

Views Controlled Vocabularies testing notes

2022-02-09

Introduction

Biodiversity Information Standards (TDWG) is an international organization that develops standards for sharing biodiversity metadata. Audubon Core is the TDWG standard for describing multimedia resources.

The Audubon Core properties *subjectPart* and *subjectOrientation* are used to categorize images for searching and display. Our working group is standardizing the values that can be provided for these two terms. Each possible value is considered a *concept* and can be denoted either by a *controlled value string* or by an *internationalized resource identifier (IRI)*. The controlled value string resembles an English word or words (e.g. femaleCone) and is easy for humans to type. The IRI (e.g. <http://rs.tdwg.org/acpart/values/p0007>) is globally unique and can be used for machine processing. Because a controlled value string and an IRI represent the same concept, they are interchangeable.

We are seeking test implementers who are willing to try to apply the two controlled vocabularies to a small, representative selection of their images, then report a summary of their experience to the task group. The feedback we receive will be compiled into a report to be published in TDWG's journal *Biodiversity Information Science and Standards (BISS)*, with test implementers as co-authors.

Reference resources

- subjectPart controlled vocabulary:
<https://github.com/tdwg/ac/blob/master/views/subjectPart.md>
- subjectOrientation controlled vocabulary:
<https://github.com/tdwg/ac/blob/master/views/subjectOrientation.md>
- Categorized lists:
 - a. https://github.com/tdwg/ac/blob/master/views/code/orient_collections.md
 - b. https://github.com/tdwg/ac/blob/master/views/code/part_collections.md
- Feedback form questions (PDF, a Google Form will be used at the end of testing):
https://github.com/tdwg/ac/blob/master/views/views_controlled_vocabularies_implementation_report.pdf

Additional resources (CSV and JSON files) are linked in the text below.

Questions? Email steve.baskauf@vanderbilt.edu or JENNIFER.GIRON@ttu.edu

Recommended process

Most test implementers will probably want to perform manual categorization, where a human looks at an image and enters a controlled value string for subjectPart and subjectOrientation for each image into a spreadsheet.

In this scenario, we recommend the following process:

- Begin by the testing team examining a small number of images (5 to 10) and working together to assign values for the two terms. The purpose of this stage is to become familiar with the reference materials and to clarify the task. The team should also review the feedback form to be aware of the kinds of questions they will be answering in their report.
- One or more testers should then categorize a larger number of images (20 to 30) that cover the range of subjectParts and subjectOrientations included in their collection.
- After the testers complete their work, another member of the testing team should review the results to identify any cases where there is disagreement.
- The tester and reviewer can then fill out the feedback form.

For more details and examples, see the "1. Manual Testing" section below.

Some test implementers may want to experiment with machine-aided entry or scripted conversion of existing data. In those cases, see the detailed instructions in parts 2 and 3 of the "Types of Testing" section below.

Terminology

View

A "view" is a non-technical term for a description of the organism part and orientation of an image.

Broader concepts

Some narrow concepts are linked to broader concepts. For example, left and right subjectOrientations have the broader concept "lateral". Because of this relationship, it is possible to infer that a "right" orientation is also a "lateral" orientation. Generally, it is preferable to provide the narrowest concept possible unless it is not possible for the user to differentiate between the narrower categories and the user needs to select the broader category.

To discover the broader relationships

- look in the skos_broader column of the two controlled value spreadsheets https://github.com/tdwg/ac/blob/master/views/code/subjectPart/subjectPart_cv.csv and https://github.com/tdwg/ac/blob/master/views/code/subjectOrientation/subjectOrientation_cv.csv
- look in the "Has broader concept" term metadata field in the two term lists <https://github.com/tdwg/ac/blob/master/views/subjectPart.md> and <https://github.com/tdwg/ac/blob/master/views/subjectOrientation.md>
- look for skos:broader values in the machine-readable JSON-LD at <https://tdwg.github.io/rs.tdwg.org/cvJson/acpart.json> and <https://tdwg.github.io/rs.tdwg.org/cvJson/acorient.json>

Regions of Interest

A region of interest (ROI) is a designated part of a media item. Each ROI in an image can be assigned its own values for subjectPart and subjectOrientation. Thus an image with several regions of interest may have multiple records for these two terms. Here is an example (region of interest outlined in yellow):

