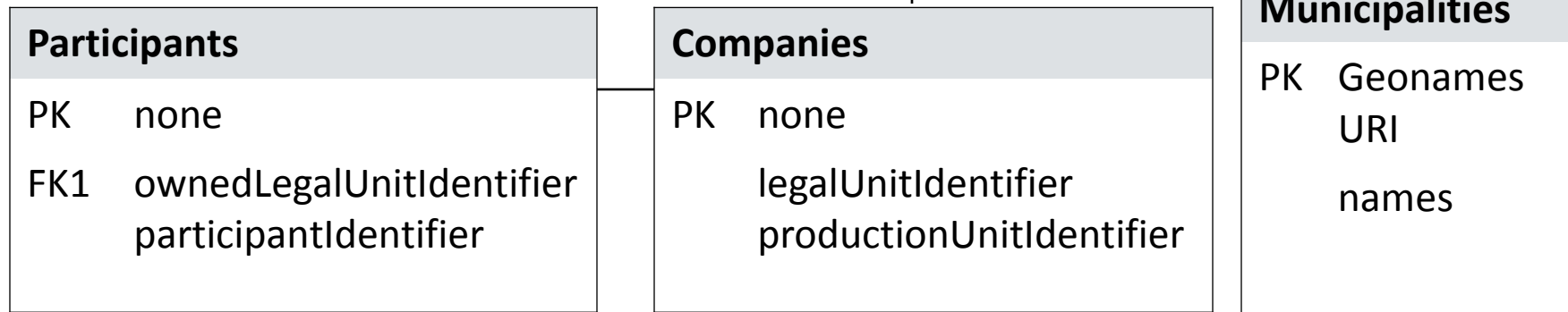
Process Overview

# 3. Use Case : Data Schema

## Agricultural Data Collection

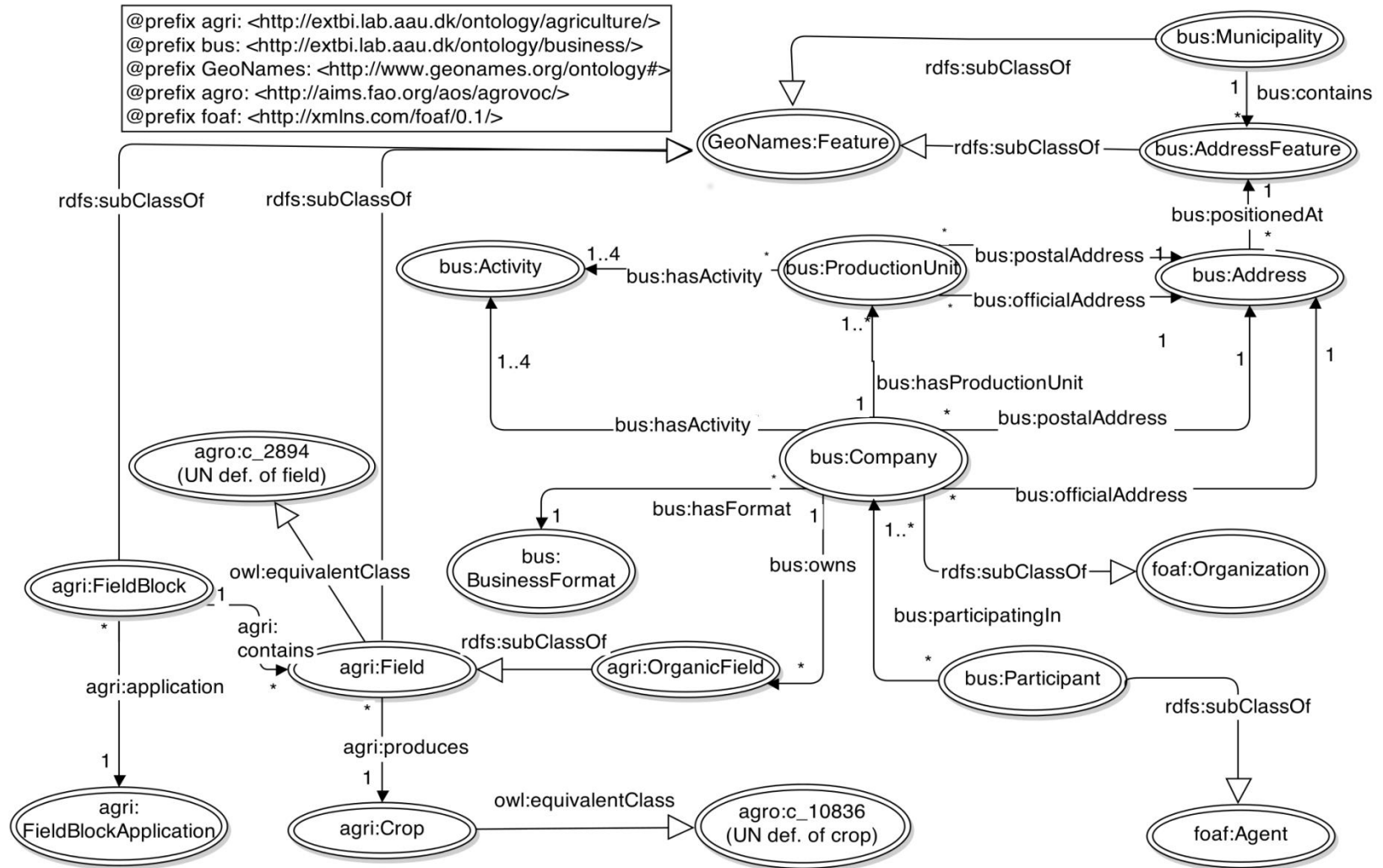


## Business Data Collection



Compact schemas for collected datasets

