

Decentralised Semantics: A Semantic Engine user perspective

Carly Huitema, Ph.D.

Manager – Agri-food Data
Strategy

SciDataCon 2023

organized by



as part of



ADC Ecosystem

Goal

- Agri-food Data Canada's vision is a research data ecosystem
 - Improving the connections between existing resources
 - Supporting data FAIRness with machine-readable context
- Agri-food Data Canada's current approach is **researcher centric**, helping them create FAIR data

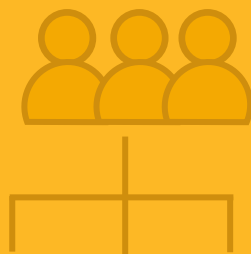
Findable, Accessible, Interoperable, Reusable



**AGRI-FOOD DATA
CANADA**

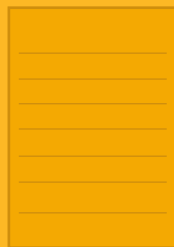
AT THE UNIVERSITY of GUELPH

Governance

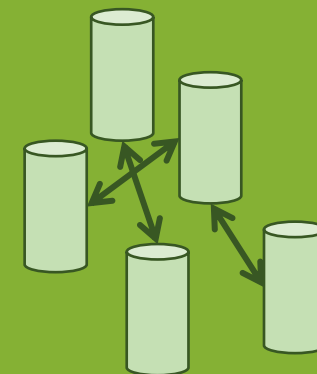


Working Groups

Ecosystem
Principles and
Governance



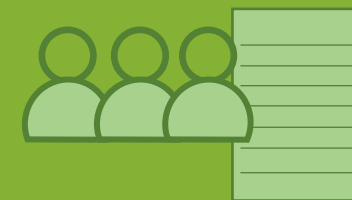
Ecosystem



Federate existing
nodes



Ecosystem Scale
Catalogues



Knowledge and
training

The Semantic Engine for context

ADC is developing a suite of tools called the Semantic Engine



The Semantic Engine helps researchers write rich contextual and machine-readable data documentation

Data Schemas

Data schemas – the **first** contextual documentation target of ADC



Semantic Engine for better schemas



**HUMAN COLOSSUS
FOUNDATION**

Adopting and adapting
overlays capture
architecture (OCA)



Developed overlays capture
architecture (OCA)

OCA is an international and open
standard for data schemas

OCA Documentation:
DOI: [10.5281/zenodo.7707466](https://doi.org/10.5281/zenodo.7707466)



