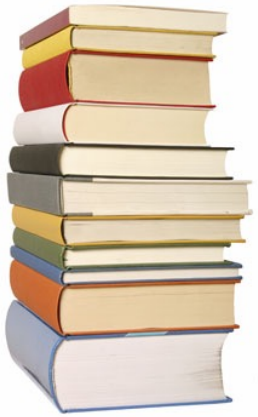
Semantic Species Description: TaxonWorks & PhenoScript

Sergei Tarasov

Beetle Curator
Finnish Museum of Natural History, Helsinki, Finland

- @tarasov_sergio
- sergei.tarasov@helsinki.fi
- https://www.tarasovlab.com

Natural Language Spesies Descriptions (~250 years of effort)



1.3 million of described species

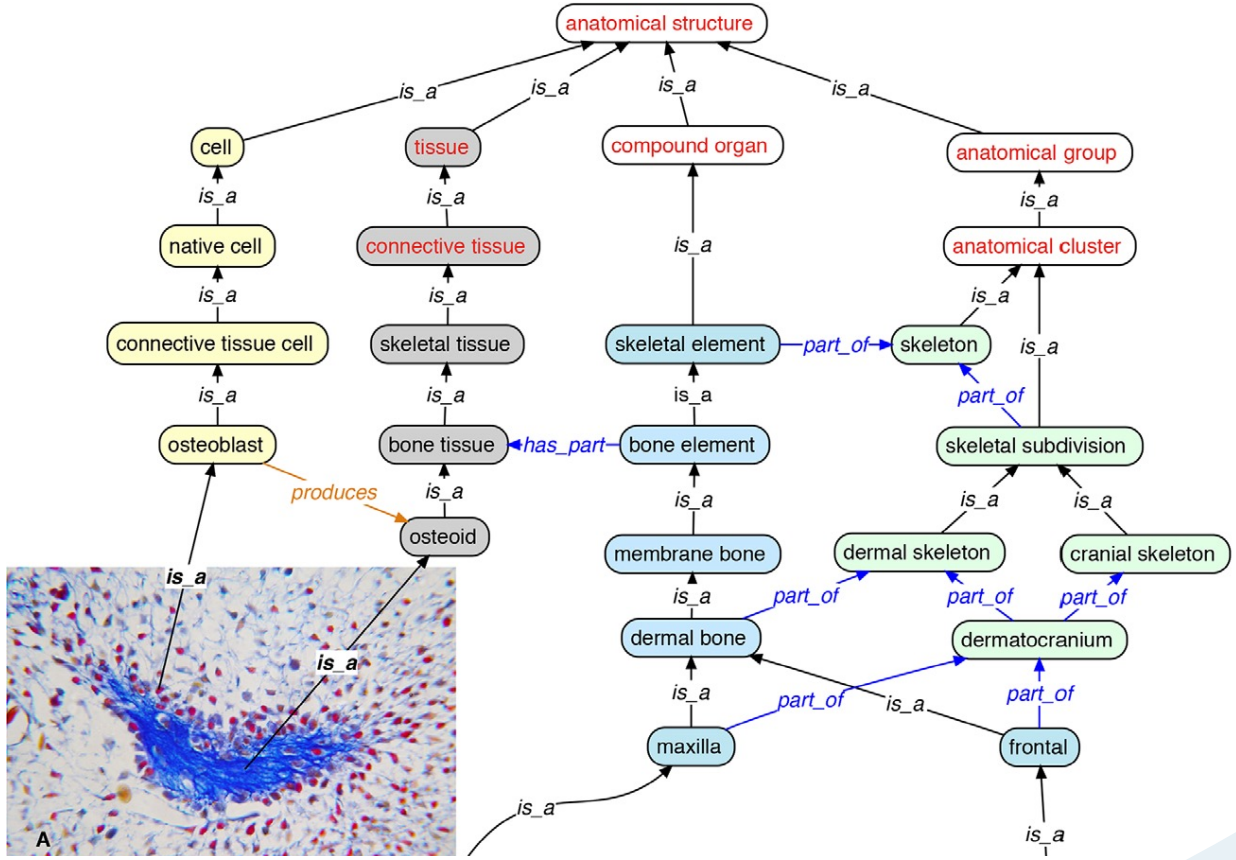
NOT COMPUTER PARSABLE

Carl von Linné, 1775

- **Rarely reused**
- **Not computer-parsable**
- **This precludes comparative phenomics in broad sense**

