

FAIRifying research data as regular business

Jan Top

Wageningen University & Research

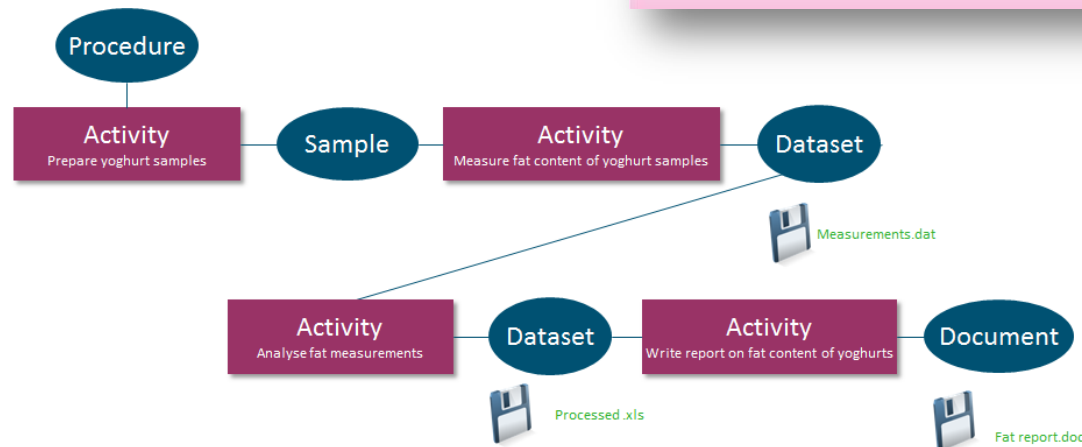
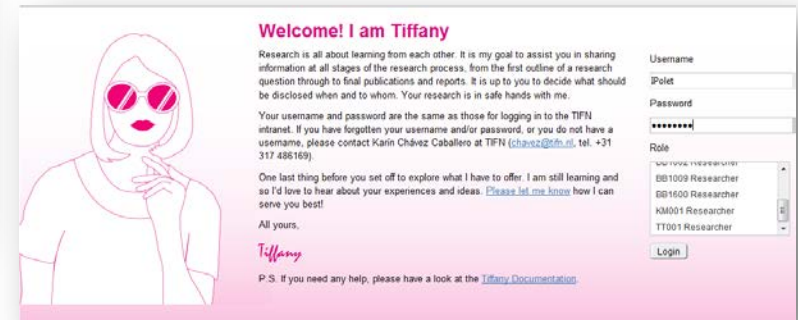
Vrije Universiteit Amsterdam



FOOD & BIOBASED RESEARCH
WAGENINGEN UR



Research data management in food research



FAIR principles

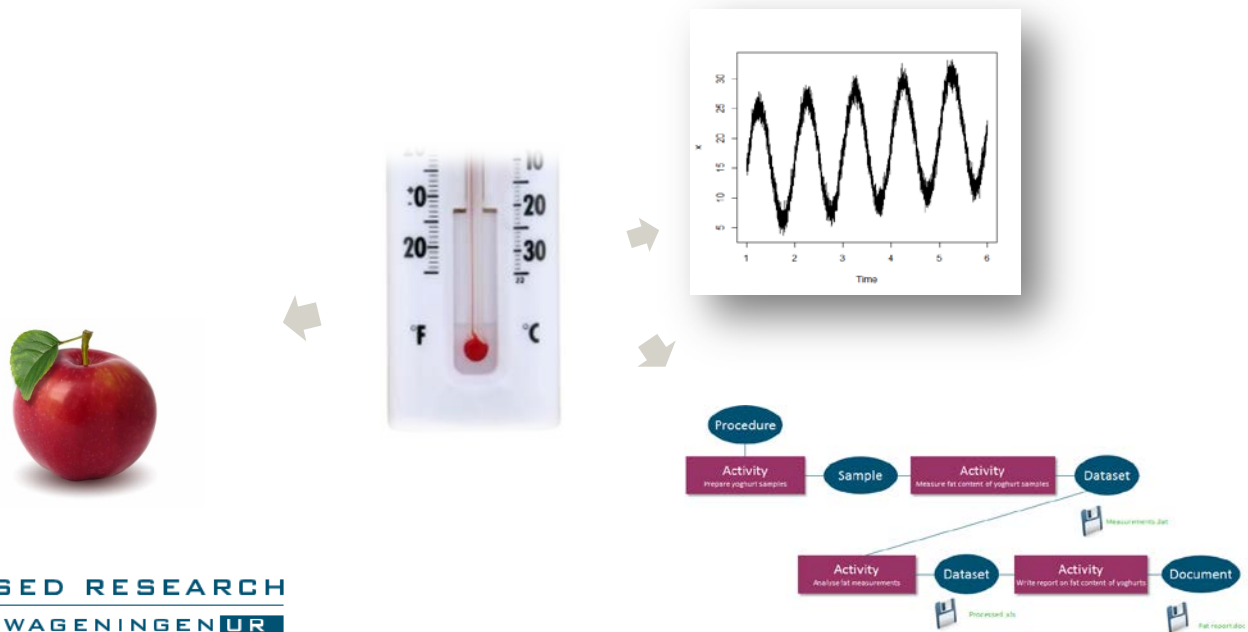
- Findable
- Accessible
- Interoperable
- Reusable