Overlays capture architecture
for rich data schemas

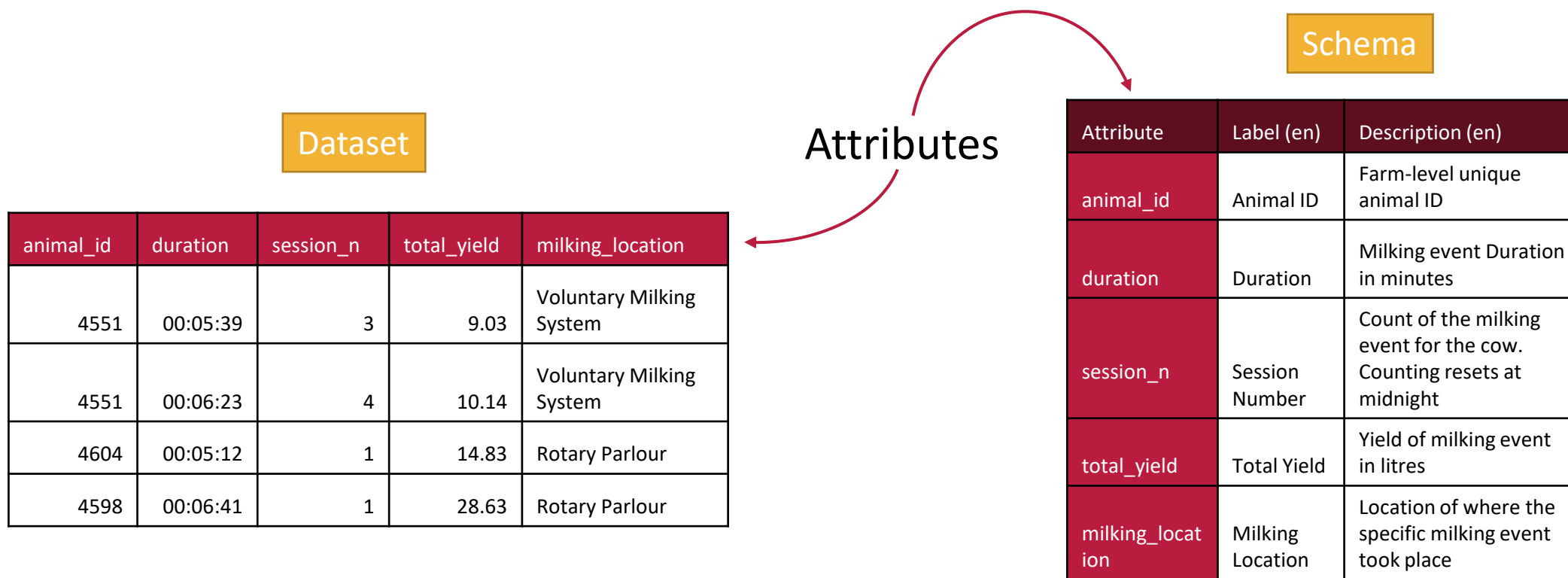
Data documentation: Schemas

- All datasets have a schema
 - Explicit or implicit
 - Contains useful details or ‘user-must-guess’
- A schema describes the attributes (variables)

animal_id	duration	session_n	total_yield	milking_location
4551	00:05:39	3	9.03	Voluntary Milking System
4551	00:06:23	4	10.14	Voluntary Milking System
4604	00:05:12	1	14.83	Rotary Parlour
4598	00:06:41	1	28.63	Rotary Parlour

← Attributes

Attributes are described in a schema



Layers of a schema

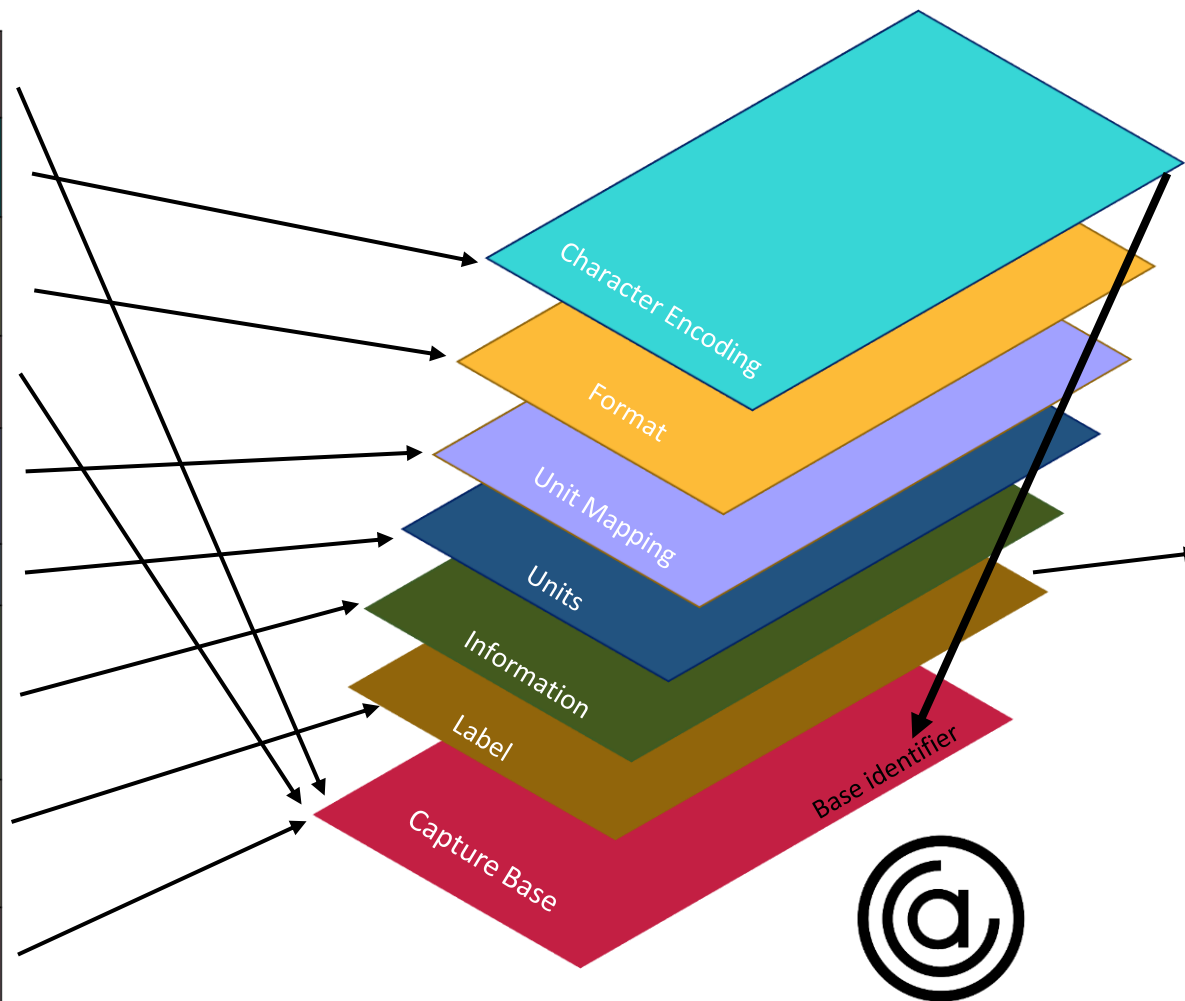
- Extend the schema with more features (details)
- Each feature is a layer of the schema in OCA
 - Advantage – can split schema layers into tasks with independent contributors

