

# Adding semantics to tabular agrifood data

Jan Top

December 12, 2018



# Observational records in agrifood

	Macro-sectors' output (%)	Total multiplier	Regional effects	Interregional spillover	Interregional feedback	ROI/Bas*	Agri food business chain/network revealed								
<b>Agriculture</b>							<b>Institutional enabling environment</b>	<b>Policy development &amp; execution, research, project funding, donor organization, educational institutes...</b>							
Cereal grains	13.2%	2.814	0.866	0.944	0.00			<b>Service providers</b>		Quality control		R&D facilities		Private training	
Horticulture	0.0%	2.473	0.705	0.765	0.00					Finance, banks, insurance ICT		Document processing Certification		Consultancy Techn. maintenance	
Permanent crops	6.5%	3.060	1.048	1.008	0.00		<b>Chain actors</b>	Fertilizer	Crop	Meat	Commodity	Store	Local		
Livestock	67.1%	2.641	0.744	0.893	0.00			Crop protection	Animals	Dairy prod.	Fresh	Kiosk	National		
Mixed	13.2%	3.068	0.963	1.101	0.00			Seeds	Fish	Flour	Processed	Market	Regional		
<b>Food industry</b>								Breeds	Milk	Juice	Cold	Hotel	Global		
Meat	6.0%	3.268	1.048	1.214	0.00			Equipment	Silk	Food prod.	Dry	Restaurant			
Fish	0.7%	2.774	0.864	0.907	0.00		<b>Input supplier</b>	<b>Producer cooperative</b>	Semi processed	<b>Trader</b>	<b>Distributor/retailer</b>	<b>Consumer</b>			
Olive oil	6.0%	3.438	1.207	1.226	0.00										
Vegetable oils, sugar, pasta	41.6%	3.154	0.967	1.182	0.00										
Vegetables and fruits	5.9%	3.268	1.066	1.197	0.00	112.29%									
Dairy products	17.4%	3.277	0.987	1.284	0.006	130.09%									
Cereals	3.2%	3.331	1.115	1.210	0.006	108.52%									
Animal feed	1.1%														
Wine	5.7%														
Water and other beverage	12.5%														
<b>Section/division/group</b>							<b>2004</b>	<b>2005</b>	<b>2006</b>	<b>2007</b>	<b>2008</b>				
<b>Food and live animals</b>							<b>3.13</b>	<b>3.06</b>	<b>3.36</b>	<b>4.53</b>	<b>3.78</b>				
Live animals							1.90	0.68	0.79	4.50	14.72				
Meat and meat preparations							4.03	5.04	9.24	10.75	4.95				
Dairy products and birds' eggs							1.44	2.20	7.48	5.93	5.31				
Fish (not marine mammals), crustaceans, molluscs and aquatic invertebrates, and preparations thereof							0.03	0.10	0.10	0.13	0.22				
Cereals and cereal preparations							4.23	10.31	11.63	15.18	11.69				

# Characteristics

- Repetitive over time, instances, conditions
- Possibly nested, irregular, incomplete
- Created for a particular purpose
- Partial and unclear annotations
- Closed boxes
- Hard to find and combine



*Here are my data, be my guest!*

01001	BUTTER,WITH SALT	15.87	717	0.85	81.11	2.11	0.06	0	0.06
01002	BUTTER,WHIPPED,WITH SALT	15.87	717	0.85	81.11	2.11	0.06	0	0.06
01003	BUTTER OIL,ANHYDROUS	0.24	876	0.28	99.48	0	0	0	0
01004	CHEESE,BLUE	42.41	353	21.4	28.74	5.11	2.34	0	0.5
01005	CHEESE,BRICK	41.11	371	23.24	29.68	3.18	2.79	0	0.51
01006	CHEESE,BRIE	48.42	334	20.75	27.68	2.7	0.45	0	0.45
01007	CHEESE,CAMEMBERT	51.8	300	19.8	24.26	3.68	0.46	0	0.46
01008	CHEESE,CARAWAY	39.28	376	25.18	29.2	3.28	3.06	0	
01009	CHEESE,CHEDDAR	37.1	406	24.04	33.82	3.71	1.33	0	0.28

