Audubon Core Maintenance Group
Online meeting
2021-04-14 20:00 UTC

Meeting notes:

Participating: Steve Baskauf, Dan Stowell, Roger Hyam, Kate Webbink, Vijay Barve
Regrets: Rich Pyle

Notes taken during the meeting are in red.

I. 3D Task Group report

II. Views Task Group report from Steve

    A. Progress on addressing remaining issues blocking testing, but got bogged down on appropriate granularity for grouping of organism parts and is a "skin" part the same as a whole organism part. Will continue discussion next week.

III. Steve report on discussion with Tanya Berger-Wolf (Wildbook project for ID whales,zebras, etc.) about interface between machine learning and segmentation/fragmentation/views

    A. Importance of clear definitions (e.g. "segmentation", "annotation", "ontology")

    B. Our work could be very useful for machine learning community

    C. Willingness of developers to spend time on standards development?

IV. ~~Segmentation~~. Fragmentation? Decided best term to use is "region of interest" (ROI)

    A. Define atomic terms for all of the pieces (from last meeting). Steve will demo with Python script examples:

        1. Specify how to assemble the pieces into valid W3C Media Fragments

        2. Document how to use the Media Fragments in annotations in general and IIIF manifests in particular

        3. (via Dan): Optionally, we could later “extend” the selector format, by providing our own conformsTo=”*” (e.g. to allow viewer implementations that support frequency bounds)

    B. Still images (Steve):

        1. For still images, everyone uses x,y,w,h . See **Appendix 2** for usage examples. No "standard" vocabulary, but similar terms used: "x", "y", "width", "height". Generally appears to be assumed that the units are pixels, not percents.

        2. To support media fragments for abstract media items (not suitable for direct generation of media fragments as in https://www.w3.org/TR/media-frags/#naming-space)

            a) xFrac

            b) yFrac

            c) widthFrac

            d) heightFrac

3. Examples in https://github.com/tdwg/ac/tree/master/fragments/transformation_examples directory of GitHub repo. Code in https://github.com/tdwg/ac/blob/master/fragments/transformation_examples/transformations.ipynb
    a) Delineate image regions using VIA web app.
    b) Create fractional values (above) using absolute pixel values of Best Quality service access point.
    c) For BQ, SAP, generate: media fragments, IIIF server URLs, W3C Web Annotations, IIIF manifest with fragments designated using "describing" W3C annotations.
    d) Example using Digirati's annotation viewer, Canvas Panel: https://canvas-panel.digirati.com/#/examples/fullpage?manifest=https://stephenwf.github.io/ocean-liners.json
4. Probably need to create term ~~ac:hasFragment or~~ ac:hasROI (analogous to ac:hasServiceAccessPoint) to support linking to fragments, also I suppose add a ~~ac:Fragment~~ ac:RegionOfInterest class to explain what we consider a "fragment" to be.
5. Example JSON-LD serializations for linking service access points and fragments to abstract media items (ultimately the source for generating the resources demonstrated above). Use cases:
    a) Multiple occurrences documented within a single media item: https://github.com/tdwg/ac/blob/master/fragments/transformation_examples/image_graph_hess.json **Issue that came up here: I think we should create ac:associatedOccurrenceReference and deprecate the separate specimen and observation terms.** Note added after meeting: solicited advice about this from the Darwin Core Maintenance Group https://github.com/tdwg/dwc/issues/333
    b) Fragments within a published figure image depict different species and orientations (from Views Task Group): https://github.com/tdwg/ac/blob/master/fragments/transformation_examples/image_graph_zookeys.json
6. Action: Steve to draft term proposals for the four atomic terms.

C. Sound (Dan)
1. "startTime" and "endTime" terms to add - see draft text in **Appendix 1**. Dan has considered two main sources (Timeline Ontology and w3c MF; see also older notes), and received extensive feedback from the bioacoustics-l mailing list about preferred formats. Dan has drafted a text based on bioacoustics-l feedback while sticking closely to w3c MF compatibility. The idea is at least partly that the fields can directly be filled into a w3c Media Fragment selector, without any calculation needed for reformatting, since the value formatting options are fully compatible.

