FAIRifying research data as regular business

Jan Top

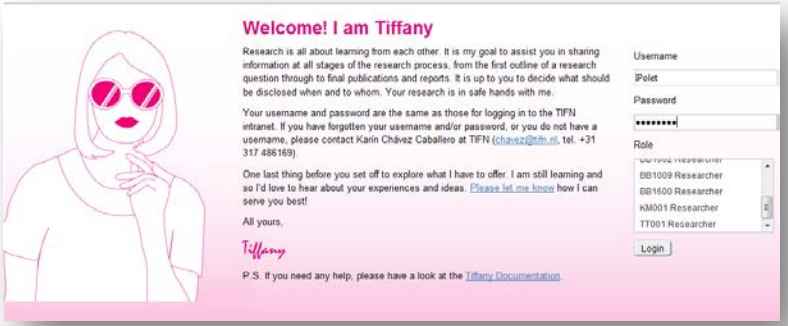
Wageningen University & Research

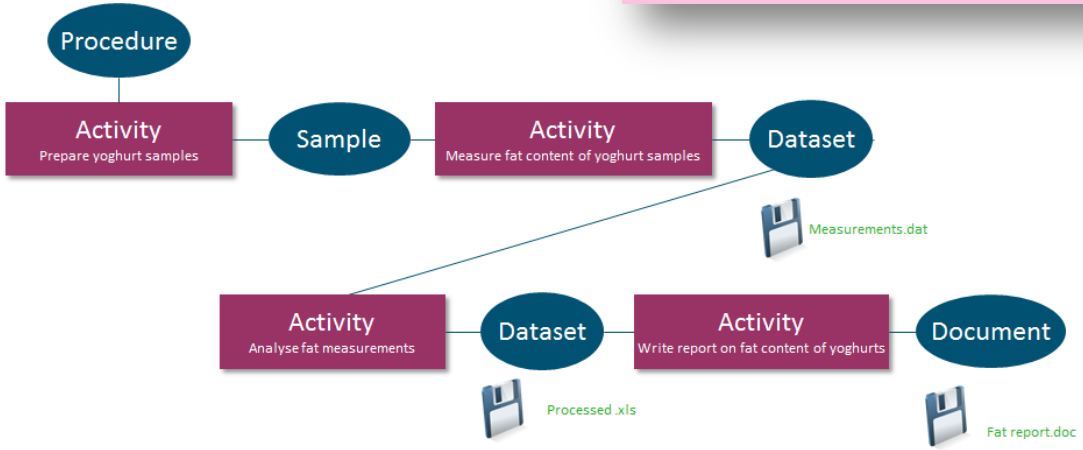
Vrije Universiteit Amsterdam

FOOD & BIOBASED RESEARCH
WAGENINGEN UR

VU VRIJE UNIVERSITEIT AMSTERDAM

Research data management in food research





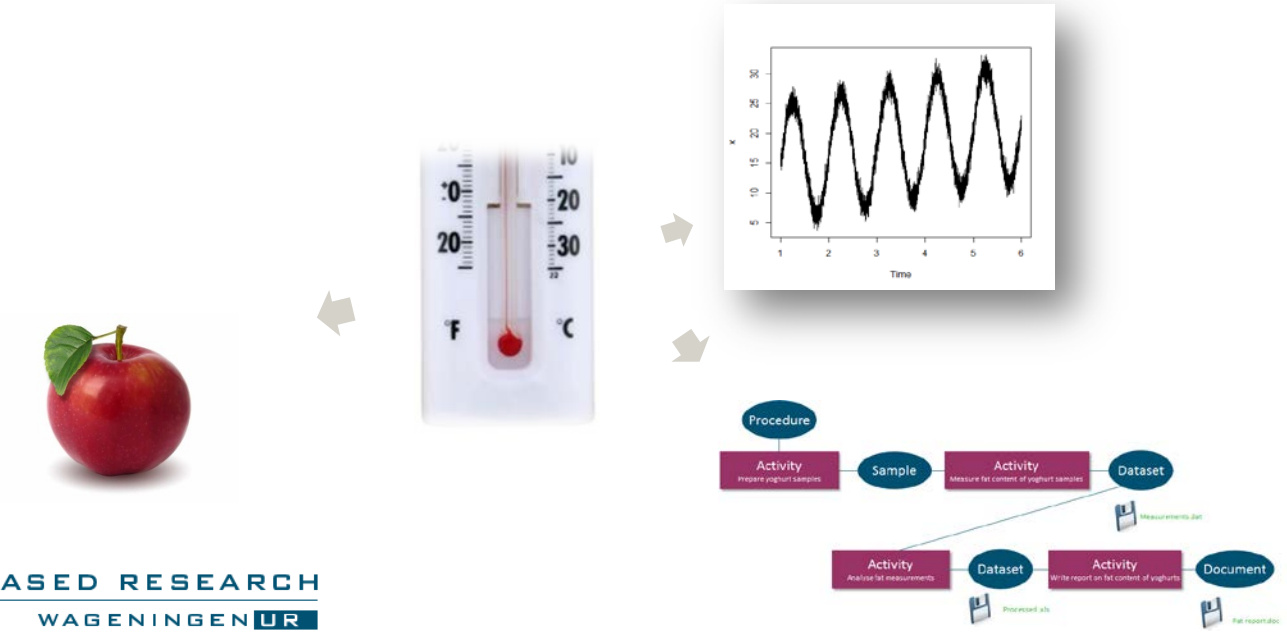
FOOD & BIOBASED RESEARCH
WAGENINGEN UR