# First layer of explanation

	A	B	C	D	E	F	G	H	I	J
1	NDB_No	Shrt_Desc	Water_(g)	Energy_Kcal	Protein_(g)	Lipid_Tot_(g)	Ash_(g)	Carbohydrt_(g)	Fiber_TD_(g)	Sugar_Tot_(g)
2	01001	BUTTER,WITH SALT	15.87	717	0.85	81.11	2.11	0.06	0	0.06
3	01002	BUTTER,WHIPPED,WITH SALT	15.87	717	0.85	81.11	2.11	0.06	0	0.06
4	01003	BUTTER OIL,ANHYDROUS	0.24	876	0.28	99.48	0	0	0	0
5	01004	CHEESE,BLUE	42.41	353	21.4	28.74	5.11	2.34	0	0.5
6	01005	CHEESE,BRICK	41.11	371	23.24	29.68	3.18	2.79	0	0.51
7	01006	CHEESE,BRIE	48.42	334	20.75	27.68	2.7	0.45	0	0.45
8	01007	CHEESE,CAMEMBERT	51.8	300	19.8	24.26	3.68	0.46	0	0.46
9	01008	CHEESE,CARAWAY	39.28	376	25.18	29.2	3.28	3.06	0	
10	01009	CHEESE,CHEDDAR	37.1	406	24.04	33.82	3.71	1.33	0	0.28

# Extracting data from tables



## RDF Spreadsheet Editor: Get (G)rid of Your RDF Data Entry Problems

Generating RDF from Tabular Data on the Web



## OpenRefine

OpenRefine is a free, open source power tool for working with messy data and improving it

Mapping Relational Data to RDF  
with Virtuoso's RDF Views

COMPUTING AND INFORMATICS, VOL 36, NO 6 (2017)