Ontology-based phenotype descriptions

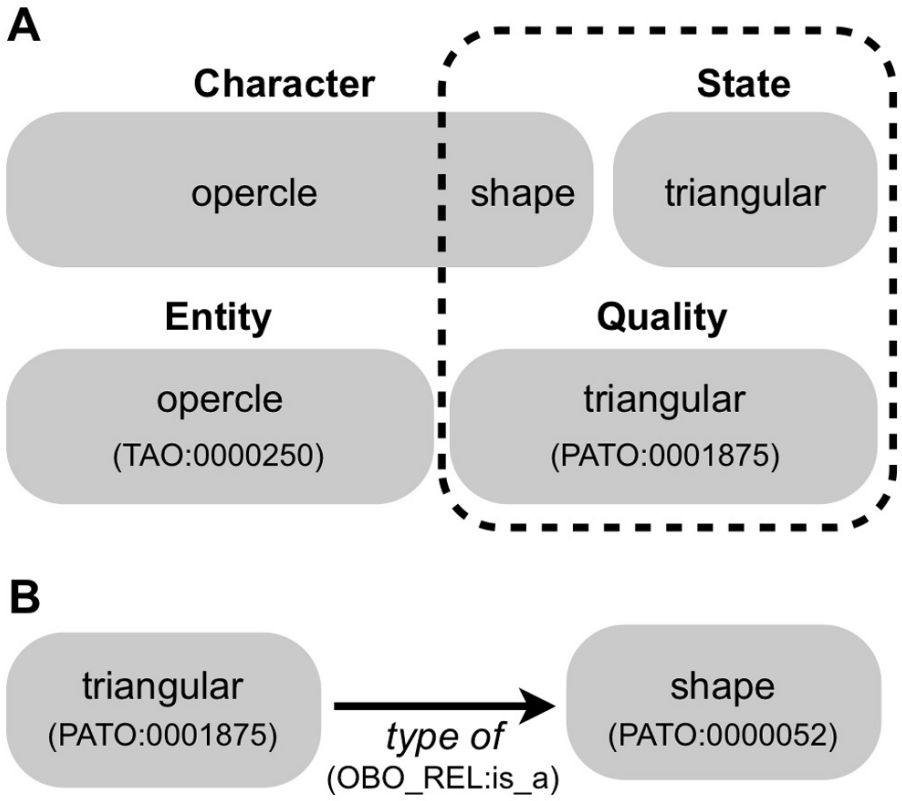
- Ontologies are networks representing our knowledge
- Part of Artificial Intelegence toolbox
- Ontology-driven approach alllows computer parsable species descriptions
- Ontology-based descriptions is an active area of research