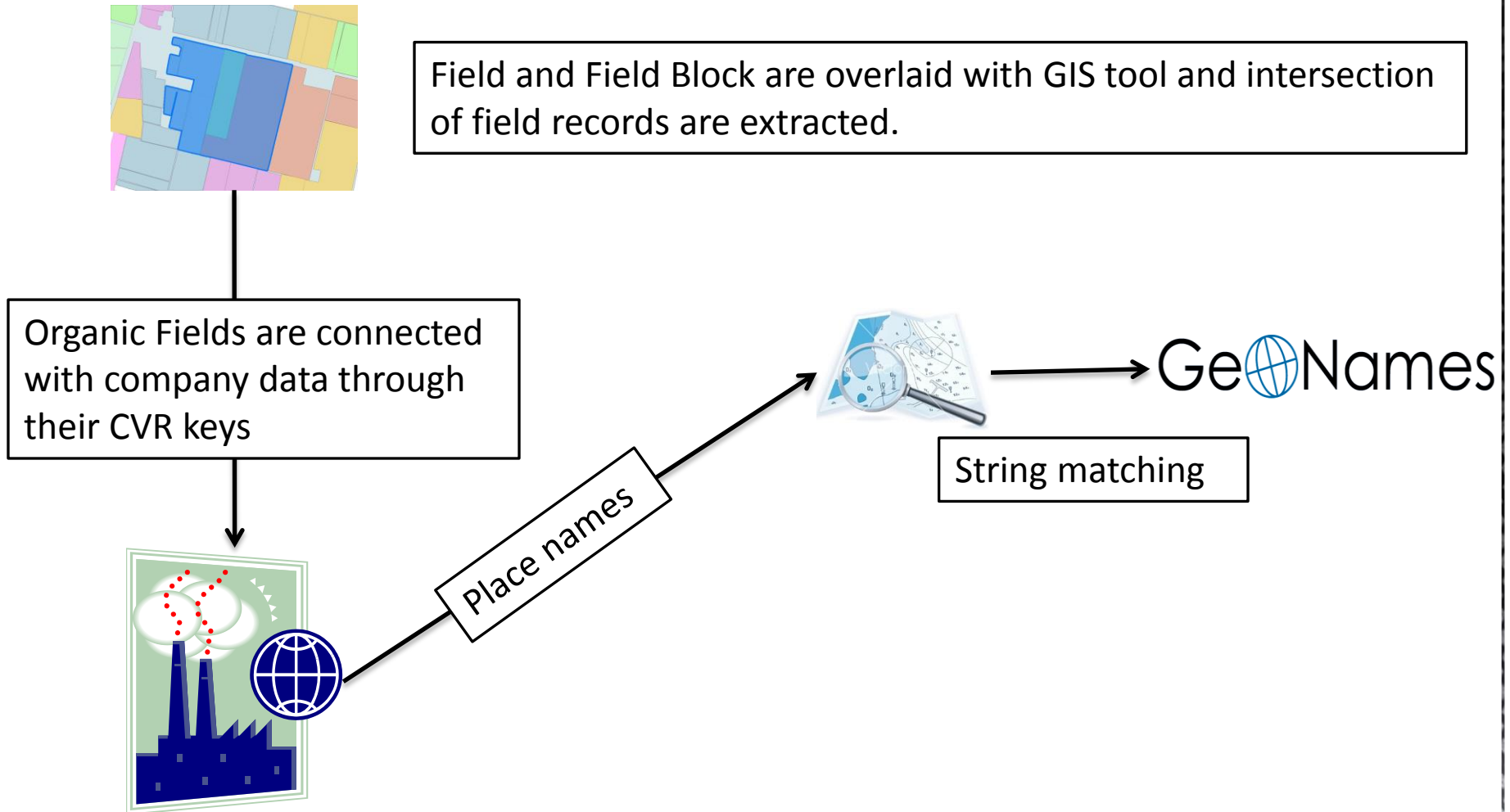
# 3. Use Case: Ontology



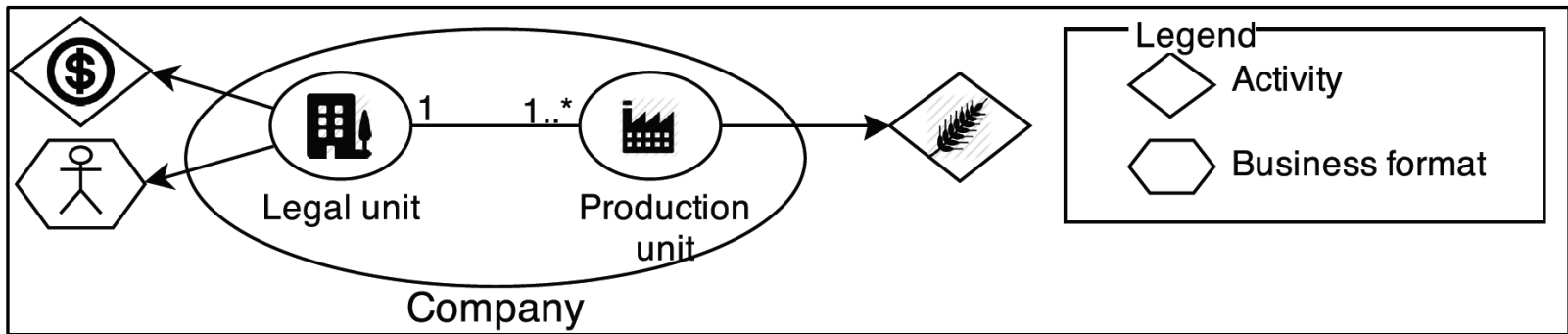
Ontology Schema



# 3. Use Case : Conversion & Linking



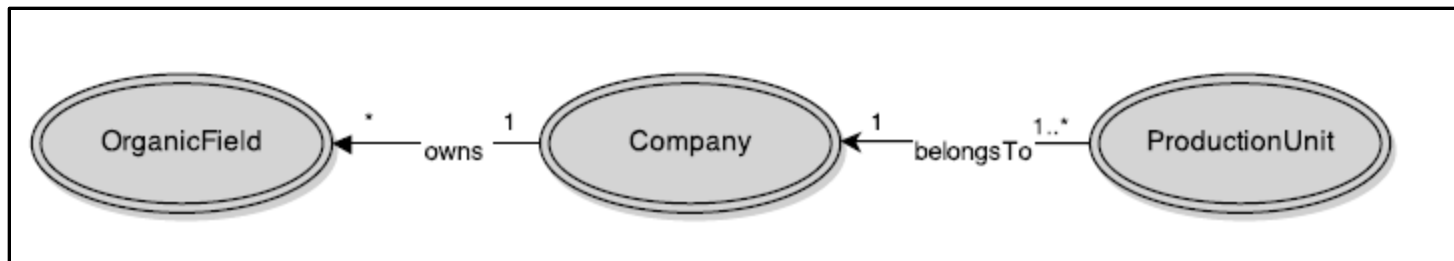
# 3. Use Case : Conversion & Linking



Units of Company

Responsible	CVR Number	Field	Field Block	Area
51149	13317186	10-1	696200-41	12869.7
51149	13317186	10-2	514269-84	35560.4

