SEMANTICS IN DAIRY FARMING

Jacco Spek

Running presentation of the semantics work in dairy farming

2016

SDF 1.0 (2011 – 2014)
SDF 2.0 (2015 – 2017)

- **Collaboration project**
  - 3 Cooperations
  - 7 SME's
  - 5 Research institutes
  - 7 Real farmers
- Timeline:
  - SDF1: 2011 – 2014
- Nothern part of the Netherlands
- Website (in Dutch):
  - http://www.smartdairyfarming.nl/nl/

- **Goal of SDF:**
  - to support dairy farmers in the care of **individual animals.**
  - with the specific goal of a **longer productive stay** at the farm due to **improvement of individual health**.

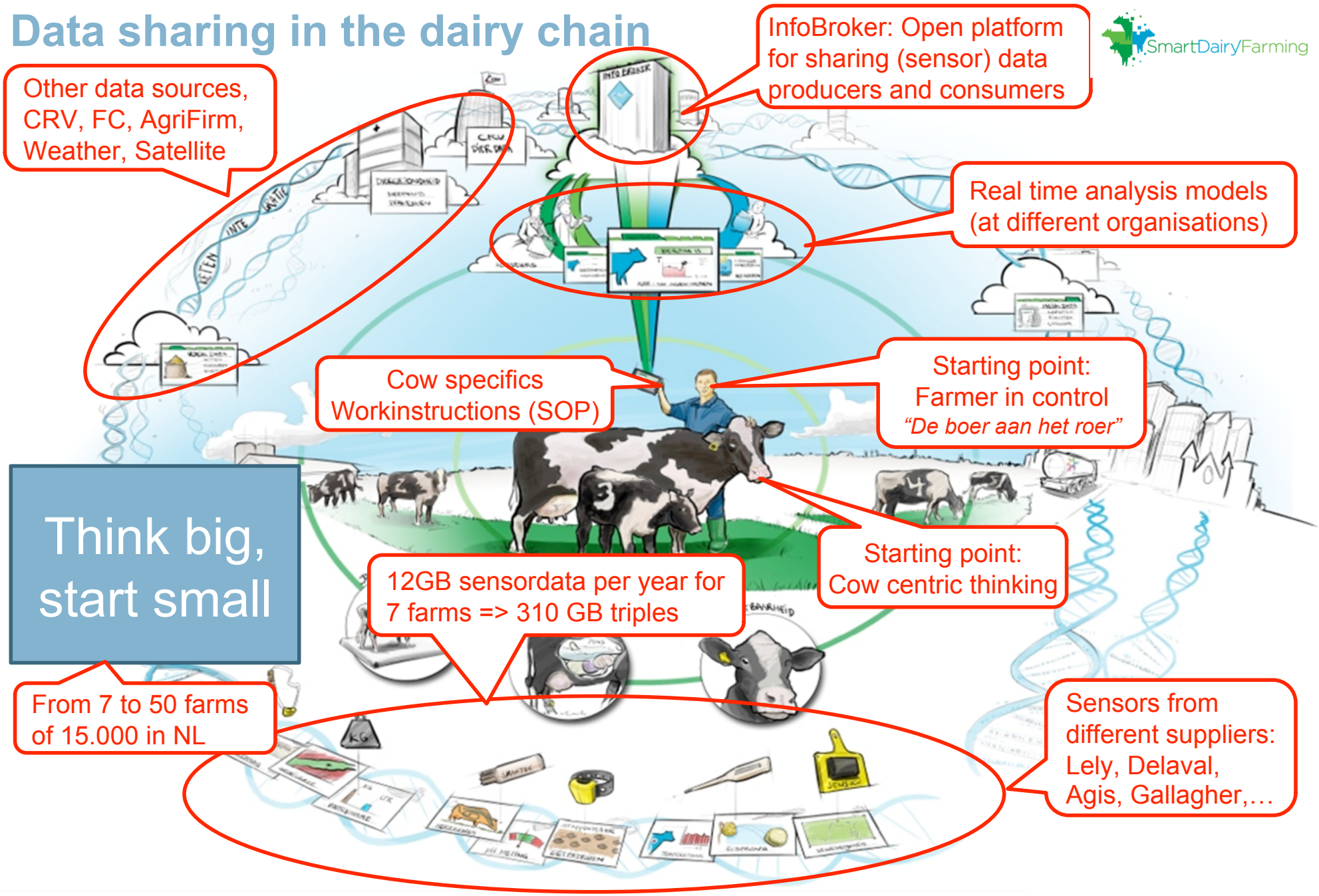
- **Challenge SDF2:**
  - more farmers: from 7 to 60 (and prepare for 2500)
  - more sensor suppliers and more data consumers
  - incorporate semantics and big data analysis

Numbers for the Dutch situation:

- 15000+ farmers
- in total more then 1.5 million milk cows
- 20 to 200+ datafields per cow
- many different stakeholders in the chain

