

A Lightweight Ontology for Describing Images

Pat Hayes

Florida IHMC
40 S. Alcaniz St.,
Pensacola, FL 32502
phayes@ihmc.us

Margaret Warren

CARMA Productions
362 Gulf Breeze Pkwy #176
Gulf Breeze, FL 32561
mm@carmapro.com

Abstract

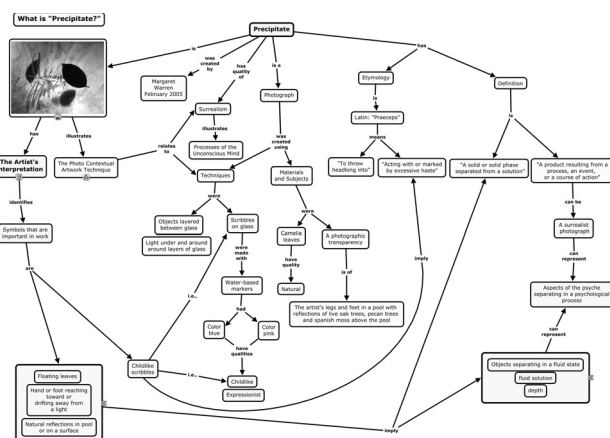
We describe the background and design of a lightweight ontology for descriptions of images made by the image creators, with a specific focus on the description of artwork.

As part of a continuing project devoted to analyzing what image creators themselves say about their images, we have designed a 'lightweight' ontology framework for describing images and their depictions, with a focus on descriptions provided by the image creators themselves.

We began by attempting to discover and semi-formalize what artists say about their own work, using concept mapping as a conceptual capture tool (Eskridge et. al 2006). Concept mapping has been widely used in this way to capture knowledge from domain experts in technical and scientific domains. The resulting concept maps represent a challenge to the expressive powers of current representation languages and ontologies, one that we cannot yet reach. As this example illustrates, they often

contain bodies of descriptive text which we could not fully analyze into a concept map node-and-arc structure. They do however illustrate the wide-ranging nature of the kinds of content that it is desirable to represent.

In another initial project, we designed and juried an art exhibit at a prominent local art gallery, in which artists were asked to write some text about the pieces they submitted, to be exhibited along with the artwork itself. Our interest was in getting a sample of the ways in which artists themselves choose to verbally describe their work. The artist's statements, not surprisingly, often focused on the process of creating the image, rather than describing the finished product. Painters write about their motivations; photographers, about details of the circumstances in which the photograph was taken. We have noted that existing vocabularies for describing finished art works, such as the Visual Resources Association VRA standard (VRA 2007) do not provide for this kind of information, being primarily oriented towards curation rather than creation. As this theme of what one might call the creative process was so ubiquitous in artists' own descriptions, we feel that an ontology for image description must provide for it.



Images and works

It is not clear how to individuate images, particularly when they can be reproduced very easily in various media. For our purposes, an image must be something that can be referred to by a URI reference, and can be transmitted as an HTML payload comprising a digital file of a recognized image file type. However, many images have passed through several stages of reproduction, each of which can be thought of either as producing a new image with its own identity, or simply as another "representation" of the first image. Is a digital photograph of a television screen showing an iconic photograph of Marilyn Monroe itself an image of the movie star? As a practical matter, the ubiquity of EXIF data determines 'the' image in such a chain to be the first point in the sequence of re-renderings which is made using a digital imaging device, usually a digital camera. This is the image to which date, location and other

information are automatically attached. We will call this the *subject image* as it will be the initial subject of our RDF descriptions, and the primary target for linking other data to image descriptions. The ontology therefore provides for descriptions which have a single digital image as the RDF subject. Any subject image is in the class *foaf:Image*.