Sample Relation of CVR and Field (Records belong to Organic Field Dataset)



Relation Links of Company with Fields

# 4. Process (Iteration)

Agricultural Datasets

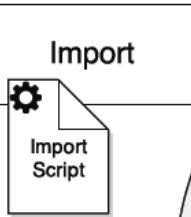
Ministeriet for Fødevarer,  
Landbrug og Fiskeri



<http://fvm.dk>

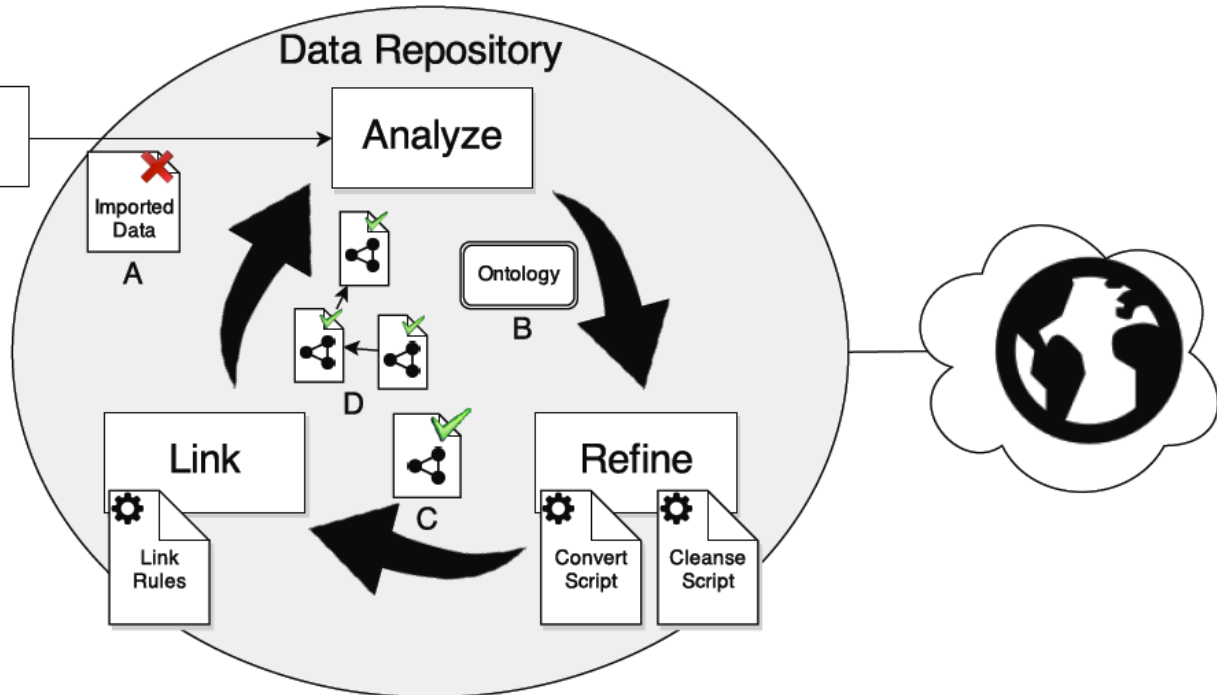


Import



<http://cvr.dk>

Business Datasets



Process Overview

# 5. Evaluation : Overview

1. Setup and Data
2. Evaluation Strategies
3. Queries
4. Results (Load times & Run times)
5. Integration (Load) Time
6. Query Evaluation (Run) Time