HTAB2RDF: MAPPING HTML TABLES TO RDF TRIPLES

*Djelloul Bouchiha, Mimoun Malki, Abdullah Alghamdi, Khalid Alnajjar*

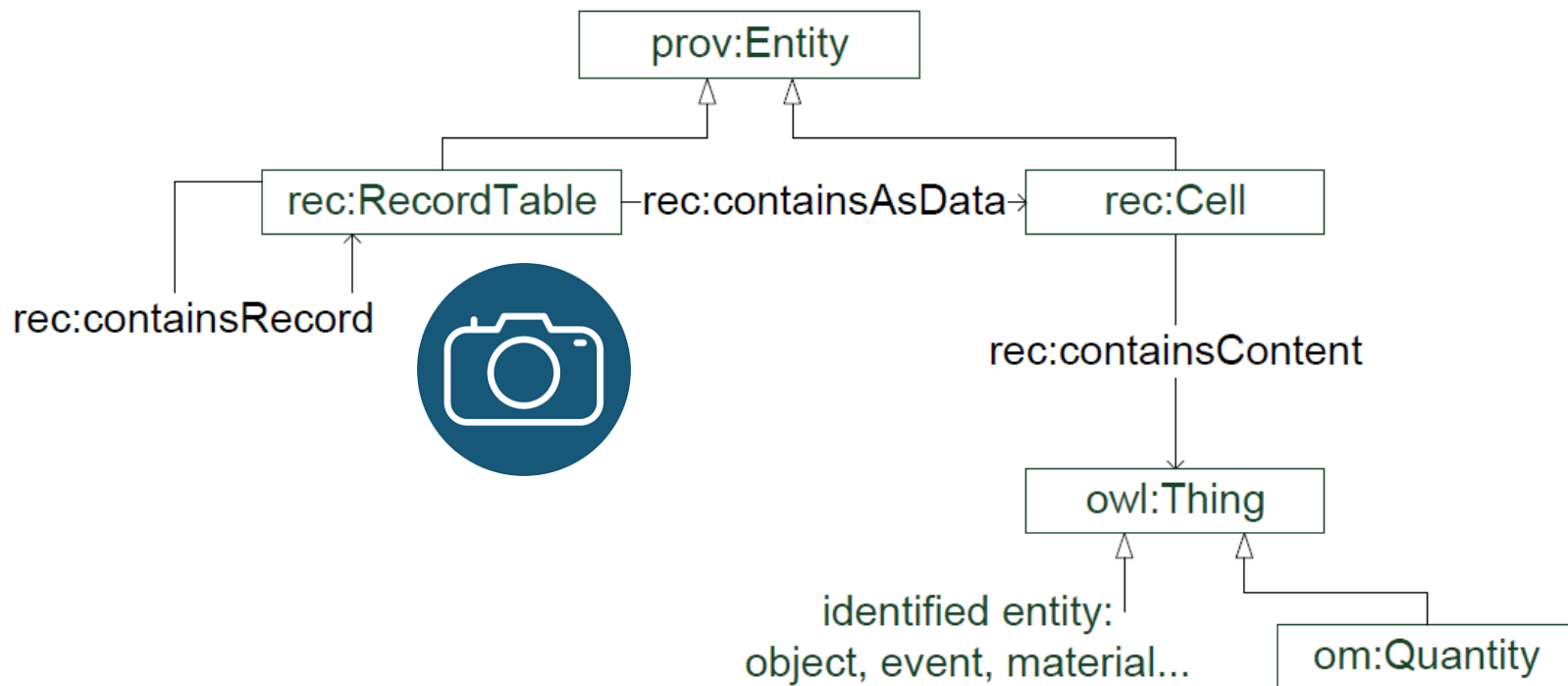
# Typical approach

- Table seen as object, rows are properties
- Require regular and similar structures
- Table headers are mapped to high-level vocabularies
- No distinction between 'phenomena' and 'quantities'

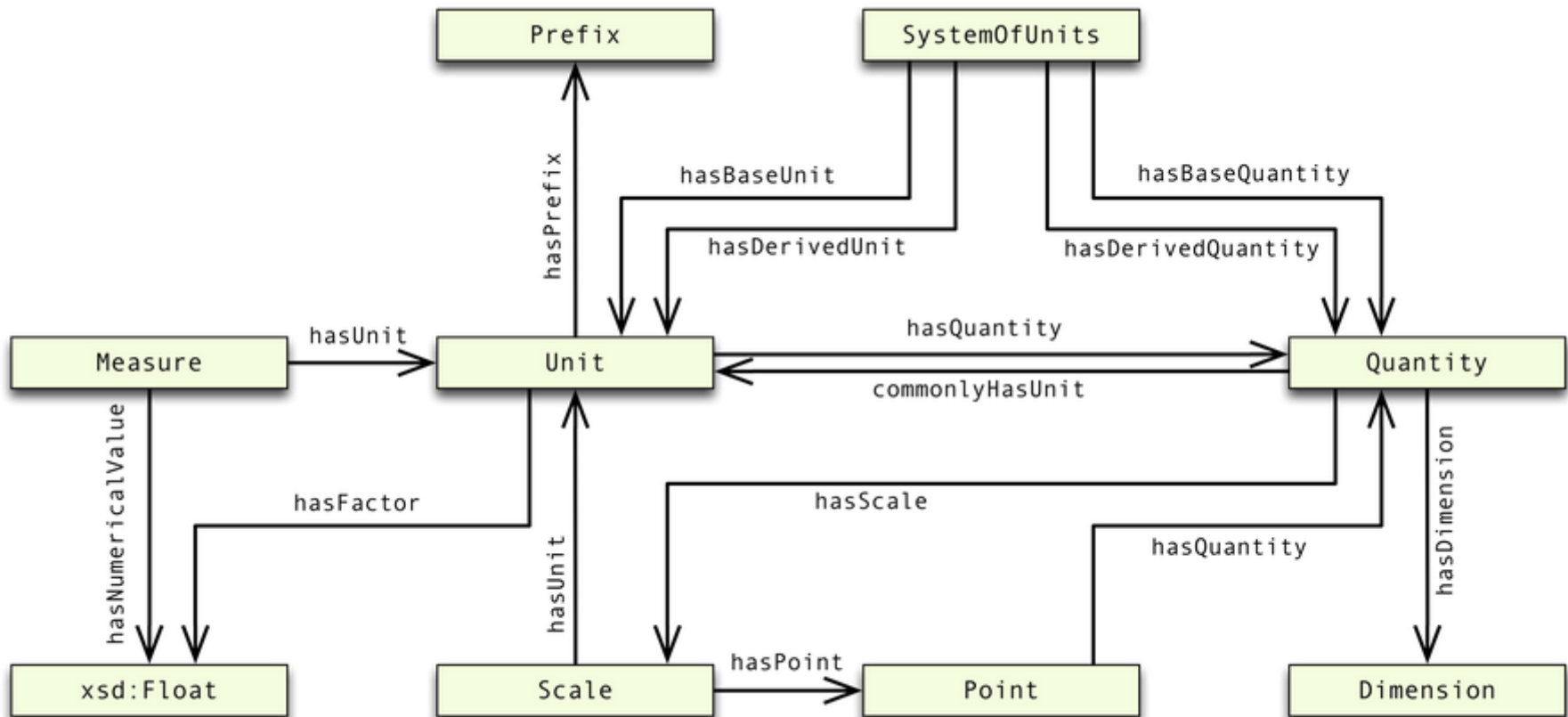
	A	B	C	D	E	F	G
1	NDB_No	Shrt_Desc	Water_(g)	Energ_Kcal	Protein_(g)	Lipid_Tot_(g)	Ash_(g)
2	01001	BUTTER,WITH SALT	15.87	717	0.85	81.11	2.11
3	01002	BUTTER,WHIPPED,WITH SALT	15.87	717	0.85	81.11	2.11
4	01003	BUTTER OIL,ANHYDROUS	0.24	876	0.28	99.48	0
5	01004	CHEESE,BLUE	42.41	353	21.4	28.74	5.11
6	01005	CHEESE,BRICK	41.11	371	23.24	29.68	3.18