In cases where one image is a re-rendering of another image or image-like object, we will use the property name *mw:facsimile* rather than 'representation' or 'reproduction'. (This might be informally referred to as 'sameImageAs', one kind of 'approximately identical' (Hayes 2009)). Take for example a drawing posted on Flickr. The original drawing, which itself can be described as an image, consists of pigmented marks on paper. The image visible at the Flickr URI is a digital photograph of the drawing, and this is the subject of our RDF description. To refer to the original drawing, we have the property *mw:facsimile* which relates the subject image to the drawing itself, to which appropriate properties may be attached. In the case of catalogs of artworks, curatorial information, etc., this 'work' image will be the primary object of interest, and will have properties typical for a 'work' such as a medium (pastel on paper), a physical size, a current location (a gallery) and a creator (the artist). This notion of a 'work' only applies in some cases; not every image has a corresponding work, and many works – sculptures, for example – are not images in our sense. In many cases the subject image itself will be the work.

Vocabularies for describing artworks for archival and curating purposes (such as the Visual Resources Association VRA standard and the Getty CDWA (CDWA 2006)) refer to works rather than images, so elements of this class provide natural points of attachment for such descriptions. In general, the subject image is either the work itself, or a facsimile of the work, or simply a depiction – a picture – of the work.

In some cases, the same properties may apply to both the image and the work, but have different values. For example, the work may be a higher-resolution image which is more tightly protected from reproduction than the subject image.

Depiction

Naturally, we wish to be able to refer to the *things* pictured in an image. In many cases, these are the primary items of interest in the description, as for example when searching for an image of a person or a place. Since images can be of just about anything, we make no claims to offer a general technique of describing these depicted things, but only a few basic distinctions between types of relationships to the image. (Contrast for example the LSCOM project (LSCOM 2007) which sets out to provide descriptive categories for everything that might appear in a news video, running to many thousands of concepts.)

We use the general property *mw:depicts* to identify what is depicted by the image. We do not specify the exact nature of what can be depicted, nor do we assume that

depicts is functional (so there can be several things depicted by an image.) The property *mw:depicts* is intended to be very broad, corresponding in scope to the CDWA element *cdwalite:subjectTerm*, described as including people, places and events, but also iconography, themes from literature or concepts or issues, and which can also include "the function or purpose of the work." (CDWA 2006). Depiction is usually understood in a more restricted and 'literal' sense. The senses of depiction used in current ontologies FOAF, Cyc and LSCOM, for example, are narrower in meaning, and are all subproperties of *mw:depicts*. Our notion of depiction is not intended to convey the real existence, or even the possibility, of the things depicted. Salvador Dali's *The Persistence of Memory* *mw:depicts* a melted watch, even though actual watches cannot possibly melt.

This choice of a very broad notion of depiction as the primary meaning may seem to violate the 'common-sense' presumptions of a lightweight ontology intended for general use. However, we believe that generality of scope is important here. By allowing images to depict themes, moods, activities and artistic intentions, we open up the ontology to a much wider range of uses, including those involved in our original project of capturing the ideas of artists.

In order to provide finer control over the description of what an image depicts, we use subproperties with more restricted ranges. Users are encouraged to create such subproperties (or *skos:narrowerTransitive* concepts) as required by particular applications, restricting their ranges and possibly also domains, so these properties could be used to define classes of images such as landscape photographs, medical Xray images, or images illustrating scenes from Shakespeare. The subproperties provided in the initial design are *mw:depictsScene*, referring to the place or area shown in a landscape or cityscape image, or more generally the overall scene or situation depicted, considered as a single unit; *mw:depictsEvent* for images showing or representing some event or activity, and *mw:depictsPerson* for images showing people; both of these latter are also subproperties of *foaf:depicts*, which can also be used to express these properties when finer distinctions are not needed. Such specialized depiction subproperties allow simple OWL definitions of classes of images in terms of the places, people or events that they depict, which is a frequent use case when searching for images.