# 5.Evaluation : Setup and Data

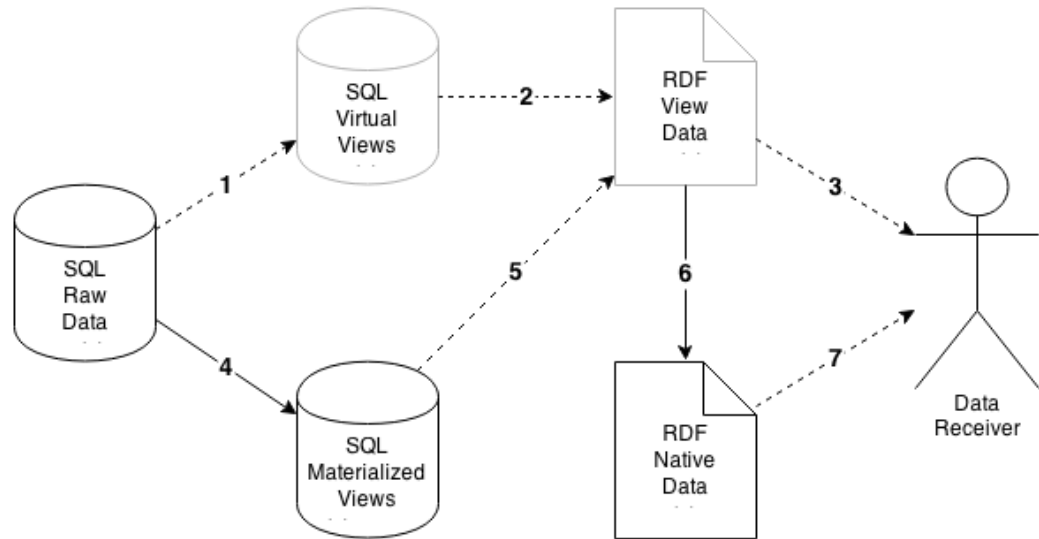
OpenLink Virtuoso 07.00.3203 server  
 3.4 GHz Intel Core i7-2600 processor  
 8 GB RAM  
 Ubuntu 13.10, saucy

SETUP

Data Collection	Sets	Entities	Triples	Date
Agricultural Data Collection	Field, Field Block, Organic Field	984,789	9,277,293	October 1 <sup>st</sup> , 2013
Business Data Collection	Companies, Production Units, Participants	1,617,235	22,819,139	October, 2013

# 5.Evaluation : Strategies

1. Virtual (1-2-3)
2. Materialized(4-5-3)
3. Native(4-5-6-7)



Data Flow for Materialization Strategies

Query Templates	Definition	Sample Constructs
SQT	Standard Query Template	SELECT, WHERE, FILTER, NOT EXISTS
AQT	Aggregate Query Template	ORDER BY, SUM, GROUP BY, COUNT