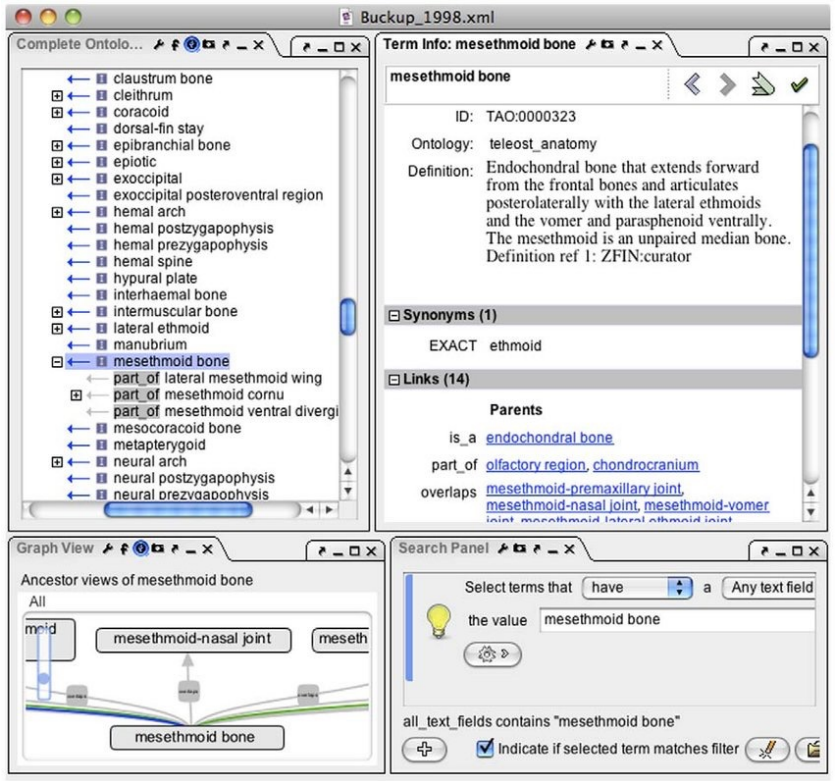


Phenoscape Project

PHENOSCAPE

https://phenoscape.org





Balhoff, James P., et al. "Phenex: ontological annotation of phenotypic diversity." PloS one 5.5 (2010): e10500.

Balhoff, James P., et al. "A semantic model for species description applied to the ensign wasps (Hymenoptera: Evaniidae) of New Caledonia." Systematic Biology 62.5 (2013): 639-659.

Phenoscape Project: character annotation

PATO terms used to annotate systematic characters

Below is a simplified graph showing a subset of quality terms from the Phenotype and Trait Ontology (PATO) used to curate systematic characters. Terms in orange represent attribute-level qualities.

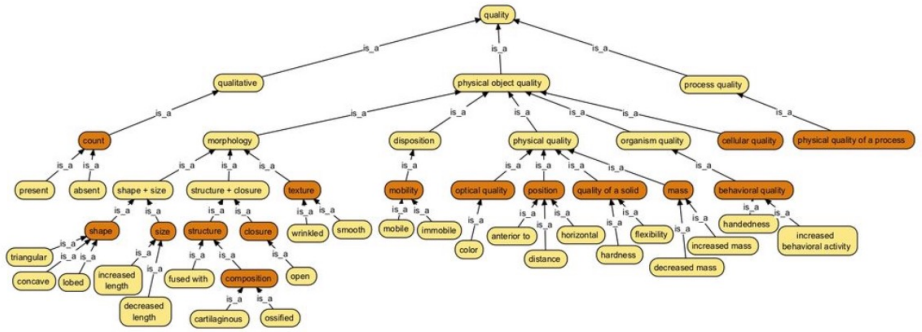


Image:1000 pixels

PATO attribute-level term	Synonyms	Examples of child terms
shape		triangular, lobed, concave, protruding into
position	placement, location	horizontal, anterior to
size		thin, large, decreased height
structure		porous, fused with
composition	composed of, content	ligamentous, cartilaginous
texture		smooth, wrinkled
optical quality		color, brightness
mass		increased mass
quality of a solid		flexible, hard
mobility		mobile, immobile
closure		open, closed
behavioral quality		
count	number, amount	present, absent

Annotation Specificity

Annotations are made using the most specific term or post-composition possible for Entity and Quality. This may not be possible for some characters involving very complex phenotypes, particularly those describing shape variation. In these cases, annotations are made to 'attribute' level PATO terms (e.g., PATO: shape).

Character Annotation Examples

The following are examples of character types commonly encountered in the systematic literature and how we annotate them using the Entity-Quality (EQ) model in Phenex. A character state can often require multiple EQs to capture all of the anatomical details described by an author. For example, a state may describe "premaxillary teeth round and multicuspidate." This is represented with the following two phenotypes:

PHENOSCAPE

https://wiki.phenoscape.org/wiki/Guide_to_Character_Annotation

The First Programming Language For Describing Species

Main aim

- provide a simple way to write semantic descriptions for species
- explores instance-based approach that is computationally cheaper

https://github.com/sergeitarasov/PhenoScript/wiki

From textual description to Species Programming

Traditional Description

Species description

***Parachorius semsanganus* Tarasov & Keith sp. n.**
urn:lsid:zoobank.org:act:E79FA444-725E-40E6-986A-9FB48B7D232A
http://species-id.net/wiki/Parachorius_semsanganus
Figs 1–5

Type locality. Laos, Xieng Khouang prov., Phou Sane Mt.

Material examined. *Holotype* (NHMB), male bearing the following labels:

LAOS-NE, Xieng Khouang prov., 19°38.20'N 103°20.20'E, Phonsavan (30 km NE): PHOU SANE Mt., 1420 m, 10.-30.v.2009, D. Hauck leg.