Data sharing in the dairy chain

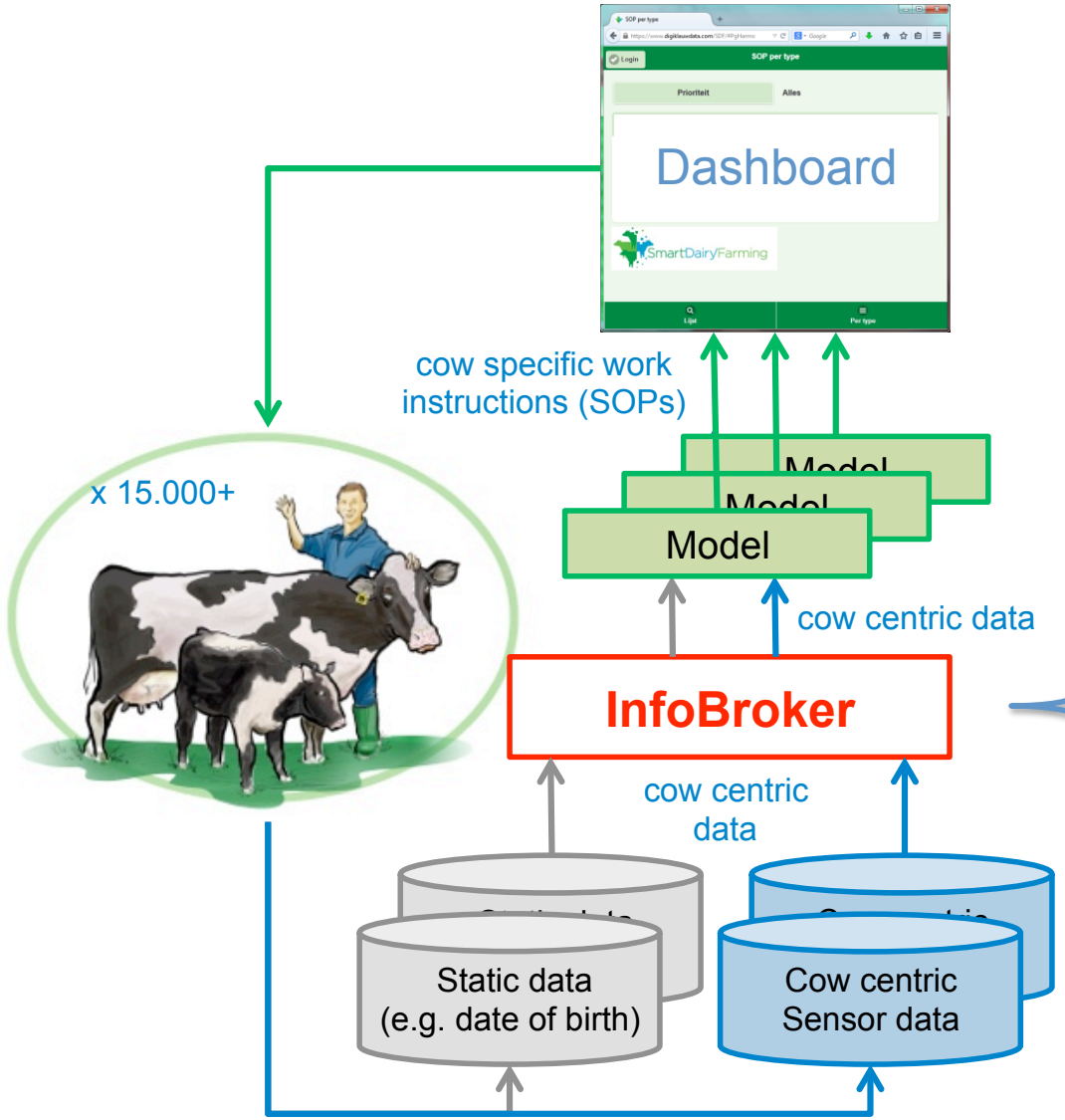
SmartDairyFarming



This project is made possible by:

provinsje fryslân provincie fryslân · provincie groningen · provincie Drenthe · Ministerie van Economische Zaken, Landbouw en Innovatie · SNN · landbouw, natuur en voedselkwaliteit · Mesdag · Productschap Zuivel · LTO Nederland

InfoBroker concept



InfoBroker functionalities:

- Open interfaces for data exchange (API)
- Authentication
  - who are you (are you allowed to login)
- Permissions
  - which data may be used by whom
  - to be set by the farmers
- Namingservice
  - location where the data can be found
    - static data
    - cow-centric sensor data
- Integration
  - combining info from different sources
- Pay-per-use
  - fixed costs (connections)
  - variable costs (used data)

So:

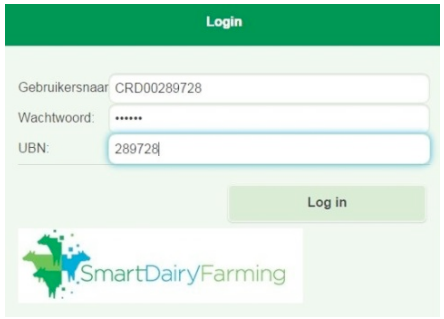
- no central datastore for (sensor)data!
- but indeed a broker
- and reduces/prevents duplication

SDF in practice

Farmer: Dairy Campus

SmartDairyFarming

Dashboard

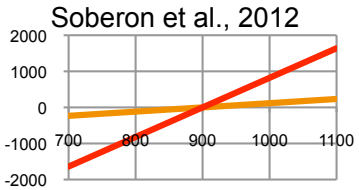


SOP-generation by models

Speenschema aanpassen
Krachtvoer aanpassen
Ruwvoer aanpassen
Pink insemineren
Kalf/pink behandelen
Voeradvies inwinnen
Selecteren voor afvoer

SOP-application
SOPs

Soberon et al., 2012



Model Fertility
Model Transition
Model Young cattle

Datacollection on the farm

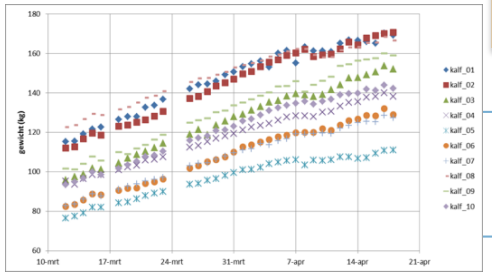
N.A. 40x

FORSTER TECHNIK
Milk intake
RFID 2x

GALLAGHER
Weight 4x
RFID 4x
Water intake 4x

Dairy Campus laptop
loggers 6x

Sensor data logistics



InfoBroker
rovecom

```
life_number,sensor,date_time,wcorr,wstable,wzero,wavg,wavgmin,wavgmax,werrors,wnousecounter,wnousetime,wusetime
NL 916075572,dc_roostervloerhok1_weegschaal1,2013-11-04 04:18:35 UTC,129.0,129.0,0.0,102.3,30.8,129.0,0,0,470,18590
NL 916075572,dc_roostervloerhok1_weegschaal1,2013-11-04 04:18:40 UTC,129.5,129.5,0.0,129.6,129.5,130.0,0,0,475,18590
NL 916075572,dc_roostervloerhok1_weegschaal1,2013-11-04 04:18:45 UTC,130.0,130.0,0.0,129.5,129.0,130.0,0,0,480,18590
NL 916075572,dc_roostervloerhok1_weegschaal1,2013-11-04 04:18:50 UTC,130.0,130.0,0.0,129.8,129.5,130.0,0,0,485,18590
```