# 5. Evaluation : Queries

```

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

# 5.Evaluation: Results

Step	Virtual	Materialized	Native
Data Cleansing	74.92	603.35	603.35
Load Ontology	1.01	1.01	1.01
Load Mappings	8.76	12.35	12.35
Dump RDF	0.00	0.00	4684.82
Load RDF	0.00	0.00	840.04
<b>Total</b>	84.68	616.70	6141.56

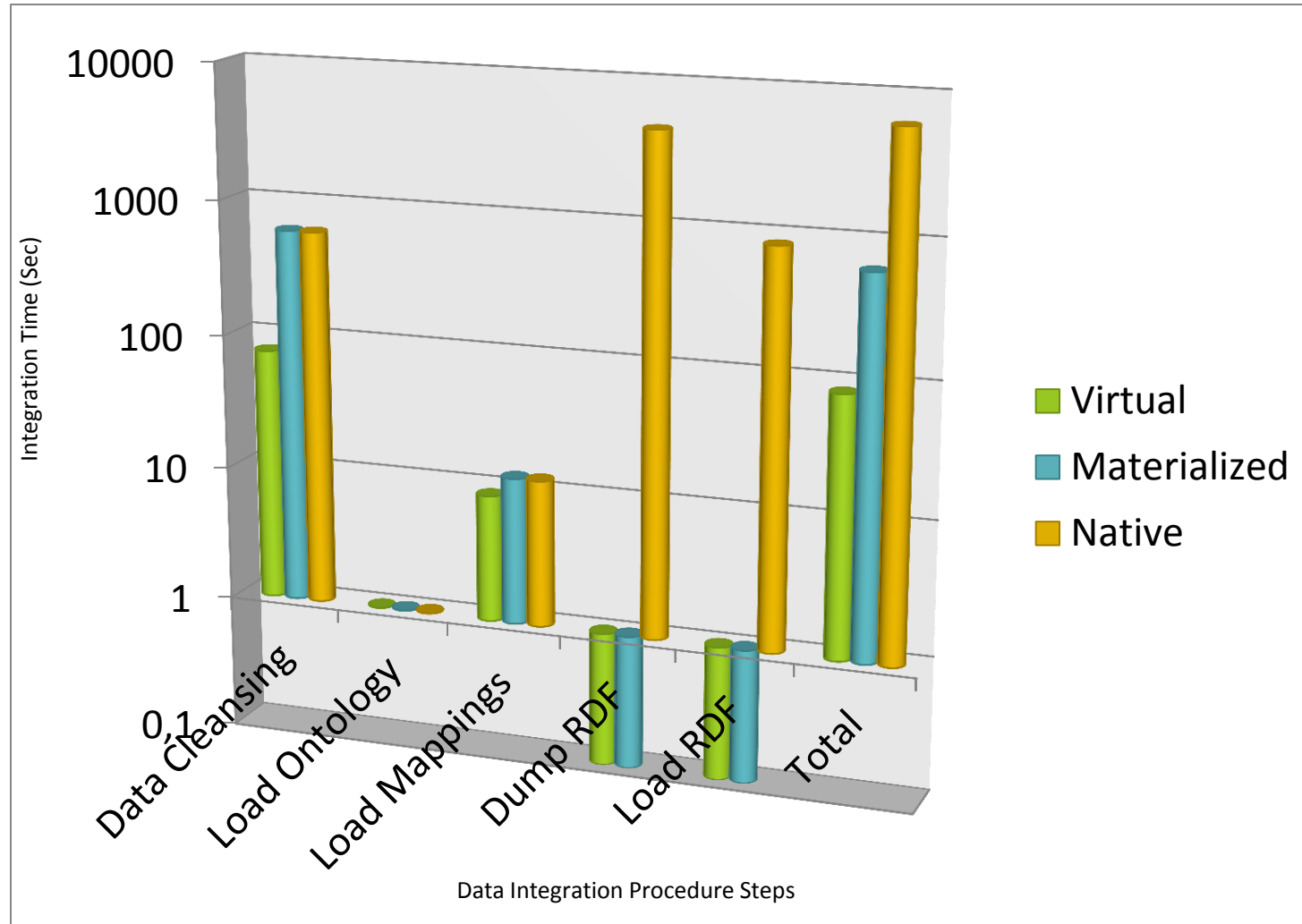
Load times in seconds

Query	Triple Patterns	Special	Virtual	Materialized	Native
AQT 1	3		5.92	3.39	1.04
AQT 2	2	SUM	13.32	7.00	0.23
AQT 3	2	ORDER BY	10.81	7.70	0.05
AQT 4	4		-	-	0.14
AQT 5	3	2 GROUP BYs	-	20.37	0.86
SQT 1	5		-	-	2.35
SQT 2	2		0.09	0.12	0.10