Feedback included: discussion of whether non-UTC values were meaningful (e.g. if people clip subsections of audio); should the different time-formats be split into multiple terms (e.g. startTimeUTC startTimeSeconds).
Action: Dan to look into writing a proposal for startTime/endTime terms split into 4 terms corresponding to ISO8601/UTC and to seconds-from-beginning-of-media.

2. Related possibilities:
   a) Duration as an alternative to endTime: (not discussed, not needed) I think it's fair to allow both options, unless you think it will create unneeded complexity in the standard. The Timeline Ontology allows both: there's tl:duration. -- I propose we do not implement duration until the startTime and endTime terms have bedded down.
   b) Time regions expressed in terms of “samples”: only meaningful if mo:sample_rate is specified. I suggest we skip this; its main use is in high-resolution work, where we allow arbitrary sub-second precision instead. This was discussed briefly, not a priority.

V. Action items and timeframes for finishing example guides: Did not get to this item.
   A. Still images
   B. Sound
   C. Video

VI. Next meeting
   A. Date: May 12, 20:00 UTC.

Appendix 1: Defining startTime and endTime: Draft text adapted from w3c Media Fragments "4.2.1 Temporal Dimension"

Temporal regions are specified as an interval with a begin time and an end time. The interval is half-open: the begin time is considered part of the interval whereas the end time is considered to be the first time point that is not part of the interval. Temporal values are specified in one of the following formats:

(a) absolute offset relative to the beginning of the presentation (i.e. Normal Play Time RFC 2326), specified as seconds, with an optional fractional part to indicate milliseconds or finer. ~~*, or as colon-separated hours, minutes and seconds (again with an optional fraction). Minutes and seconds must be specified as exactly two digits, hours and fractional seconds can be any number of digits. The hours, minutes and seconds specification for NPT is a convenience only, it does not signal frame accuracy*~~*.*

(b) real-world clock time ISO 8601 timestamps, using UTC timezone, with an optional fractional part to indicate milliseconds or finer.

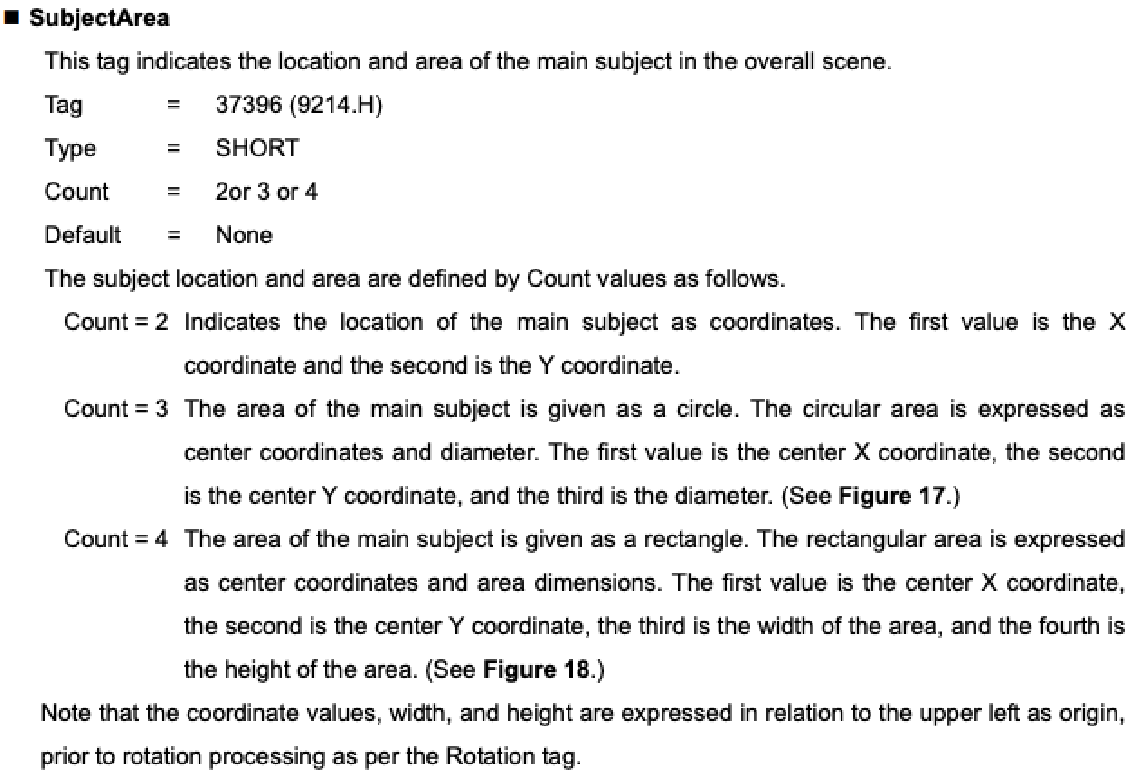
Begin and end times are always specified in the same format. The recommended default format is NPT seconds. There is no limit on the number of decimal places for the decimal fraction. Begin or end time may each be omitted, with omission implying the value cannot be determined (either because they are un-discernible or because they occur before/after the duration of the source media).