# RDF Record Table

01001	BUTTER,WITH SALT	15.87	717	0.85	81.11
01002	BUTTER,WHIPPED,WITH SALT	15.87	717	0.85	81.11
01003	BUTTER OIL,ANHYDROUS	0.24	876	0.28	99.48
01004	CHEESE,BLUE	42.41	353	21.4	28.74



# Ontology of Units of Measure



# Using a shared vocabulary

<http://www.foodvoc.org/resource/FoodTaxonomy/Blue>

<http://www.ontology-of-units-of-measure.org/resource/om-2/Mass>

	A	B	C	D	E	F	G	H	I	J
1	NDB_No	Shrt_Desc	Water_(g)	Energy_Kcal	Protein_(g)	Lipid_Tot_(g)	Ash_(g)	Carbohydrt_(g)	Fiber_TD_(g)	Sugar_Tot_(g)
2	01001	BUTTER,WITH SALT	15.87	717	0.85	81.11	2.11	0.06	0	0.06
3	01002	BUTTER,WHIPPED,WITH SALT	15.87	717	0.85	81.11	2.11	0.06	0	0.06
4	01003	BUTTER OIL,ANHYDROUS	0.24	876	0.28	99.48	0	0	0	0
5	01004	CHEESE,BLUE	42.41	353	21.4	28.74	5.11	2.34	0	0.5
6	01005	CHEESE,BRICK	41.11	371	23.24	29.68	3.18	2.79	0	0.51
7	01006	CHEESE,BRIE	48.42	334	20.75	27.68	2.7	0.45	0	0.45
8	01007	CHEESE,CAMEMBERT	51.8	300	19.8	24.26	3.68	0.46	0	0.46
9	01008	CHEESE,CARAWAY	39.28	376	25.18	29.2	3.28	3.06	0	
10	01009	CHEESE,CHEDDAR	37.1	406	24.04	33.82	3.71	1.33	0	0.28