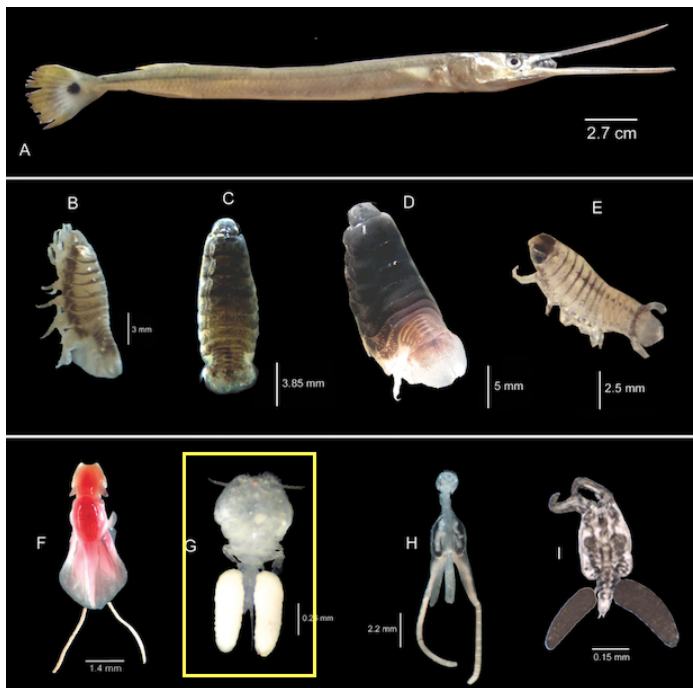


Image IRI: <https://zenodo.org/record/959321>

Image file name: big_34616.jpg

ROI IRI: <https://zenodo.org/record/959321#g>

subjectPartLiteral: entireOrganism

subjectOrientationLiteral: dorsal

Types of testing.

1. Manual entry.

The users examine an image and use the lists or spreadsheets of concepts to determine the controlled value appropriate for the image, then enter the value in a spreadsheet containing metadata about that image. Here are some examples:

Image of a live fish



The user refers to the spreadsheet of controlled values for subjectPart at https://github.com/tdwg/ac/blob/master/views/code/subjectPart/subjectPart_cv.csv and selects the appropriate controlled value string for an entire organism: "entireOrganism".

	A	B	J	type
1	term_localName	label	controlled_value_string	
2	p	subject part concept scheme		http://www.w3.org/2
3	p0000	unspecified part	unspecifiedPart	http://www.w3.org/2
4	p0001	entire organism	entireOrganism	http://www.w3.org/2
5	p0002	bark	bark	http://www.w3.org/2
6	p0003	twig	twig	http://www.w3.org/2
7	p0004	stem	stem	http://www.w3.org/2
8	p0005	leaf	leaf	http://www.w3.org/2
9	p0006	strobilis (cone)	strobilis	http://www.w3.org/2
10	p0007	inflorescence	inflorescence	http://www.w3.org/2
11	p0008	fruit	fruit	http://www.w3.org/2
12	p0009	seed	seed	http://www.w3.org/2
13	p0010	male cone	maleCone	http://www.w3.org/2
14	p0011	female cone	femaleCone	http://www.w3.org/2
15	p0012	flower	flower	http://www.w3.org/2
16	p0013	head	head	http://www.w3.org/2
17	p0014	berry	berry	http://www.w3.org/2

The user enters this value into the spreadsheet:

	A	B	C	D
1	image	ac:subjectPartLiteral	ac:subjectOrientationLiteral	
2	https://zenodo.org/record/907755	entireOrganism		
3				
4				

The user refers to the spreadsheet of controlled values for subjectOrientation at https://github.com/tdwg/ac/blob/master/views/code/subjectOrientation/subjectOrientation_cv.csv and selects the appropriate controlled value string for the right side of the organism: "right".

subjectOrientati

Liberation Sans 10 pt B / U A

A7:AMJ7 fx Σ = r0004

	A	B	J	
1	term_localName	label	controlled_value_string	type
2	r	subject orientation concept scheme		http://www.w
3	r0000	unspecified orientation	unspecifiedOrientation	http://www.w
4	r0001	anterior side	anterior	http://www.w
5	r0002	posterior side	posterior	http://www.w
6	r0003	lateral side	lateral	http://www.w
7	r0004	right side	right	http://www.w
8	r0005	left side	left	http://www.w
9	r0006	dorsal side	dorsal	http://www.w
10	r0007	ventral side	ventral	http://www.w
11	r0008	adaxial side	adaxial	http://www.w
12	r0009	abaxial side	abaxial	http://www.w
13	r0010	apical side	apical	http://www.w
14	r0011	basal side	basal	http://www.w
15				

The user enters this value into the spreadsheet:

Liberation Sans 10 pt B / U A

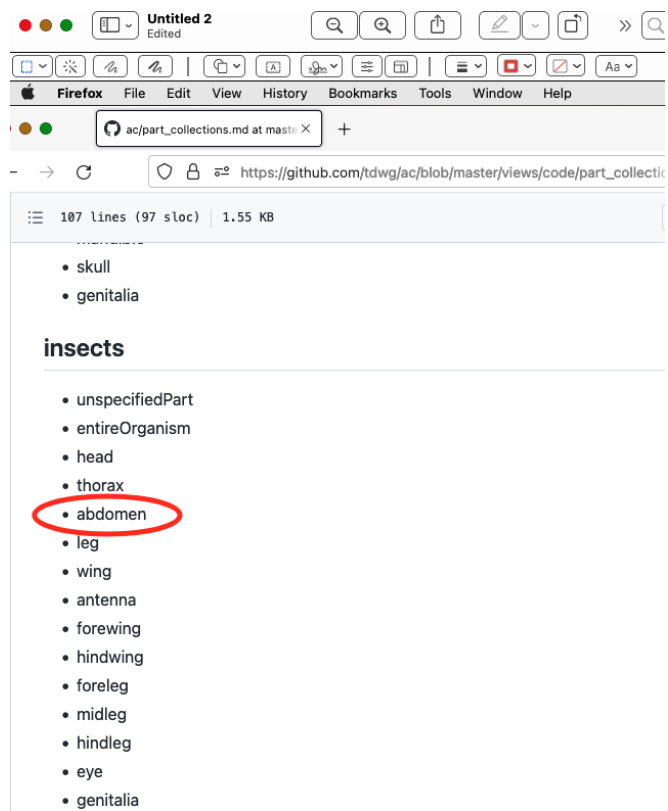
C2 fx Σ = right

	A	B	C
1	image	ac:subjectPartLiteral	ac:subjectOrientationLiteral
2	https://zenodo.org/record/907755	entireOrganism	right
3			
4			

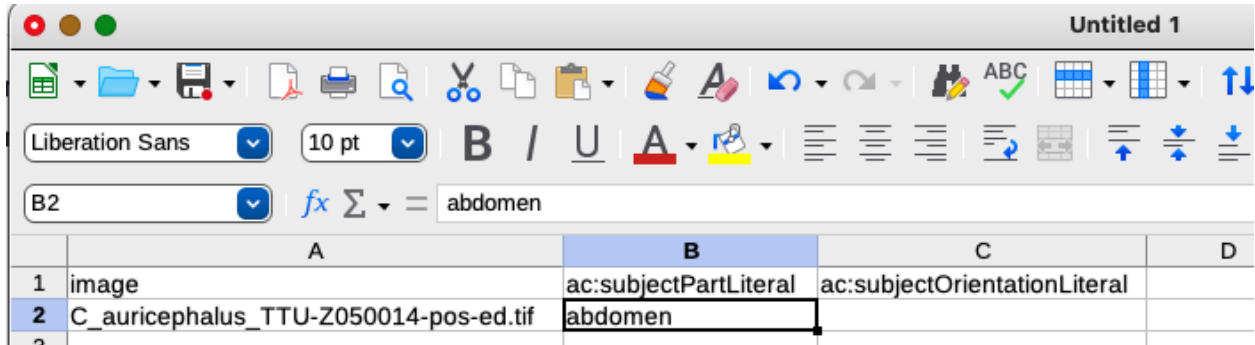
Image of an insect specimen



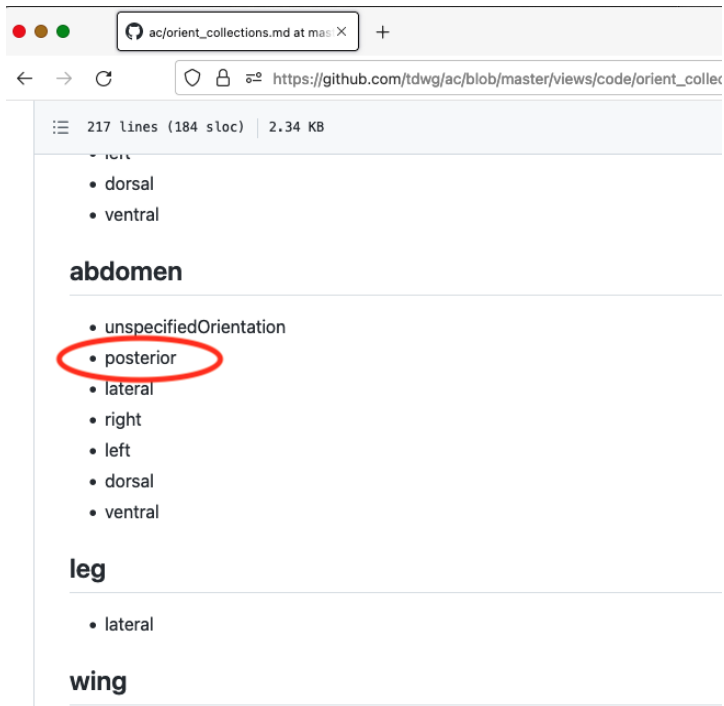
The user refers to the web page listing controlled value strings appropriate for different organism groups at https://github.com/tdwg/ac/blob/master/views/code/part_collections.md and finds the list for insects.