Metadata in an open world



- explanation and qualification
- provenance and method



Metadata requires meta-effort

Get closer to the individual researcher, making the task easier



Possible directions

- Generic, shared ontologies and task specific ontologies.
- Annotation embedded in existing tools
- Automated annotation
- Allow for flexible annotation – but limit data footprint

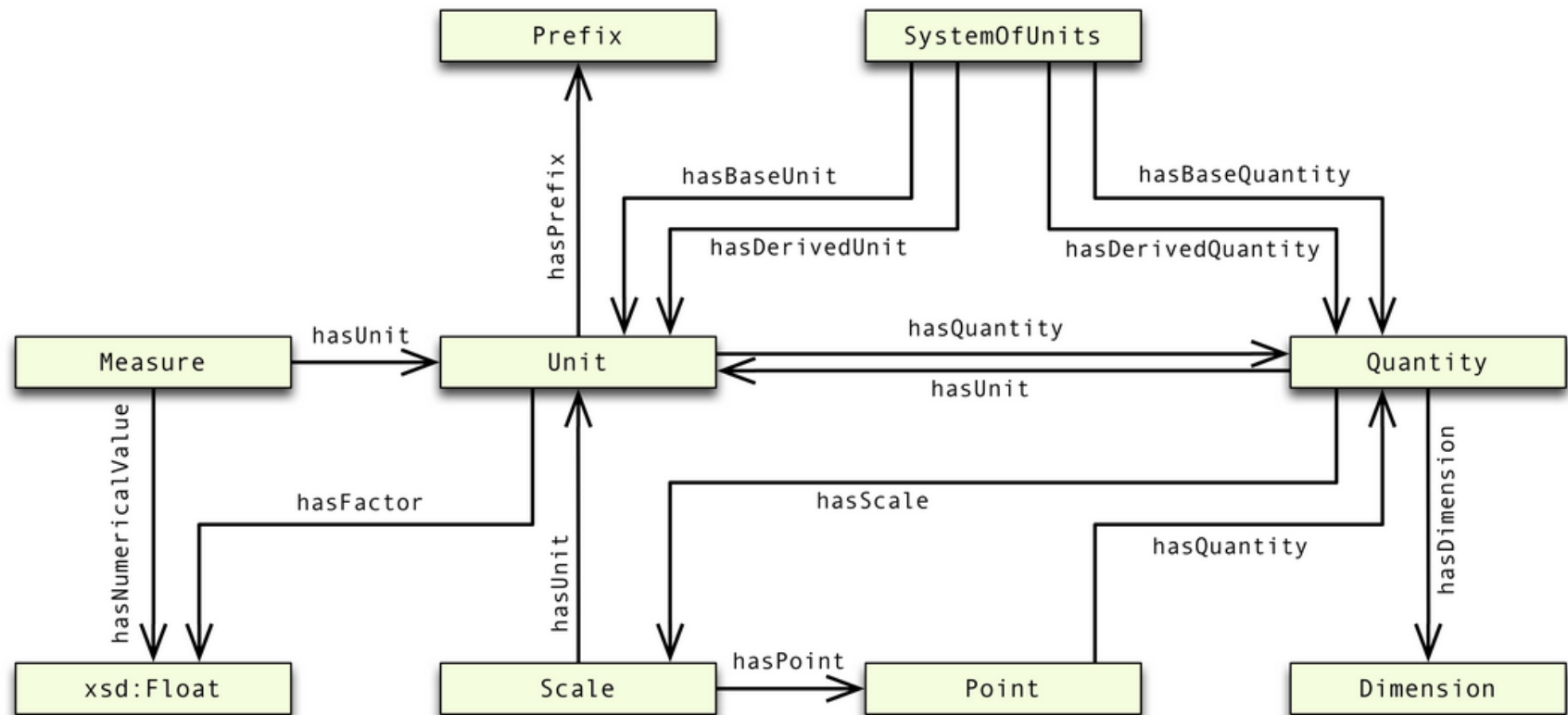


Expressive but general ontologies

- Agrovoc, Gene Ontology, see for example AgroPortal project
- Basic annotation of numerical data: units of measure and quantities



OM – Ontology of units of Measure



Use Cases and Suitability Metrics for Unit Ontologies. Markus D. Steinberg, Sirko Schindler and Jan Martin Keil.

Table 2: The presence of features within the examined ontologies.

(●...feature modeled; ○...feature not modeled)

	MUO	OBOE	OM	QU	QUDT	SWEET	UO
unit of measurement (unit)	●	●	●	●	●	●	●
kind of quantity (qk)	●	●	●	●	●	●	●
field of application (app)	○	○	●	●	○ ¹⁰	●	○
dimension (dim)	○	○	●	○	●	○	○