Appendix 2: Research on bounding box formats and metadata used in annotation

JPEG notes from http://www.cipa.jp/std/documents/e/DC-008-2012_E.pdf

Tests:
https://www.robots.ox.ac.uk/~vgg/software/via/via_demo.html
https://mirador-dev.netlify.app/__tests__/integration/mirador/
https://json-ld.org/playground/
https://zenodo.org/record/4683346
https://zenodo.org/record/4683346/files/trco--fr040529-17e5384.jpg
https://zenodo.org/api/iiif/v2/2180af36-6fea-424a-b6b4-b59f3c862c51:b535ccd4-2869-4e84-9e19-1df68c7abcb4:trco--fr040529-17e5384.jpg/full/856,/0/default.jpg

https://zenodo.org/record/959321



■ **SubjectArea**

This tag indicates the location and area of the main subject in the overall scene.

Tag	=	37396 (9214.H)
Type	=	SHORT
Count	=	2or 3 or 4
Default	=	None

The subject location and area are defined by Count values as follows.

Count = 2 Indicates the location of the main subject as coordinates. The first value is the X coordinate and the second is the Y coordinate.

Count = 3 The area of the main subject is given as a circle. The circular area is expressed as center coordinates and diameter. The first value is the center X coordinate, the second is the center Y coordinate, and the third is the diameter. (See **Figure 17**.)

Count = 4 The area of the main subject is given as a rectangle. The rectangular area is expressed as center coordinates and area dimensions. The first value is the center X coordinate, the second is the center Y coordinate, the third is the width of the area, and the fourth is the height of the area. (See **Figure 18**.)

Note that the coordinate values, width, and height are expressed in relation to the upper left as origin, prior to rotation processing as per the Rotation tag.