Forster.csv

pass trough
S&S Systems

Match & Merge
rovecom

milkintake.csv
weight.csv
waterintake.csv

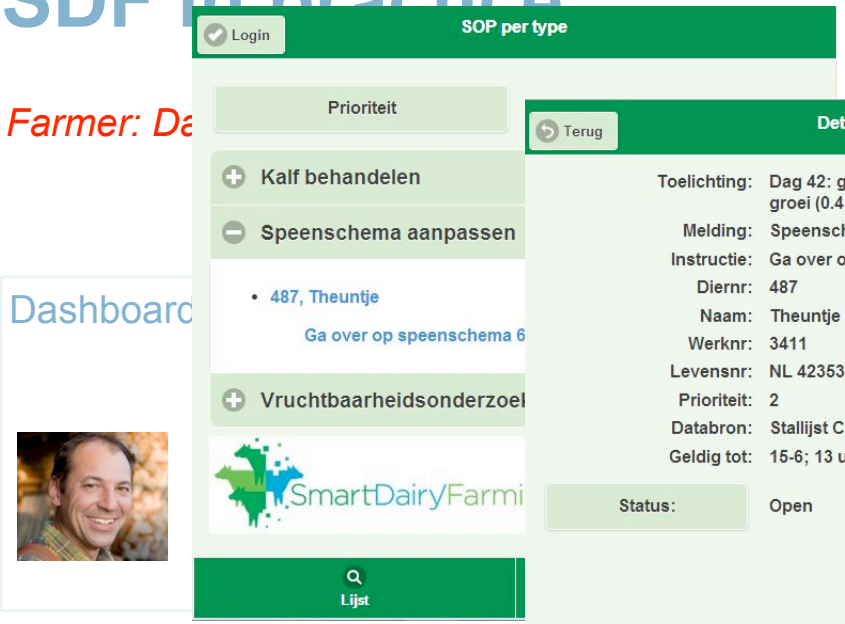
RFID -> lifenr

Realtime cow centric sensor data
sftp
AnySense
TNO

Static data
CRV_Animalregistr.
CRV_Diagnosys
CRV_Treatment

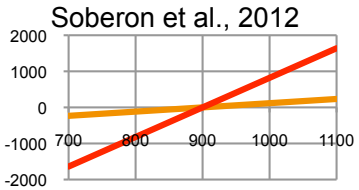
SDF in practice

Farmer: Da



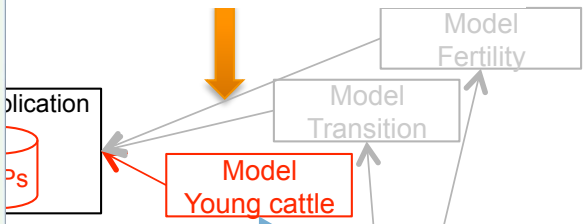
SmartDairyFarming

Soberon et al., 2012

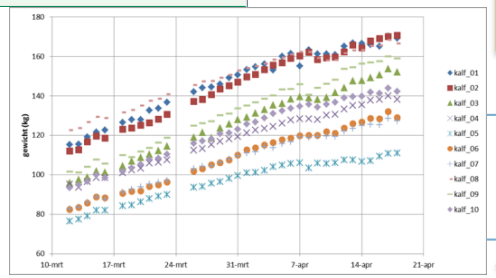


Dashboard

Models



Datacollection on the farm



InfoBroker

rovecom

```
life_number,sensor,date_time,wcorr,wstable,wzero,wavg,wavgmin,wavgmax,werrors,wnousecounter,wnousetime,wusetime
NL 916075572,dc_roostervloerhok1_weegschaal1,2013-11-04 04:18:35 UTC,129.0,129.0,0.0,102.3,30.8,129.0,0,0,470,18590
NL 916075572,dc_roostervloerhok1_weegschaal1,2013-11-04 04:18:40 UTC,129.5,129.5,0.0,129.6,128.5,130.0,0,0,475,18590
NL 916075572,dc_roostervloerhok1_weegschaal1,2013-11-04 04:18:45 UTC,130.0,130.0,0.0,129.5,128.0,130.0,0,0,480,18590
NL 916075572,dc_roostervloerhok1_weegschaal1,2013-11-04 04:18:50 UTC,130.0,130.0,0.0,129.8,128.5,130.0,0,0,485,18590
```

FÖRSTER TECHNIK — Milk intake, RFID 2x

GALLAGHER — Weight 4x

GALLAGHER — RFID 4x

Water intake 4x

Dairy Campus

Forster.csv

loggers

pass trough

Match & Merge

weight.csv

waterintake.csv

sftp

Realtime cow centric sensor data

AnySense

TNO

RFID -> lifenr

Static data

CRV_Animalregistr.
CRV_Diagnosys
CRV_Treatment

In SDF 1.0 a temporary route and storage

- to make it possible with existing/new sensor systems

Direct coupling with InfoBroker in SDF 2.0

- storage at the sensor system/provider
- and registration at the InfoBroker

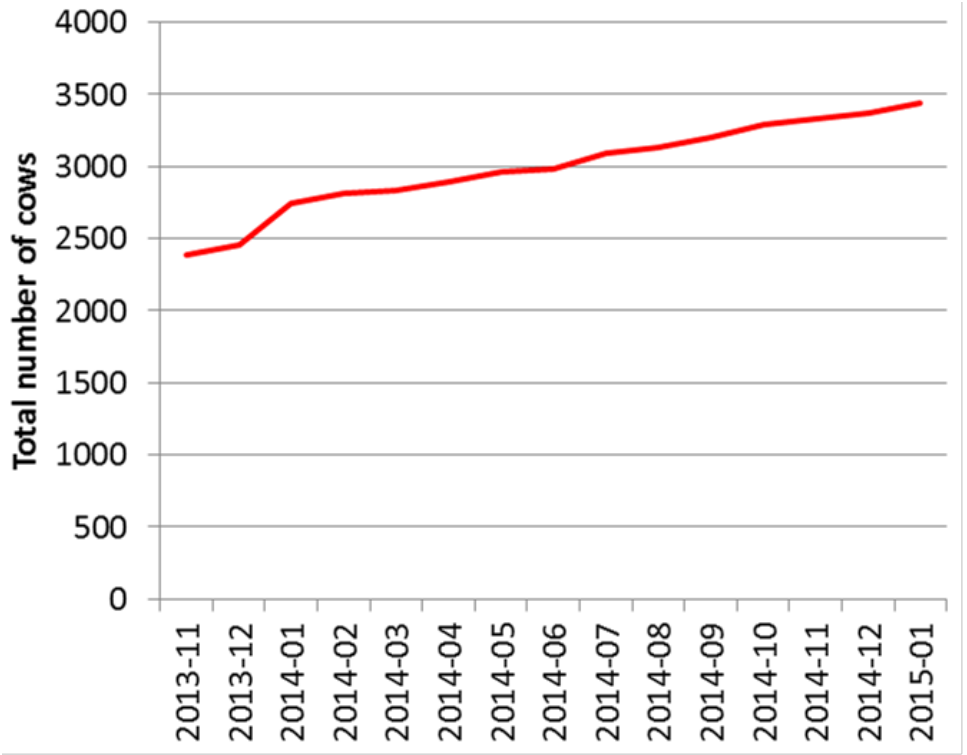
InfoBroker – Facts & Figures

	Farm 1	Farm 2	Farm 3	Farm 4	Farm 5	Farm 6	Farm 7
# cows/calves	459	186	315	239	706	202	351
Behaviour	x				x		
Temperature	x				x		
Activity	x	x	x	x	x	x	
Milk production	x	x			x	x	x
Food intake		x				x	x
Weight	x	x	x	x	x	x	x
Water intake			x	x			
Milk intake			x	x			