dimension vector (vector)	○	○	●	○	●	○	○
system of units (system)	○	○	●	○	●	○	○
phenomenon (phen)	○	●	●	○	○	○	○
measurement (meas)	●	●	●	○	○	○	○
conversion (conv)	○	●	●	●	●	●	○
prefix (prefix)	●	○	●	●	●	●	●
unit ↔ system	○	○	●	○	●	○	○
unit ↔ qk	●	●	●	●	●	●	●
unit ↔ dim	○	○	●	○	●	○	○
unit ↔ vector	○	○	●	○	●	○	○
unit ↔ prefix	○	○	●	●	○	●	○
unit ↔ app	○	○	●	○	○	○	○
qk ↔ app	○	○	●	●	○	●	○
meas ↔ phen	○	●	●	○	○	○	○
meas ↔ qk	○	●	●	○	○	○	○
meas ↔ unit	●	●	●	○	○	○	○
meas ↔ value	●	●	●	○	○	○	○
symbols for units	●	○	●	●	●	●	○ ¹⁰
symbols for qks	○	○	●	○	●	○	○
typ. values per units and apps	○	○	○	○	○	○	○
typ. values per qks and apps	○	○	○	○	○	○	○
allowed values per units	○	○	○	○	○	○	○
allowed values per qks	○	○	○	○	○	○	○
precision of conversion	○	○	○	○	○	○	○
number of diff. lang. tags ¹¹	1	1 ¹²	3	0	0	0 ¹²	0
unit composition	○	○	●	○	○	○	○
quantity composition	○	○	○	○	○	○	○
everyday lang. designators	○	○	●	○	○	○	○
resolvable URIs	○ ¹³	○ ¹³	●	○	○	○ ¹³	●



Table 2: The presence of features within the examined ontologies.