SQT 3	2	Un-index FILTER	2188.85	1.81	0.40
SQT 4	3	NOT EXISTS	6.57	2.35	1.63
SQT 5	5		-	23.79	3.29
<b>Average</b>			370.93	8.31	1.01

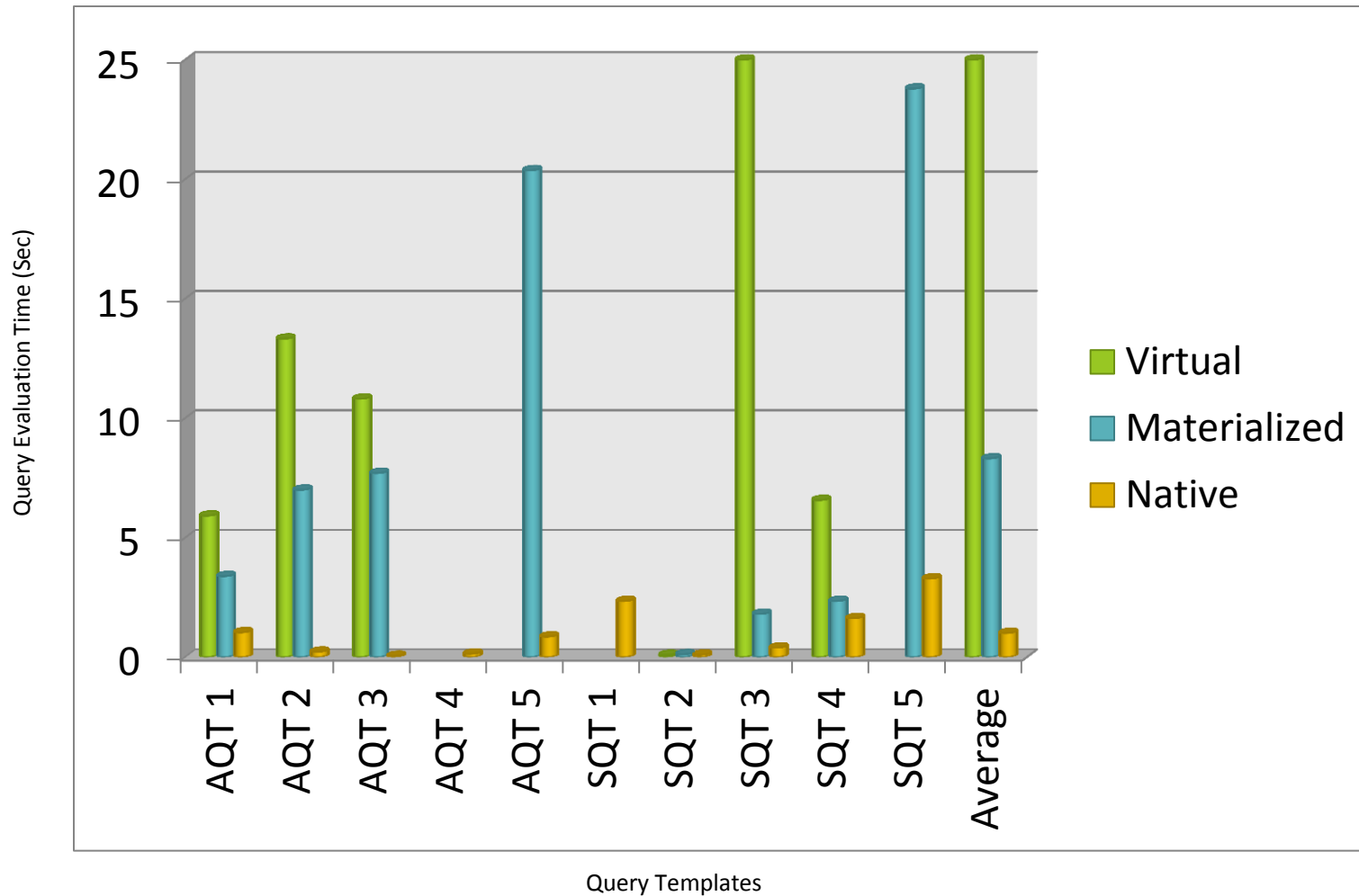
Runtimes in seconds



# 5.Evaluation : Integration Time

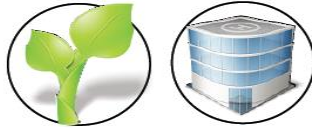


# 5.Evaluation : Query Evaluation Time

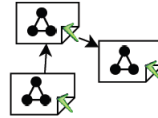


# 6. Conclusion : Summary

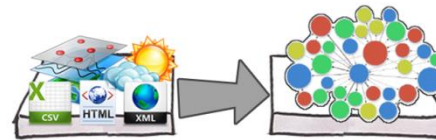
- Use-Case introduction



- Designated Ontology



- Conversion and Linking of Danish Agricultural and Business Data



- Iterative Process Implementation



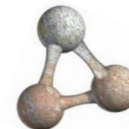
- Sample Queries