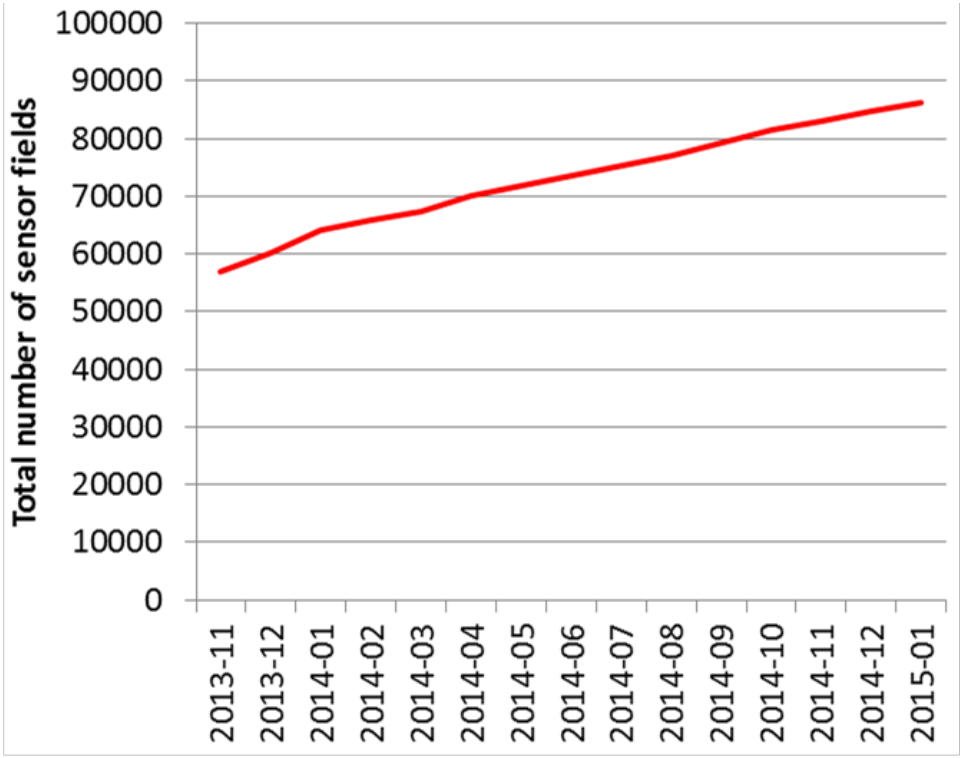
Date: february 2015
NB1: this are “sensor data categories” at a farm
NB2: not all animals are monitored for SDF (e.g. 3 and 4 only calves)

InfoBroker – Facts & Figures

Number of cows vs time



Number of sensorfields vs time



WHY LINKED DATA AND SEMANTICS?

1. To make the various data sets accessible in an automatically linkable manner for easier integration

2. To enrich the semantics of the datasets in isolation as well as in combination using ontologies

3. To enable any possible question to be queried on the datasets for better analysis

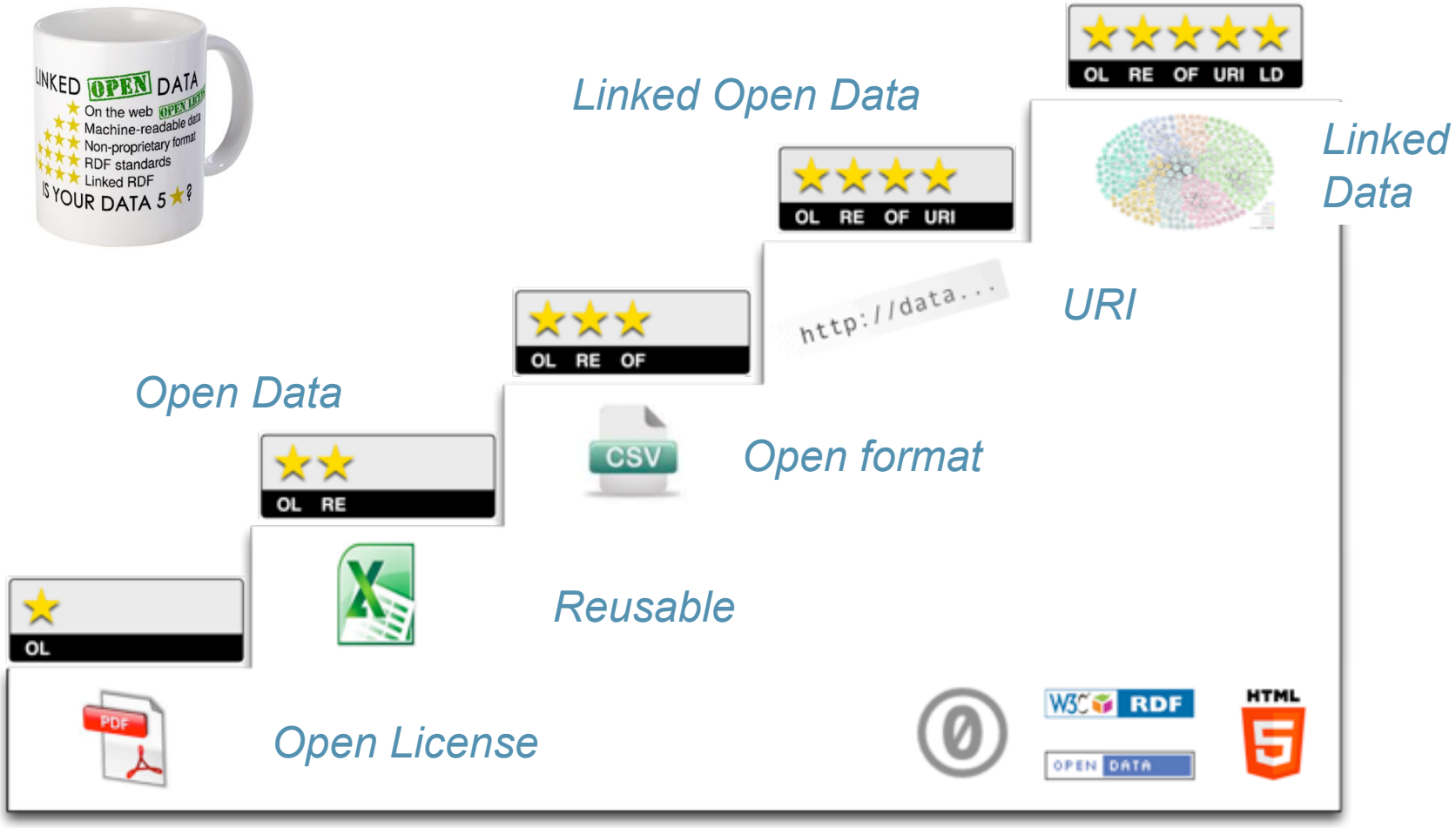
WHAT IS LINKED DATA?