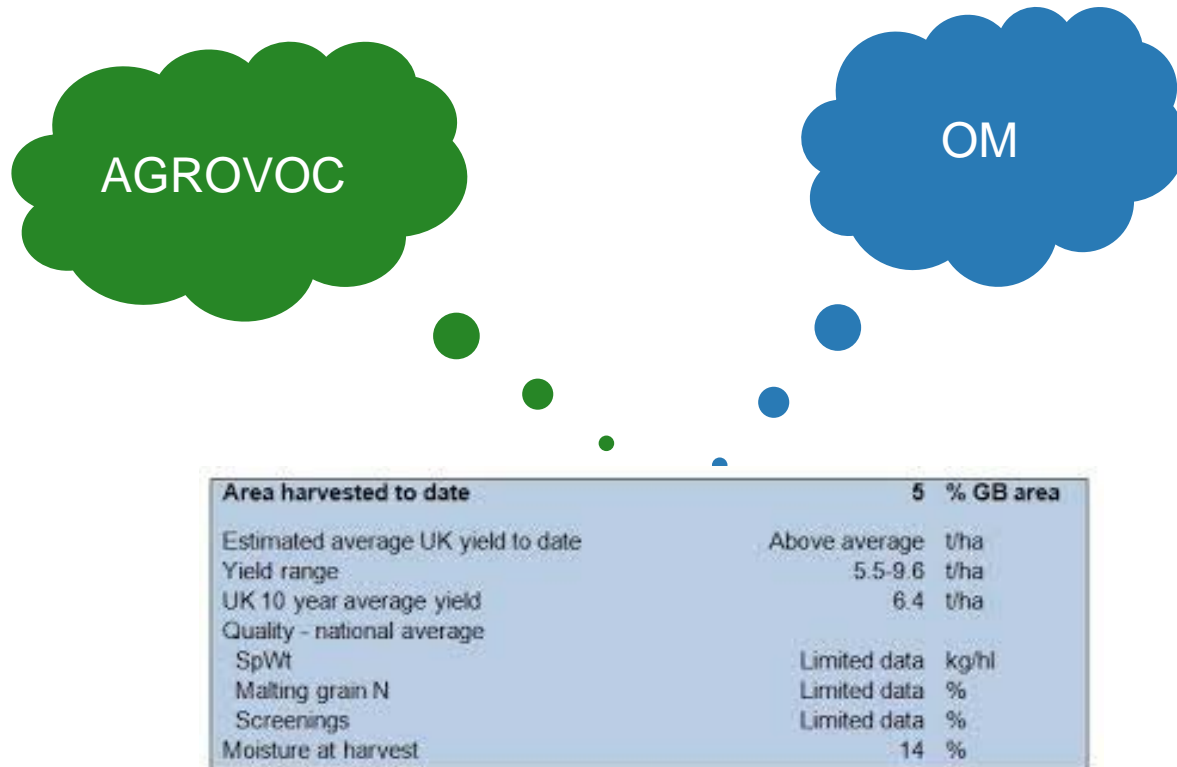
(●...feature modeled; ○...feature not modeled)

	MUO	OBOE	OM	QU	QUDT	SWEET	UO
unit of measurement (unit)	●	●	●	●	●	●	●
kind of quantity (qk)	●	●	●	●	●	●	●
field of application (app)	○	○	●	●	○ ¹⁰	●	○
dimension (dim)	○	○	●	○	●	○	○
dimension vector (vector)	○	○	●	○	●	○	○
system of units (system)	○	○	●	○	●	○	○
phenomenon (phen)	○	●	●	○	○	○	○
measurement (meas)	●	●	●	○	○	○	○
conversion (conv)	○	●	●	●	●	●	○
prefix (prefix)	●	○	●	●	●	●	●
unit ↔ system	○	○	●	○	●	○	○
unit ↔ qk	●	●	●	●	●	●	●
unit ↔ dim	○	○	●	○	●	○	○
unit ↔ vector	○	○	●	○	●	○	○
unit ↔ prefix	○	○	●	●	○	●	○
unit ↔ app	○	○	●	○	○	○	○
qk ↔ app	○	○	●	●	○	●	○
meas ↔ phen	○	●	●	○	○	○	○
meas ↔ qk	○	●	●	○	○	○	○
meas ↔ unit	●	●	●	○	○	○	○
number of dim. lang. tags	1	1 ¹³	5	0	0	0 ¹³	0
unit composition	○	○	●	○	○	○	○
quantity composition	○	○	○	○	○	○	○
everyday lang. designators	○	○	●	○	○	○	○
resolvable URIs	○ ¹³	○ ¹³	●	○	○	○ ¹³	●