Depiction is not universal. Some images, such as abstract paintings or technical drawings, may not depict anything. Moreover, depiction is not the appropriate property to use when talking about the two-dimensional lines and shapes in the image itself, rather than in the (even imaginary) three-dimensional world that the image is an image of. We discuss this distinction in more detail later.

Depiction vs. Facsimile

There is no sharp line between an image being a *depiction* or a *facsimile* of another image. Since we allow things

(such as a drawing) which are both a physical object and an image, then either assertion may be made. The intention is that *facsimile* should be preferred when the purpose of the described image is primarily to reproduce, if only approximately, the appearance of the original image. Thus, a thumbnail version of a high-resolution digital image would be a *facsimile* of the image, whereas a photograph of a museum wall with a painting hanging on it would *depict* the work, even if the resolution and color accuracy were high.

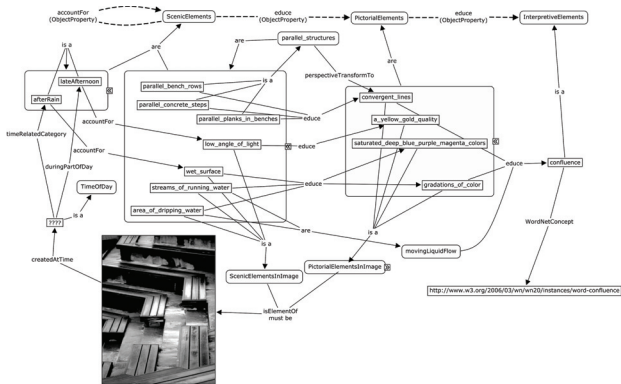
Subject

Something depicted in an image is a *subject* when it is one of the primary things that the image is said to be "about". This need not be the most visually prominent object depicted, but if present, must be depicted: *mw:subject* is a subproperty of *mw:depicts*. The subject may, for example, be the most newsworthy or most famous person in the picture, even when this person is hardly visible. Some images have no subject:



Pictorial Elements

In our earlier work with artists building concept maps, a conceptual distinction emerged between things 'in' the picture and things 'in' the scene depicted. For example, in the following figure, the parallel rows of bleacher benches *in the depicted scene* produce converging lines that are *in the image*.



Ordinary English usage, however, commonly uses the construction "in the image" to refer to things depicted,

rather than to shapes or visual aspects of the image surface itself, so to speak of shapes or lines being *in the image* as opposed to in the scene, creates a conflict. We have found it almost impossible to avoid this usage in our own discussions. We have therefore adopted the convention that anything said to be 'in' an image is something depicted, and that surface pictorial elements such as shapes, lines, colors and textures, are described as being *part of* the image rather than being *in* it. This usage follows what seems to be a common metaphor where an image, constructed of pictorial elements on the image surface, is thought of as a *container* of the depicted scene.

We have a central class *PictorialElement*, and the property pair *hasVisualPart* from images to such elements, and its inverse *visualPartOf*. A pictorial element is any aspect or visual property of any part or parts of the image, or the composition, which may be relevant to its interpretation; what art educators call "elements of design". (Delahunty 1996). It is impossible to give a sharp definition of this class, but it can include geometric elements such as lines or areas, or merely collinearities between scattered pieces of the image, colors and qualities of colors, color harmonies, shapes or properties of shapes, contrasts, textures, etc.. Some pictorial elements, such as an overall color cast or a quality of sharpness, may be impossible to localize within the image by any of the region-geometric devices currently being developed (Muda 2008); we will still refer to them as *part of* the image, however. Thus, *visualPartOf* should not be thought of in a strict mereological sense.

Although we have described depiction as having images as its domain, we allow its extension to pictorial elements, so that a particular element can be said to depict a thing in the depicted scene. This allows quite detailed analyses of the ways that parts or aspects of the image refer to or indicate entities in the depicted scene, for example that the shape of a line depicts the movement of an arm. The simplest way to do this is to treat *PictorialElements* as a subclass of images.