Schema

Attribute	Label (en)	Information (en)	Units	Unit mapping (UCUM)	Type	Format	Character encoding	Sensitive
animal_id	Animal ID	Farm-level unique animal ID			Numeric	^[1-9]\d*\$	UTF-8	Y
duration	Duration	Milking event Duration in minutes	min	min	DateTime	hh:mm:ss	UTF-8	
session_n	Session Number	Unique count of the milking event per cow, per day, per milking system. Resets at midnight.			Numeric	^[1-9]\d*\$	UTF-8	
total_yield	Total Yield	Yield of milking event in litres	L	l	Numeric	^\d*\.\d+\$	UTF-8	
milking_location	Milking Location	Location of where the specific milking event took place			Text	^.{0,250}\$	UTF-8	

OCA layer structure

Attribute	Label (en)	Information (en)	Units	Unit mapping (UCUM)	Type	Format	Character encoding	Sensitive
animal_id	Animal ID	Farm-level unique animal ID			Numeric	^[1-9]\d*\$	UTF-8	Y
duration	Duration	Milking event Duration in minutes	min	min	DateTime	hh:mm:ss	UTF-8	
session_n	Session Number	Unique count of the milking event per cow, per day, per milking system. Resets at midnight.			Numeric	^[1-9]\d*\$	UTF-8	
total_yield	Total Yield	Yield of milking event in litres	L	l	Numeric	^\d*\.?\d+\$	UTF-8	
milking_location	Milking Location	Location of where the specific milking event took place			Text	^[0-250]\$	UTF-8	