Challenge in OM: fill in application areas



Specific tasks require specific ontologies



Creating task-specific ontologies



Firefox ROC+ www.fbresearch.nl/rocplus/RelateConcepts.aspx

ROC+

COMMIT/

Select project Add terms Identify synonyms Get suggestions Create taxonomy Indicate relations

Select Project Add Terms Identify Synonyms Get Suggestions Create Taxonomy Indicate Relations Visualisation Export Ontology

Create Taxonomy

Drag and drop terms into the tree

☐ Hide concepts with 'RelatedTo' relations

Search term Search Reset

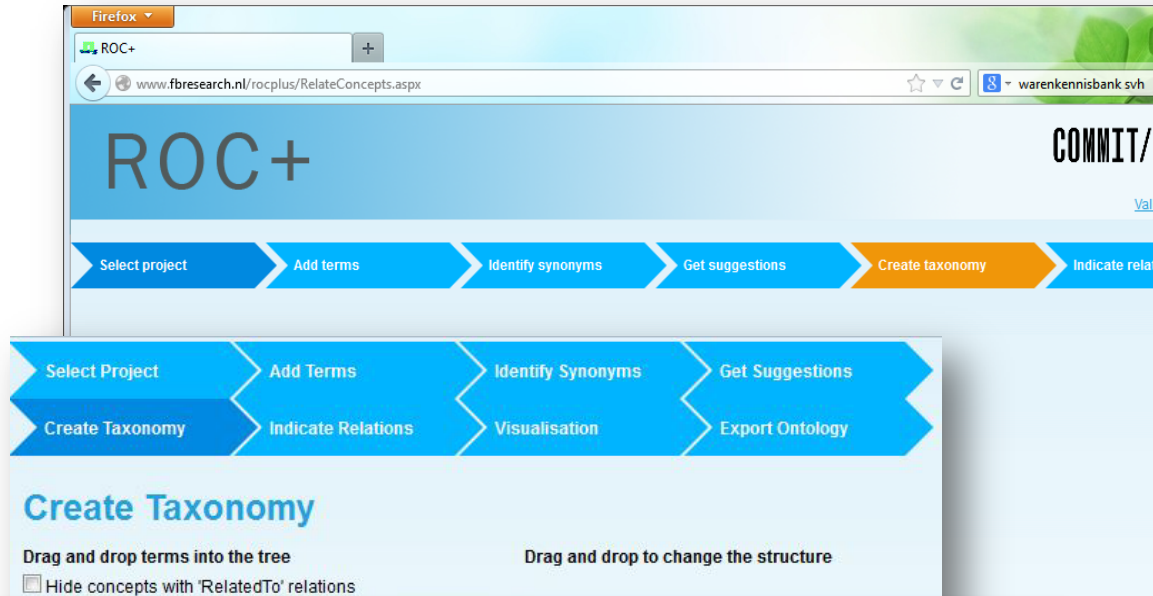
- microbial activity
- microbial activity
- microbial population
- Microbial process
- microbial properties
- microfauna
- microflora
- milking machine
- millet
- mineral

Drag and drop to change the structure

- biennial
- broad-leaved weed
- grass weed
- parasitic weed
- perennial
- perennial grass weed
- perennial weed
- pest weed
- black grass
 - non resistant black grass
 - resistant black grass