```
SELECT ?crop COUNT (*) AS ?cnt
FROM <http://extbi.lab.aau.dk/resource/agriculture>
WHERE {
  ? field agri:produces ?crop .
```

- Performance evaluation of converted data with three different strategies



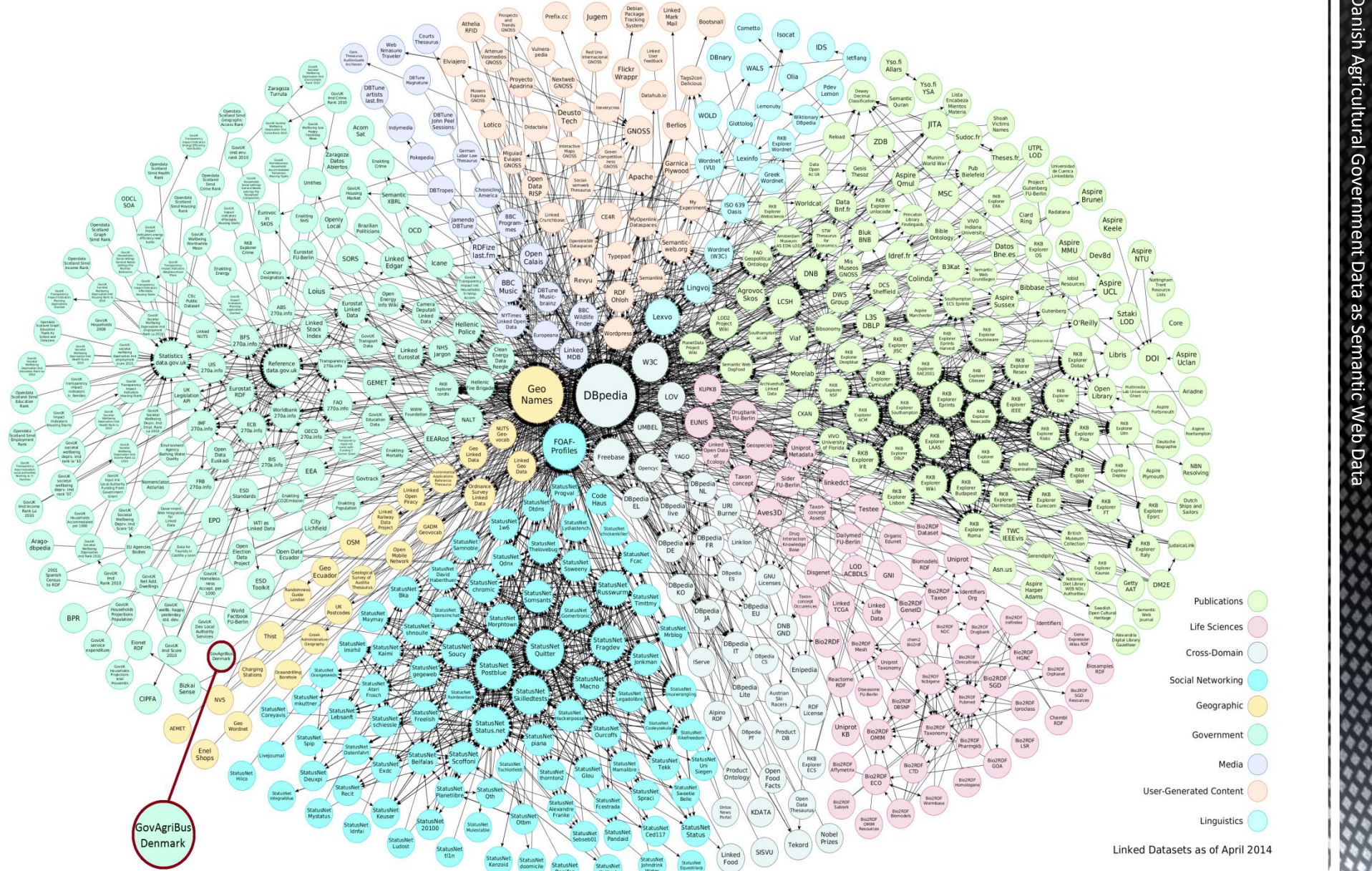
- *Future work: Spatio-Temporal LOD towards Business Intelligence*



# 6. Conclusion

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

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



Linked Datasets as of April 2014



# Thank you for your attention

## Questions?

[nurefsan@cs.aau.dk](mailto:nurefsan@cs.aau.dk)

[nurefsan.gur@ulb.ac.be](mailto:nurefsan.gur@ulb.ac.be)

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

This Project is funded by EM – Erasmus Mundus with an additional fund from AAU – Aalborg University.



ECOLE  
POLYTECHNIQUE  
DE BRUXELLES



AALBORG UNIVERSITY  
DENMARK