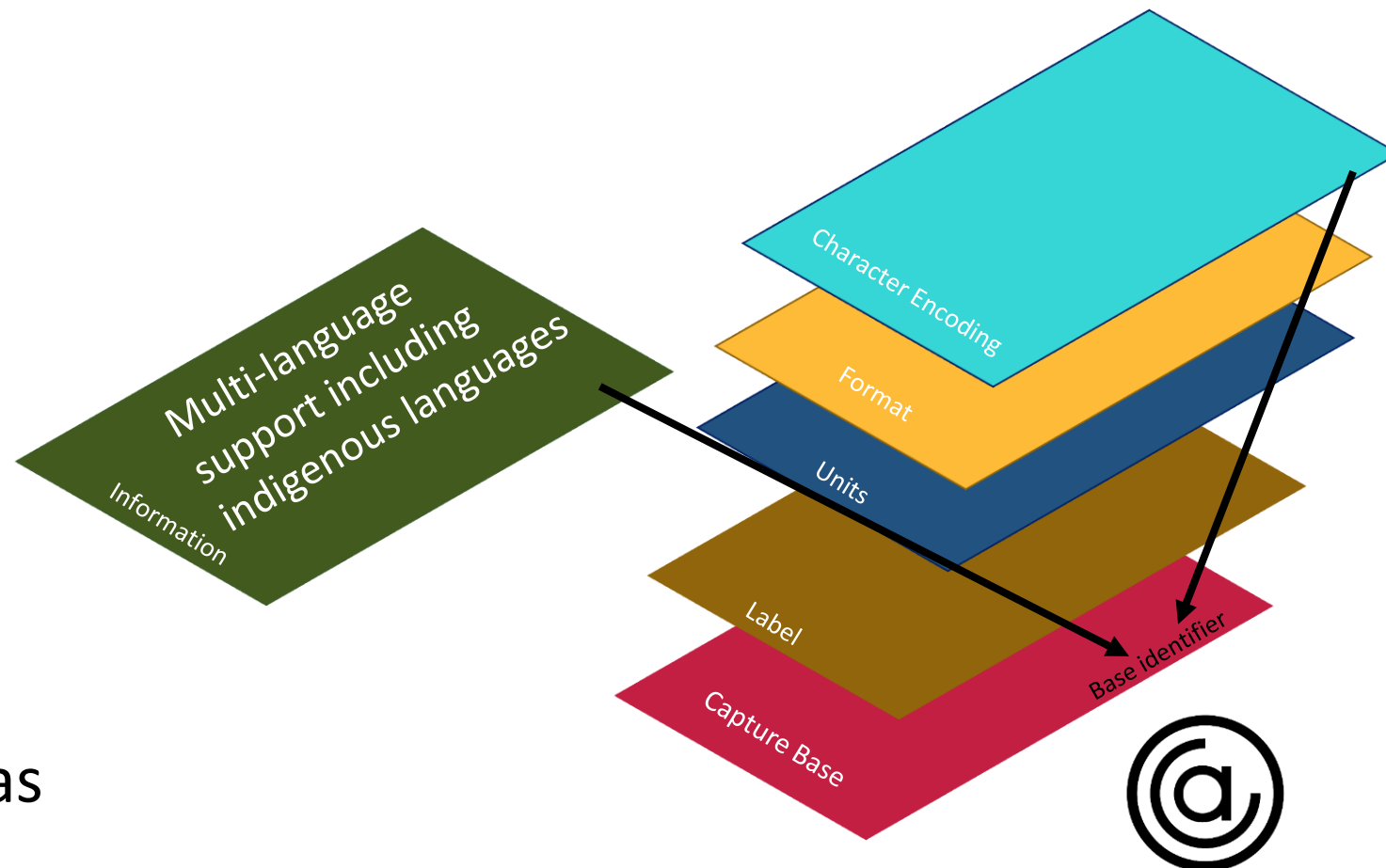


OCA is a machine-readable format

```
{
  "capture_base": "EzB01m3oiJ7-tNwHipu-KSPdqVQpbizTXFuu4CdcCQbA",
  "digest": "EF4w6K-SMVZ4VFxkc-jqNNGUXKb0587o3MtcpBU_zeko",
  "type": "spec/overlays/label/1.0",
  "language": "po",
  "attribute_labels": {
    "animal_id": "ID do animal",
    "begin_time": "Hora de início",
    "date": "Data de início",
    "dim": "Dias em leite",
    "duration": "Duração",
    "end_date": "Data de término",
    "end_time": "Hora de término",
    "lact_n": "Número da lactação",
    "milking_location": "Local de ordenha",
    "session_n": "Número da sessão",
    "total_yield": "Produção total"
  },
  "attribute_categories": [
  ],
  "category_labels": {
  }
}
```

Identifiers glue the schema together

- Identifiers enable a decentralized schema architecture
- OCA supports multiple contributors for a schema
 - OCA uses self-addressing identifiers to link and track content
- Enables community improvement of OCA schemas



Who is using OCA

- Research projects at the University of Guelph
- University of Guelph research barns and stations
- Government of British Columbia using OCA for multi-language credential presentation
 - Gaining interest in other provinces, early pilots with Quebec
- EU project – cardiovascular consortium of 21 organizations (including universities and institutes) across Europe, CH, UK, and USA.
- IOT platform in supply chain - monitoring and telemetry data (multinational organization)