Creating task-specific ontologies



Challenge in ROC+: inherit partial structures from general ontologies



Not yet another software solution please

Rosanne: extending Excel with semantics

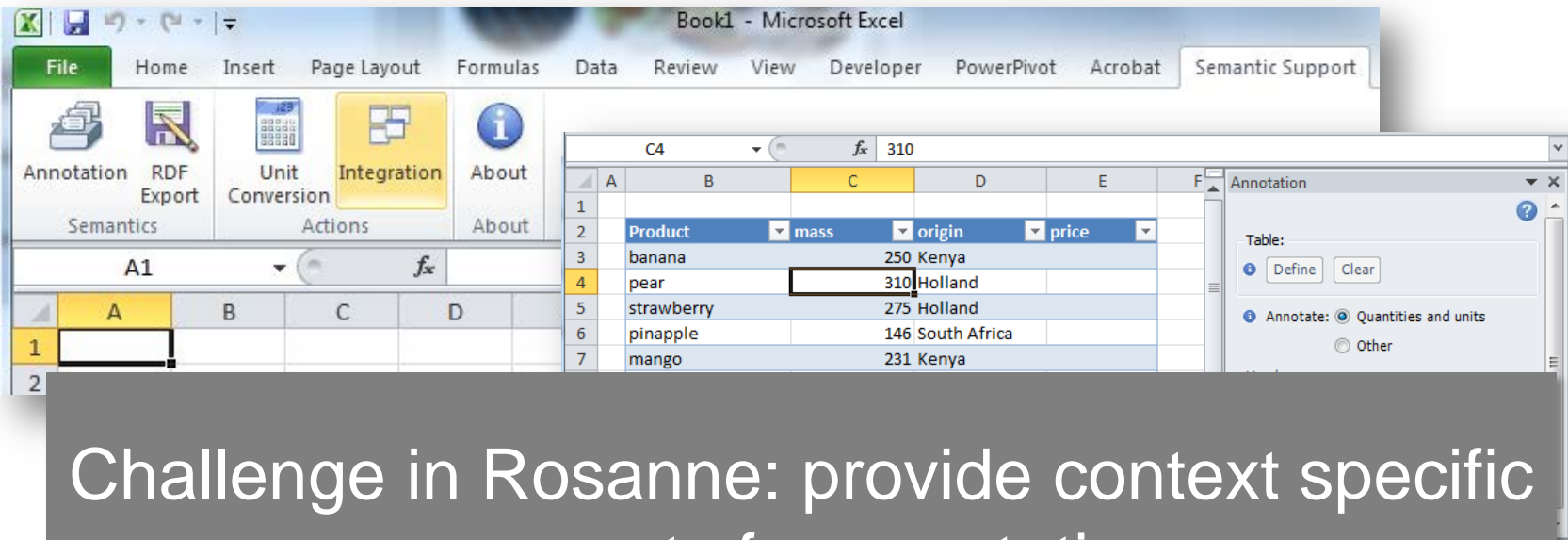
The screenshot shows the Microsoft Excel interface with the 'Semantic Support' ribbon active. The ribbon contains sections for 'Semantics' (Annotation, RDF Export) and 'Actions' (Unit Conversion, Integration, About). The main worksheet, 'Sheet1', contains a table of fruit data. The 'Annotation' dialog box is open, showing the 'Table' tab with 'Define' and 'Clear' buttons. The 'Annotate' section has 'Quantities and units' selected. The 'Header' section shows 'Quantity' as 'mass' and 'Unit' as 'gram'. The 'Options' button is at the bottom right of the dialog.

Product	mass	origin	price
banana	250	Kenya	
pear	310	Holland	
strawberry	275	Holland	
pinapple	146	South Africa	
mango	231	Kenya	
apple	520	Holland	
orange	430	Spain	



Not yet another software solution please

Rosanne: extending Excel with semantics



The screenshot shows the Microsoft Excel interface with the 'Semantic Support' ribbon tab selected. The ribbon includes buttons for 'Annotation', 'RDF Export', 'Unit Conversion', 'Integration', and 'About'. The 'Integration' button is highlighted. The main worksheet area displays a table with the following data:

Product	mass	origin	price
banana	250	Kenya	
pear	310	Holland	
strawberry	275	Holland	
pinapple	146	South Africa	
mango	231	Kenya	

The 'Annotation' pane is open on the right, showing a 'Table' section with 'Define' and 'Clear' buttons. Below it, the 'Annotate' section has two radio buttons: 'Quantities and units' (selected) and 'Other'.

Challenge in Rosanne: provide context specific concepts for annotation



Flexible, but with limited footprint

missing values, merged data, different record lengths,
inhomogeneous, different units, accuracies ...