## Type

Concept

## Literals

Property	Value	Language
preferred label	Blue	en
comment	It is a white cheese with blue veins and sometimes crumbly interior. This cheese usually has tangy, piquant, spicy and peppery flavor. Use in salad dressings with cream cheese for spreads.	en
preferred label	Blue	nl
	Hard	en

/om-2/Mass

Dow

			Literals		Property	Value						Language
					comment	Mass is the amount of matter of a phenomenon. It is a base quantity in the International System of Units. Mass is force divided by acceleration.						en
	A	B			label	mass						en
1	NDB_No	Shrt_Desc			label	massa						nl
2	01001	BUTTER,WITH SALT			label	质量						zh
3	01002	BUTTER,WHIPPED,WITH SALT			symbol	m						en
4	01003	BUTTER OIL,ANHYDROUS			unofficial label	weight						en
5	01004	CHEESE,BLUE										
6	01005	CHEESE,BRICK										
7	01006	CHEESE,BRIE	48.42	334	20.75	27.68	2.7	0.45	0	0.45		
8	01007	CHEESE,CAMEMBERT	51.8	300	19.8	24.26	3.68	0.46	0	0.46		
9	01008	CHEESE,CARAWAY	39.28	376	25.18	29.2	3.28	3.06	0			
10	01009	CHEESE,CHEDDAR	37.1	406	24.04	33.82	3.71	1.33	0	0.28		

# Data, metadata and semantics

<http://www.foodvoc.org/resource/FoodTaxonomy/Blue>

Product	Leverancier	prijs/kg
Brie	AH	14.59
Gouda	Jumbo	12.29
Leerdam	Jumbo	13.33
Maasland	Spar	13.27
Danish Blue	AH	13.89

	A	B	C	D	E	F	G	H	I	J
1	NDB_No	Shrt_Desc	Water_(g)	Energ_Kcal	Protein_(g)	Lipid_Tot_(g)	Ash_(g)	Carbohydrt_(g)	Fiber_TD_(g)	Sugar_Tot_(g)
2	01001	BUTTER,WITH SALT	15.87	717	0.85	81.11	2.11	0.06	0	0.06
3	01002	BUTTER,WHIPPED,WITH SALT	15.87	717	0.85	81.11	2.11	0.06	0	0.06
4	01003	BUTTER OIL,ANHYDROUS	0.24	876	0.28	99.48	0	0	0	0
5	01004	CHEESE,BLUE	42.41	353	21.4	28.74	5.11	2.34	0	0.5
6	01005	CHEESE,BRICK	41.11	371	23.24	29.68	3.18	2.79	0	0.51
7	01006	CHEESE,BRIE	48.42	334	20.75	27.68	2.7	0.45	0	0.45
8	01007	CHEESE,CAMEMBERT	51.8	300	19.8	24.26	3.68	0.46	0	0.46
9	01008	CHEESE,CARAWAY	39.28	376	25.18	29.2	3.28	3.06	0	
10	01009	CHEESE,CHEDDAR	37.1	406	24.04	33.82	3.71	1.33	0	0.28

Get(File\_X\_sheet\_1\_B5, File\_X\_sheet\_1\_C5, File\_Y\_sheet\_1\_C10)



Get(blue cheese, water\_content, price)

# Rosanne

- Excel implementation
- Annotate
- Convert
- Find
- Merge

Proteine CE 2014-6.xlsx [Compatibility Mode] - Microsoft Excel

File Home Insert Page Layout Formulas Data Review View Develop PowerPivot Semantics Tiffany ?