NHMB Basel, NMPC Prague Laos 2009 Expedition: M. Brancucci, M. Geiser, Z. Kraus, D. Hauck, V. Kubáň

HOLOTYPE Parachorius semsanganus S. Tarasov & D. Keith det. 2011

Paratypes. 9♀, same data as holotype; 2♂, same data as holotype but Z. Kraus leg.; 4♂, 3♀, LAOS-NE, Xieng Khouang prov., 19°37-8.'N 103°20'E, 30 km NE Phonsavan: Ban Na, Lam Phou Sane Mt., 1300–1500 m, 10.-30.v.2009, M. Brancucci leg.

Description. Oval, convex, black, entirely shiny; mouthparts, antennae and legs red-brown. Dorsal body side covered with two types of punctures: larger (normal) punctures and very tiny punctures (which can be observed only under higher magnification of 40x or more) Dorsal and ventral body sides glabrous. Length 8.2–10.6

Male (Fig. 4). Head flat, punctation fine; anterior margin notched medially; notch delimited by 2 prominent triangular obtuse teeth; clypeus laterad of each tooth very slightly notched; eyes completely divided by canthus into lower and upper lobes; lower lobes significantly larger than upper ones; genae and clypeus not distinctly separated from frons; genae rounded and protruding; antennae with 9 segments, antennal club with 3 segments.

Phenoscript

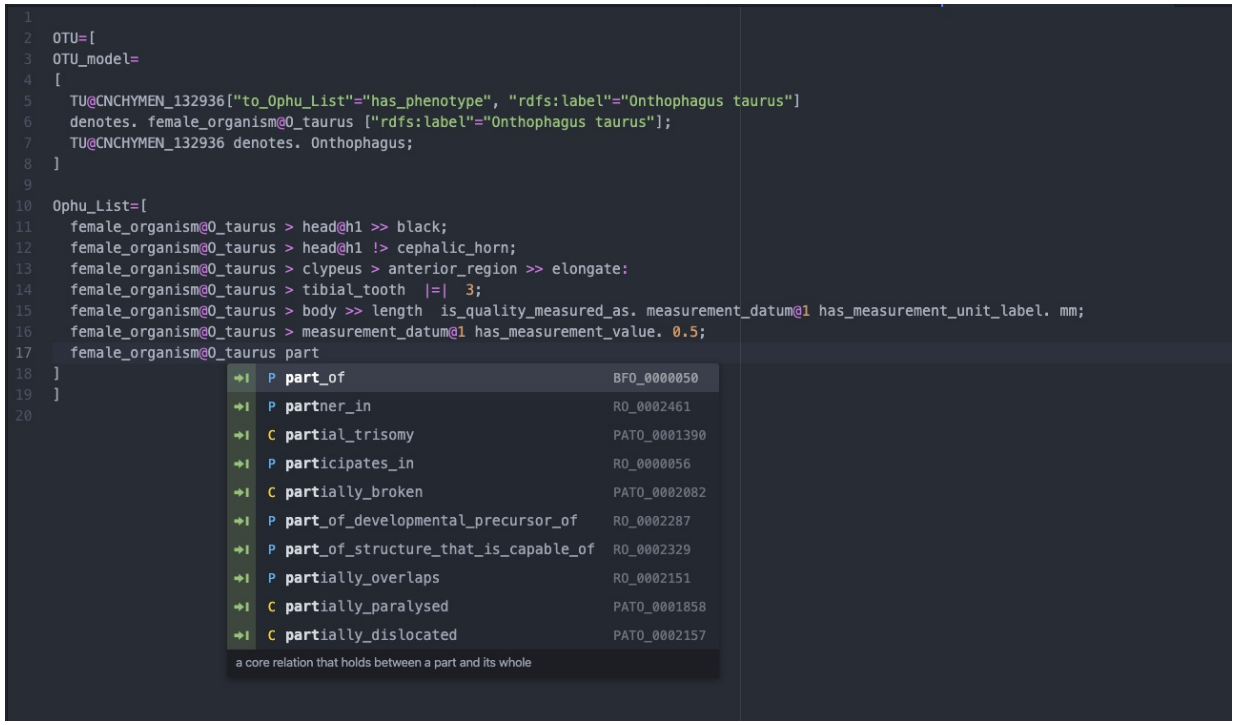
```
OTU=[
OTU_model=
[
  TU@CNCHYMEN_132936["to_Ophu_List"="has_phenotype", "rdfs:label"="Onthophagus taurus"]
  denotes. female_organism@O_taurus ["rdfs:label"="Onthophagus taurus"];
  TU@CNCHYMEN_132936 denotes. Onthophagus;
]

Ophu_List=[
  female_organism@O_taurus > head@h1 >> black;
  female_organism@O_taurus > head@h1 !> cephalic_horn;
  female_organism@O_taurus > clypeus > anterior_region >> elongate:
  female_organism@O_taurus > tibial_tooth  |=|  3;
  female_organism@O_taurus > body >> length  is_quality_measured_as. measurement_datum@1 has_measurement_unit_label. mm;
  female_organism@O_taurus > measurement_datum@1 has_measurement_value. 0.5;
  female_organism@O_taurus part
]
]
```

```
P part_of                                  BFO_0000050
P partner_in                               RO_0002461
C partial_trisomy                          PATO_0001390
P participates_in                          RO_0000056
C partially_broken                         PATO_0002082
P part_of_developmental_precursor_of       RO_0002287
P part_of_structure_that_is_capable_of     RO_0002329
P partially_overlaps                       RO_0002151
C partially_paralysed                      PATO_0001858
C partially_dislocated                     PATO_0002157
a core relation that holds between a part and its whole
```