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

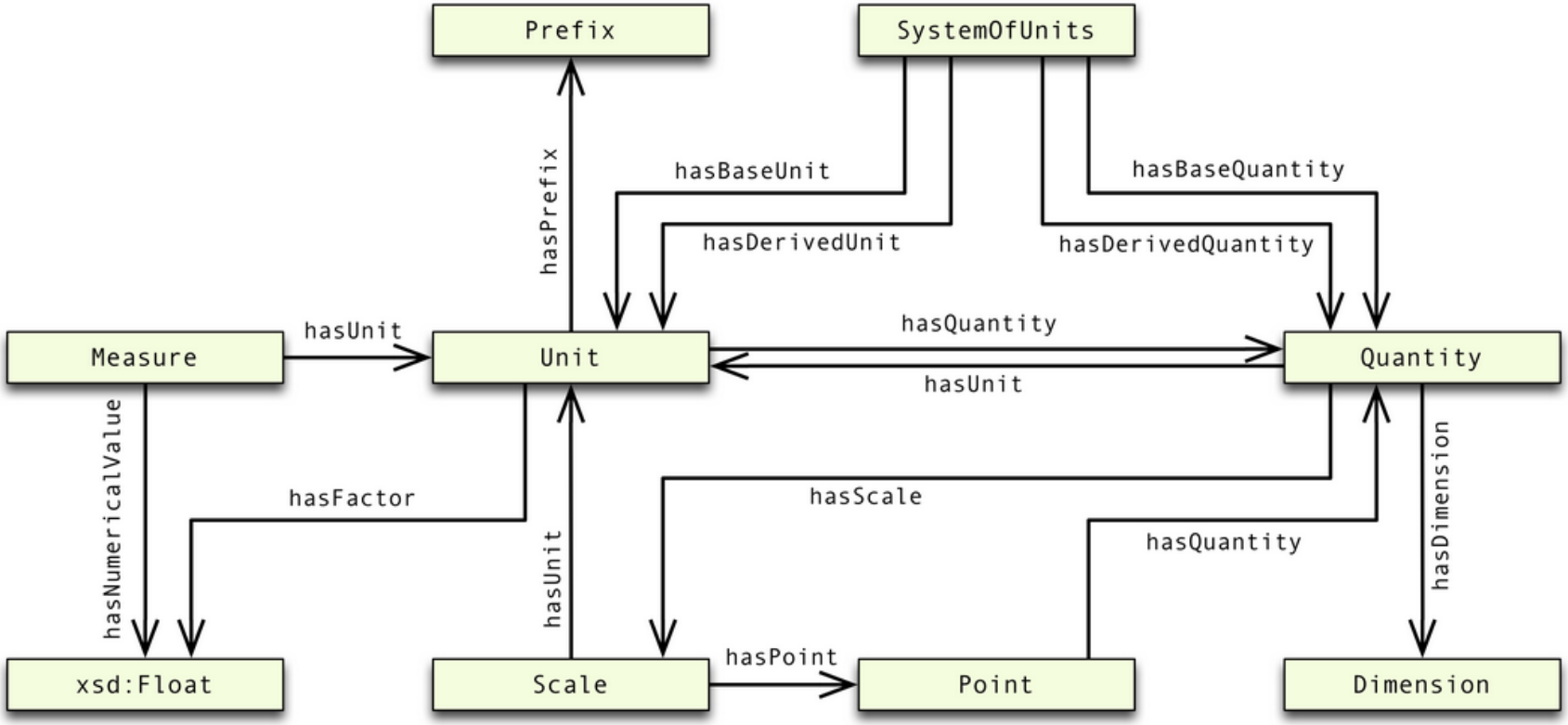
- Generic, shared ontologies and task specific ontologies.
- Annotation embedded in existing tools
- Automated annotation
- Allow for flexible annotation – but limit data footprint

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]	●

FOOD & BIOBASED RESEARCH
WAGENINGEN UR

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)