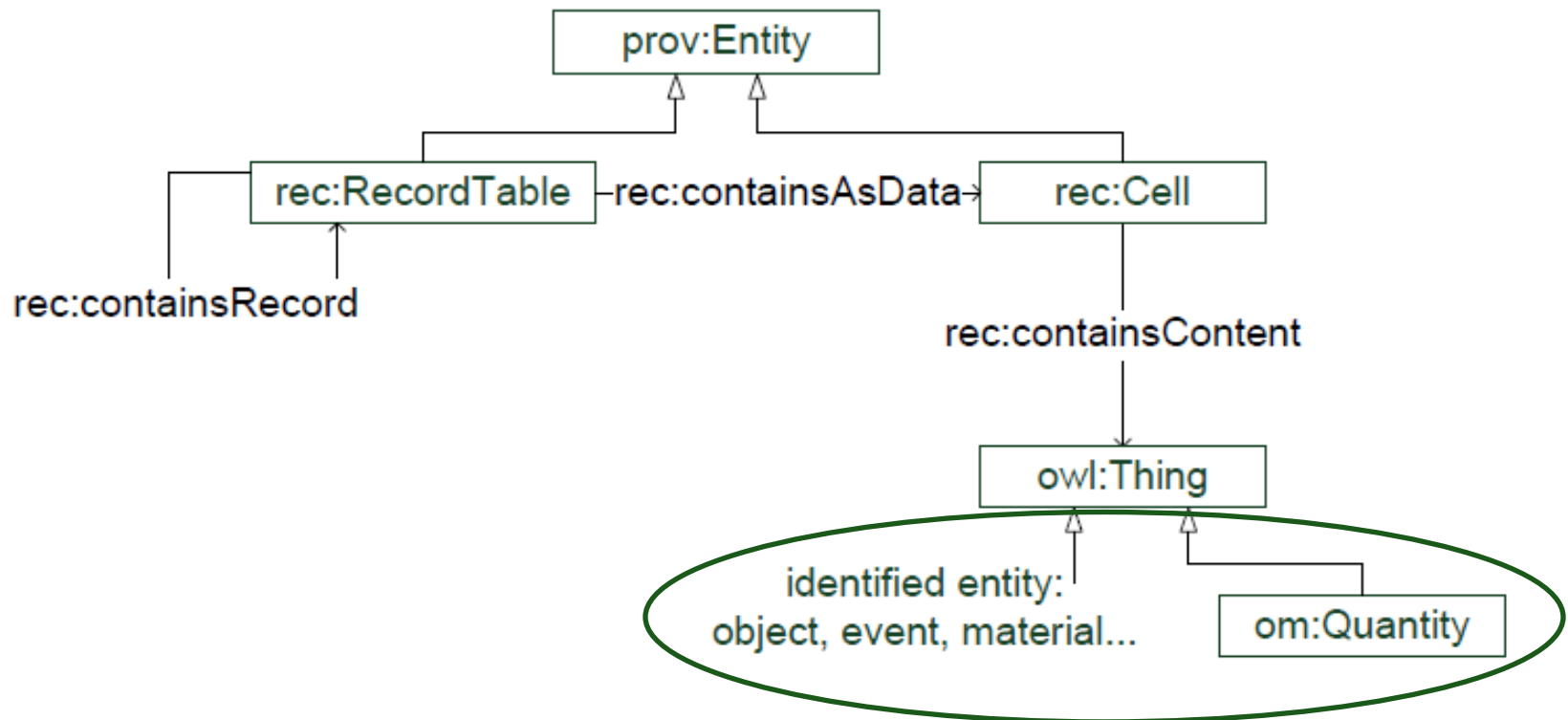
		Emissions	Costs
		Mtonne CO ₂ -eq	Meuro
Crop	Corn	28.0	50.0
	Soybean	70.0	75.0
	Canola	38.0	120.0
	Total	136.0	245.0