From textual description to Species Programming

- PhS parser and compiler are written in R & Python
- PhS offers a simple syntax to write semantic statements and export them into OWL (Web Ontology Language)
- It’s like constructing a semantic graph
- Inspired by DOT language
- PhS does not run semantic inferences
- OWL statements can be subjected to semantic reasoning for automatic comparison and construction of character matrices (still underexplored field)

Phenoscript



PhenoScript Syntax: writing phenotype semantic graphs

Atom Editor

```
OTU=[
OTU_model=
[
  TU@CNCHYMEN_132936["to_Ophu_List"="has_phenotype", "rdfs:label"="Onthophagus taurus"]
  denotes. female_organism@O_taurus ["rdfs:label"="Onthophagus taurus"];
  TU@CNCHYMEN_132936 denotes. Onthophagus;
]

Ophu_List=[
  female_organism@O_taurus > head@h1 >> black;
  female_organism@O_taurus > head@h1 !> cephalic_horn;
  female_organism@O_taurus > clypeus > anterior_region >> elongate:
  female_organism@O_taurus > tibial_tooth  |=|  3;
  female_organism@O_taurus > body >> length  is_quality_measured_as. measurement_datum@1 has_measurement_unit_label. mm;
  female_organism@O_taurus > measurement_datum@1 has_measurement_value. 0.5;
  female_organism@O_taurus part
]
]
```

P	**part**_of	BFO_0000050
P	**part**ner_in	RO_0002461
C	**part**ial_trisomy	PATO_0001390
P	**part**icipates_in	RO_0000056
C	**part**ially_broken	PATO_0002082
P	**part**_of_developmental_precursor_of	RO_0002287
P	**part**_of_structure_that_is_capable_of	RO_0002329
P	**part**ially_overlaps	RO_0002151
C	**part**ially_paralysed	PATO_0001858
C	**part**ially_dislocated	PATO_0002157

a core relation that holds between a part and its whole

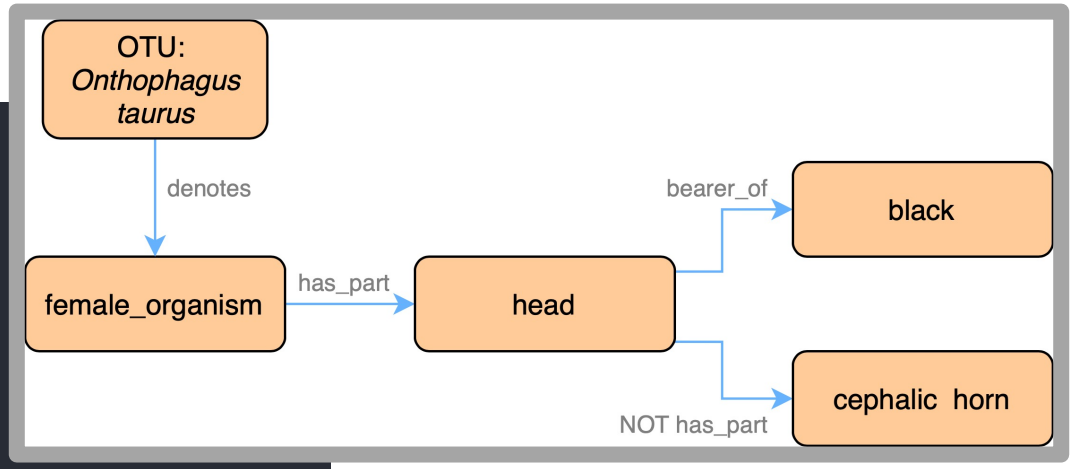
- Atom Editor
- Syntax Highlight
- Aliases (>,>>…) for commonly used terms
- Snippets to guide term selection