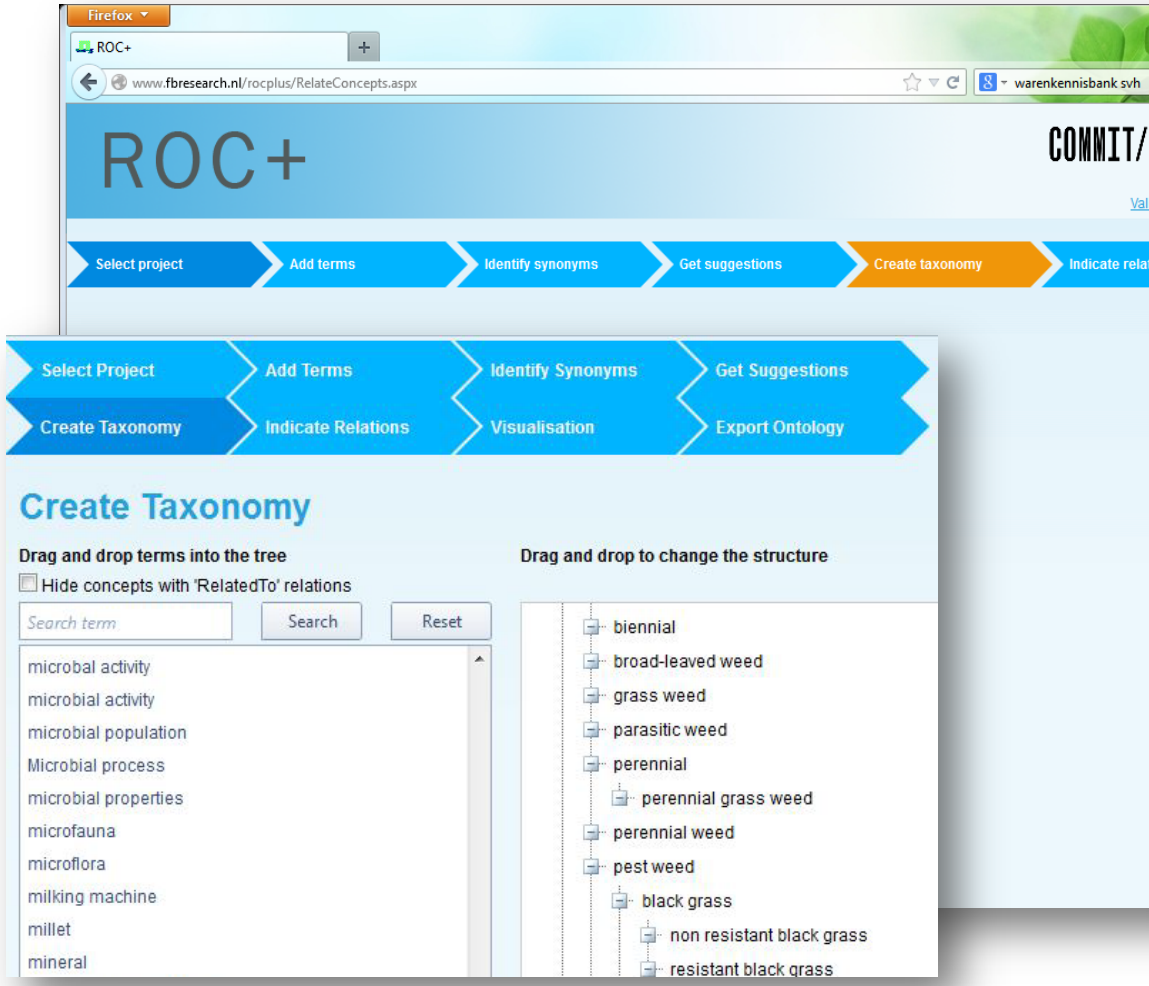
	MUO	OBOE	OM	QU	QUDT	SWEET	UO
unit of measurement (unit)	●	●	●	●	●	●	●
kind of quantity (qk)	●	●	●	●	●	●	●
field of application (app)	○	○	●	●	○[10]	●	○
dimension (dim)	○	○	●	○	●	○	○
dimension vector (vector)	○	○	●	○	●	○	○
system of units (system)	○	○	●	○	●	○	○
phenomenon (phen)	○	●	●	○	○	○	○
measurement (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	●	●	●	○	○	○	○
							○
							○[10]
							○
							○
							○
							○
							○
number of dim. lang. tags	1	1[12]	3	0	0	0[13]	0
unit composition	○	○	●	○	○	○	○
quantity composition	○	○	○	○	○	○	○
everyday lang. designators	○	○	●	○	○	○	○
resolvable URIs	○[13]	○[13]	●	○	○	○[13]	●