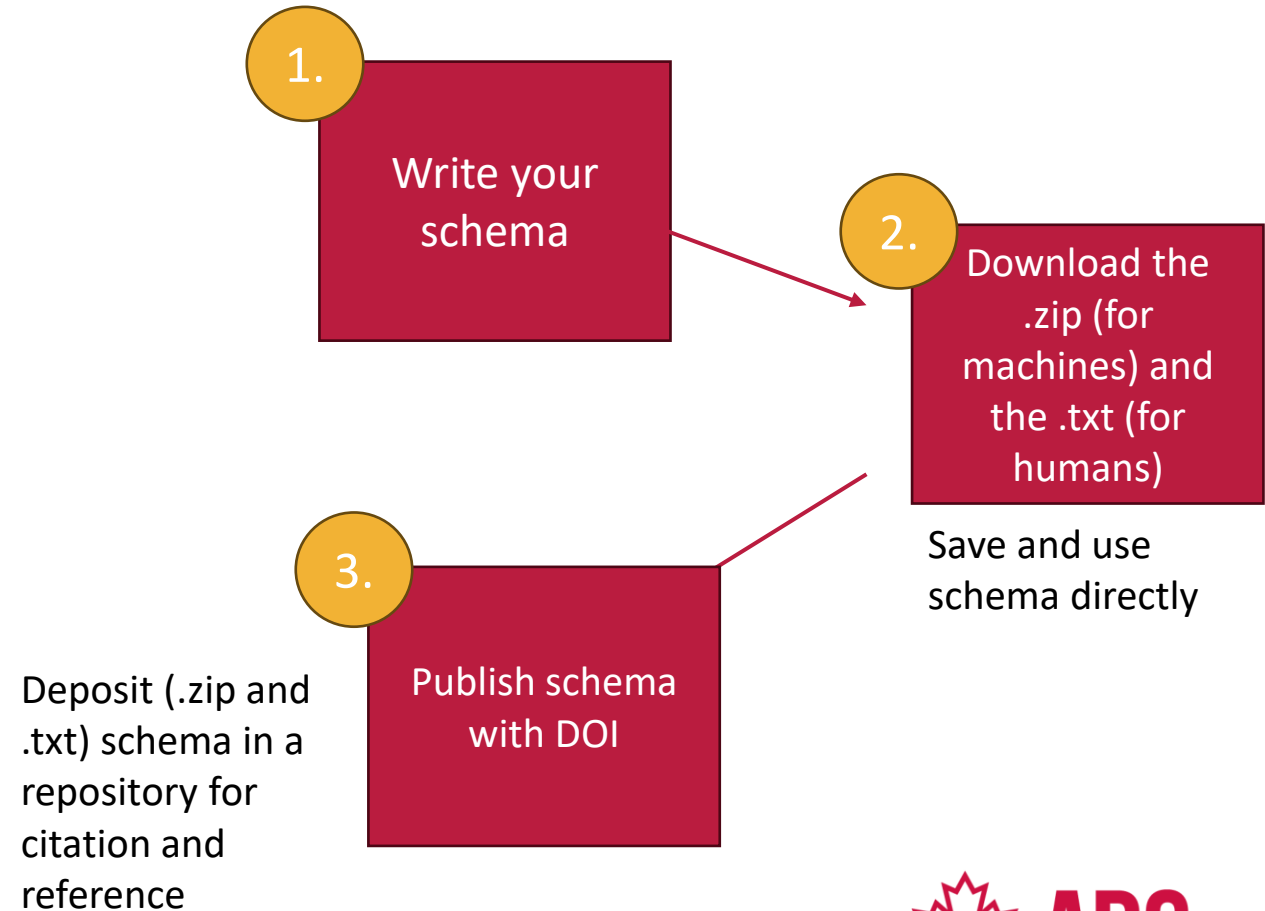


How anyone can use OCA schemas

semanticengine.org

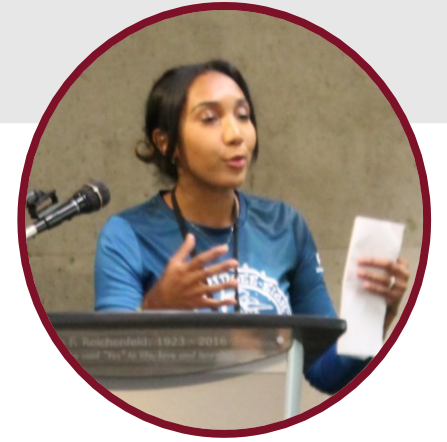
The screenshot shows the 'Semantic Engine' website. At the top, it says 'Semantic Engine' and 'AGRI-FOOD DATA CANADA AT THE UNIVERSITY OF GUELPH'. Below this, a section titled 'Schemas add value to data' explains that schemas make data more valuable and provides a 'Learn more' link. A video player shows a thumbnail for 'Agri-food Data Canada Metadata Expl...'. Below the video, a 'Quick-start (with Quick Links)' section lists three steps: 1. Write your schema and generate the schema .zip bundle and a simple .txt Readme. 2. Use your .zip schema here to view, validate, and generate an Excel sheet for data input aligned with your schema. 3. Store your schema files with your data, put them in a repository, or collaborate by sharing them with others. At the bottom, there are two expandable sections: 'What is a Schema' and 'Write a Schema'. The 'Write a Schema' section is expanded, showing a 'Quick Links' area with a 'Write a schema' link and a button to 'Upload schema bundle (.zip) Or drag and drop one'.

General flow



Example usage

- Beef carcass data from the Meat Lab at the University of Guelph
 - All research farm products are sent to the campus meat lab
 - Meat lab is a small-scale meat packing plant operating under federal inspection regulations
 - Meat lab grades animals and collects data
 - Information on beef carcasses (multidimensional)
 - Schema documentation to help researchers use the data collected



Raw data collection

	A	B	C	D	E	F	G	H	I	J
1	date	animal_id	hotcarcass_wt	rib_fat1	rib_fat2	rib_fat3	grade_fat	ribeyearea_sqin	ribeyearea_sqmm	marbling
2	19970401	46F	244	9	12	10	8	9.2	.	4.5
3	19970421	82F	323	17	7	9	9	12.1	.	5
4	19970429	35F	310	17	14	10	10	10.5	.	6