PhenoScript Syntax: writing phenotype semantic graphs

```
OTU=[
OTU_model=
[
  TU@CNCHYMEN_132936["to_Ophu_List"="has_phenotype", "rdfs:label"="Onthophagus taurus"]
  denotes. female_organism@O_taurus ["rdfs:label"="Onthophagus taurus"];
  TU@CNCHYMEN_132936 denotes. Onthophagus;
]

Ophu_List=[
  female_organism@O_taurus > head@h1 >> black;
  female_organism@O_taurus > head@h1 !> cephalic_horn;
  female_organism@O_taurus > clypeus > anterior_region >> elongate:
  female_organism@O_taurus > tibial_tooth  |=|  3;
  female_organism@O_taurus > body >> length  is_quality_measured_as. measurement_datum@1 has_measurement_unit_label. mm;
  female_organism@O_taurus > measurement_datum@1 has_measurement_value. 0.5;
  female_organism@O_taurus part
]
]
```

P	**part_of**	BFO_0000050
P	**partner_in**	RO_0002461
C	**partial_trisomy**	PATO_0001390
P	**participates_in**	RO_0000056
C	**partially_broken**	PATO_0002082
P	**part_of_developmental_precursor_of**	RO_0002287
P	**part_of_structure_that_is_capable_of**	RO_0002329
P	**partially_overlaps**	RO_0002151
C	**partially_paralysed**	PATO_0001858
C	**partially_dislocated**	PATO_0002157

a core relation that holds between a part and its whole



- Atom Editor
- Syntax Highlight
- Aliases (>,>>…) for commonly used terms
- Snippets to guide term selection

Integrating TW & PhS

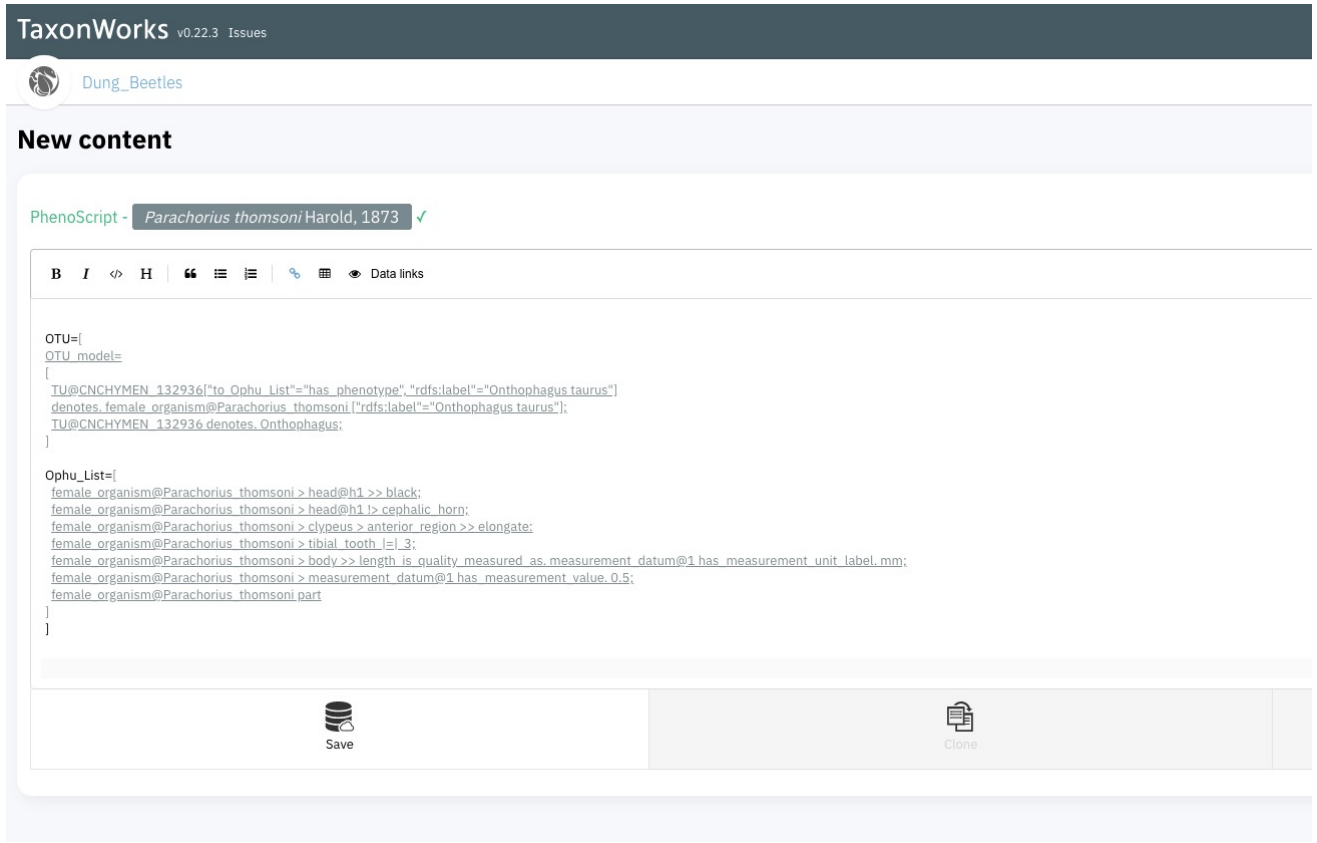
(currently under development)