Challenge in OM: fill in application areas

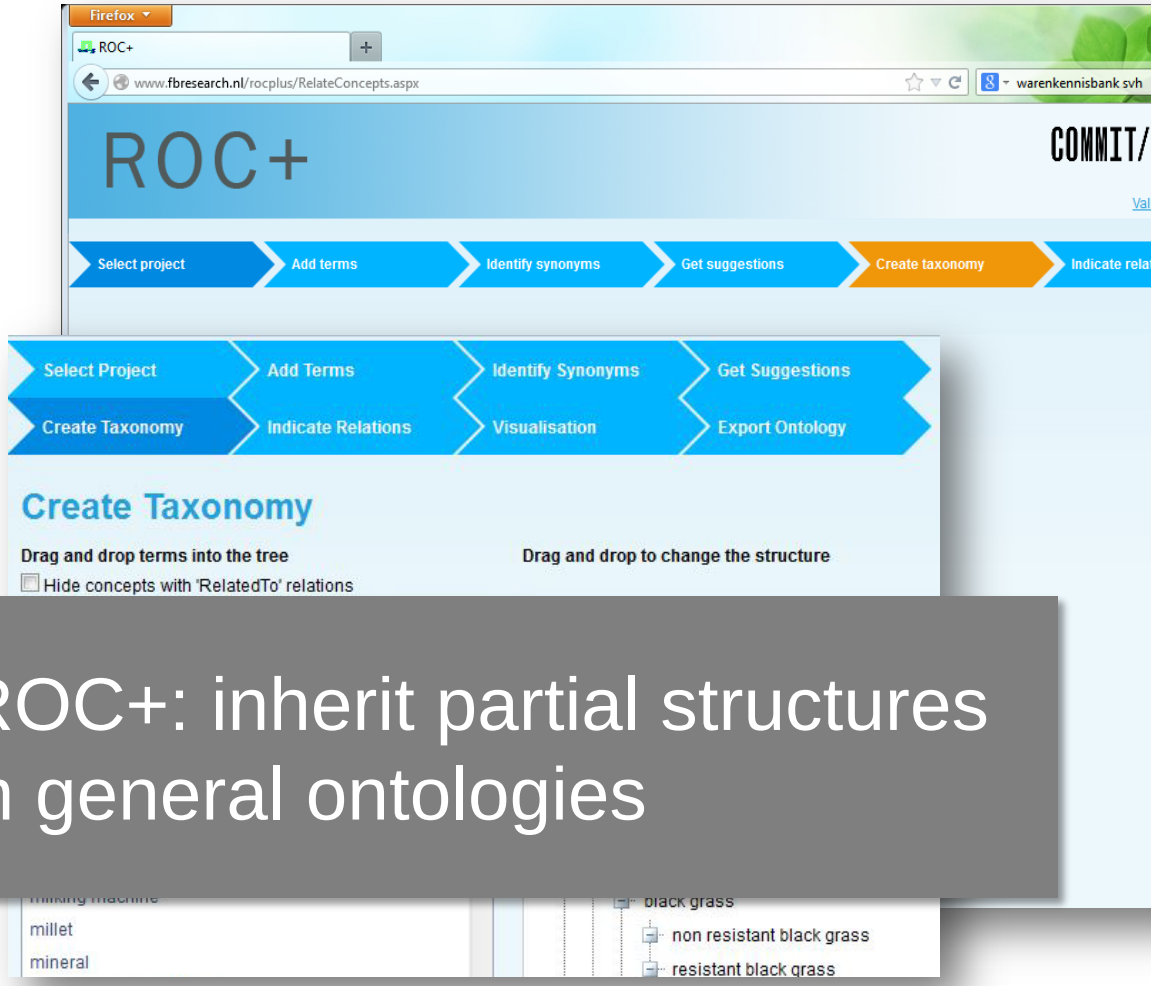
Specific tasks require specific ontologies

Area harvested to date	5	% GB area
Estimated average UK yield to date	Above average	t/ha
Yield range	5.5-9.6	t/ha
UK 10 year average yield	6.4	t/ha
Quality - national average		
SpWt	Limited data	kg/hl
Malting grain N	Limited data	%
Screenings	Limited data	%
Moisture at harvest	14	%