Annotation RDF Export Unit Conversion Integration About

Semantics Actions About

OUM\_annotatedWit... fx

	A	B	C	D	E	F	G
	Protein	Time time (s)	torque (mN m)	temperatur e (°C)	shear stress (Pa)	Frequenc y frequenc y (Hz)	Storage Modulus storage modulus (Pa)
1	Sample1-1	22.5	4.12E-06	60.1	3.18E-04	1	1.59E-04
2	Sample1-1	45	3.87E-06	59.9	2.99E-04	1	1.49E-04
3	Sample1-1	67.4	3.70E-06	59.6	2.86E-04	1	1.43E-04
4	Sample1-1	89.9	3.82E-06	59.4	2.95E-04	1	1.47E-04
5	Sample1-1	112	3.24E-06	59.2	2.50E-04	1	1.25E-04
6	Sample1-1	135	3.57E-06	59	2.76E-04	1	1.38E-04
7	Sample1-1	157	4.18E-06	58.8	3.23E-04	1	1.61E-04
8	Sample1-1	180	3.87E-06	58.6	3.00E-04	1	1.50E-04
9	Sample1-1	202	3.67E-06	58.5	2.84E-04	1	1.42E-04
10	Sample1-1	225	3.85E-06	58.3	2.97E-04	1	1.48E-04
11	Sample1-1	247	3.49E-06	58.1	2.70E-04	1	1.35E-04
12	Sample1-1	270	3.99E-06	57.9	3.09E-04	1	1.54E-04
13	Sample1-1	292	3.89E-06	57.7	3.01E-04	1	1.50E-04
14	Sample1-1	315	3.65E-06	57.5	2.82E-04	1	1.41E-04
15	Sample1-1	337	3.70E-06	57.3	2.86E-04	1	1.43E-04
16	Sample1-1	360	3.90E-06	57.1	3.02E-04	1	1.51E-04
17	Sample1-1	382	2.68E-06	56.9	2.07E-04	1	1.04E-04
18	Sample1-1	405	3.87E-06	56.7	3.00E-04	1	1.50E-04
19	Sample1-1	427	3.94E-06	56.6	3.04E-04	1	1.52E-04
20	Sample1-1	449	3.66E-06	56.4	2.83E-04	1	1.41E-04
21	Sample1-1	472	3.63E-06	56.2	2.81E-04	1	1.40E-04
22	Sample1-1						

Expected IntermediateRawD...

Ready

Annotation

Table: Define Clear

Annotate: ☒ Quantities and units ☐ Other

Header:

Quantity: torque

Unit: millinewton metre

Apply Clear

Options

COMMIT/

WAGENINGEN UR

For quality of life

70%

# Dealing with legacy

		Emissions	Costs
		Mtonne CO2-eq	Meuro
Crop	Corn	28.0	50.0
	Soybean	70.0	75.0
	Canola	38.0	120.0
	Total	136.0	245.0

# Blocks

		context block	
		Emissions	Costs
		Mtonne CO2-eq	Meuro
Crop	Corn	28.0	50.0
	Soybean	70.0	75.0
	Canola	38.0	120.0
	Total	136.0	245.0
context block		table body	

# Concepts

		Emissions	Costs	om:Quantity
		Mtonne CO2-eq	Meuro	om:Unit
Crop	Corn	28.0	50.0	
	Soybean	70.0	75.0	
	Canola	38.0	120.0	
	Total	136.0	245.0	

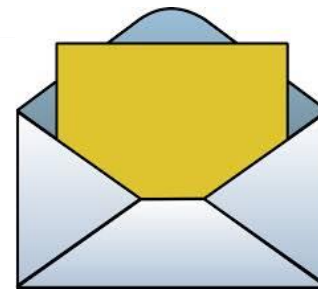
om:Phenomenon

om:Measure

# Interpretation

		om:'Mass'	
val:Emission		<u>Emissions</u>	Costs
		<u>Mtonne CO2-eq</u>	Meuro
Crop	Corn	28.0	50.0
val:Soybean	<u>Soybean</u>	70.0	75.0
	Canola	38.0	120.0
Total		136.0	245.0

# Message



- Improve data sharing: move from **separate** to **linked** tabular data
- Distinguish between
  - **phenomena identification**
  - **quantity value**
- Use relatively **specific** ontologies, including OM for units and quantities
- Embed in **popular tools**, for example Rosanne in Excel
- Legacy data can be **annotated automatically** to some extent

# Thank you

[jan.top@wur.nl](mailto:jan.top@wur.nl)