Certain special cases of pictorial elements may be defined in terms of what they are used to depict. We will refer to these as *renderings*. For example, we might define the notion of a "rendering of an eye" to refer to any pictorial element which could be naturally interpreted as depicting an eye. Note that this does not necessarily imply that such a pictorial element depicts what it is a rendering of, in any given image. The distinction between rendering and depiction allows the description of visual metaphors of various kinds.

Just as an image may depict many things but only a few of them are said to be its actual subject, it may be that a small number of the pictorial elements making up an image are more salient than the rest. We allow the use of the special property *mw:primaryElement* to indicate such elements. Just as with *mw:subject*, the use of this property is subjective and imprecise, but we believe that its utility outweighs potential problems with its use.

Relationships between pictorial elements can of course be extremely complex, and would require a rich ontology of compositional, geometric and pictorial ideas to be fully described. We plan to begin with tentative steps towards such a pictorial ontology in future work.

Interpretation

A great deal of what image creators say about their images is concerned with aspects variously described as emotional, interpretive, stylistic or conceptual, having to do with what the image suggests rather than what it depicts. A blue period Picasso might be described as conveying a *mood of sadness*. We will loosely describe these all as *interpretive* aspects of the image, or as being concerned with its *interpretation*.

Again, no fully adequate ontology, or even a satisfactory basis for one, is yet available for such descriptions. However, some utility can be found in the structured keyword tags developed within the Getty image corpus (Getty 2009), which has evolved a semi-structured catalog of recommended image tags, many of which seem to be concerned with interpretation in this sense; and the corpus itself is a large source of images which are publicly (although not freely) accessible by using these tags. We therefore propose, as an initial step, to utilize these tags in our descriptions by allowing them to be attached to images by the datatype property *mw:gettyTag*. Although they are simple text literals, their status as recommended Getty image tags allows them to be used to retrieve images from the Getty corpus, so in a sense they provide a very weak form of "link" between an interpretive description and images.

It is not conventional to speak of a tag as a link, but if we define "link" to mean some textual object which supports access, to a resource by computational means, then a Getty image tag can be viewed as a link in this broad sense; and a collection of such tags is an even better link.

We have experimented with more structured descriptions in which actual entities referred to as *moods*, *interpretations*, etc. – collectively, as *concepts* – are associated with an image as actual entities, corresponding to nodes in the RDF description. This is similar to the approach towards datatyping described in (Berners-Lee 1998), and similar arguments can be given for its superiority as a modeling technique. However, as we have not yet found a clear use case for this level of sophistication, this option is open for future use.

Links in tags

Descriptions of an image using our ontology are stored as RDFa which links to the image URL, allowing the descriptions to inhabit a conventional image corpus rendered in HTML, for example in an online image gallery. The reverse direction, however, requires that a link be attached to the image file itself. While current photo metadata standards such as IPTC (<http://www.iptc.org/>)

and Adobe XMP (<http://www.adobe.com/products/xmp>) allow for such links, these may be modified when images are uploaded to sites such as Flickr. A general solution seems to be to incorporate a link as the text in an image tag. Most image corpora allow tags to be attached to images.

In order to be able to recognize such a URI-containing tag, the URI is prefixed with a character sequence '*mw*' for some alphanumeric character *. Very few of these sequences are currently used as image tags. For example, the tag

```
1mw1bit.ly/7JSIDd
```

encodes a link (via redirection) to the URL. <http://purl.org/net/mw/image/surfingWithPsyche>.

Precision, vagueness and subcategories

When designing an ontology for a topic this far away from any attempt at a complete formalization, one is obliged to use terminology which is underspecified. Our class of *pictorial elements* is probably the most glaring example here. There are several ways to approach this situation. One is to try to identify at least the beginnings of an upper-level categorization or taxonomy of the classes in question, and provide users with a vocabulary to register these distinctions. One sees this approach for example in the VRA Core vocabulary, which distinguishes *conceptTopic*, *descriptiveTopic* and *iconographicTopic* relationships between a work and various possible themes or topics it might have. We have, however, deliberately not followed this course, for example by leaving the class *PictorialElement* without any named subclasses.