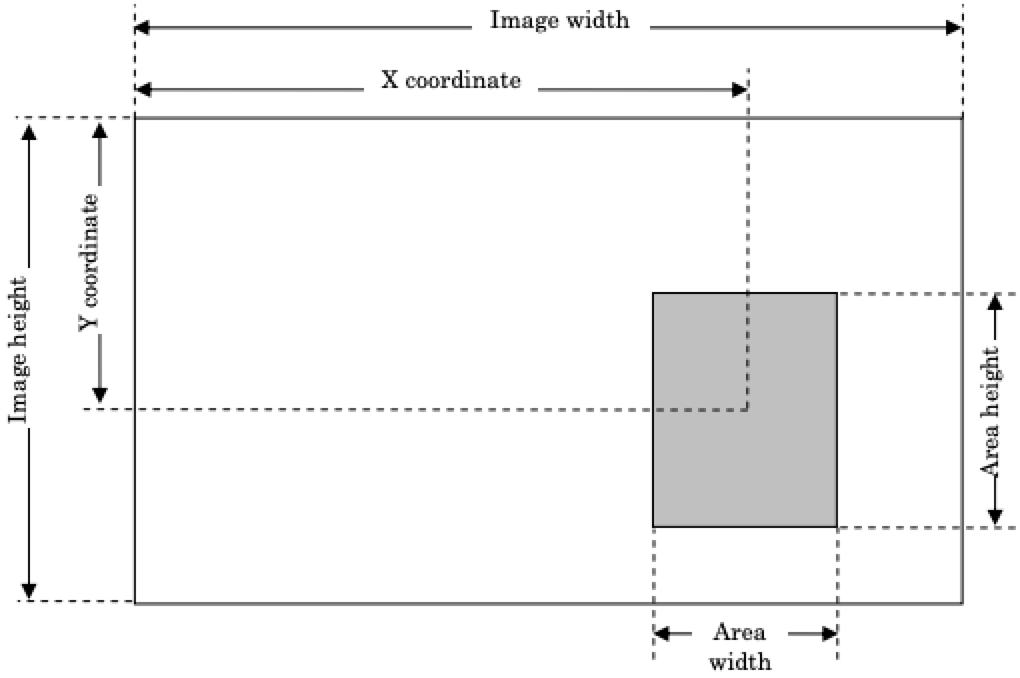


Figure 18 Rectangular area designation

58

Image annotation formats:
https://towardsdatascience.com/image-data-labelling-and-annotation-everything-you-need-to-know-86ede6c684b1

COCO results format for bounding boxes:
https://cocodataset.org/#format-results

```
[{
"image_id": int, "category_id": int, "bbox": [x,y,width,height], "score": float,
}]
```

Pascal VOC: Pascal VOC stores annotation in XML file

```
<annotation>
 <folder>Train</folder>
 <filename>01.png</filename>
 <path>/path/Train/01.png</path>
 <source>
```

```
   <database>Unknown</database>
 </source>
 <size>
   <width>224</width>
   <height>224</height>
   <depth>3</depth>
 </size>
 <segmented>0</segmented>
 <object>
   <name>36</name>
   <pose>Frontal</pose>
   <truncated>0</truncated>
   <difficult>0</difficult>
   <occluded>0</occluded>
   <bndbox>
     <xmin>90</xmin>
     <xmax>190</xmax>
     <ymin>54</ymin>
     <ymax>70</ymax>
   </bndbox>
 </object>
</annotation>
```

Yolo format stores data in files with numbers only.