- *“Linked data is a set of design principles for sharing machine-readable data on the Web for use by public administrations, business and citizens.”*

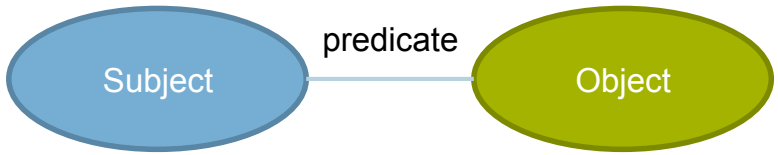
The **four design principles** of Linked Data *(by Tim Berners Lee)*:

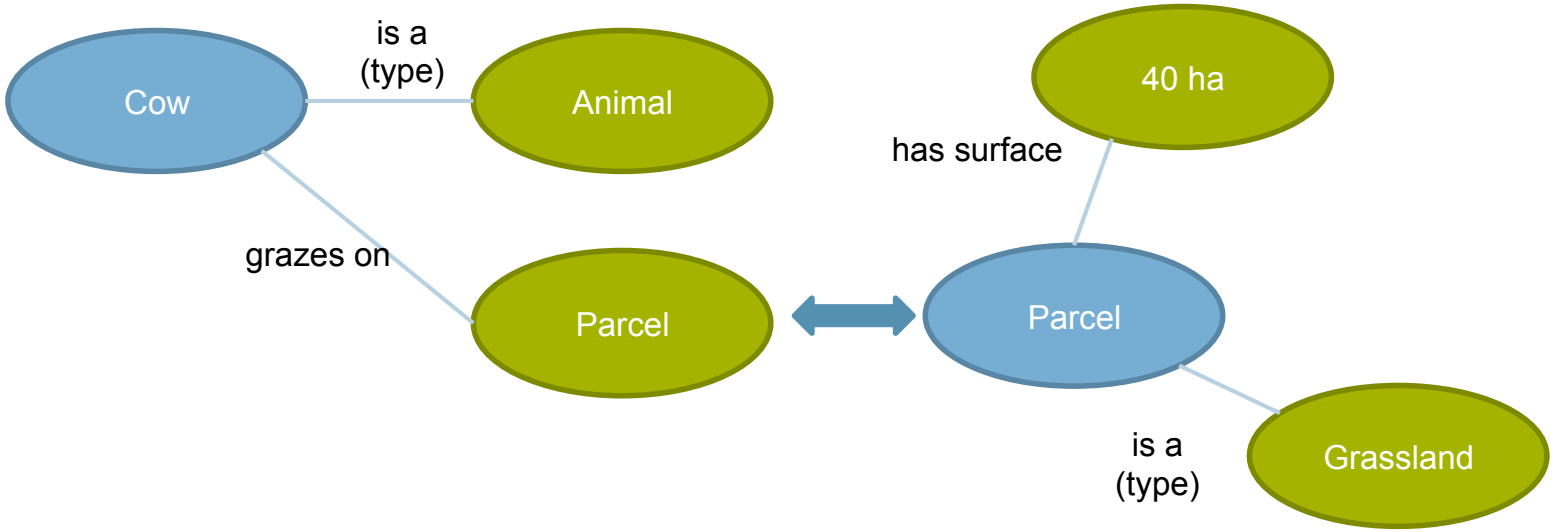
1. Use Uniform Resource Identifiers (URIs) as names for things.
2. Use HTTP URIs so that people can look up those names.
3. When someone looks up a URI, provide useful information, using the standards (RDF*, SPARQL).
4. Include links to other URIs so that they can discover more things.

5 STAR MODEL


LINKED OPEN DATA
On the web
Machine-readable data
Non-proprietary format
RDF standards
Linked RDF
IS YOUR DATA 5★?
Linked Open Data
OL RE OF URI LD
Linked Data
OL RE OF URI
http://data...
URI
OL RE OF
Open Data
CSV
Open format
OL RE
Reusable
OL
PDF
Open License
RDF
OPEN DATA
HTML