The provision of ways to make taxonomic distinctions, without clear or widely accepted rationales for putting items into those categories, is more likely to cause harm than good. Our strategy has therefore been to 'retreat upwards' where possible, introducing large classes with highly inclusive criteria, expecting that distinctions between cases will emerge in the future as a byproduct of actual usage. The fact that such classes and properties have no exact defining criteria for membership is in line with a general attitude towards ambiguity explained in (Hayes & Halpin 2008).

The creative process, and the need for blank nodes

In order to attach artists' descriptions of their own creative processes to an image, we need a notion like *the circumstances surrounding the creation of the image*, that would then be the primary subject of such descriptions of when, where, by whom and for what reasons, etc., the image was created. However, this raises a difficulty. The current recommended practices for using RDF in linked data applications (Bizer, Cyganiak & Heath, 2007) suggest that blank nodes not be used, but their use here seems to be almost required. It is asking too much, surely, to expect

people to coin a unique URI to refer to the circumstances of creation of every image.

In a more expressive logic, the natural course to take here would be to use a function, which as well as implicitly asserting the existence of the function value, also automatically creates a name for it, by application of the function name to the argument name. But such a name - strictly, a logical term - would not be a URI.

We could imitate functional terms by the systematic use of a URI naming convention. For example, we might use the URI formed by adding *"creation"* to the root URI of the main description. But this kind of naming convention goes outside any existing semantic standards, and encodes significant meaning in a form where it is invisible to SWeb reasoning engines. The RDF blank node construction was explicitly intended to be used in this way, so it seems reasonable to continue to use it here, in spite of the negative recommendations. We therefore propose to allow RDF with blank nodes as the 'connector' nodes denoting the creation event of an image or work, linked from the image or work URI by the special property *mw:creation*. This then makes possible markup of the following sort, inspired by the concept map in the earlier figure:

```
http://farm1.static.flickr.com/
30/88603499_6c09fc71e4_o.jpg mw:creation
_:ev1 .
_:ev1 rdf:type http://www.carmapro.com/
concepts/PhotoContextualArt/ .
_:ev1 http://www.w3.org/2006/
time#hasDateTimeDescription _:d .
_:d http://www.w3.org/2006/time#year
"2005"^^xsd:gYear .
_:d http://www.w3.org/2006/time#month
"--02"^^xsd:gMonth .
_:ev1 :locatedIn http://sws.geonames.org/
4168228/ .
```

Exactly what counts as a creation is not tightly specified; it may be thought of as a combination of a place and a time-period, or more generally as a 'situation' in the sense described by Situation Semantics, i.e. roughly, a partially described set of circumstances defined by some information (Kratzer 2009). It may be partially defined by the mental activity or motivations of the creator, and it may be thought of as an extended process or a singular occurrence. All of these possibilities are deliberately left open.

Summary of ontology

Classes

Image A marked surface of some kind, including virtual surfaces. The class includes digital images but also

drawings, paintings on 2-dimensional surfaces, etc.. The subject image of an MW description is always a digital image, typically with EXIF data attached.

Work An individual work of art depicted by an image, or which an image is a facsimile of. Not all images are facsimiles of a Work, and a Work may or may not be an Image. Works may be physical objects with attributes such as sizes, location, etc..

PictorialElement Any aspect or property of an image which is relevant to its interpretation. Examples include geometric elements (lines, shapes), colors and color effects, textural aspects, etc.. Some pictorial elements *may* be locatable in an image by image regions. This is a very broad class which is hard to define, but it includes geometric shapes, lines, relationships between them such as parallelism and inclusion, color and textural properties, and so on.