5	19970429	44F	278	11	18	8	7	9.1	.	5.5
6	19970429	35F	291	17	15	8	8	11.9	.	3
7	19970506	OCA 54F	324	22	10	10	10	12.5	.	5.5
8	19970506	7258	318	28	12	12	11	10.7	.	4.5
9	19970513	7240	316	16	16	11	11	11.6	.	4.5
10	19970513	OCA 21F	291	27	10	8	8	11.5	.	5.5
11	19970513	86F	347	11	12	8	8	12.1	.	3.5
12	19970521	OCA 3F	381	21	10	9	9	14.3	.	4.5
13	19970521	11F	373	8	17	8	8	13.6	.	4.5
14	19970521	F105	326	11	20	10	10	10.6	.	4.5
15	19970521	7259	313	25	8	9	9	9.3	.	4.5
16	19970527	119F	364	7	19	9	9	12	.	4.5
17	19970527	F78	280	10	18	11	9	10.5	.	4
18	19970527	23F	353	10	24	12	12	11.6	.	5.5
19	19970527	9F	373	10	25	9	9	12.2	.	5
20	19970603	OCA 18F	356	8	16	9	9	13	.	6
21	19970603	F54	339	9	21	11	10	11.2	.	5.5
22	19970603	28F	372	8	15	8	8	11.8	.	5
23	19970603	OCA 51F	294	19	15	11	11	10.9	.	5
24	19970610	93F	309	25	13	12	10	11.8	.	5
25	19970610	F442	310	25	10	8	8	11.2	.	5



Author Schema using Semantic Engine

What is a Schema



Write a Schema



Using our [schema writer](#) you can easily create human- and machine-readable versions of your schema.

You will either be documenting an existing dataset or you will start with an idea of what columns or variables you will be collecting for your dataset. [Learn more about how to design a dataset.](#)

[Write a schema](#)

Quick Links

[Write a schema](#)

Upload schema bundle (.zip) Or
drag and drop one

VIEW SCHEMA

EDIT SCHEMA

GENERATE README

Author Schema using Semantic Engine



Review Schema ?

HELP WITH THIS PAGE

En

< BACK

EDIT SCHEMA

Schema Language ?

ENGLISH

English

Schema Metadata ?

Name of Schema

Beef carcass grading and rib dissection

Description

Schema for beef carcass grading and rib dissection data. Originally designed to document Schema2010_Jan-Nov_CGRD

Classification

Animal and dairy science

Schema Details ?