SIMPLE EXAMPLE OF LINKED DATA



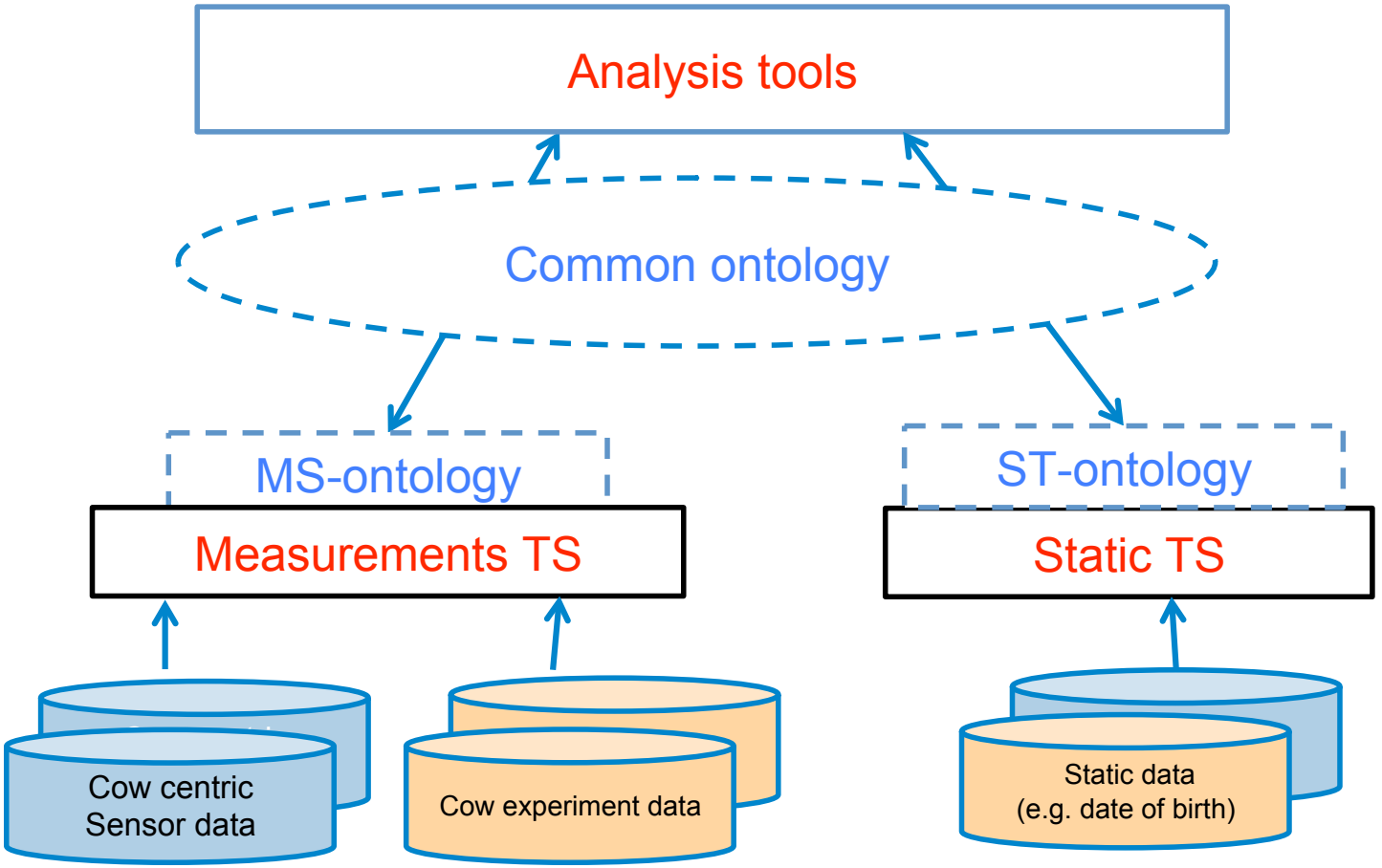


BIG DATA ANALYSIS QUESTIONS

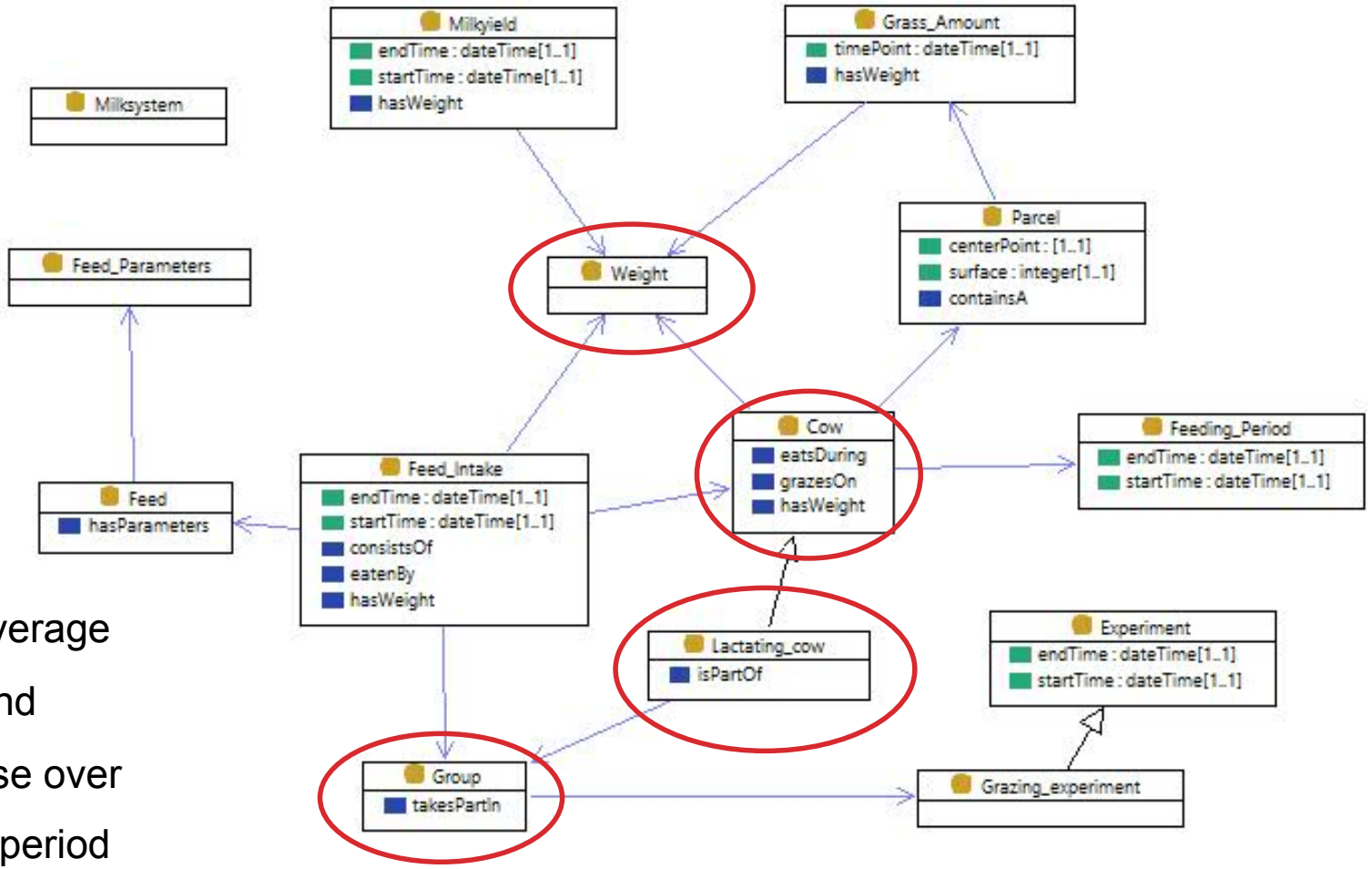
“How much feed did a group of cows at a dairy farm take in a certain feeding period at a specific parcel?”

“What was the average weight per day over the last lactation period of a cow and what was the weight in/decrease over that period?