Properties (Domain, Range)

facsimileOf (Image, Image) The subject is a more or less precise re-rendering, reproduction or visual representation of the object. A photograph of a drawing taken for exhibition in a catalog would be the typical example. It does not imply exactness of reproduction: a lower-resolution thumbnail can be a facsimile, for example.

depicts (Image, Thing) Primary relation between an image or a pictorial element in an image, and anything it can be said to picture, show, exhibit or exemplify. The thing depicted may be an event, theme or intention of an artist in creating an image. This is intended to be used in a very general, broad sense. (Superproperty of *foaf:depicts*.)

depictsPerson (Image, Foaf:Person) Subproperty of *depicts* and of *foaf:depicts*

depictsScene (Image, Thing) The entire place or location visible in an image, including all its visible contents. Applies to images of a place such as landscape photographs, street scenes, etc., but may be applied to other images if an ontology admits imaginary places. Scenes may be loosely identified with geographical places. An example of a scene might be *Paris, view to the east from Mont St. Michel*. Subproperty of *depicts*.

depictsEvent (Image, Thing) Subproperty of *depicts*.

subject (Image, Thing) The depicted entity which the image can be said to be 'about'. Not all images have a subject.

gettyTag (Image, literal) Value is a string which could be used as an image tag according to the Getty

image corpus guidelines for image tagging. Note that this is a restricted set of strings.

renderingOf (PictorialElement, Thing) Relationship between pictorial element and a class of things roughly corresponding to the English phrase "looks like".

hasVisualPart (PictorialElement, Image) Relationship between a pictorial element and the image containing or exhibiting the element.

visualPartOf (Image, PictorialElement) inverse of *hasVisualPart*.

creation (Work, creation event) relates a work to the event or process of its creation.

A. Kratzer, 2009. Situations in Natural Language Semantics. *Stanford Encyclopedia of Philosophy*. <http://plato.stanford.edu/entries/situations-semantics/>

LSCOM 2007. *Large Scale Concept Ontology for Multimedia*. <http://www.lsc.com.org/>

Z. Muda, 2008. Ontological Description of Image Content Using Regions Relationships. *Proc. ESWC Phd Symposium, European Semantic Web Conference (ESWC2008)* Tenerife, Spain. <http://eprints.ecs.soton.ac.uk/16612/>

VRA 2007. *VRA Core 4.0 Element Outline*. Visual Resources Association Data Standards Committee. <http://www.vraweb.org/projects/vracore4/>

References

T. Berners-Lee, 1998. *Interpretation Properties*. <http://www.w3.org/DesignIssues/Interpretations.html>

C. Bizer, R. Cyganiak & T. Heath, 2007. *How to Publish Linked Data on the Web*. <http://www4.wiwiss.fu-berlin.de/bizer/pub/LinkedDataTutorial/>

CDWA 2006. *CDWA Lite: Specification for an XML Schema for Contributing Records via the OAI Harvesting Protocol*, J. Paul Getty Trust, 17 July 2006. http://www.getty.edu/research/conducting_research/standards/cdwa/cdwalite/cdwalite.pdf

M. Delahunt, 1996. *The Art Dictionary*. <http://www.artlex.com>

T. Eskridge, P. Hayes, R. Hoffman & M. Warren, 2006. Formalizing the Informal: A Confluence of Concept Mapping and the Semantic Web. *Concept Maps: Theory, Methodology, Technology. Proc. of the Second Int. Conference on Concept Mapping*. A. J. Cañas, J. D. Novak, Eds. San José, Costa Rica. <http://cmc.ihmc.us/cmc2006Papers/cmc2006-p199.pdf>

Getty 2009. *Getty Images Keyword Guide*. http://imagery.gettyimages.com/getty_images_keyword_guide/usa/keyword_guide.en-us.pdf

P. Hayes & H. Halpin, 2008. In Defense of Ambiguity. *International Journal on Semantic Web and Information Systems* 4(2) pp 1-18. <http://www.ibiblio.org/hhalpin/homepage/publications/indefenseofambiguity.html>