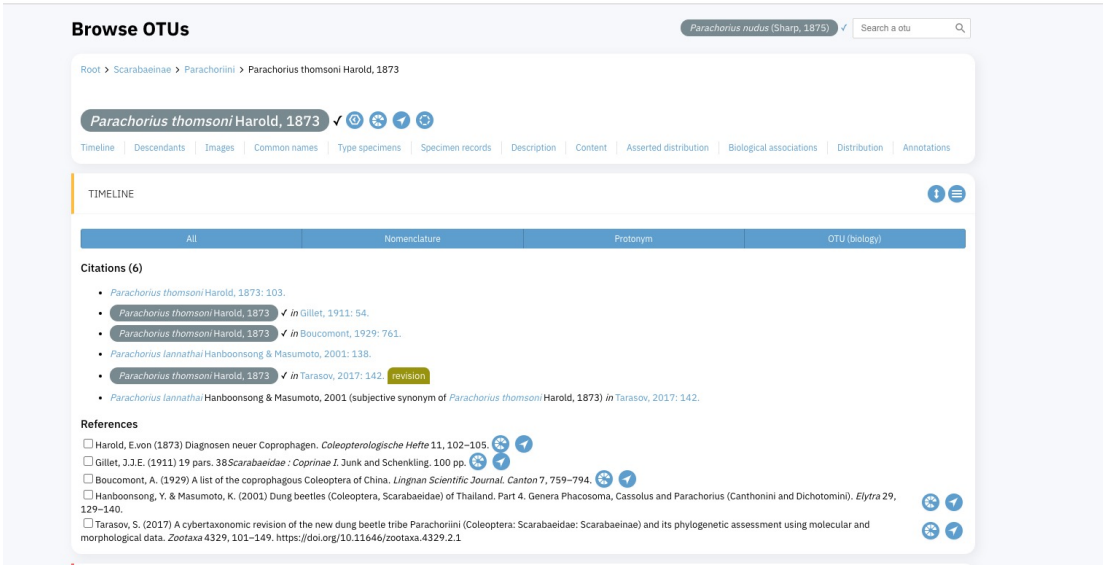
Expectations:

- Simple interaction between ontologies and users
- Facilitate automated comparative phenomics
- Link OTUs with phenotypes
- Facilitate reuse of the description across various biological fields

Integrating

TW Content Editor tab





Acknowledgement

Thanks to:

- Mat Yoder
- Dmitry Mozzherin

for a fruitful discussion about integration of TW &PhenoScript

- and Academy of Finland for funding

Relative Materials:

Watch Recent talk by *Jennifer Giron*
“Unlocking the power of morphological data with the Coleoptera Anatomy Ontology”
https://www.youtube.com/watch?v=Nwpk3pBHn1U