Creating task-specific ontologies



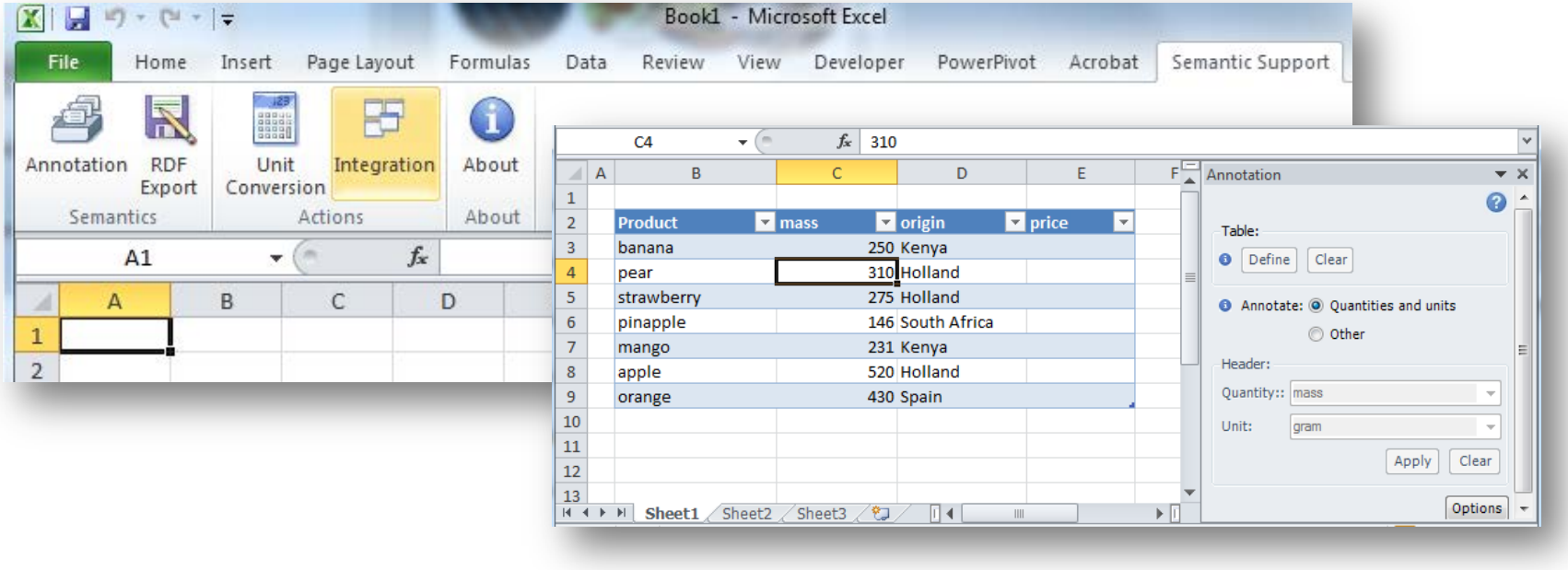
Creating task-specific ontologies



Challenge in ROC+: inherit partial structures from general ontologies

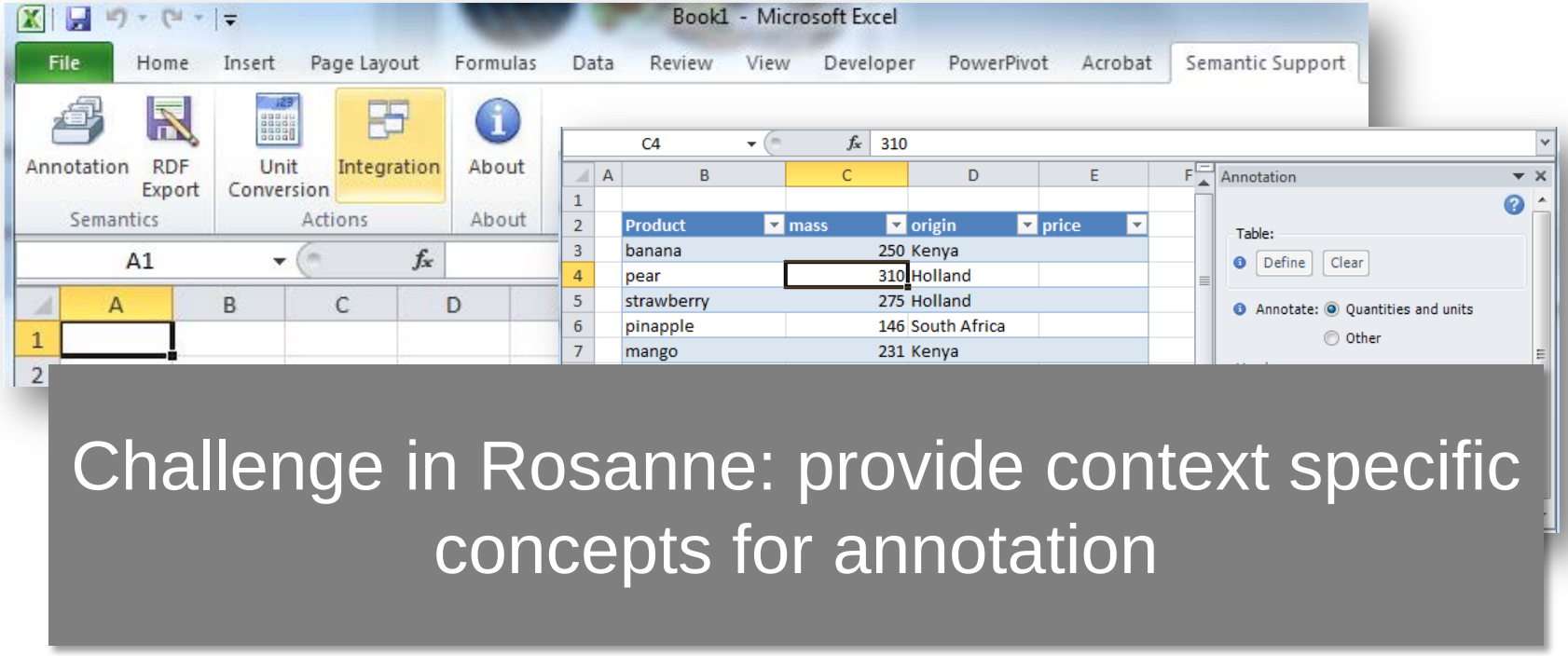
Not yet another software solution please