STEP 1: ONTOLOGY MODELING



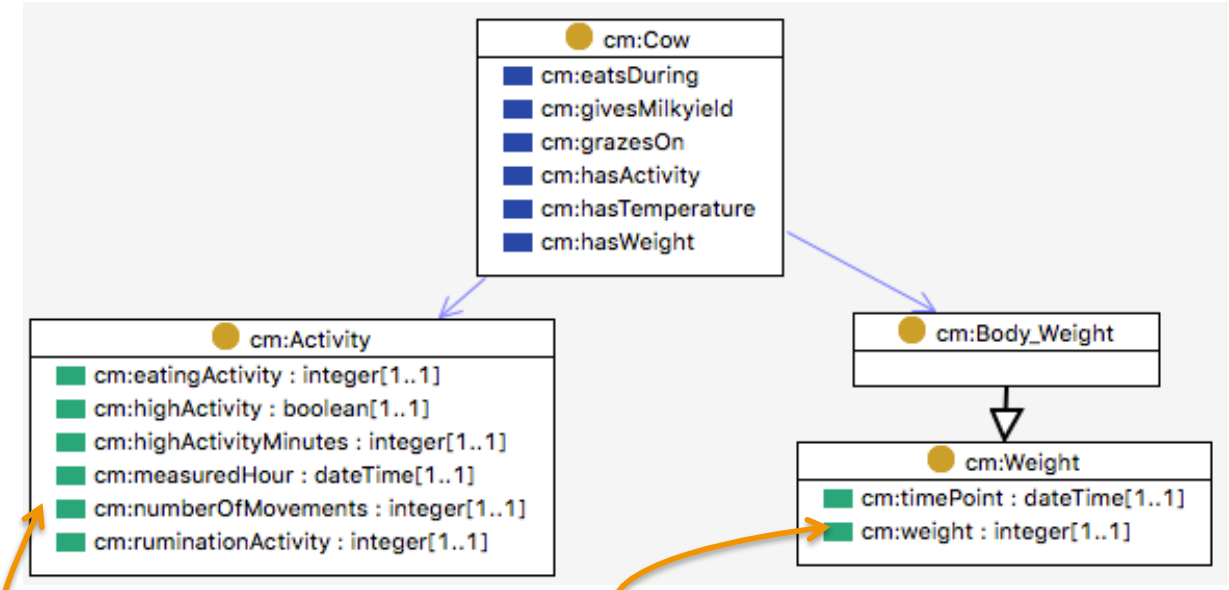
STEP 2: COMMON ONTOLOGY



"What was the average weight per day and weight in/decrease over the last lactation period of a cow in a group ?"

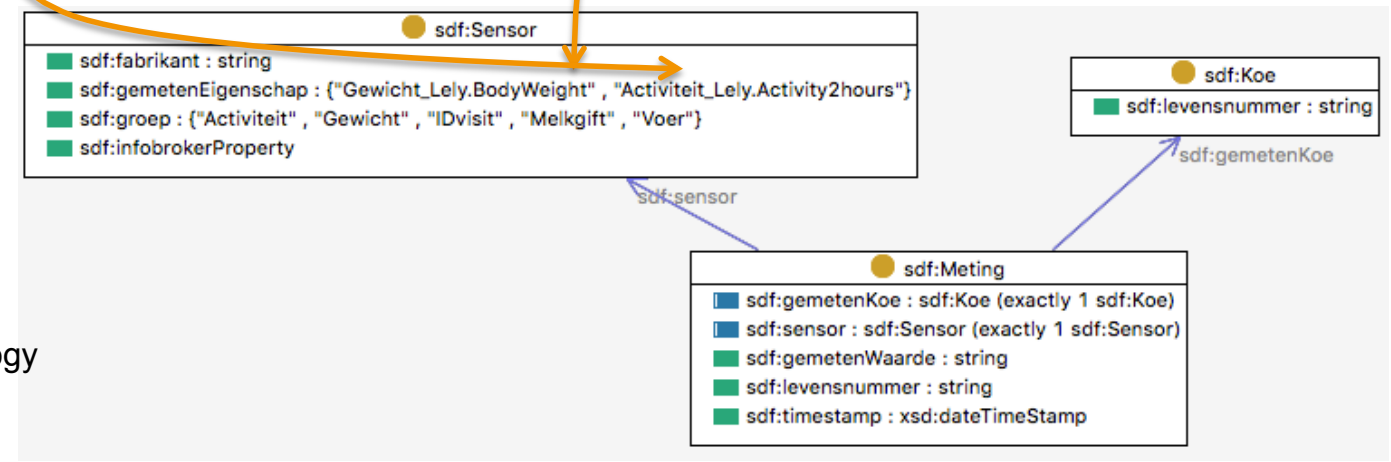
STEP 3: ONTOLOGY MAPPING

Common ontology



rdfs:label = “Activity2hours”

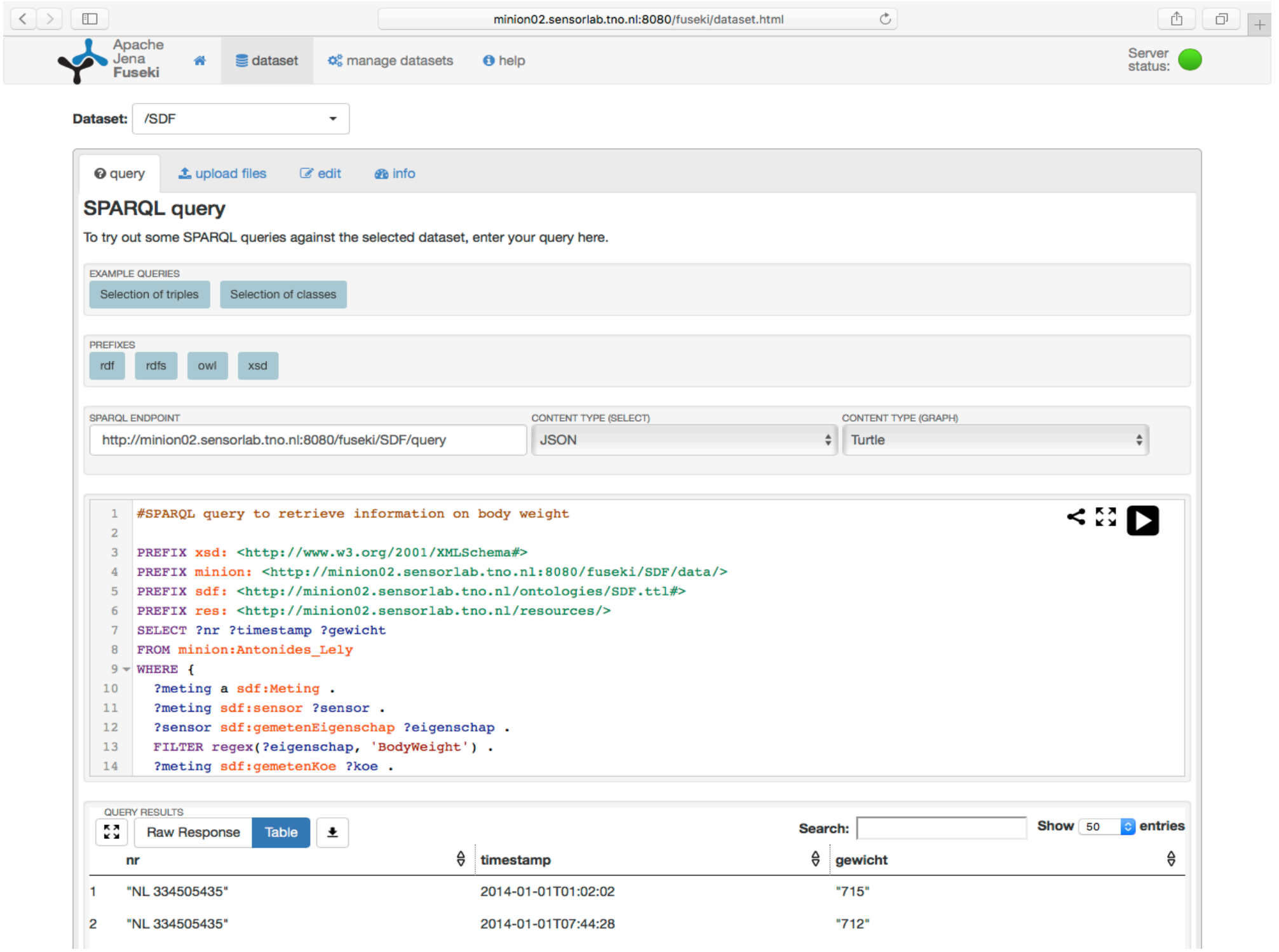
rdfs:label = “BodyWeight”