The crop Canola leads to an emission of 38.0 Mtonne CO₂-eq

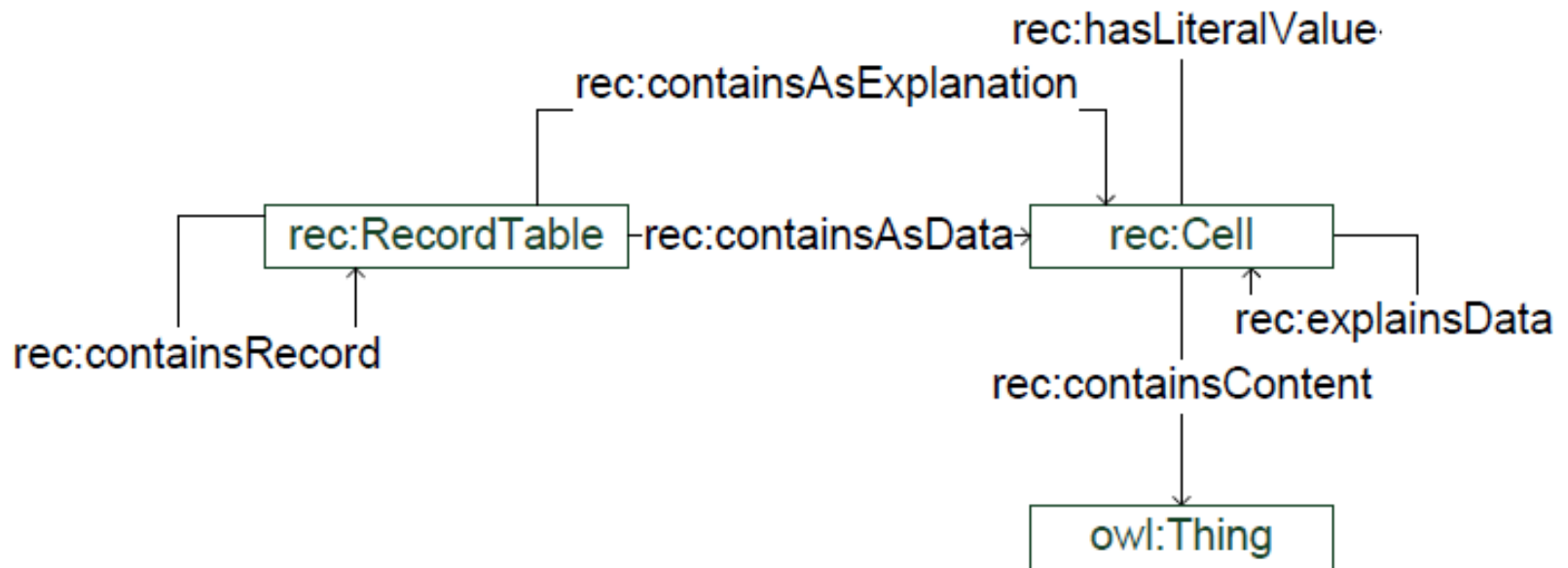


RDF Record Table

basic ideas: record as a snapshot & entities versus quantities



Removing redundancy in tables



Removing redundancy in tables

Challenge: evaluate the model in different situations

rec:containsRecord

rec:containsContent
rec:explainsData

Challenge: how to link to compact formats for homogeneous data



Annotating legacy data

MARTINE DE VOS

INTERPRETING NATURAL SCIENCE
SPREADSHEETS

		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

Identify initial
block structure
(approach 1)

		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

Semantically
categorize blocks
(approaches 2,3,4a)

		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

om:Quantity

om:Unit

om:Phenomenon

om:Measure

Annotate
individual terms
(approaches 2,3,4a)

		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

val:Emission

om:'Mass'

om:tonne

val:Soybean



FOOD & BIOBASED RESEARCH

WAGENINGEN UR

Annotating legacy data

MARTINE DE VOS
INTERPRETING NATURAL SCIENCE
SPREADSHEETS

Identify initial
block structure
(approach 1)

		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

Semantically
categorize blocks
(approaches 2,3,4a)

		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	

om:Quantity

om:Unit

om:tonne

val:Soybean

Challenge: implement heuristics in commonly used software



Message

- Proper metadata requires additional effort
- Turn semantic annotation into regular business
- Do this by contextualizing software



Message

- Proper metadata requires additional effort
- Turn semantic annotation into regular business
- Do this by contextualizing software



Can you help bringing the presented models and tools further into the agrifood research infrastructure?



Thank you



contact: jan.top@wur.nl



FOOD & BIOBASED RESEARCH
WAGENINGEN **UR**