Rosanne: extending Excel with semantics



Not yet another software solution please

Rosanne: extending Excel with semantics



Challenge in Rosanne: provide context specific concepts for annotation

Flexible, but with limited footprint

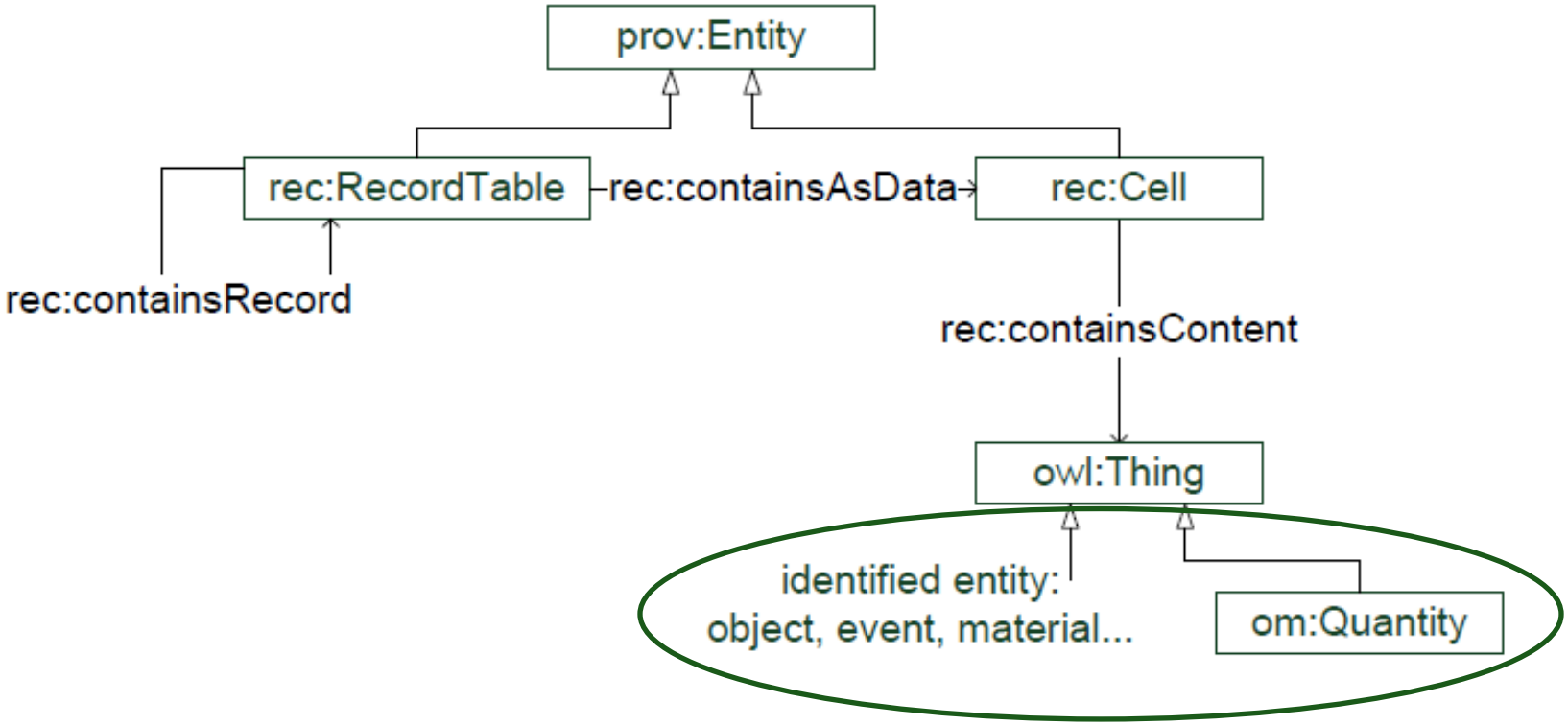
missing values, merged data, different record lengths, inhomogeneous, different units, accuracies ...

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

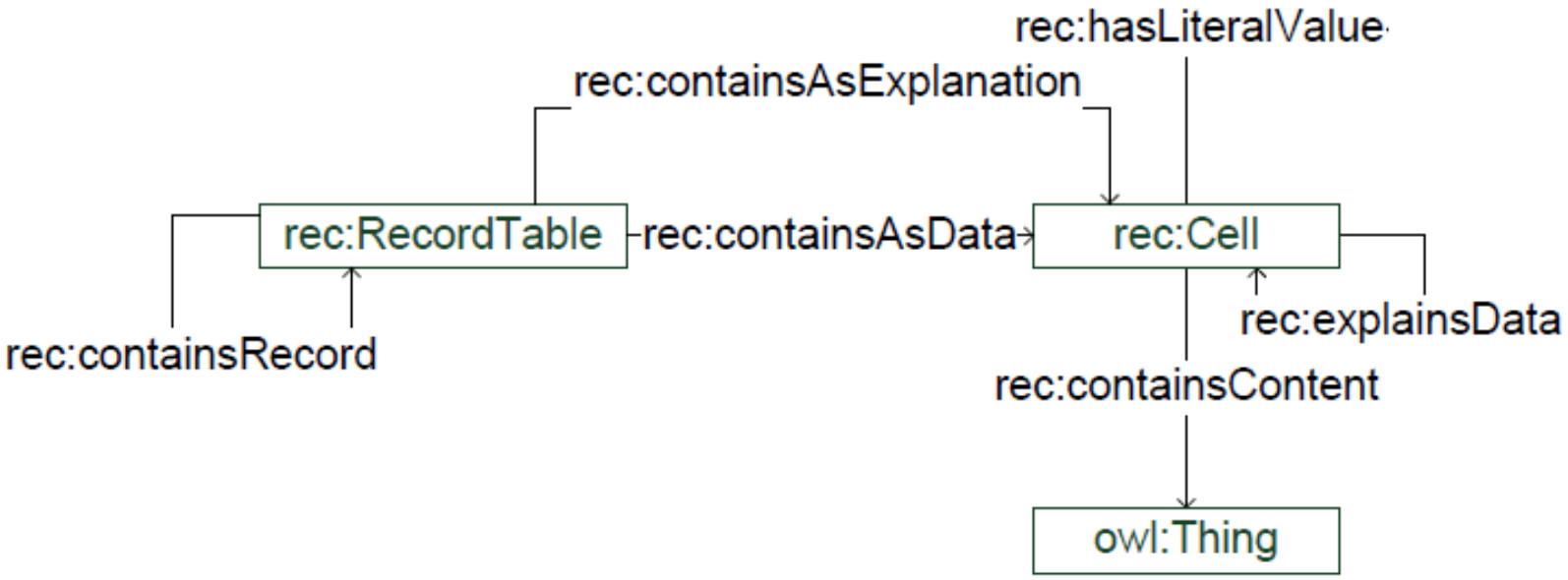
The crop Canola leads to an emission of 38.0 Mtonne CO_2-eq