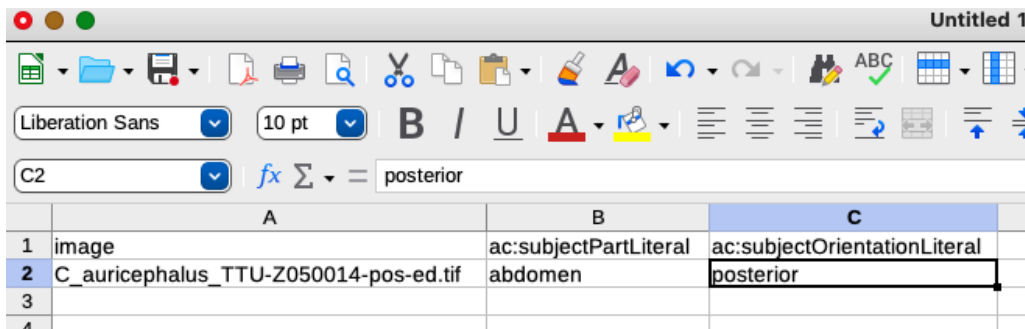
The user selects the value "abdomen" and enters it into a spreadsheet.



The user refers to the web page listing controlled value strings appropriate for different subject parts at https://github.com/tdwg/ac/blob/master/views/code/orient_collections.md and finds the list for abdomen:



The user selects the value "posterior" and enters it into a spreadsheet.



2. Machine-guided entry

The users examine images that are presented by a content management system (CMS). The CMS guides the user to select concepts that are appropriate for a particular organism group or subjectPart using dropdown menus or some other system. Here is an example:

The screenshot displays the Bioimages Collection Manager - Data Entry interface. The main window shows a large image of a yellow flower. To the right of the image is a metadata form with fields for Organism ID, Organism Name, and various remarks. Below the main image is a search and selection interface with a red box highlighting the 'Specimen Part' dropdown menu, which is open and showing 'inflorescence' selected. To the right of the main form is a grid of thumbnail images. At the bottom right, there is a table with columns for Selected Filename, Image Date, Time, and TZ, containing data for a specific image.

Selected Filename	Image Date	Time	TZ
e5-flyellow-close39194.jpg	2005-04-10	14:32:14	-05:00

The user selected the organism group label "herbaceous angiosperms" and the subjectPart collections data were used to determine which subjectParts were appropriate for that organism group. The CMS used this selection to assign the value

`ac:subjectPart = http://rs.tdwg.org/acpart/values/p0007`

to assign the IRI for "inflorescence" to the property `ac:subjectPart`, whose values are intended to be IRIs. The user would then be presented with a picklist of subjectOrientations that are appropriate for inflorescences, based on the subjectOrientation collections data. In this case, the user would select "lateral" from the dropdown and the CMS would assign the value

`ac:subjectOrientation = http://rs.tdwg.org/acorient/values/r0003`

since the property `ac:subjectOrientation` is expected to have an IRI value.

The following JSON files may be useful for setting up the CMS to generate the dropdowns:

subjectPart collection providing parts appropriate for different organism groups:

https://tdwg.github.io/rs.tdwg.org/cvJson/acpart_collection.json

metadata about subjectPart concepts (labels, definitions, controlled value strings, broader relationships, ontology links): <https://tdwg.github.io/rs.tdwg.org/cvJson/acpart.json>

subjectOrientation collection providing orientations appropriate for different organism parts:

https://tdwg.github.io/rs.tdwg.org/cvJson/acorient_collection.json

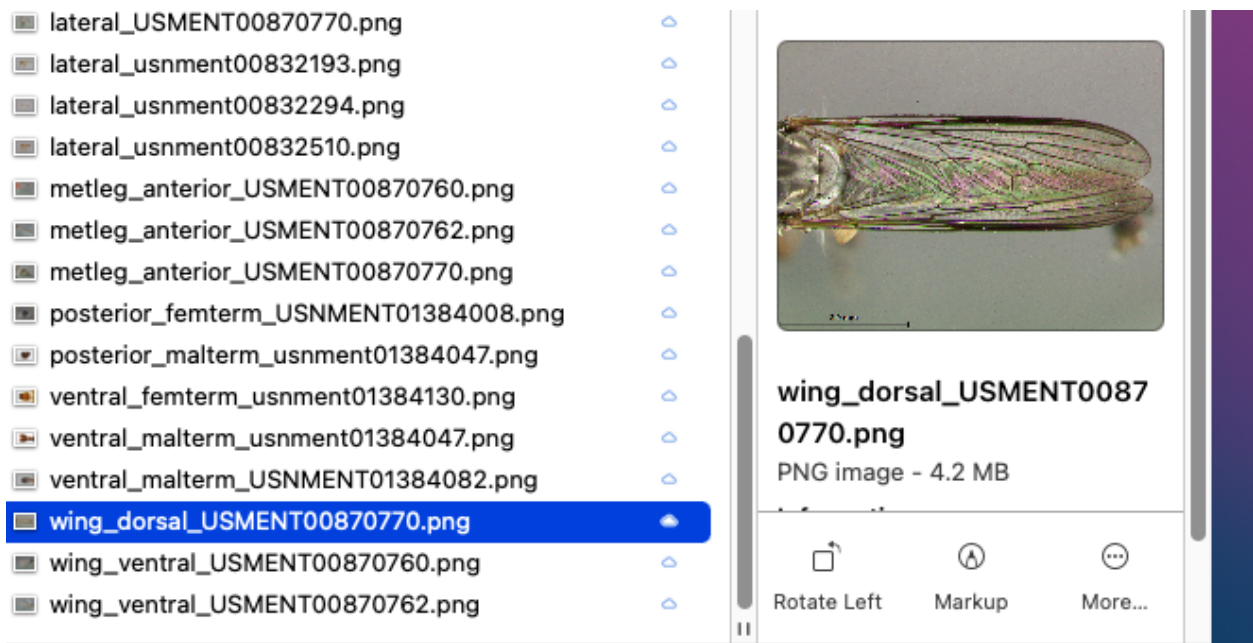
metadata about subjectOrientation concepts (labels, definitions, controlled value strings, broader relationships, ontology links): <https://tdwg.github.io/rs.tdwg.org/cvJson/acorient.json>

3. Machine processing

Data managers map the controlled value concepts to existing text-based schemes (e.g. text embedded in filenames). Values for subjectPart and subjectOrientation are generated for images using these mappings.

Example:

The filenames of these images embed information about the body parts:



The screenshot shows a file explorer interface. On the left, a list of image files is displayed, each with a small thumbnail icon and a blue arrow icon to its right. The files are:

- lateral_USMENT00870770.png
- lateral_usnment00832193.png
- lateral_usnment00832294.png
- lateral_usnment00832510.png
- metleg_anterior_USMENT00870760.png
- metleg_anterior_USMENT00870762.png
- metleg_anterior_USMENT00870770.png
- posterior_femterm_USNMENT01384008.png
- posterior_malterm_usnment01384047.png
- ventral_femterm_usnment01384130.png
- ventral_malterm_usnment01384047.png
- ventral_malterm_USNMENT01384082.png
- wing_dorsal_USMENT00870770.png** (highlighted in blue)
- wing_ventral_USMENT00870760.png
- wing_ventral_USMENT00870762.png

On the right, a preview of the selected image is shown. The image is a dorsal view of a fly wing, showing the intricate venation pattern. Below the image, the filename **wing_dorsal_USMENT00870770.png** is displayed, along with the text "PNG image - 4.2 MB". At the bottom of the preview area, there are three icons: a square with a curved arrow for "Rotate Left", a person icon for "Markup", and a three-dot menu icon for "More...".

The file name wing_dorsal_USMENT00870770.png can be split based on underscore positions:

part: "wing"

orientation: "dorsal"

rest of string: USMENT00870770.png

Mappings:

"wing" = <http://rs.tdwg.org/acpart/values/p0017>

"dorsal" = <http://rs.tdwg.org/acorient/values/r0006>

Metadata for image:

filename: wing_dorsal_USMENT00870770.png

ac:subjectPart: <http://rs.tdwg.org/acpart/values/p0017>

ac:subjectOrientation: <http://rs.tdwg.org/acorient/values/r0006>

Feedback

After users perform the tests, someone from the organization needs to assess the results (were human-selected choices correct? did software correctly assign values?). The results need to be reported using the feedback form: <https://forms.gle/6dDb3VAwV4CDHRZQ6>