Via project: https://gitlab.com/vgg/via

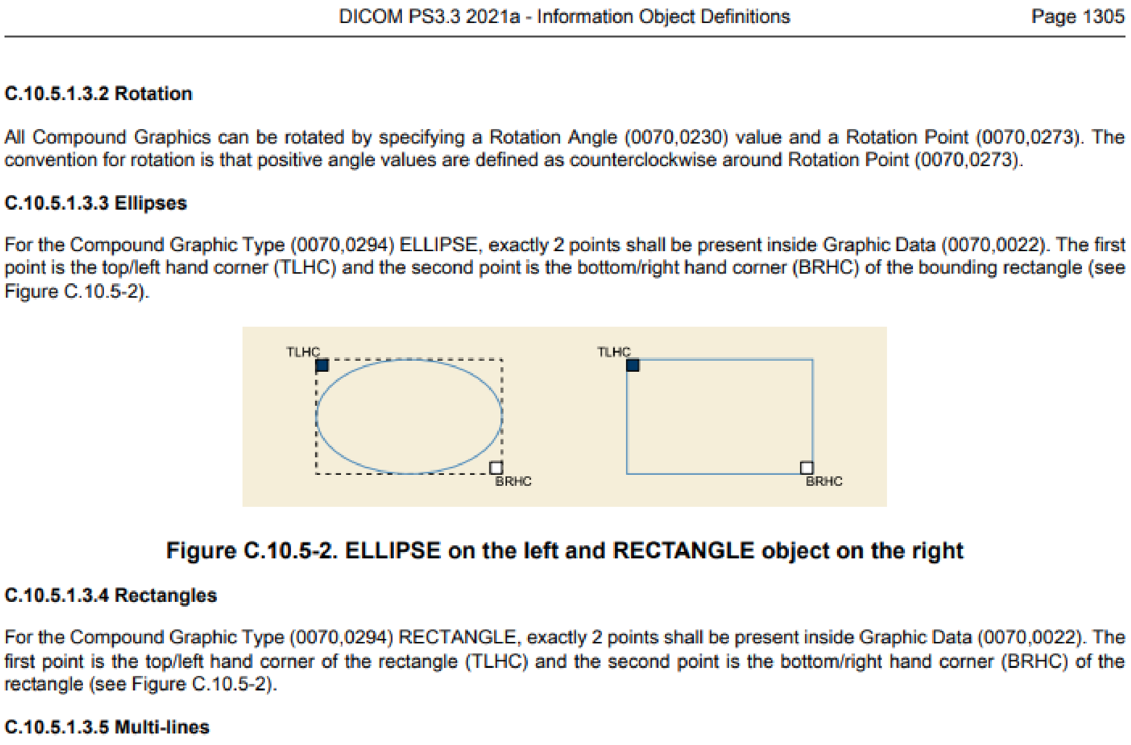
Output from rectangular annotation:

```
{
  "shape_attributes": {
    "name": "rect",
    "x": 59,
    "y": 75,
    "width": 420,
    "height": 230
  },
  "region_attributes": {
    "name": "not_defined",
    "type": "bird",
        "image_quality": {
           "good": true,
           "frontal": true,
           "good_illumination": true
        }
     }
  }
  ],
```

Python tool: https://github.com/tzutalin/labelImg outputs in YOLO format.

DICOM standard:

http://dicom.nema.org/medical/dicom/current/output/pdf/part03.pdf



DICOM PS3.3 2021a - Information Object Definitions Page 1305

C.10.5.1.3.2 Rotation

All Compound Graphics can be rotated by specifying a Rotation Angle (0070,0230) value and a Rotation Point (0070,0273). The convention for rotation is that positive angle values are defined as counterclockwise around Rotation Point (0070,0273).

C.10.5.1.3.3 Ellipses

For the Compound Graphic Type (0070,0294) ELLIPSE, exactly 2 points shall be present inside Graphic Data (0070,0022). The first point is the top/left hand corner (TLHC) and the second point is the bottom/right hand corner (BRHC) of the bounding rectangle (see Figure C.10.5-2).

Figure C.10.5-2. ELLIPSE on the left and RECTANGLE object on the right

C.10.5.1.3.4 Rectangles

For the Compound Graphic Type (0070,0294) RECTANGLE, exactly 2 points shall be present inside Graphic Data (0070,0022). The first point is the top/left hand corner of the rectangle (TLHC) and the second point is the bottom/right hand corner (BRHC) of the rectangle (see Figure C.10.5-2).

C.10.5.1.3.5 Multi-lines

SVG example included in semantic annotation example at

https://www.w3.org/2005/Incubator/mmsem/XGR-image-annotation/#solution_NASA

```
<j.1:svgOutline>
```

```
      <svg:svg xml:space="preserve" width="451" heigth="640" viewBox="0 0 451 640">
        <image xlink:href="http://grin.hq.nasa.gov/IMAGES/SMALL/GPN-2000-001171.jpg" x="0" y="0"  width="451" height="640" />
        <rect x="242.0" y="79.0" width="46.0" height="236.0" style="fill:none; stroke:yellow; stroke-width:1pt;"/>
      </svg:svg>
    </j.1:svgOutline>
```