RDF Record Table

basic ideas: record as a snapshot & entities versus quantities

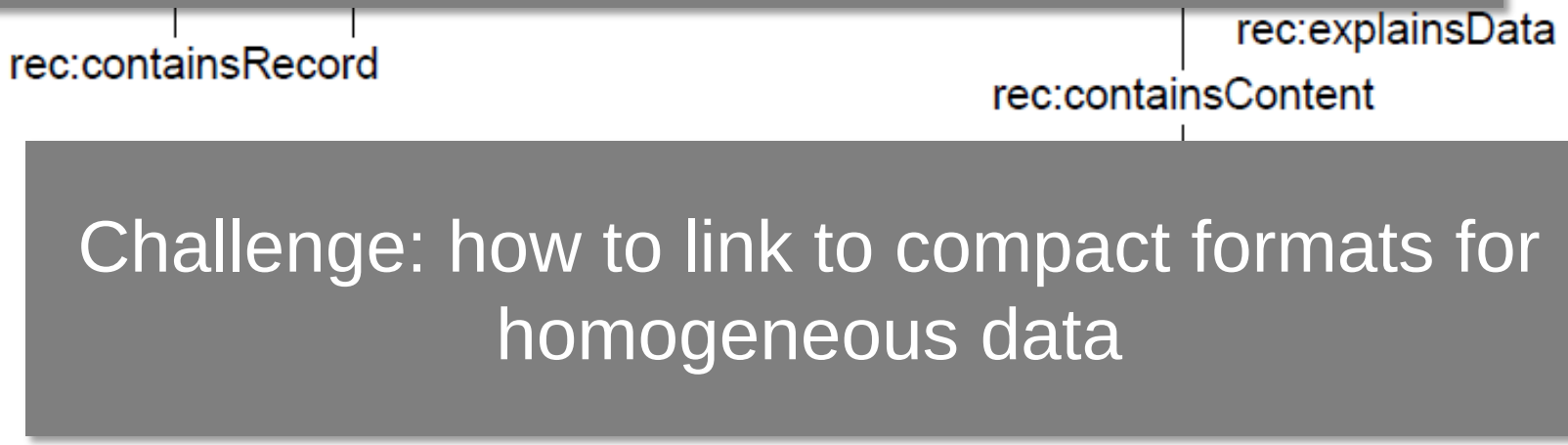


Removing redundancy in tables



Removing redundancy in tables

Challenge: evaluate the model in different situations

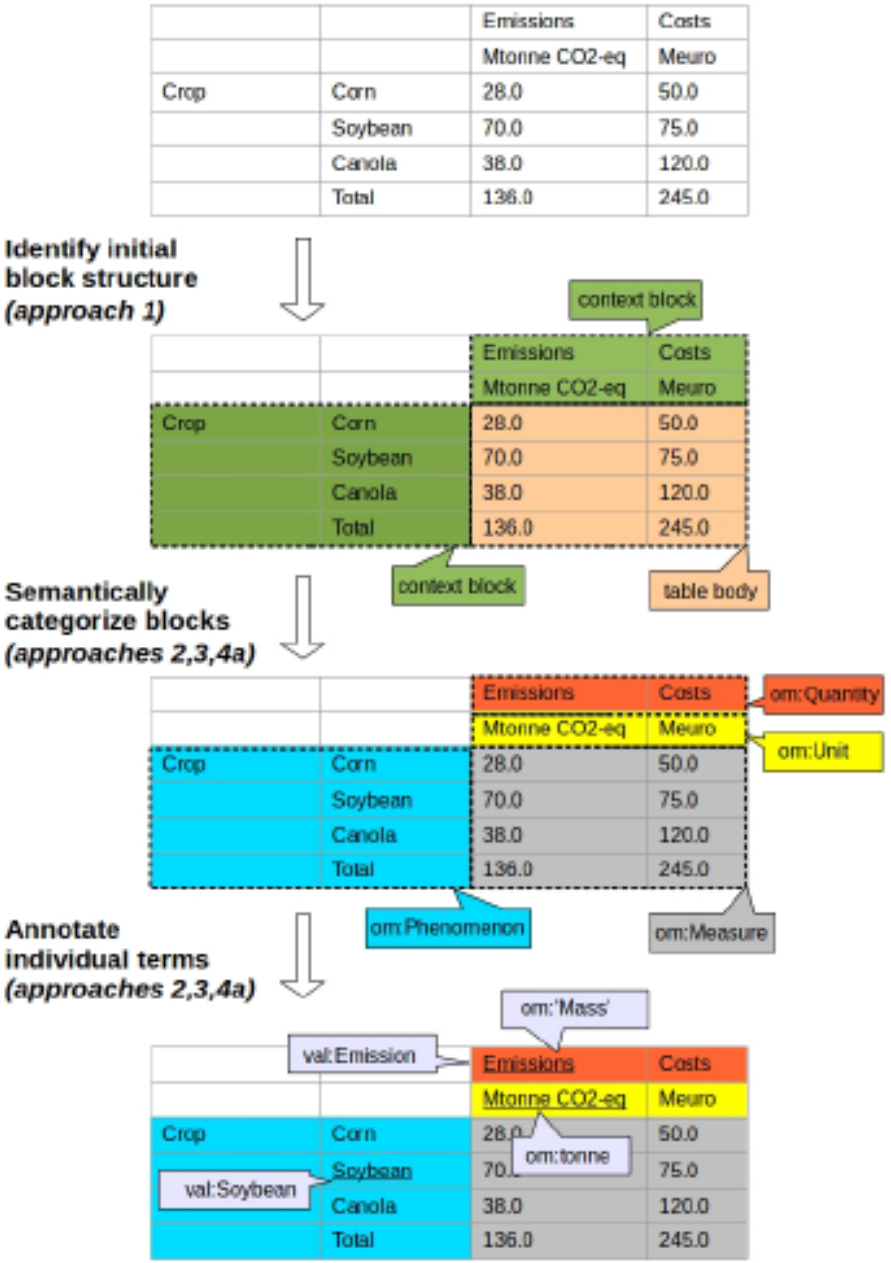


Challenge: how to link to compact formats for homogeneous data