Attribute	Sensitive	Unit	Type	Label (max 50 chars)	Description (max 200 chars)	List
animal_id	<input type="checkbox"/>		Text	Alpha-numeric ID of carcass		Not a List
body_fat	<input type="checkbox"/>	kg	Numeric	Amount of adipose tissue in the live animal		Not a List
bone_wt	<input type="checkbox"/>	kg	Numeric	Weight of bones in carcass	After removal of meat and other tissues	Not a List
date	<input type="checkbox"/>		Numeric	Date of sampling		Not a List



Schema in .txt and .zip (JSON) formats

Readme.txt

```
|
BEGIN_REFERENCE_MATERIAL
*****
OCA_READ_ME/1.0
This is a human-readable schema, based on the OCA schema standard.

Reference for Overlays Capture Architecture (OCA):
https://doi.org/10.5281/zenodo.7707467

Reference for OCA_READ_ME/1.0:
https://github.com/agrifooddatacanada/OCA_README

A schema describes details about a dataset.
In OCA, a schema consists of a capture_base which documents the attributes and their most basic features.
A schema may also contain overlays which add details to the capture_base.
For each overlay and capture_base, a hash of their original contents has been calculated and is reported here as the SAID value.

This README format documents the capture_base and overlays that were associated together in a single OCA Bundle.
OCA_MANIFEST lists all components of the OCA Bundle.
For the OCA_BUNDLE, each section between rows of ****'s contains the details of one "layer type/version" of the OCA Bundle.
*****
END_REFERENCE_MATERIAL

BEGIN_OCA_MANIFEST
*****
Bundle SAID: unavailable

"capture_base" SAID: "ECVOLOYXAiPyT41U0P1V_Pm40jp7e401v61IFoWjaUUw",
"character_encoding" SAID: "EjttH2d4NN8LU9qrly914IfuFkesoWs50uaL1fiJFnY",
"information (en)" SAID: "EY25fT-ckXLWW_RjZ92LC_TBbQt2d9KoYaJTwl0Lmg-k",
"label (en)" SAID: "EIp_dI-md43kpMTHWSMw0C0OURcpqu_2yINFs-V4wTto",
"meta (en)" SAID: "EpW1GBud80JlwoQ222tpHY_3PW5getsCTbWHM4zvvgx38",
"unit" SAID: "EAU0J5BG1zgp63nk7XnRVhVq1tUTw2Kzm4tHBgOk2dko"
*****
END_OCA_MANIFEST

BEGIN_OCA_BUNDLE
*****
Layer name: capture_base/1.0
SAID: ECVOLoyXAiPyT41U0P1V_Pm40jp7e401v61IFoWjaUUw
classification: RDF402

Schema attribute: data type
animal_id: Text
body_fat: Numeric
bone_wt: Numeric
date: Numeric
decl_bull: Text
decl_cow: Text
decl_heifer: Text
decl_steer: Text
```

JSON Information overlay (viewer)

▼ object {5}


```
capture_base : ECVOLoyXAiPyT41U0P1V_Pm40jp7e401v61IFoWja
               UUw
digest : EY25fT-ckXLWW_RjZ92LC_TBbQt2d9KoYaJTwl0Lmg-k
type : spec/overlays/information/1.0
language : en
```

▼ attribute_information {39}

```
body_fat : value
bone_wt : After removal of meat and other tissues
date : value
decl_bull : Animals declared bull
decl_cow : Animals declared cow
decl_heifer : Animals declared heifer
decl_steer : Animals declared steer
discrepancies : Animals with discrepancies between
                input sheet and tracing data
```




Schema now available in repository



Search ▾ User Guide Support English ▾ Log In

Beef carcass grading and rib dissection data schema

Version 3.0



Muhammad, Fatimatzahra, 2023, "Beef carcass grading and rib dissection data schema", <https://doi.org/10.5683/SP3/GHC9Q9>, Borealis, V3

[Cite Dataset ▾](#) [Learn about Data Citation Standards.](#)

Access Dataset ▾

Contact Owner Share

Dataset Metrics ?

20 Downloads ?

Description ?

Schema for beef carcass grading and rib dissection datasets. This schema is applicable to all datasets collecting beef carcass and rib dissection data. (2023-05-29)


Subject ?