Measurement ontology

OUR BIG, LINKED DATA PLATFORM

- Server: 128GB memory, 5TB HDD
- Marmotta Triplestore with RDB
- Jena Fuseki Triplestore with TDBVirtuoso RDB +TDB
- OpenRefine with RDF extension


minion02.sensorlab.tno.nl:8080/fuseki/dataset.html
Apache Jena Fuseki
dataset
manage datasets
help
Server status:
Dataset: /SDF
query
upload files
edit
info
SPARQL query
To try out some SPARQL queries against the selected dataset, enter your query here.
EXAMPLE QUERIES
Selection of triples
Selection of classes
PREFIXES
rdf
rdfs
owl
xsd
SPARQL ENDPOINT
http://minion02.sensorlab.tno.nl:8080/fuseki/SDF/query
CONTENT TYPE (SELECT)
JSON
CONTENT TYPE (GRAPH)
Turtle
#SPARQL query to retrieve information on body weight
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX minion: <http://minion02.sensorlab.tno.nl:8080/fuseki/SDF/data/>
PREFIX sdf: <http://minion02.sensorlab.tno.nl/ontologies/SDF.ttl#>
PREFIX res: <http://minion02.sensorlab.tno.nl/resources/>
SELECT ?nr ?timestamp ?gewicht
FROM minion:Antonides_Lely
WHERE {
?meting a sdf:Meting .
?meting sdf:sensor ?sensor .
?sensor sdf:gemetenEigenschap ?eigenschap .
FILTER regex(?eigenschap, 'BodyWeight') .
?meting sdf:gemetenKoe ?koe .
QUERY RESULTS
Raw Response
Table
Search:
Show 50 entries
nr
timestamp
gewicht
1 "NL 334505435" 2014-01-01T01:02:02 "715"
2 "NL 334505435" 2014-01-01T07:44:28 "712"

```
PREFIX res: <http://minion02.sensorlab.tno.nl/resources/>
SELECT ?nr ?timestamp ?gewicht
FROM minion:Antonides_Lely
WHERE {
  ?meting a sdf:Meting .
  ?meting sdf:sensor ?sensor .
  ?sensor sdf:gemetenEigenschap ?eigenschap .
  FILTER regex(?eigenschap, 'BodyWeight') .
  ?meting sdf:gemetenKoe ?koe .
  ?meting sdf:timestamp ?timestamp .
  ?meting sdf:gemetenWaarde ?gewicht .
  ?koe sdf:levensnummer ?nr
}
LIMIT 100
```

QUERY RESULTS

Raw Response | Table

Search: Show 50 entries

	nr	timestamp	gewicht
1	"NL 334505435"	2014-01-01T01:02:02	"715"
2	"NL 334505435"	2014-01-01T07:44:28	"712"
3	"NL 334505435"	2014-01-01T14:34:21	"693"
4	"NL 334505435"	2014-01-01T19:58:32	"736"
5	"NL 334505435"	2014-01-02T04:14:39	"708"
6	"NL 334505435"	2014-01-02T09:59:15	"714"
7	"NL 334505435"	2014-01-02T15:46:39	"715"
8	"NL 334505435"	2014-01-03T00:38:08	"707"
9	"NL 334505435"	2014-01-03T07:12:51	"708"
10	"NL 334505435"	2014-01-03T12:33:23	"701"
11	"NL 334505435"	2014-01-03T18:38:00	"723"
12	"NL 334505435"	2014-01-04T02:20:28	"718"
13	"NL 334505435"	2014-01-04T08:12:15	"687"
14	"NL 334505435"	2014-01-04T13:47:00	"712"
15	"NL 334505435"	2014-01-04T19:44:03	"717"
16	"NL 334505435"	2014-01-05T02:16:01	"713"

CHALLENGE WITH CURRENT SET-UP:

- Large amount of sensor-data: Inefficient data-storage of RDF
  - 12GB of CSV data turned into +/- 310 GB of RDF (in fuseki TDB)
- Performance issues
  - Our relatively powerful setup is not able to answer more complex questions in acceptable time,
  - traditional datastorage (RDB) **is** able to answer these questions in acceptable time.

NEW APPROACH: COMBINE BEST OF BOTH

- **Openlink virtuoso:**
  - Has both a relational database and a triplestore
  - Efficiently stores RDB data
  - Ability to define a mapping between DB schema and ontology
    - Query Relational DB using SPARQL
  - Provides a SPARQL endpoint

NEW APPROACH: COMBINE BEST OF BOTH

- **Hypothesis:**
  - Because the data is stored more efficiently, and SQL can query this data more efficiently than a complete RDF+SPARQL implementeation; the performance of a hybrid solution will be faster than pure RDF storage.

- **Open questions:**
  - How efficient is the Virtuoso translation from SPARQL to SQL?
  - How will this setup perform with more complex ontologies or database schemas?
  - Are there any unforeseen limitations of this setup?