Annotating legacy data

MARTINE DE VOS

INTERPRETING NATURAL SCIENCE SPREADSHEETS

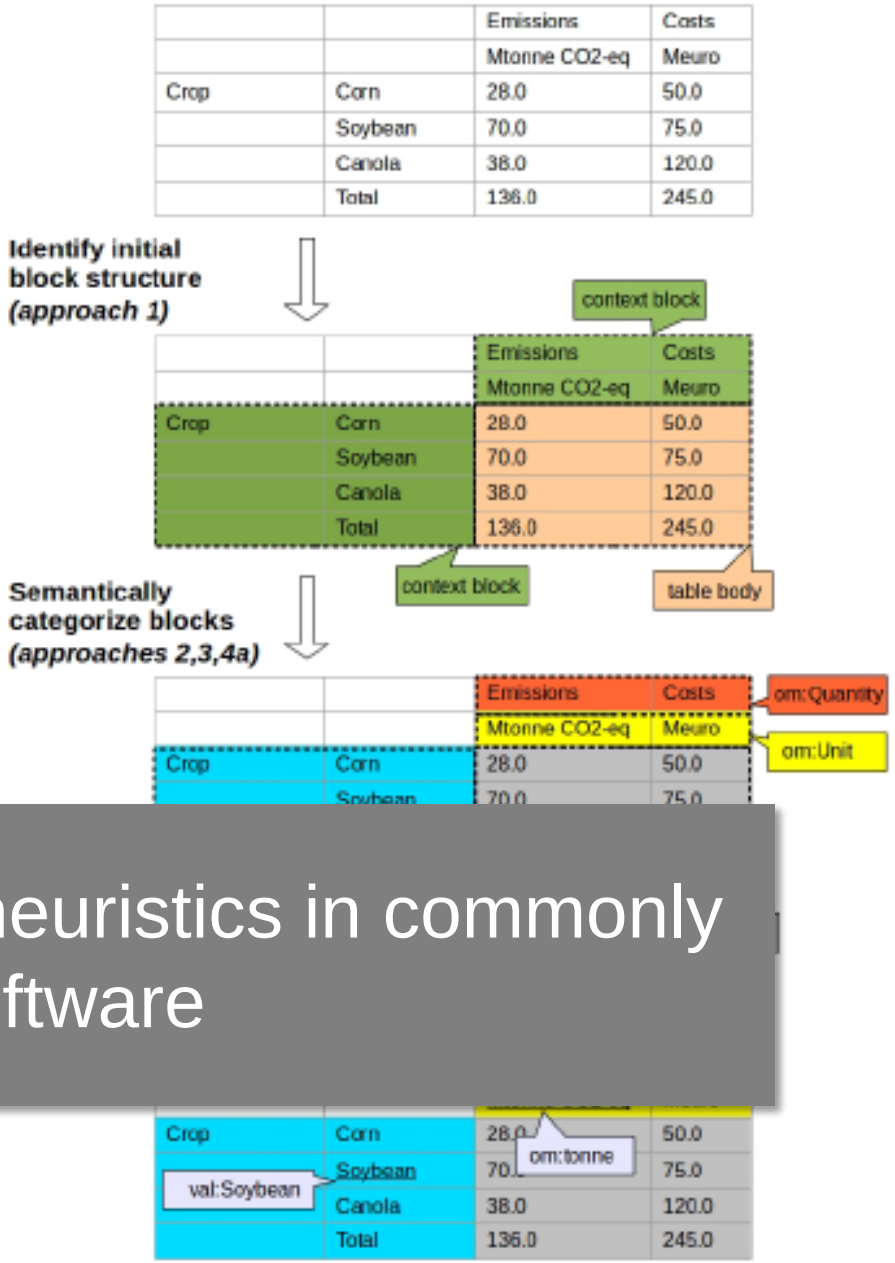


		Emissions	Costs
		Mtonne CO2-eq	Meuro
Crop	Corn	28.0	50.0
	Soybean	70.0	75.0
	Canola	38.0	120.0
	Total	136.0	245.0

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

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