Agricultural Sciences


Keyword ?

schema, beef carcass, grading, rib dissection

License/Data Use Agreement


 CC-BY 4.0

Files Metadata Terms Versions


Search this dataset... 



Filter by
File Type: All ▾ Access: All ▾


1 to 2 of 2 Files




[SCHEMA_beef_carcass_rib_dissection.txt](#)



Plain Text - 10.0 KB
Published Oct 6, 2023
0 Downloads
MD5: d25...d9b 






[SCHEMA_beef_carcass_rib_dissection.zip](#)

ZIP Archive - 4.6 KB
Published Oct 6, 2023
0 Downloads
MD5: c4e...96e 





Download

Sort ▾




[SCHEMA_beef_carcass_rib_dissection.txt](#)


Plain Text - 10.0 KB
Published Oct 6, 2023
0 Downloads
MD5: d25...d9b 



[SCHEMA_beef_carcass_rib_dissection.zip](#)

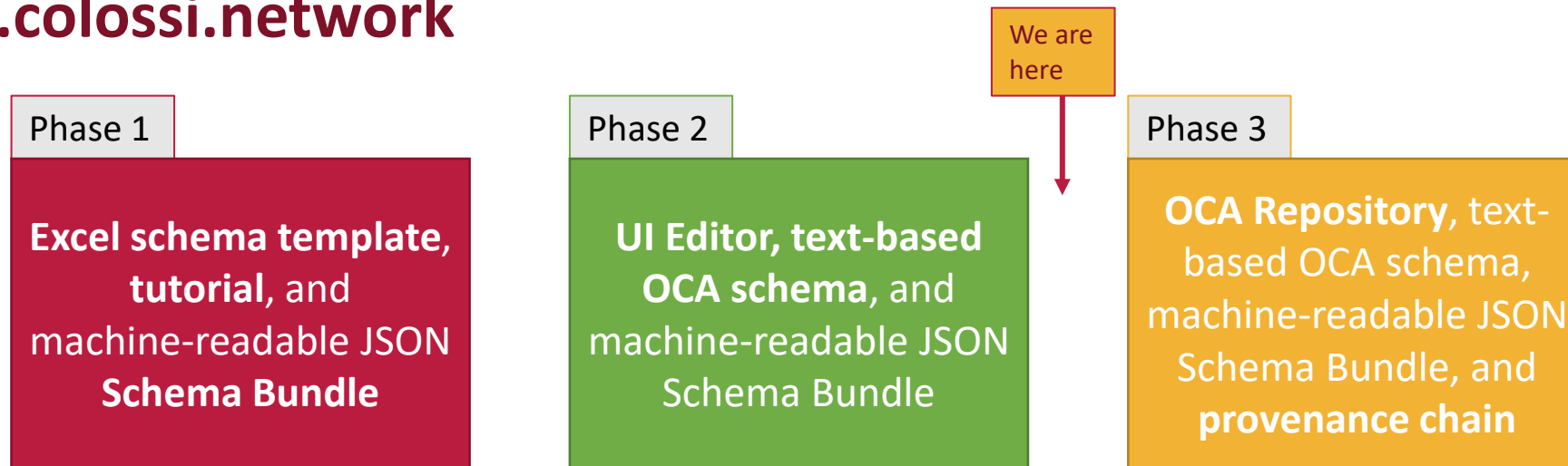
ZIP Archive - 4.6 KB
Published Oct 6, 2023
0 Downloads
MD5: c4e...96e 

18 | Decentralised Semantics: A Semantic Engine user perspective



How to follow our work

- **Agrifooddatacanada.ca** – the root of all our work
- Tools available at **semanticengine.org**
- Reach out to us at adc@uoguelph.ca with ideas, bugs, comments, collaborations
- Participate in the OCA standard and more at **humancolossus.org** and **oca.colossi.network**



THANK YOU

Agri-food Data Canada at the University of Guelph is an innovation platform for Canada's agriculture and food sectors.



**AGRI-FOOD DATA
CANADA**

AT THE UNIVERSITY *of* GUELPH

50 Stone Road East, Guelph, ON N1G 2W
adc@uoguelph.ca

agrifooddatacanada.ca

**UNIVERSITY
of GUELPH**